

RUI JORGE TRAMONTIN JUNIOR

**UM MODELO BASEADO EM CONTEXTO PARA EXPANSÃO
DE CONSULTAS SEMÂNTICAS EM REDES COLABORATIVAS
DE ORGANIZAÇÕES**

**FLORIANÓPOLIS
2011**

UNIVERSIDADE FEDERAL DE SANTA CATARINA

**PROGRAMA DE PÓS-GRADUAÇÃO
EM ENGENHARIA ELÉTRICA**

**UM MODELO BASEADO EM CONTEXTO PARA EXPANSÃO
DE CONSULTAS SEMÂNTICAS EM REDES COLABORATIVAS
DE ORGANIZAÇÕES**

Tese submetida à
Universidade Federal de Santa Catarina
como parte dos requisitos para a
obtenção do grau de Doutor em Engenharia Elétrica.

RUI JORGE TRAMONTIN JUNIOR

Florianópolis, junho de 2011.

Catálogo na fonte pela Biblioteca Universitária
da
Universidade Federal de Santa Catarina

T771m Tramontin Junior, Rui Jorge

Um modelo baseado em contexto para expansão de consultas semânticas em redes colaborativas de organizações [tese] / Rui Jorge Tramontin Junior ; orientador, Ricardo José Rabelo. - Florianópolis, SC, 2011.

283 p.: il., grafs., tabs.

Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Engenharia Elétrica.

Inclui referências

1. Engenharia elétrica. 2. Empresas - Redes sociais on-line. 3. Recuperação da informação. 4. Ontologia. 5. Web semântica. 6. Serviços da Web. I. Rabelo, Ricardo José. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação em Engenharia Elétrica. III. Título.

CDU 621.3

**UM MODELO BASEADO EM CONTEXTO PARA EXPANSÃO
DE CONSULTAS SEMÂNTICAS EM REDES
COLABORATIVAS DE ORGANIZAÇÕES**

Rui Jorge Tramontin Junior

‘Esta Tese foi julgada adequada para obtenção do Título de Doutor em Engenharia Elétrica, Área de Concentração em *Automação e Sistemas*, e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina.’

Prof. Ricardo José Rabelo, Dr.
Orientador

Prof. Chihab Hanachi, Dr.
Co-orientador

Prof. Roberto de Souza Salgado, Ph. D.
Coordenador do Programa de Pós-Graduação em Engenharia Elétrica

Banca Examinadora:

Florianópolis, 6 de junho de 2011.

Prof. Ricardo José Rabelo, Dr.
Presidente

Prof. Celson Pantoja Lima,
Doutor

Prof. Luiz Márcio Spinosa,
Dr. ès Sci.

Prof. Renato Fileto,
Doutor

Prof. Ricardo Triska,
Dr. Eng.

AGRADECIMENTOS

Ao longo da escrita desta tese pude contar com a ajuda de muitas pessoas que de alguma forma deram a sua contribuição e que, em menor ou maior escala, foram fundamentais para o término bem sucedido deste trabalho.

Primeiramente, agradeço aos meus pais, que nunca pouparam esforços para me ajudar e sempre me apoiaram no caminho que escolhi.

Agradeço também ao meu orientador, Ricardo José Rabelo, pela dedicação, competência e, sobretudo, paciência durante a orientação deste trabalho.

Aos companheiros do GSIGMA, Carlos Eduardo Gesser, Fabiano Baldo, Leandro Loss, Saulo Popov Zambiasi, Maiara Heil Cancian, Marcus Vinicius Drissen Silva, Omir Alves, Daniel Mayer, Edmilson Rampazo Klen, Alexandra Pereira Klen e Alexandre Perin de Souza. O ambiente oferecido pelo grupo bem como a ajuda individual foram muito importantes durante o desenvolvimento do trabalho.

Ao meu co-orientador, Chihab Hanachi, que me guiou durante meu estágio no exterior, e cujas ideias foram a semente para o modelo proposto nesta tese. Agradeço também aos amigos que fiz durante esse período e que ajudaram a enriquecer a minha experiência: Wassim Bouaziz, Eric Andonoff, Stéphanie Combettes, Ines Thabet, Mohamed Amine Chaabane, Omar Tahir, Christophe Sibertin-Blanc, Abdelaziz Berrais e Karim Djemal. Agradeço também à Chantal Morand, coordenadora do grupo de pesquisas, que me recebeu muito bem e que me ajudou quando tive necessidade.

À Fabiana Ávila, que de maneira incondicional me deu apoio e forças para continuar durante a “reta final” do trabalho. Mesmo nos momentos mais difíceis, teve paciência e dedicação para me manter no caminho e não me permitiu desistir. Sua ajuda foi fundamental e devo muito a ela por conseguir terminar a tese com sucesso.

À Eliane Pozzebon, pela amizade e pelas dicas que me ajudaram muito durante meu estágio na França.

Aos amigos e colegas que de alguma forma me ajudaram ou deram apoio, em particular a Marco Antônio Floriano de Oliveira, André Fortunato, Rodrigo Panchiniak Fernandes, Juliana Ávila e Anderson Fernandes

Perez.

À Marília Andrea Rathje, que esteve ao meu lado durante grande parte do trabalho, e cujo apoio e companheirismo foram fundamentais, especialmente nos preparativos para a viagem e o período de vivência no exterior.

Aos novos amigos e colegas da UDESC, que me apoiaram e acompanharam o final da escrita deste trabalho.

Agradeço aos membros do exame de qualificação e da banca examinadora desta tese, que contribuíram enormemente para o crescimento da qualidade final do trabalho.

Aos professores e funcionários do PGEEL, por proporcionarem as condições necessárias ao meu crescimento ao longo do curso; em particular, agradeço ao Wilson e ao Marcelo, que nunca pouparam esforços em resolver os problemas administrativos.

Finalmente, agradeço ao CNPq, pelos recursos financeiros, incluindo os recursos para o estágio no exterior.

Resumo da Tese apresentada à UFSC como parte dos requisitos necessários para obtenção do grau de Doutor em Engenharia Elétrica.

UM MODELO BASEADO EM CONTEXTO PARA EXPANSÃO DE CONSULTAS SEMÂNTICAS EM REDES COLABORATIVAS DE ORGANIZAÇÕES

Rui Jorge Tramontin Junior

Junho, 2011

Orientador: Prof. Ricardo José Rabelo, Dr.

Co-orientador: Prof. Chihab Hanachi, Dr.

Área de Concentração: Automação e Sistemas

Palavras-chave: Redes Colaborativas de Organizações, Recuperação de Informação, Ontologias, Contexto, Expansão da Consulta.

Número de Páginas: 283

RESUMO: As novas tendências do mundo globalizado levaram organizações e profissionais a focarem em estratégias baseadas em trabalho colaborativo. Tais estratégias têm sido enquadradas no conceito mais geral de Rede Colaborativa (RC), onde organizações e indivíduos trabalham juntos para aumentar o acesso a novas oportunidades de negócio, compartilhar riscos, reduzir custos e atingir metas que seriam inalcançáveis individualmente. A implantação de RCs depende da existência de infraestruturas computacionais que provejam funcionalidades de suporte à colaboração, incluindo compartilhamento e busca de informações, integração de sistemas, gestão de segurança, entre outros. O foco desta tese está na funcionalidade relacionada à busca de informação, requisito fundamental considerando-se o fato de que os parceiros de tal rede compartilham informações que precisam ser recuperadas. Além disso, a busca de informação se justifica pelo seu uso potencial em diversas outras aplicações necessárias a RCs, tais como: suporte à gestão de conhecimento, seleção de indicadores, busca de parceiros, auxílio no suporte à decisão, entre outras. Nesse sentido, este trabalho propõe um arcabouço que define uma infraestrutura de serviços de suporte à busca de informação em RCs. A estratégia adotada neste trabalho foi dividida em dois passos: em primeiro lugar utilizou-se ontologias para o enriquecimento das fontes de informação, com base na definição de anotações semânticas. Ontologias foram também usadas como base para a definição de consultas semânticas. O segundo passo envolveu a utilização do contexto do usuário visando a melhoria dos resultados da busca. No âmbito de uma RC, o contexto pode ser definido por diversos elementos, incluindo processo,

tarefa e papel desempenhado pelo usuário. A abordagem utilizada nessa etapa consistiu na definição de um modelo do contexto, que é associado à ontologia da RC, e de um conjunto de regras que, com base no contexto atual do usuário, efetuam uma expansão na consulta original. Em suma, a abordagem proposta usa o contexto do usuário para sugerir novos tópicos a serem buscados aplicando-se restrições à consulta definida pelo usuário. A avaliação deu-se a partir de experimentos com coleções de teste, onde medidas baseadas em precisão e cobertura foram utilizadas na comparação do modelo proposto com um sistema baseado em palavras-chave e com outro baseado em ontologias.

Abstract of Thesis presented to UFSC as a partial fulfillment of the requirements for the degree of Doctor in Electrical Engineering.

**A CONTEXT-BASED MODEL FOR SEMANTIC QUERY
EXPANSION IN COLLABORATIVE NETWORKED
ORGANIZATIONS**

Rui Jorge Tramontin Junior

June, 2011

Advisor: Prof. Ricardo José Rabelo, Dr.

Co-advisor: Prof. Chihab Hanachi, Dr.

Area of Concentration: Automation and Systems

Keywords: Collaborative Networked Organizations, Information Retrieval, Ontologies, Context, Query Expansion.

Number of Pages: 283

ABSTRACT: The new trends of the globalized world have led organizations and professionals to focus on strategies based on collaboration. Such strategies have been classified in the more general concept of collaborative network (CN), where organizations and individuals work together in order to increase their access to new business opportunities, share risks, reduce costs and achieve goals that would be impossible to achieve individually. The realization of CNs depends on computational infrastructures which provide supporting functionalities for collaboration, including information sharing and search, systems integration, security management, among others. The focus of this thesis is the information search functionality, considered a fundamental requirement as partners of these networks share information that should be retrieved. Besides that, information search can be integrated as part of several CN-related applications, such as: tools for supporting knowledge management, performance indicators search, partners search and decision supporting systems. In this sense, this work proposes a framework composed of services for supporting information search in CNs. The approach adopted in this work was divided in two steps: firstly, ontologies were used to enrich the information sources by means of semantic annotations. Ontologies were also used to support the definition of semantic queries. The second step involved the use of the user context in order to improve the quality of the search results. In the scope of a CN, the context can be defined by several elements, including the process, task and role performed by the user. The approach used in this step consisted of a context model, which is linked to the ontology used by the CN partners, and a set of rules. The

rules take the current user context in order to expand the original query. In sum, the proposed approach takes the user context into account to suggest new restrictions to be included in the original query. The evaluation was performed using some test collections, based on standard precision and recall measures. The proposed model was compared with a keyword-based system and an ontology-based system.

Sumário

| | | |
|----------|---|-----------|
| 1 | Introdução | 1 |
| 1.1 | Tema de Pesquisa e Motivação | 2 |
| 1.1.1 | Características da Busca em Redes Colaborativas | 4 |
| 1.1.2 | Problemas relacionados à Busca de Informação em Redes Colaborativas | 5 |
| 1.2 | Problema de Pesquisa | 8 |
| 1.2.1 | Perspectiva Semântica | 8 |
| 1.2.2 | Perspectiva de Contexto | 8 |
| 1.2.3 | Visão Geral | 10 |
| 1.3 | Objetivos | 10 |
| 1.3.1 | Objetivo Geral | 10 |
| 1.3.2 | Objetivos Específicos | 11 |
| 1.4 | Justificativa do Trabalho | 11 |
| 1.4.1 | Adequação às Linhas de Pesquisa do Curso | 12 |
| 1.5 | Delimitação da Pesquisa | 12 |
| 1.6 | Contribuições do Trabalho | 13 |
| 1.7 | Procedimentos Metodológicos | 14 |
| 1.7.1 | Classificação da Pesquisa | 14 |
| 1.7.2 | Etapas da Pesquisa | 14 |
| 1.7.3 | Assuntos selecionados para a pesquisa bibliográfica | 16 |
| 1.8 | Estrutura do Documento | 17 |
| 2 | Redes Colaborativas e Busca Empresarial | 19 |
| 2.1 | Conceitos Básicos de Redes Colaborativas | 19 |
| 2.1.1 | Classificação das Redes Colaborativas | 20 |
| 2.2 | Infraestruturas de Colaboração | 23 |
| 2.2.1 | Infraestrutura de Referência para RCs | 25 |
| 2.3 | Busca Empresarial | 27 |
| 2.3.1 | Requisitos | 27 |
| 2.3.2 | Abordagens e Tecnologias | 29 |

| | | |
|----------|---|-----------|
| 3 | Ontologias e Web Semântica | 31 |
| 3.1 | Introdução | 31 |
| 3.2 | Representação de Ontologias | 32 |
| 3.3 | Tipos de Ontologias | 33 |
| 3.4 | Áreas de Pesquisa e Aplicação | 34 |
| 3.5 | Web Semântica | 35 |
| 3.5.1 | Arquitetura da Web Semântica | 35 |
| 3.5.2 | Ferramentas | 42 |
| 3.6 | Anotações Semânticas | 43 |
| 3.6.1 | Extração de Informação | 43 |
| 3.6.2 | Anotação Semântica | 45 |
| 3.6.3 | Estado da Arte | 47 |
| 3.6.4 | Comentário sobre Plataformas para Anotação Semântica | 49 |
| | | |
| 4 | Recuperação de Informação | 51 |
| 4.1 | Introdução | 51 |
| 4.2 | Conceitos | 52 |
| 4.2.1 | O Conceito de Índice | 52 |
| 4.2.2 | Visão Lógica dos Documentos e Operações de Texto | 52 |
| 4.2.3 | O Processo de Recuperação de Informação | 53 |
| 4.3 | Modelos para RI | 55 |
| 4.4 | Avaliação de Sistemas de Recuperação de Informação (RI) | 56 |
| 4.4.1 | Medidas Básicas: Precisão e Cobertura | 57 |
| 4.4.2 | Curva de Precisão x Cobertura | 59 |
| 4.4.3 | Valores Detalhados | 61 |
| 4.4.4 | Outras Medidas | 62 |
| 4.4.5 | Coleções de Referência | 63 |
| 4.4.6 | Comentários a Respeito das Coleções de Teste | 64 |
| 4.5 | Problemas relacionados à RI | 65 |
| 4.6 | RI Baseada de Ontologias | 66 |
| 4.6.1 | Estado da Arte | 68 |
| 4.7 | RI Baseada em Contexto | 69 |
| 4.7.1 | Definição de Contexto | 69 |
| 4.7.2 | Captura do Contexto | 71 |
| 4.7.3 | Representação do Contexto | 71 |
| 4.7.4 | Uso do Contexto | 72 |
| 4.7.5 | Expansão da Consulta | 73 |
| 4.7.6 | Estado da arte na expansão de consultas usando ontologias | 77 |
| 4.7.7 | Outras abordagens e Comentários sobre o Estado da Arte | 78 |

| | | |
|----------|---|------------|
| 5 | Modelo Proposto | 81 |
| 5.1 | Visão Geral da Abordagem | 81 |
| 5.2 | Perspectiva Semântica | 82 |
| 5.2.1 | Ontologia para RCs | 84 |
| 5.3 | Perspectiva de Contexto | 88 |
| 5.4 | Modelo de Contexto | 90 |
| 5.4.1 | Ligação do Modelo de Contexto à Ontologia | 92 |
| 5.5 | As Regras de Customização | 94 |
| 5.5.1 | Selecionando conceitos relacionados pelos conceitos da consulta | 95 |
| 5.5.2 | Selecionando conceitos que têm relação com os conceitos da consulta | 96 |
| 5.5.3 | Selecionando conceitos que pertencem à hierarquia dos conceitos da consulta | 96 |
| 5.6 | Posicionamento numa Arquitetura de Sistema | 98 |
| 5.6.1 | Casos de Uso | 100 |
| 5.7 | Considerações | 101 |
| | | |
| 6 | Protótipo Computacional | 103 |
| 6.1 | Arquitetura de Implementação | 103 |
| 6.1.1 | Tecnologias e Ferramentas Utilizadas | 103 |
| 6.2 | Implementação dos Serviços do <i>K-search</i> | 106 |
| 6.2.1 | Serviço para Navegação de Ontologias | 107 |
| 6.2.2 | Serviço para Gestão de Documentos | 108 |
| 6.2.3 | Serviço Motor de Busca | 109 |
| 6.2.4 | Serviço Capturador de Contextos | 110 |
| 6.2.5 | Serviço Motor de Regras | 112 |
| 6.3 | Integração com a Ontologia PROTON | 114 |
| 6.4 | Demais Elementos da Arquitetura | 117 |
| 6.4.1 | Aplicativos | 117 |
| 6.4.2 | Elementos Externos | 117 |
| 6.5 | Considerações sobre o Protótipo | 119 |
| | | |
| 7 | Avaliação do Modelo | 121 |
| 7.1 | Avaliação da Perspectiva Semântica | 121 |
| 7.2 | Avaliação da Perspectiva de Contexto | 123 |
| 7.2.1 | Metodologia de Avaliação | 124 |
| 7.3 | Preparação dos Experimentos | 125 |
| 7.3.1 | Instâncias para a Base de Conhecimento | 126 |
| 7.3.2 | Documentos para a Coleção de Teste | 132 |
| 7.3.3 | Consultas | 133 |
| 7.3.4 | Instâncias do Modelo de Contexto | 136 |
| 7.3.5 | CrITÉrios de Relevância | 137 |

| | | |
|-------------------|---|------------|
| 7.4 | Análise dos Resultados | 139 |
| 7.4.1 | Busca por Indicadores de Desempenho | 140 |
| 7.4.2 | Busca por Documentos | 147 |
| 7.5 | Considerações | 153 |
| 8 | Conclusões | 155 |
| 8.1 | Contribuições | 156 |
| 8.2 | Produção Bibliográfica | 157 |
| 8.3 | Limitações da Abordagem Proposta | 157 |
| 8.4 | Sugestões para Trabalhos Futuros | 158 |
| Apêndice A | Interfaces dos Serviços do <i>K-Search</i> | 159 |
| A.1 | Serviço para Navegação de Ontologias | 159 |
| A.2 | Serviço para Gestão de Documentos | 169 |
| A.3 | Serviço Motor de Busca | 178 |
| Apêndice B | Diagramas UML do Protótipo | 185 |
| B.1 | Serviço para Navegação de Ontologias | 185 |
| B.2 | Serviço para Gestão de Documentos | 186 |
| B.3 | Serviço Motor de Busca | 187 |
| B.4 | Serviço Capturador de Contextos | 188 |
| B.5 | Serviço Motor de Regras | 188 |
| B.6 | Implantação do Protótipo | 192 |
| Apêndice C | Base de conhecimento | 193 |
| Apêndice D | Modelo e Regras na sintaxe do Jess | 203 |
| Apêndice E | Coleções de Teste | 209 |
| E.1 | Indicadores de Desempenho | 209 |
| E.2 | Documentos | 215 |
| Apêndice F | Ferramentas de Suporte | 241 |
| F.1 | Geração de Relatórios de Desempenho de Sistemas de RI | 241 |
| F.1.1 | Avaliação Simples | 241 |
| F.1.2 | Comparação entre Sistemas | 244 |
| Apêndice G | Questionário para Definição dos Critérios de Relevância | 249 |
| Apêndice H | Respostas do Questionário | 261 |

| | | |
|-----------------------------------|---|------------|
| Apêndice I | Produção Bibliográfica | 263 |
| I.1 | Artigos publicados em congressos internacionais | 263 |
| I.2 | Artigos publicados em revistas internacionais | 263 |
| Referências Bibliográficas | | 264 |

Lista de Figuras

| | | |
|-----|---|----|
| 1.1 | Foco do Trabalho. | 10 |
| 1.2 | Delimitação da Pesquisa. | 13 |
| 2.1 | Classificação das Redes Colaborativas. | 21 |
| 2.2 | Infraestrutura de TIC e serviços verticais. | 24 |
| 2.3 | Arquitetura de Referência da ICT-I. | 26 |
| 3.1 | Arquitetura da Web Semântica. | 36 |
| 3.2 | Exemplo de Documento XML. | 38 |
| 3.3 | Exemplo de declaração RDF. | 39 |
| 3.4 | RDF representado pela <i>Notation3</i> | 39 |
| 3.5 | RDF representado na forma de uma grafo. | 39 |
| 3.6 | Classes e propriedades definidas em RDFS. | 40 |
| 3.7 | Anotações semânticas apontando para algumas instâncias (em branco) de um fragmento de ontologia (em cinza). | 45 |
| 3.8 | Classificação das plataformas para anotação semântica. | 46 |
| 3.9 | Plataformas para anotação semântica e suas dependências. | 47 |
| 4.1 | Arquitetura conceitual de um sistema de RI. | 54 |
| 4.2 | Documentos relevantes para uma dada consulta. | 58 |
| 4.3 | Precisão nos 11 níveis de cobertura padrão. | 60 |
| 4.4 | Curvas médias de precisão x cobertura entre dois sistemas. | 61 |
| 4.5 | Busca baseada em ontologia. | 67 |
| 4.6 | Tipos e abordagens para expansão da consulta. | 75 |
| 5.1 | Posicionamento da abordagem em relação aos elementos de busca em Rede Colaborativas (RCs). | 82 |
| 5.2 | Perspectiva Semântica. | 83 |
| 5.3 | Ontologia para RCs. | 85 |
| 5.4 | Fragmento da ontologia que define indicadores de desempenho. | 86 |
| 5.5 | Fragmento da ontologia que define os papéis de um participante. | 87 |
| 5.6 | Níveis da ontologia de RCs. | 89 |

| | | |
|------|--|-----|
| 5.7 | Perspectiva de Contexto. | 89 |
| 5.8 | Modelo de contexto. | 91 |
| 5.9 | Ligação entre os elementos de contexto com a ontologia. . . | 93 |
| 5.10 | Fragmento de uma ontologia. | 95 |
| 5.11 | Arquitetura Conceitual. | 99 |
| 5.12 | Casos de uso do sistema de busca. | 101 |
| 5.13 | Níveis da ontologia em relação às regras. | 102 |
| | | |
| 6.1 | Arquitetura de Implementação do <i>K-search</i> | 104 |
| 6.2 | Diagrama de classes do <i>k-search</i> , mostrando as principais classes e pacotes. | 107 |
| 6.3 | Diagrama de objetos para uma consulta semântica. | 110 |
| 6.4 | Arquivo “ <i>context.properties</i> ”. | 111 |
| 6.5 | <i>Templates</i> para estrutura do modelo de contexto. | 112 |
| 6.6 | Trecho do metamodelo <i>Resource Description Framework (RDF)/Web Ontology Language (OWL)</i> , definido na forma de triplas. . . | 113 |
| 6.7 | Trecho da Ontologia de Indicadores de Desempenho, codificada na forma de triplas. | 114 |
| 6.8 | Regra 1 (associação da consulta ao contexto) codificada na sintaxe do Jess. | 114 |
| 6.9 | <i>Templates</i> para estrutura do modelo de contexto, importadas de classes Java. | 115 |
| 6.10 | Classes utilizadas para a integração do <i>k-search</i> como o Jess. | 116 |
| 6.11 | <i>Portlet</i> para publicação de documentos. | 117 |
| 6.12 | <i>Portlet</i> para busca semântica: (a) definição da consulta; (b) resultados; (c) visualização de um documento. | 118 |
| | | |
| 7.1 | Metodologia de Avaliação. | 126 |
| 7.2 | Fragmento da base de conhecimento, representado na forma de grafo. | 130 |
| 7.3 | Fragmento da base de conhecimento, representado na forma de predicados. | 130 |
| 7.4 | Representação de uma instância e seus dois apelidos. | 131 |
| 7.5 | Instâncias representando o Ambiente de Criação de Organizações Virtuais (ACV) “Helice” e seus dois apelidos. | 132 |
| 7.6 | Documento usado na coleção de testes. | 133 |
| 7.7 | Curvas de Precisão x Cobertura para a consulta 1. | 141 |
| 7.8 | Curvas de Precisão x Cobertura para a consulta 2. | 143 |
| 7.9 | Curvas de Precisão x Cobertura para a consulta 3. | 144 |
| 7.10 | Curvas de Precisão x Cobertura para a consulta 4. | 146 |
| 7.11 | Curvas de Precisão x Cobertura médias para a busca por indicadores. | 147 |
| 7.12 | Curvas de Precisão x Cobertura para a consulta 5. | 149 |

| | | |
|------|--|-----|
| 7.13 | Curvas de Precisão x Cobertura para a consulta 6. | 150 |
| 7.14 | Curvas de Precisão x Cobertura para a consulta 7. | 152 |
| 7.15 | Curvas de Precisão x Cobertura médias para a busca por documentos. | 153 |
| | | |
| B.1 | Diagrama de classes do serviço de navegação na ontologia. . . | 186 |
| B.2 | Diagrama de classes do serviço de gestão de documentos. . . | 187 |
| B.3 | Diagrama de classes do serviço motor de busca. | 189 |
| B.4 | Diagrama de sequência da busca usando o contexto. | 190 |
| B.5 | Diagrama de classes do serviço de captura de contextos. . . | 191 |
| B.6 | Diagrama de classes do serviço motor de regras. | 191 |
| B.7 | Diagrama de implantação do <i>k-search</i> | 192 |
| | | |
| F.1 | Tela da ferramenta de avaliação (modo de avaliação simples). . | 241 |
| F.2 | Formato de arquivo para os resultados esperados. | 242 |
| F.3 | Formato de arquivo para os resultados obtidos. | 243 |
| F.4 | Formato de arquivo para o relatório de desempenho (fragmento). | 244 |
| F.5 | Relatório de avaliação de um sistema, após a transformação com <i>Extensible Stylesheet Language (XSL)</i> | 245 |
| F.6 | Tela da ferramenta de avaliação (modo de comparação entre sistemas). | 245 |
| F.7 | Relatório de comparação entre sistemas, visualizado num navegador Web. | 247 |

Lista de Tabelas

| | | |
|------|---|-----|
| 3.1 | Comparação do desempenho entre algumas plataformas. . . | 48 |
| 4.1 | Coleções de teste disponíveis publicamente (Universidade de Glasgow). | 64 |
| 4.2 | Estado da arte na busca usando contexto. | 79 |
| 5.1 | Modelo de contexto definido como predicados. | 92 |
| 6.1 | Pontos de integração entre a ontologia de RCs e a KIMO. . . | 116 |
| 7.1 | Ligação entre as instâncias do contexto e a base de conhecimento. | 138 |
| 7.2 | Consulta 1 e seu contexto. | 140 |
| 7.3 | Diferentes expressões para a consulta 1. | 140 |
| 7.4 | Medidas de desempenho para a consulta 1. | 141 |
| 7.5 | Consulta 2 e seu contexto. | 142 |
| 7.6 | Diferentes expressões para a consulta 2. | 142 |
| 7.7 | Medidas de desempenho para a consulta 2. | 143 |
| 7.8 | Consulta 3 e seu contexto. | 143 |
| 7.9 | Diferentes expressões para a consulta 3. | 144 |
| 7.10 | Medidas de desempenho para a consulta 3. | 144 |
| 7.11 | Consulta 4 e seu contexto. | 145 |
| 7.12 | Diferentes expressões para a consulta 4. | 145 |
| 7.13 | Medidas de desempenho para a consulta 4. | 146 |
| 7.14 | Medidas de desempenho médio para o busca de indicadores. . . | 147 |
| 7.15 | Consulta 5 e seu contexto. | 148 |
| 7.16 | Diferentes expressões para a consulta 5. | 148 |
| 7.17 | Medidas de desempenho para a consulta 5. | 148 |
| 7.18 | Consulta 6 e seu contexto. | 149 |
| 7.19 | Diferentes expressões para a consulta 6. | 149 |
| 7.20 | Medidas de desempenho para a consulta 6. | 150 |
| 7.21 | Consulta 7 e seu contexto. | 151 |

| | | |
|------|--|-----|
| 7.22 | Diferentes expressões para a consulta 7. | 151 |
| 7.23 | Medidas de desempenho para a consulta 7. | 151 |
| 7.24 | Medidas de desempenho médio para a busca por documentos. | 152 |
| C.1 | Instâncias da base de conhecimento. | 193 |
| C.2 | Relações da base de conhecimento. | 193 |
| H.1 | Documentos relevantes para as consultas 1 e 2. | 261 |
| H.2 | Documentos relevantes para as consultas 3 e 4. | 261 |
| H.3 | Documentos relevantes para as consultas 5 e 6. | 262 |
| H.4 | Documentos relevantes para a consulta 7. | 262 |

Acrônimos

ACV Ambiente de Criação de Organizações Virtuais

API *Application Programming Interface*

BLD *Basic Logic Dialect*

BPM *Business Process Management*

CSCW *Computer Supported Cooperative Work*

CVP Comunidade Virtual de Profissionais

DAML *DARPA Agent Markup Language*

ECOLEAD *European Collaborative Networked Organization LEADership Initiative*

EI Extração de Informação

EV Empresa Virtual

HTML *HyperText Markup Language*

OIL *Ontology Inference Layer*

OV Organização Virtual

OWL *Web Ontology Language*

PRD *Production Rules Dialect*

RC Rede Colaborativa

RCO Rede Colaborativa de Organizações

RDF *Resource Description Framework*

RDFS *RDF Schema*

RI Recuperação de Informação

RIF *Rule Interchange Format*
RMI *Remote Method Invocation*
SOA *Service Oriented Architecture*
SGML *Standard Generalized Markup Language*
SHOE *Simple HTML Ontology Extensions*
SVG *Scalable Vector Graphics*
TIC *Tecnologia de Informação e Comunicação*
TREC *Text REtrieval Conference*
UML *Unified Modeling Language*
URI *Uniform Resource Identifier*
URL *Uniform Resource Locator*
URN *Uniform Resource Name*
W3C *World Wide Web Consortium*
WSDL *Web Service Definition Language*
XML *Extensible Markup Language*
XSL *Extensible Stylesheet Language*

Capítulo 1

Introdução

A economia mundial está sofrendo mudanças fundamentais ocasionadas tanto pela globalização quanto pela revolução da Tecnologia de Informação e Comunicação (TIC) (Helaakoski et al., 2007), levando organizações a focarem em novas estratégias para obterem vantagens competitivas no mercado. Tais estratégias são baseadas principalmente em *formas de colaboração*, habilitadas por TIC, onde organizações trabalham juntas para compartilhar riscos, reduzir custos e aumentar o seu acesso a novas oportunidades de negócio e/ou em novos mercados.

O trabalho colaborativo tem se manifestado em diferentes formas, incluindo cadeias de fornecimento, empresas virtuais, organizações virtuais, e comunidades de profissionais virtuais. Tendências similares podem ser encontradas em contextos sem fins lucrativos, tais como gestão de desastres, redes de saúde, entre outros; e no meio acadêmico, onde laboratórios e pesquisadores visam alcançar excelência científica para desenvolvimentos inovadores (Camarinha-Matos e Afsarmanesh, 2006). Todas essas formas de colaboração têm sido identificadas por projetos de pesquisa como manifestações do conceito mais geral de Rede Colaborativa (RC). Alguns projetos relevantes nessa área são: VOMap (VOMap, 2004), MyFashion.eu (MyFashion.eu, 2004), THINKcreative (THINKcreative, 2003), ECOLEAD (ECOLEAD, 2004) e COIN (COIN, 2008).

Uma RC é constituída por uma variedade de entidades (organizações e indivíduos) que são autônomas, largamente distribuídas geograficamente e heterogêneas em termos de ambiente operacional, cultura, capital social e metas. Essas entidades colaboram para melhor atingir objetivos comuns ou compatíveis e suas interações são suportadas por redes de computadores. Ao contrário de outras redes, a colaboração em RCs é uma propriedade intencional que deriva da crença comum de que, juntos, os membros da rede podem atingir metas inalcançáveis ou que teriam um custo muito alto individualmente. Para tal, seus membros compartilham recursos (humanos, máquinas, computadores, informação), riscos e ganhos (Camarinha-Matos et al., 2005).

O estudo de RC vem emergindo como uma nova disciplina científica (Camarinha-Matos et al., 2007b), recebendo contribuições de múltiplas áreas, tais como ciência e engenharia da computação, comunicação e redes, gestão, economia, ciências sociais, direito e ética, entre outras. Todavia, apesar de estar em crescente desenvolvimento, ainda apresenta uma série de questões em aberto, pois se trata de uma área nova onde estas formas de colaboração não são totalmente conhecidas e onde muitos conceitos carecem ainda de consenso e maior consolidação.

Portanto, de modo a aproveitar os potenciais benefícios desse tipo de colaboração, os participantes de RCs, normalmente chamados de *parceiros*, devem lidar com uma série de requisitos desafiadores, advindos de múltiplas perspectivas: tecnológica, organizacional, de integração de sistemas, social, jurídica e de modelo de negócios (Chituc e Azevedo, 2005). Tais desafios incluem, dentre outras coisas: a manutenção de níveis de privacidade e visibilidade (segurança), pois parceiros de RCs são autônomos; barreiras culturais e legais, pois esses parceiros trabalham de maneiras diferentes e podem estar localizados em países diferentes; e aspectos de integração, envolvendo semântica e interoperabilidade, pois parceiros e sistemas computacionais são heterogêneos.

No entanto, apesar da pesquisa em todas essas perspectivas ser essencial para o estabelecimento de RCs, esta tese foca-se exclusivamente na perspectiva tecnológica, conforme é descrito na próxima seção.

1.1 Tema de Pesquisa e Motivação

Trabalhar colaborativamente num cenário de RCs significa compartilhar informações, recursos e responsabilidades entre entidades que, de maneira conjunta, planejam, implementam e avaliam um programa de atividades para atingir uma meta em comum (Camarinha-Matos e Afsarmanesh, 2006). Exemplos de informações que são intercambiadas incluem melhores práticas, histórico de desempenho de parceiros em negócios passados, histórico de negociações, manifestações de parceiros ao longo dos processos durante o ciclo de vida da rede e lições aprendidas que podem ter impacto na evolução da rede. O compartilhamento de informações entre parceiros de RCs tem se tornado de crucial importância na colaboração, sendo útil não apenas no nível operacional – ajudando parceiros a atingir seus objetivos –, mas também nos níveis tático e estratégico - ajudando gestores a tomarem decisões e a planejarem ações futuras (Loss et al., 2007).

Segundo Zara (2005), o problema do compartilhamento de informações e conhecimento cobre tanto o aspecto **humano**, experimentado como perda de poder, quanto o aspecto **técnico**, isto é, não é preciso apenas tornar disponível a informação; pessoas e aplicações devem conseguir recuperá-la. No que diz respeito ao aspecto técnico, dois subproblemas devem ser, por-

tanto, tratados: i) compartilhamento de informação, e ii) busca da informação. O tema deste trabalho é a busca da informação em RCs, abordada sob o ponto de vista técnico.

O problema do compartilhamento de informação pode ser tratado por ferramentas para Trabalho Cooperativo com Suporte de Computadores¹. As aplicações típicas desse tipo de ferramenta incluem *e-mail*, sistemas de notificação, vídeo conferência, salas de bate papo (*chats*), sistemas de mediação, entre outros (Bohanec, 2003). Tais ferramentas dão suporte à interação entre os parceiros que trabalham de forma distribuída (como em uma RC), propiciando a troca de informação entre eles e armazenando o conteúdo de tais interações em várias fontes, como documentos, fóruns, redes sociais, entre outros. Como exemplos de ferramentas colaborativas, temos o *Google Docs* e *Google Spreadsheet*², *weblogs*³ e *wikis*⁴

Já a busca de informação normalmente é tratada com técnicas de Recuperação de Informação (RI). RI é uma área bastante abrangente e, em geral, o seu objetivo principal é prover meios para buscar informações em documentos, buscar os próprios documentos e buscar os metadados que descrevem os documentos (Salton, 1997; Salton et al., 1997). As implementações mais comuns de tais técnicas são as máquinas de busca da Web, como por exemplo, *Google* e *Yahoo!*. Adicionalmente, ferramentas de CSCW em geral oferecem interfaces de busca para a posterior recuperação e uso da informação que foi uma vez compartilhada. Por outro lado, essa busca se restringe ao tipo de ferramenta utilizada, ou seja, em uma busca em uma lista de discussão, por exemplo, apenas a informação compartilhada nessa ferramenta poderá ser recuperada. No caso de ferramentas colaborativas implementadas na forma de aplicações da Web (*wikis*, *blogs*, listas de discussão, etc.) esse problema pode ser parcialmente solucionado através das máquinas de busca da Web. Assim, uma consulta por um determinado assunto no Google por exemplo, pode ter resultados apontando para a Wikipedia, *blogs*, *sites* de notícias, entre outros.

Entretanto, as máquinas de busca da Web conseguem indexar apenas a parte pública destes *sites*. Em outras palavras, o conteúdo disponível em um *site*, e que é restrito a um grupo de usuários, não pode ser acessado e procurado por tais máquinas de busca. Quando se considera um ambiente

¹Do inglês *Computer Supported Cooperative Work (CSCW)*

²Ferramentas para edição de textos e planilhas que permitem compartilhamento e edição de documentos de modo colaborativo (<http://docs.google.com/>).

³Um *weblog* (ou *blog*) é uma ferramenta para compartilhamento de informação que permite que pessoas publiquem conteúdo na Web de forma cronológica, tal como em um diário (Du e Wagner, 2005).

⁴Uma *wiki* é um *site* da Web onde todos os usuários podem editar páginas existentes e adicionar novas páginas usando apenas um navegador Web (Rick e Guzdial, 2006). O exemplo mais conhecido deste tipo de aplicação é a *Wikipedia*, uma enciclopédia onde os próprios usuários contribuem para o seu conteúdo.

corporativo, tais aspectos se tornam essenciais. Segundo Mukherjee e Mao (2004), mesmo que uma organização tenha acesso a conteúdo relevante, há casos onde tal conteúdo não pode ser indexado – por exemplo, por questões de segurança – ou é proibido de ser indexado devido a questões legais.

Essa questão relativa à segurança é apenas um dos vários aspectos a serem considerados quando da busca de informação em uma RC. Há também outros aspectos, como a *semântica*, pois a informação distribuída pode posuir diferentes significados. Adicionalmente, os usuários de uma organização ou de uma RC estão inseridos num *contexto* de negócios ou de resolução de problema bem definidos (Demartini, 2007). O contexto do usuário pode ser definido por diversos aspectos, tais como os seus interesses individuais, a tarefa em execução e o papel desempenhado (Hawking et al., 2005; Demartini, 2007). Em outras palavras, aspectos como semântica e contexto têm influencia no resultado da busca: o primeiro diz respeito a como a informação é interpretada; o segundo tem a ver com a motivação do usuário, que é o processo de negócios onde está trabalhando. Estes aspectos são, portanto, essenciais no âmbito da busca em organizações e em RCs, visto que permitem uma melhoria significativa nos resultados e, conseqüentemente, no potencial de qualidade das decisões a serem tomadas em função disso.

Como um simples – mas representativo – exemplo disso, basta se imaginar um dado parceiro, de um dado país, baseado num dado contexto de negócios, quando deseja disponibilizar uma certa melhor prática de processo de fabricação de um certo bem. Com que infraestrutura, ferramenta, formato, terminologia, semântica e segurança deverá fazê-lo, na medida que somente alguns membros da RC podem ter acesso a tal informação estratégica? E do ponto de vista destes membros que poderão ter tal acesso, como fazê-lo? Igualmente, com que infraestrutura, ferramenta, formato, terminologia, semântica e segurança deverão fazê-lo? Suportar este cenário, como mencionado, envolve inúmeros aspectos, de variada complexidade, para os quais soluções tecnológicas e abordagens científicas ainda carecem de resultados plenamente adequados à realidade empresarial.

1.1.1 Características da Busca em Redes Colaborativas

Do ponto de vista do usuário, além das dificuldades naturais (por exemplo, como expressar suas necessidades e em seguida como filtrar o que é irrelevante), o cenário de RCs apresenta algumas características particulares: as informações estão armazenadas de maneiras diferentes e estão distribuídas; a informação pode ter diferentes significados (pois empresas são autônomas); dependendo da empresa, pode haver diferentes políticas de segurança; e empresas estão envolvidas em vários processos de negócio – ou mesmo em várias RCs – simultaneamente (isto é, múltiplos contextos de interesse).

Portanto, busca por informação em uma rede de organizações não

é o mesmo que uma simples consulta em uma fonte de dados estruturada usando algoritmos tradicionais baseados em palavras. De fato, ela envolve o tratamento de alguns aspectos desafiadores, herdados da área de Busca Empresarial⁵ (Abrol et al., 2001; Raghavan, 2001; Hawking, 2004; Mukherjee e Mao, 2004; Demartini, 2007; Belkin, 2008). Tais características podem ser organizadas em dois grupos principais, detalhados a seguir:

1. Fontes de Informação:

- São distribuídas entre múltiplas organizações e RCs;
- Podem ser tanto estruturadas (bancos de dados) ou não estruturadas (documentos);
- São heterogêneas em termos de *formato* (html, pdf, doc, etc.), *tipo / estrutura* (artigo, e-mail, wiki, etc.) e *conectividade* (HTTP, POP, conexão com banco de dados, etc);
- Não são parte de uma rede (*web*), ou seja, há poucas ou nenhuma referência (na forma de *links*) entre elas. A razão para isto é porque estão distribuídas entre múltiplas organizações (sistemas independentes).
- Têm diferentes níveis de acesso (segurança) e;
- Têm diferentes significados.

2. Contexto: usuários estão associados a um contexto de trabalho explícito, pois:

- Usuários desempenham algum *papel* (ou seja, têm alguma função) durante a execução de *tarefas* de um dado *processo de negócio*;
- Usuários podem participar em várias RCs simultaneamente, ou seja, esse é mais um elemento que fez parte do seu contexto.
- Usuários também têm interesses individuais (perfis do usuário);
- Contexto também pode ser associado às fontes de informação (contexto quando usuários geram e consomem a informação).

1.1.2 Problemas relacionados à Busca de Informação em Redes Colaborativas

As características apresentadas impõem desafios para a implementação de sistemas de busca em RCs. Vários problemas foram identificados na literatura:

⁵Do inglês *Enterprise Search*.

1. *Como fazer a classificação adequada entre coleções heterogêneas de fontes de informação?* A heterogeneidade pode ser abordada de diversas maneiras:
 - (a) *Como unificar a busca entre fontes de informação estruturadas e semi-estruturadas* (Abrol et al., 2001; Hawking, 2004; Demartini, 2007), ou seja, documentos, e-mail, bases de dados, entre outras;
 - (b) *Como interpretar o significado das informações.* Isso é a consequência da natureza das próprias RCs, que são formadas por entidades heterogêneas, ou seja, cada parceiro tem a sua própria forma de representar sua semântica. Esse problema pode ser mais complexo, visto que um parceiro pode estar envolvido em mais de uma RC. Os termos utilizados em uma dada RC podem ter significados diferentes em outra.
 - (c) *Como realizar buscas federadas,* ou seja, buscas que envolvem várias organizações e/ou RCs. Uma busca federada (também conhecida como *recuperação de informação distribuída*) liga múltiplas máquinas de busca em um único sistema virtual (Si e Callan, 2005), permitindo que usuários façam buscas em múltiplos sistemas com uma única requisição. Por exemplo, um parceiro envolvido em três RCs pode fazer uma busca em todas elas ou apenas em uma.
2. *Como estimar a relevância de documentos que não fazem parte de uma rede, ou seja, que têm pouca ou nenhuma referência hipertextual entre si?* (Abrol et al., 2001; Hawking, 2004). Isso significa que algoritmos típicos para busca na Web, tal como o *PageRank* (Brin e Page, 1998), não seriam completamente adequados visto que a estrutura hipertextual entre os documentos não é tão importante. Outro ponto relacionado, e que enfatiza a diferença em relação às técnicas aplicadas para busca na Web, é que o conteúdo disponível numa intranet é essencialmente livre de *spam* (Fagin et al., 2003).
3. *Como lidar com diferentes políticas de segurança?* (Abrol et al., 2001; Mukherjee e Mao, 2004). Mecanismos de segurança são essenciais para controlar o acesso à informação por parte dos parceiros da RC, bem como do sistema de busca. É um requisito ortogonal que deve ser considerado em vários desses problemas. Por exemplo, o sistema de busca deve estar ciente de que um dado usuário, por participar de duas RCs ao mesmo tempo, tem permissão para fazer uma busca (*federada*) em ambas, se necessário.
4. *Como o contexto do usuário pode ser usado para personalizar a busca?* (Abrol et al., 2001; Hawking, 2004; Demartini, 2007) Segundo De-

martini (2007), na busca empresarial as consultas são relacionadas ao trabalho do usuário, ou seja, na tarefa que este executa. Portanto, o sistema deveria usar esse contexto para melhorar os resultados da busca. O mesmo autor também fala que os resultados da busca podem ser customizados de acordo com o papel do usuário, pois papéis diferentes (tais como gerente, suporte, desenvolvedor) têm necessidades de informação diferentes.

5. *Como avaliar sistematicamente sistemas de busca empresarial?* (Demartini, 2007). Este é um problema difícil de se resolver dada a quantidade e complexidade dos requisitos envolvidos. Hawking (2004) também enumera como um problema relacionado à definição de uma coleção de teste apropriada para tal avaliação.

Além desses problemas, Demartini (2007) levanta a seguinte questão: “Como se beneficiar da *Web Semântica* e de técnicas de RI na busca empresarial?” Essa é uma questão interessante, pois o termo “Web Semântica” está relacionado à utilização de ontologias para enriquecer o conteúdo das páginas da Web de modo que não apenas pessoas, mas também sistemas, possam acessar o conteúdo da Web de forma mais precisa (Berners-Lee et al., 2001). Ontologias são estruturas que definem os conceitos de um determinado domínio (Noy e McGuinness, 2001). Barrows e Traverso (2006) reforçam que a adoção de ontologias – e em particular, alguns formatos adotados pela Web Semântica – representa alguns dos elementos que permitem a implementação de sistemas de busca em empresas.

De fato, uma tendência que se estabeleceu nos últimos anos na área de RI consiste no uso de ontologias no intuito de aumentar a precisão da busca por informações (Kiryakov et al., 2004; Stojanovic, 2005; Köhler et al., 2006; Zhuhadar e Nasraoui, 2008). Ontologias vêm sendo usadas em uma variada gama de aplicações para representar a semântica de uma maneira formal. Em outras palavras, técnicas de RI têm evoluído de buscas baseadas em palavras-chave para buscas baseadas em semântica, definida através de ontologias.

Ontologias desempenham um papel central da solução de alguns dos problemas enumerados acima. Por exemplo, ontologias são usadas na definição de *anotações semânticas* em fontes de informação, permitindo um processamento e busca mais precisos (Uren et al., 2006). A ideia de anotar o conteúdo de documentos com informação semântica definida em ontologias foi proposta pela iniciativa da *Web Semântica* e tem sido implementada por diversas aplicações.

Outro aspecto abordado consiste no suporte a *traduções semânticas* durante *buscas federadas* com o uso de *mapeamentos entre ontologias*. Mapeamento entre ontologias é uma solução promissora para o problema da heterogeneidade semântica e visa encontrar correspondências entre entidades

semanticamente relacionadas de diversas ontologias, permitindo a interoperação entre informações associadas a diferentes ontologias (Uschold, 2003; de Bruijn et al., 2006; *OntologyMatching*, 2006).

1.2 Problema de Pesquisa

Considerando os problemas apresentados na seção anterior, este trabalho tem como foco dois aspectos: *semântica* (problema 1a) e *contexto* (problema 4). Em outras palavras, esta tese trata dos seguintes problemas: (i) como prover uma semântica mais precisa para as fontes de informação; e (ii) como combinar essa semântica e o contexto do usuário para melhorar o desempenho da busca. Essas duas perspectivas são abordadas de maneira sequencial, onde a primeira serve de base para a segunda.

1.2.1 Perspectiva Semântica

No que diz respeito à perspectiva semântica, será utilizada uma ontologia para melhorar a busca, em dois passos:

- Definição de **anotações semânticas** em fontes de informação. Com uma semântica bem definida expressa de uma forma interpretável por máquinas, fontes de informação podem ser processadas e buscadas mais precisamente.
- Definição de **buscas semânticas**. Por ser baseada em ontologias, a consulta pode expressar precisamente o tipo da informação que se deseja buscar e que é recuperada de acordo com as anotações semânticas previamente definidas.

A abordagem adotada para este trabalho tem como pressuposto a existência de uma ontologia comum adotada pelos parceiros da RC. Embora seja um pressuposto forte assumir que toda rede de organizações tenha sua própria ontologia, pesquisas nessa área apontam a adoção de ontologias em aplicações que dão suporte a esse tipo de rede. Exemplos desse tipo podem ser encontrados em Loss (2007); Ermilova e Afsarmanesh (2008); Baldo et al. (2009).

Esta perspectiva pode ser considerada um passo preparatório para a implementação da busca baseada em contexto. Para esta parte, serão usadas técnicas e ferramentas existentes, tanto para anotação semântica automática, quanto para busca semântica. Portanto, a contribuição principal deste trabalho está na perspectiva de contexto.

1.2.2 Perspectiva de Contexto

Como mencionado, parceiros de uma RC estão envolvidos em um contexto explícito de trabalho e tal contexto deve ser levado em conta durante

a busca. Portanto, a busca baseada em contexto em RCs representa um desafio de grande relevância visto que pode melhorar a qualidade da busca de informação pelos parceiros e conseqüentemente reduzir o tempo e melhorar a qualidade do seu trabalho.

Há diversas definições do termo “contexto” na literatura, mas, considerando o termo de uma maneira geral, tais definições são semelhantes entre si. Contexto pode ser definido como qualquer informação que possa ser usada para caracterizar a situação de uma entidade que é considerada relevante para a interação entre um usuário e uma aplicação (Dey, 2001). Além disso, o contexto tem um impacto significativo na forma como humanos e aplicações se comportam e como interpretam as informações; uma mudança de contexto causa uma transformação na experiência do usuário (Bolchini et al., 2007).

Baseado nessas definições, pode ser dito que contexto é a informação usada por usuários e aplicações para fazer customizações no comportamento dos sistemas. Renda e Straccia (2005) definem customização como o ato de levar o contexto em consideração de modo a adaptar informações e serviços de uma maneira específica às necessidades particulares de um usuário ou comunidade de usuários. Este assunto tem sido bastante explorado em diversas aplicações, sendo que na literatura é também referenciado como *personalização* ou *adaptação*, realizada por sistemas baseados em contexto⁶. Particularmente no escopo da gestão da informação, sistemas baseados em contexto servem para determinar que partes da informação são relevantes em relação às condições do ambiente (Bolchini et al., 2007).

As definições apresentadas até aqui dão uma ideia geral do que é contexto e sua utilidade. Todavia, quando avaliado dentro de um domínio ou aplicação mais específica, o termo “contexto” pode ser interpretado de diferentes formas, pode modelar diferentes aspectos e pode ter um impacto na aplicação igualmente diverso. De acordo com a análise da literatura, o contexto pode representar diferentes aspectos, dependendo dos requisitos da aplicação ou problema: (i) *Condições Ambientais*, tais como aspectos físicos (localização, tempo, dispositivo, etc.) (Fahy e Clarke, 2004) e organizacionais (processos, tarefas, papéis, etc.) (Hawking, 2004); (ii) *Perfil Individual*: tópicos de interesse ou histórico do usuário (Tamine-Lechani et al., 2007); (iii) *Aspectos Linguísticos ou de Estrutura do Documento*, tais como o contexto de um termo sendo o conjunto de palavras na sua vizinhança (Bhogal et al., 2007), ou os *hyperlinks* entre páginas da Web (Yu et al., 2001).

No escopo deste trabalho, contexto significa a **situação dinâmica** de um usuário no ambiente de uma RC, particularmente dos seus **aspectos organizacionais**. Tais compreendem os seguintes elementos: a rede na qual participa, a organização para a qual trabalha, o processo e a atividade em execução e o papel desempenhado. Esses aspectos vêm sendo identificados como

⁶Do inglês *context-aware systems*.

fundamentais para a obtenção de bons resultados na busca por informação em organizações (Hawking et al., 2005; Demartini, 2007; Belkin, 2008).

1.2.3 Visão Geral

De acordo com as características apresentadas na seção 1.1.2, a figura 1.1 apresenta o foco do trabalho (na área tracejada). É importante mencionar que a ontologia adotada pela RC foi posicionada como uma característica existente no ambiente da RC. Além disso, aspectos como distribuição, heterogeneidade e segurança são transparentes, ou seja, assume-se que são tratados por serviços específicos.

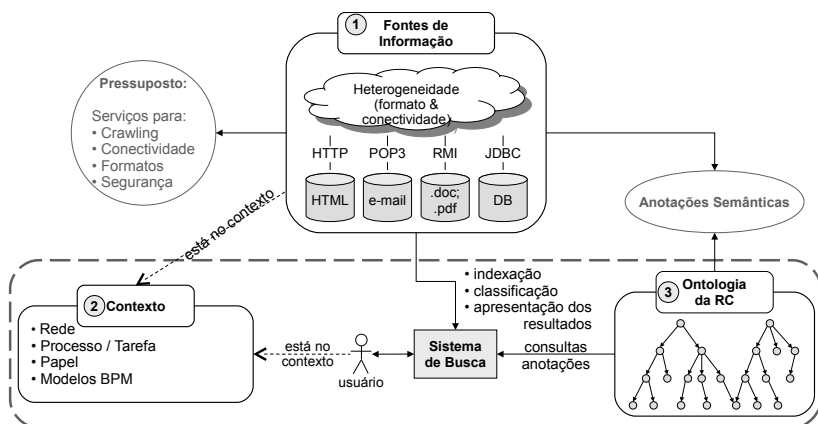


Figura 1.1: Foco do Trabalho.

O problema de pesquisa a ser abordado neste trabalho é, portanto, a modelagem e uso do contexto durante a RI, bem como a medição do seu benefício relativo ao desempenho da busca contrabalanceado pelo seu custo computacional. O problema de pesquisa pode ser sintetizado pela seguinte pergunta:

Qual é a influência do contexto na qualidade da resposta da busca por informação em uma RC?

1.3 Objetivos

1.3.1 Objetivo Geral

Visando responder à pergunta da tese, o objetivo geral deste trabalho é conceber um **modelo de contexto e regras para expansão de consultas semânticas em RCs**. Além disso, tal modelo será posicionado dentro de um

arcabouço para busca de informação baseada em ontologias, que permitirá avaliar os benefícios oferecidos pelo modelo de contexto.

1.3.2 Objetivos Específicos

Os objetivos específicos deste trabalho são:

- Especificar uma arquitetura conceitual que defina as funcionalidades necessárias à busca de informação em RCs. As funcionalidades incluem a gestão de ontologias, anotação semântica automática e indexação de documentos, e busca semântica baseada em contexto. Uma ontologia para RC também será desenvolvida;
- Instanciar a arquitetura conceitual na forma de um conjunto de serviços de *software* que implementem as funcionalidades dessa arquitetura. Serão utilizadas ferramentas existentes para a implementação da perspectiva semântica.
- Desenvolvimento da perspectiva de contexto: (i) definição do modelo para o contexto em uma RC; (ii) definição da abordagem para busca baseada em contexto. Para essa última atividade, a técnica adotada consiste na utilização de regras para inferir novos tópicos (contextualizados), visando a expansão da consulta do usuário;
- Definição de uma coleção de testes para avaliação do sistema.

1.4 Justificativa do Trabalho

Técnicas que exploram o contexto para a customização são utilizadas em uma vasta gama de aplicações, tais como bibliotecas digitais (Renda e Straccia, 2005), gestão de informações (Bouslimi et al., 2008) e escalonamento em RCs (Almeida et al., 2007).

Além de ser um aspecto bastante explorado em outros domínios, a utilização do contexto no âmbito da *busca de informação* é um tema bastante relevante. Isto porque o contexto tem um impacto na forma como o usuário busca e acessa a informação, pois a mesma consulta definida por usuários em contextos diferentes geram resultados diferentes. Em resumo, o benefício tangível de uma busca personalizada é a redução do tempo que as pessoas levam para encontrar informações Pitkow et al. (2002).

Além disso, uma infraestrutura de suporte à busca de informação em RCs se justifica pelas aplicações que podem se beneficiar com o seu uso, aplicações estas essenciais a esse tipo de rede:

1. **Aprendizado de RCs:** pode ser usada para organizar e recuperar a informação que está distribuída em documentos, auxiliando parte do

processo de gestão de conhecimento. Como aplicações desse tipo no contexto de RCs temos o trabalho desenvolvido por Loss et al. (2007).

2. **Formação de Organizações Virtuais:** uma Organização Virtual (OV) é um tipo de RC que pode ser criada dinamicamente a partir de um outro tipo de RC, chamado Ambiente de Criação de Organizações Virtuais (ACV). Uma infraestrutura de busca é útil no processo de formação de OVs, sendo utilizada na busca e seleção dos membros da ACV que farão parte da nova OV, tal como no sistema desenvolvido por Camarinha-Matos et al. (2007c). Outro aspecto particular à criação de OVs que é a busca e seleção de indicadores de desempenho, como definido por Baldo et al. (2009).
3. **Suporte à decisão / gestão de RCs:** pode dar suporte a funcionalidades de sistemas de apoio à decisão como, por exemplo, nos sistemas apresentados por Graser et al. (2005); Crave et al. (2006); Silva (2010).
4. **Busca de recursos para gestão de ACVs:** como por exemplo, nos sistemas desenvolvidos por Ermilova e Afsarmanesh (2006); Msanjila e Afsarmanesh (2006).

1.4.1 Adequação às Linhas de Pesquisa do Curso

O trabalho descrito nesta tese está inserido no contexto da Área de Concentração em Automação e Sistemas do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Santa Catarina. Este trabalho está integrado com os demais trabalhos de pesquisa sobre Integração de Sistemas e Redes Colaborativas, e com as atividades do Curso de Pós-Graduação em Engenharia Elétrica desta Universidade.

1.5 Delimitação da Pesquisa

Com base no enquadramento apresentado, o escopo da pesquisa está delimitado entre as áreas de Recuperação de Informação, Ontologias e Sistemas baseados em Contexto, tal como ilustrado na figura 1.2.

É preciso salientar ainda que, por estar no nível de infraestrutura de TICs, este trabalho não aborda problemas da camada de aplicação ou que lidem com o aspecto humano, tais como abordagens que visam a gestão de conhecimento ou interfaces homem-máquina.

Além disso, é importante deixar explícito o significado das palavras-chave que compõem o título deste trabalho, pois alguns termos podem gerar ambiguidades. Assim, para o título “Um Modelo Baseado em **Contexto** para **Expansão de Consultas Semânticas** em **Redes Colaborativas de Organizações**”, temos:

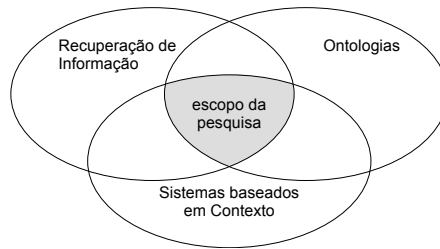


Figura 1.2: Delimitação da Pesquisa.

- **Contexto:** neste trabalho, este termo significa a situação na qual um determinado usuário se encontra, com base em elementos organizacionais (Hawking et al., 2005; Demartini, 2007; Belkin, 2008). Dentro do escopo de uma RC, tais elementos são definidos pela rede onde está participando, pela organização para a qual trabalha, pelo seu papel, e pelo processo e atividade que está trabalhando no momento. Quaisquer outras definições e interpretações que o contexto pode ter não são abordadas neste trabalho.
- **Expansão de Consulta:** trata-se de técnicas que adicionam novos termos à consulta original submetida pelo usuário. Neste trabalho, o contexto do usuário é o elemento que serve como base para tal transformação. É importante enfatizar que, por se tratar de consultas, este trabalho está relacionado à busca, ou seja, envolve técnicas de RI. Portanto, não são abordados tópicos relativos à descoberta de conhecimento, tais como mineração de dados e de texto.
- **Consultas Semânticas:** significa que as consultas são expressas em termos de conceitos definidos em ontologias, ao invés de palavras-chave definidas em linguagem natural.
- **Redes Colaborativas de Organizações:** significa que tanto o modelo de contexto quanto a ontologia e o escopo das buscas são associados a este tipo de rede em particular. No entanto, o modelo de contexto proposto pode ser estendido ou modificado para atuar em outros domínios.

1.6 Contribuições do Trabalho

A principal contribuição científica do trabalho é o **modelo de contexto e as regras para expansão de consultas semânticas**, visando melhorar a busca de informação em RCs. A definição de tal modelo foi o resultado da convergência entre RI, ontologias e contexto, aplicados à área de RCs. Além disso, outras contribuições podem ser listadas:

- Arcabouço para busca de informação em RCs, do qual o modelo faz parte, que oferece funcionalidades especificadas na forma de uma arquitetura orientada a serviços;
- Protótipo computacional, que implementa tal arcabouço, bem como uma base de conhecimento e coleção de documentos de teste;
- Ferramentas de suporte ao protótipo computacional, incluindo indexador e buscador em lote, populador da base de conhecimento e gerador de relatórios para a medição de desempenho das buscas;

1.7 Procedimentos Metodológicos

1.7.1 Classificação da Pesquisa

Segundo a classificação proposta por Cervo e Bervian (2002); da Silva e Menezes (2005), a pesquisa pode ser enquadrada quanto: i) a sua natureza; ii) os seus objetivos; iii) a abordagem do problema; e iv) os procedimentos técnicos para a sua execução. Segundo essa classificação, esta tese se enquadra da seguinte maneira:

- **Quanto à Natureza da Pesquisa:** Aplicada, pois seu objetivo tem aplicação prática definida, que é a busca de informação em RCs.
- **Quanto aos Objetivos da Pesquisa:** Exploratória, já que ela consiste basicamente em uma pesquisa bibliográfica seguida da proposta de uma arquitetura dentro de um tema não perfeitamente conhecido e consolidado, sua implementação e avaliação.
- **Quanto à Forma de Abordagem do Problema:** Quantitativa e qualitativa. Por um lado, os resultados são avaliados com base em medidas de desempenho para sistemas de RI. Por outro lado, o usuário é que define os critérios de relevância.
- **Quanto aos Procedimentos Técnicos:** Pesquisa bibliográfica, elaborada a partir de material já publicado, constituído principalmente de livros, artigos de periódicos e conferências e normas de padronização de tecnologias.

1.7.2 Etapas da Pesquisa

A pesquisa segue uma metodologia incremental onde, uma vez identificado o problema de pesquisa, o desenvolvimento é feito em dois passos principais, cobrindo as duas perspectivas identificadas (semântica e de contexto). Tal abordagem é semelhante à metodologia adotada em *Design Research*, que

é uma metodologia iterativa, adequada ao desenvolvimento e avaliação de sistemas de informação (Vaishnavi e Kuechler, 2004). As etapas da pesquisa são apresentadas a seguir:

1. **Identificação do Problema:** esta etapa inclui a definição do tema de pesquisa, revisão bibliográfica, identificação dos requisitos e definição da proposta;
2. **Desenvolvimento da perspectiva semântica:** representa a base na qual o modelo proposto será desenvolvido. Compreende os seguintes passos:
 - (a) Pesquisa por ferramentas de suporte existentes (anotação e busca semânticas);
 - (b) Modelagem e implementação das funcionalidades básicas (serviços);
 - (c) Avaliação e integração do modelo e protótipo. Esta etapa foi realizada no âmbito de um projeto de pesquisa europeu, o projeto ECOLEAD, onde o protótipo foi validado tanto em termos de arquitetura (serviços) quanto em termos de funcionais, provendo suporte a outros serviços e aplicações desenvolvidos nesse projeto;
3. **Desenvolvimento da perspectiva de contexto:** esta etapa compreende o desenvolvimento do modelo de contexto com base nos resultados obtidos na etapa anterior (modelo e implementação parciais). Ela é dividida nos seguintes passos:
 - (a) Definição do modelo de contexto e regras para expansão da consulta;
 - (b) Implementação do modelo e regras, como uma extensão do protótipo desenvolvido anteriormente;
 - (c) Testes preliminares para verificação da consistência das regras e correção de erros;
4. **Avaliação do modelo proposto:** feita com base na execução do protótipo em alguns cenários de teste. Tais cenários são instâncias reduzidas que representam situações típicas de busca dentro de uma RC. É importante mencionar aqui a dificuldade desse tipo de avaliação devido ao esforço para a implantação e análise em um ambiente real de uma RC⁷. A avaliação é composta pelos seguintes passos:

⁷Embora fosse possível utilizar recursos do projeto ECOLEAD para esta etapa, este já havia se encerrado na altura dos experimentos.

- (a) Preparação da coleção de teste, base de conhecimento, instâncias de contextos e indexação dos documentos;
 - (b) Execução das consultas e captura dos resultados;
 - (c) Análise dos resultados, com base em relatórios comparativos entre três configurações de busca: por palavras-chave, semântica e contexto.
5. **Escrita da tese:** documentando todas as etapas, resultados e conclusões do trabalho.

1.7.3 Assuntos selecionados para a pesquisa bibliográfica

Com base no problema abordado neste trabalho, é preciso realizar uma pesquisa bibliográfica das áreas envolvidas na busca de informação em RCs, a saber:

1. Área de RCs, focando aspectos de infraestruturas de TIC, bem como os requisitos para busca em RC (e busca empresarial).
2. Ontologias e tópicos relacionados, tais como anotações semânticas e Web Semântica. Esta última envolve também o estudo de tecnologias e padrões correlatos.
3. Técnicas de Recuperação de Informação (RI) e o uso de ontologias e anotações semânticas para dar suporte a buscas semânticas.
4. Contexto e tópicos relacionados: modelagem, captura e abordagens para sua utilização vinculada a técnicas de RI.

Para a realização da pesquisa bibliográfica, os seguintes *sites* de indexação de artigos científicos foram pesquisados:

- Portal CAPES – <http://www.periodicos.capes.gov.br/>
- Portal ACM – <http://portal.acm.org/>
- Science Direct – <http://sciencedirect.com/>
- SpringerLink – <http://www.springerlink.com/>
- IEEE Xplore – <http://ieeexplore.ieee.org/>
- Cite Seer – <http://citeseer.ist.psu.edu/>
- Scopus - <http://www.scopus.com/>
- Google Scholar – <http://scholar.google.com/>

1.8 Estrutura do Documento

Este documento está organizado da seguinte maneira. O capítulo 2 apresenta os conceitos básicos acerca de RCs, cobrindo os requisitos e tecnologias necessárias para a implementação de infraestruturas de suporte. Esse capítulo também fala sobre a busca em RCs, os principais problemas, requisitos, abordagens e tecnologias, com aspectos relacionados à área de busca empresarial.

Os capítulos 3 e 4 trazem a fundamentação teórica referente às áreas envolvidas neste trabalho: ontologias, Web Semântica e anotações semânticas são vistas no capítulo 3 e o capítulo 4 apresenta conceitos referentes à Recuperação de Informação, incluindo busca semântica e baseada em contexto. Para cada tópico, uma análise do estado da arte é realizada.

O capítulo 5 apresenta a abordagem adotada e os detalhes do modelo proposto neste trabalho, incluindo as perspectivas semântica e de contexto. O modelo de contexto e as regras são explicados e são posicionados em uma arquitetura orientada a serviços.

No capítulo 6 são mostrados os detalhes do desenvolvimento do protótipo computacional, incluindo as ferramentas existentes utilizadas na implementação das suas funcionalidades.

O capítulo 7 descreve os experimentos realizados para a avaliação do modelo proposto. Este capítulo mostra as técnicas e medidas típicas para avaliação de sistemas de RI, bem como a metodologia adotada para a avaliação; são apresentados os aspectos preparatórios e a condução propriamente dita dos experimentos. Por fim, os resultados são apresentados e discutidos.

O capítulo 8 apresenta as conclusões do trabalho e discute as propostas para trabalhos futuros.

Capítulo 2

Redes Colaborativas e Busca Empresarial

Este capítulo apresenta a área de aplicação deste trabalho, bem como requisitos e aspectos tecnológicos referentes a infraestruturas de colaboração e busca empresarial. Os seguintes assuntos são cobertos:

Redes Colaborativas

É a área de aplicação do trabalho. Alguns conceitos aqui apresentados serão particularmente úteis para o entendimento do tipo de informação que será buscada.

Infraestruturas de suporte a RCs

Apresenta os requisitos funcionais e tecnológicos para infraestruturas de suporte a RCs, servindo para embasar algumas decisões na definição do modelo proposto. Por exemplo, o modelo é definido e implementado como uma arquitetura orientada a serviços, que é uma abordagem que vêm sendo utilizada para a implementação de tais infraestruturas.

Busca Empresarial

Esta seção apresenta o conceito de busca empresarial, as principais características e problemas, bem como as abordagens utilizadas para a sua implementação. O propósito aqui é definir alguns dos tópicos abordados na revisão bibliográfica e análise do estado da arte.

2.1 Conceitos Básicos de Redes Colaborativas

Uma Rede Colaborativa (RC) é constituída por uma variedade de entidades, que podem ser organizações ou indivíduos, em grande parte autônomos, distribuídos e heterogêneos, que colaboram para melhor atingir objetivos comuns ou compatíveis, e cujas interações são suportadas por redes de computadores. Ao contrário de outras redes, a colaboração em RCs é uma propriedade intencional que deriva da crença comum de que, juntos, os membros da rede podem atingir metas inalcançáveis ou que teriam um custo muito alto individualmente (Camarinha-Matos et al., 2005).

Esta seção visa apresentar alguns conceitos referentes a Redes Colaborativas (RCs) que serão relevantes para este trabalho. Primeiramente, na seção 2.1.1, os tipos de RCs são apresentados, no intuito de mostrar as manifestações mais comuns deste tipo de rede. Como foco deste trabalho está no nível de infraestrutura, a seção 2.2 apresenta os requisitos que uma infraestrutura de suporte a RCs deve cumprir. Finalmente, a seção 2.2.1 apresenta infraestrutura de TIC desenvolvida no Projeto ECOLEAD na qual o arcabouço proposto neste trabalho está enquadrado.

2.1.1 Classificação das Redes Colaborativas

RCs manifestam-se das mais variadas formas, incluindo organizações virtuais, empresas virtuais, cadeias de fornecimento dinâmicas, *clusters* industriais, entre outras (Camarinha-Matos et al., 2005). Tais manifestações são classificadas conforme pode ser visto na figura 2.1, onde o conceito mais geral de RC é especializado para **Rede Colaborativa de Organizações (RCO)**, onde as manifestações desse tipo apresentam alguma *organização* nas atividades dos seus constituintes, identificação de papéis para os participantes, e algumas regras de governança; e **Processos de Colaboração Ad-hoc**, onde constituem-se formas de colaboração que não são orientadas a negócios, por exemplo em comunidades virtuais de cidadãos que contribuem em caso de um desastre natural, ou trabalhando por uma causa social. Neste segundo tipo, não há um plano pré-definido e/ou uma estrutura que define os papéis dos participantes e como suas atividades devem ser executadas.

Dentre as RCOs, há as redes que são **orientadas a objetivos**, onde há intensa colaboração visando um objetivo em comum. Por outro lado, há as **redes estratégicas de longa duração**, que visam oferecer o ambiente necessário à configuração rápida de redes colaborativas (Camarinha-Matos e Afsarmanesh, 2006).

Redes orientadas a objetivos podem, por sua vez, ser **dirigidas a produção contínua**, envolvendo redes que têm uma duração mais longa e que se mantêm relativamente estáveis durante esse tempo. Exemplos típicos desse tipo de rede incluem:

- **Cadeias de Suprimento:** rede de empresas onde cada uma delas tem um papel claro no processo de manufatura, cobrindo desde os passos iniciais de projeto do produto e obtenção da matéria prima, passando pela produção, distribuição e armazenamento, até entrega do produto final ao consumidor (Pires, 2004).
- **Governos Virtuais:** aliança de organizações governamentais que combinam os seus serviços através do uso de redes de computadores para proverem serviços integrados aos cidadãos (Camarinha-Matos e Afsarmanesh, 2006).

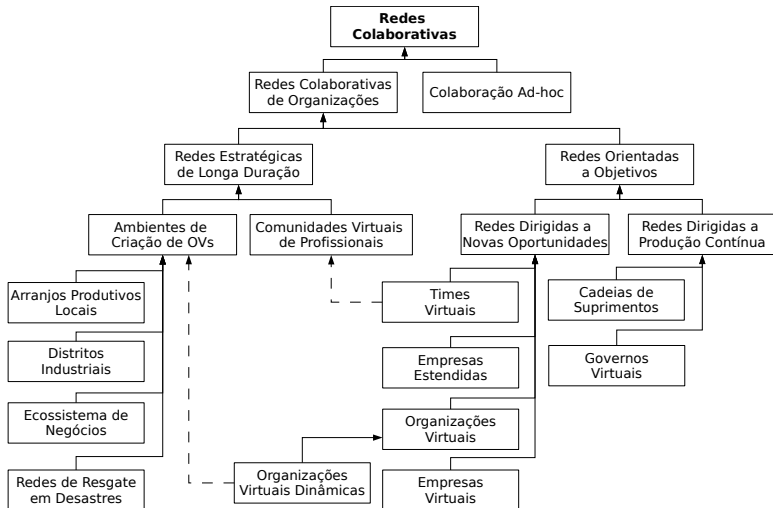


Figura 2.1: Classificação das Redes Colaborativas.

Fonte: Camarinha-Matos e Afsarmanesh (2006).

O outro caso de redes orientadas a objetivos são as **redes dirigidas a oportunidades**, pois são formadas dinamicamente para responder a uma oportunidade de negócios específica e se dissolvem assim que sua missão é cumprida. Exemplos desse tipo de rede incluem:

- **Empresas Virtuais (EV):** aliança temporária de empresas que se unem para compartilhar riscos, competências e recursos de modo a melhor responder a oportunidades de negócios (Camarinha-Matos e Afsarmanesh, 1999c).
- **Organizações Virtuais (OV):** conceito similar ao de EV, mas que pode envolver outros tipos de organizações que não sejam empresas com fins lucrativos. Uma EV é portanto, um caso particular de OV (Camarinha-Matos e Afsarmanesh, 2004b).
- **Organizações Virtuais Dinâmicas:** tipo particular de OV que é estabelecida em um curto espaço de tempo para responder a uma oportunidade competitiva de mercado. Estas redes têm um ciclo de vida curto, dissolvendo-se assim que seus objetivos são cumpridos (Goranson, 1999).
- **Empresas Estendidas:** são redes onde uma empresa dominante “estende” os seus domínios para todos ou alguns de seus fornecedores.

Uma empresa estendida pode ser vista como um caso particular de EV (Camarinha-Matos e Afsarmanesh, 1999c).

- **Times Virtuais:** são similares a OV's, mas formados por pessoas ao invés de empresas. Um time virtual é um grupo temporário de profissionais que trabalham juntos visando um objetivo comum, tal como um trabalho de consultoria por exemplo. Como nos outros tipos de RC's, o principal ambiente de interação são as redes de computadores.

O termo “virtual” nas definições acima vem do fato de que estas redes aparentam ser uma entidade única do ponto de vista de um observador externo, embora não sejam uma entidade única do ponto de vista legal, não tenham um local físico e são tipicamente distribuídas geograficamente. Isto se deve graças aos seus mecanismos de comunicação e de coordenação habilitados por redes de computadores.

Além das redes orientadas a objetivos, um outro tipo de RCO são as **redes estratégicas de longa duração**, que visam oferecer as condições e o ambiente de apoio à configuração rápida de redes colaborativas, quando oportunidades surgem. Como exemplos dessas redes temos os *ambientes de criação de OV's* e as *Comunidades Virtuais de Profissionais*.

Um **Ambiente de Criação de Organizações Virtuais (ACV)** compreende uma associação de organizações que seguem um acordo de cooperação de longa duração e adotam infraestruturas e princípios de operação em comum, com o intuito de estarem preparadas para a rápida configuração de organizações virtuais (Afsarmanesh e Camarinha-Matos, 2005). Em outras palavras, quando uma oportunidade de negócios é identificada, um subconjunto de seus membros pode ser selecionado para formar uma EV/OV. Uma situação semelhante acontece na chamada **Comunidade Virtual de Profissionais (CVP)**, mas neste caso a rede é formada por profissionais e o objetivo é fornecer um ambiente para facilitar a rápida formação de *Times Virtuais*, de maneira similar ao que o ACV visa prover para as OV's. Exemplos de ACV são:

- **Arranjos Produtivos Locais:** estes são os chamados *clusters* industriais, e representam uma das primeiras manifestações de ACVs. São constituídos grupos de empresas, normalmente localizadas na mesma região e que operam no mesmo setor de atividade.
- **Distritos Industriais:** é um termo usado normalmente na Itália, e representa um conceito muito similar ao de arranjo produtivo local. Ele também foca empresas localizadas na mesma região, mas pode envolver tanto empresas do mesmo setor de atividade quanto de setores diferentes.

- **Ecosistema de Negócios:** também chamado de ecossistema digital, é similar a um arranjo produtivo ou um distrito industrial, mas tende a cobrir os principais setores de uma dada região.
- **Redes de Laboratórios Virtuais:** são alianças de organizações de pesquisa autônomas, onde seus pesquisadores podem ser selecionados para formar redes colaborativas para a solução de problemas específicos.

Segundo a definição de RCs, as interações entre os parceiros deste tipo de rede são suportadas por redes de computadores. Em outras palavras, é preciso que exista uma infraestrutura de TIC que proveja suporte para as aplicações utilizadas na RCs, independente da forma como esta se manifesta. A próxima seção trata desse assunto.

2.2 Infraestruturas de Colaboração

Segundo Camarinha-Matos e Afsarmanesh (2004c), a implantação de redes colaborativas em larga escala depende da existência de infraestruturas de TIC transparentes, fáceis de usar, de baixo custo e que sejam escaláveis. Tais infraestruturas permitem que organizações possam estabelecer relações com outras de maneira ágil, além de ser adaptáveis de acordo com as condições do ambiente de negócios e os níveis de autonomia da organização.

O desenvolvimento desse tipo de infraestrutura é um elemento fundamental para a criação de uma cultura organizacional onde a colaboração torna-se parte do processo e não apenas uma opção de trabalho. Outro aspecto fundamental está relacionado aos diferentes tipos e tamanhos das companhias que tipicamente são membros de RCs. A maioria das pequenas e médias empresas têm enormes dificuldades para ter acesso aos principais produtos do mercado, pois exigem um alto investimento em *software* e *hardware* de suporte e em especialistas em TI (Rabelo et al., 2008).

Uma infraestrutura para RCs deve fundamentalmente prover funcionalidades associadas a cinco tipos de elementos, dando suporte à (Rabelo et al., 2008):

- Colaboração e negociação entre pessoas;
- Execução e adaptação de sistemas e serviços;
- Intercâmbio e recuperação de informação;
- Descoberta e compartilhamento de recursos computacionais e humanos;
- Interconexão e sincronização de processos.

Além disso, *segurança e interoperabilidade* são dois elementos tecnológicos que devem permear a infraestrutura, considerando que *modelos de negócio* podem ser aplicados à infraestrutura na medida que serviços e recursos em geral estão sendo acessados por outros membros da RC. A gestão eficiente e transparente desses elementos é o desafio a ser superado por infraestruturas de RCs.

Esse tipo de infraestrutura tem um papel intermediário no processo de interoperação entre os membros de uma RC, oferecendo funcionalidades gerais a qualquer tipo de RC e escondendo os detalhes de implementação de maneira análoga a um sistema operacional. Tal infraestrutura dá suporte a serviços verticais, que consistem em uma série de ferramentas e aplicações verticais, ou seja, que oferecem funcionalidades específicas ao tipo de RC sendo suportado. Por exemplo, os serviços de suporte a uma organização virtual serão diferentes dos utilizados por uma comunidade virtual de profissionais. Estes serviços são conectados à infraestrutura horizontal, conforme pode ser visto na figura 2.2.

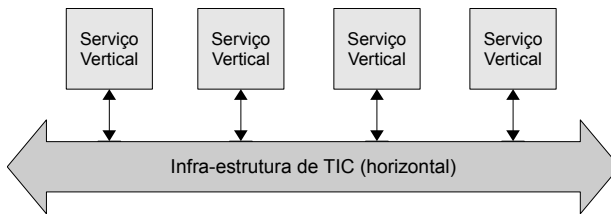


Figura 2.2: Infraestrutura de TIC e serviços verticais.

Fonte: Camarinha-Matos e Afsarmanesh (2004c).

A disponibilidade de TICs, novos paradigmas e modelos inovadores constituem um conjunto de fatores facilitadores para o desenvolvimento de infraestruturas para RCs. Tais fatores incluem: tecnologias e padrões para Web, em particular para *Serviços Web*, projetos baseados em componentes, arquitetura orientada a serviços, modelos de *Software* como Serviço (SaaS), sistemas multi-agentes, computação em malha¹, dispositivos móveis e computação pervasiva e ontologias como forma de representação de conhecimento (Camarinha-Matos e Afsarmanesh, 2004c; Camarinha-Matos et al., 2004; Rabelo et al., 2008). Por outro lado, várias destas tecnologias e conceitos ainda não atingiram maturidade suficiente para o desenvolvimento de infraestruturas adequadas para RCs (Camarinha-Matos e Afsarmanesh, 2004c).

Devido à complexidade que o desenvolvimento de tal infraestrutura representa, as soluções atuais não suprem esses requisitos e nem oferecem suporte adequado aos processos de negócio relacionados a RCs (Rabelo et al.,

¹Do inglês “*grid computing*”.

2008). Daí a necessidade desse tipo de infraestrutura, que é devotada especificamente para RCs.

É preciso considerar também que as primeiras manifestações de RCs a serem estudadas por projetos de pesquisa foram as Empresas Virtuais (EVs) e as Organizações Virtuais (OVs). Dentre estes projetos, destacam-se PRODNET II (PRODNET, 1999; Camarinha-Matos e Afsarmanesh, 1999b), DAMASCOS (DAMASCOS, 2002) e MyFashion.eu (MyFashion.eu, 2004). Embora nem todos objetivassem o desenvolvimento de infraestruturas horizontais, a falta de um modelo e arquitetura de referência comuns e largamente aceitos forçou vários projetos a definirem e implementarem sua própria infraestrutura, desviando assim alguns recursos do seu foco principal, e gerando algo que normalmente era aplicável apenas no escopo do projeto (Camarinha-Matos e Afsarmanesh, 2004c). Além disso, tais infraestruturas não eram suficientes para cumprir os requisitos de outras formas de RCs.

Com o passar do tempo, outras formas de colaboração foram sendo identificadas (apresentadas na seção anterior) e posteriormente todas foram enquadradas no conceito mais geral de Rede Colaborativa (RC). É importante destacar os projetos VOMap (VOMap, 2004) e THINKcreative (THINKcreative, 2003), que definiram as diretrizes e desafios para a pesquisa e desenvolvimento envolvendo estas novas formas de colaboração. Tais diretrizes serviram como ponto de partida para o projeto ECOLEAD (Camarinha-Matos e Afsarmanesh, 2004b), cujo objetivo geral foi definir a fundamentação teórica e arquiteturas de referência que, ao contrário de projetos anteriores, visam ser sustentáveis e servir de base para futuras RCs.

2.2.1 Infraestrutura de Referência para RCs

O projeto ECOLEAD desenvolveu fundamentos e tecnologias para dar suporte a três tipos de RCs: i) **Ambientes de Criação de Organizações Virtuais (ACVs)**; ii) **Organizações Virtuais OVs Dinâmicas**; e iii) **Comunidades Virtuais de Profissionais (CVPs)**. Cada uma destas áreas tem necessidades específicas. Por exemplo, os requisitos de um ACV estão relacionados a: mecanismos para dar suporte ao estabelecimento de confiança entre membros, serviços para a criação de *OVs dinâmicas* (seleção de parceiros, negociação, etc.), manutenção de informações históricas, entre outros. Já um CVP tem funcionalidades análogas a um ACV, mas visa a criação de *times virtuais* (ao invés de OVs). Finalmente, os requisitos de OVs dinâmicas estão vinculados à sua gestão, incluindo mecanismos avançados para medição de desempenho e suporte à decisão. Por serem específicos ao tipo de RC, estes requisitos são chamados de *requisitos verticais*.

De maneira complementar, existem outros requisitos, chamados *requisitos horizontais*, que são necessários a qualquer tipo de RC, independente da sua natureza. Tais requisitos incluem: comunicação entre pessoas, acesso

a dados, busca de informação, monitoramento e controle de processos de negócios, segurança, entre outros. Dessa forma, tal como apresentado na seção anterior (figura 2.2) os requisitos horizontais são suportados por uma infraestrutura de TIC, e os requisitos verticais são suportados por serviços que são conectados a ela. Esta infraestrutura, chamada *ICT-I (Information and Communication Technology Infrastructure)*, foi implementada com base em especificações independentes de plataforma na forma de uma arquitetura orientada a serviços (Rabelo et al., 2006, 2008). De modo a prover um modelo aberto e escalável, a *ICT-I* é definida por uma arquitetura de referência que pode ser instanciada de acordo com o tipo de RC. A figura 2.3 mostra a arquitetura de referência da *ICT-I*, bem como os serviços verticais que fazem uso dela.

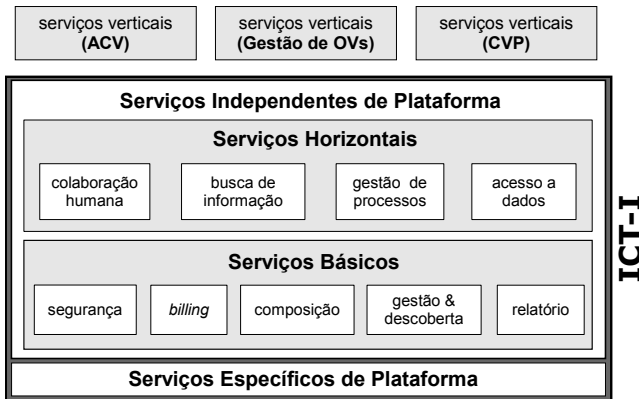


Figura 2.3: Arquitetura de Referência da *ICT-I*.

Fonte: Rabelo et al. (2006).

Para ilustrar o enquadramento da *ICT-I*, a figura mostra também os serviços verticais relacionados a ACVs, OVs e CVPs, responsáveis por lidar com os requisitos específicos a estes tipos de RCs. A *ICT-I*, por sua vez, oferece uma série de serviços de suporte aos serviços verticais. Os serviços da *ICT-I* podem ser divididos em *horizontais*, ou seja, serviços necessários a qualquer tipo de RC, e serviços *básicos*, que tratam de aspectos de nível mais baixo, tais como segurança, composição (orquestração) e gestão de serviços.

Dentre os serviços horizontais da *ICT-I*, destaca-se os serviços relacionados à busca de informação, foco desta tese. É preciso salientar, no entanto, que o modelo proposto neste trabalho foi desenvolvido de maneira independente, sendo integrado à *ICT-I* para que pudesse ser avaliado junto aos demais serviços oferecidos.

2.3 Busca Empresarial

Hawking (2004) define o termo “busca empresarial”, para qualquer organização que possua conteúdo textual em formato eletrônico, como a busca:

- No *website* externo da empresa;
- Nos *websites* internos da empresa (sua intranet) e;
- Por outros textos eletrônicos mantidos pela organização, na forma de *e-mail*, registros da banco de dados e documentos em geral.

Busca empresarial é diferente da busca na Web em vários aspectos (Raghavan, 2001; Hawking, 2004; Mukherjee e Mao, 2004). Em primeiro lugar, a noção de uma “boa” resposta para um consulta é bem diferente. Na Web, ela é definida de maneira vaga. Como muitos documentos são tipicamente relevantes a uma consulta, o usuário está geralmente procurando o “melhor” documento, ou seja, o mais relevante. Em uma intranet, a noção de uma “boa” resposta é geralmente definida como a resposta “certa”. Nesse caso, a consultas tendem a ter poucos resultados (em alguns casos apenas um), e esse resultados não são necessariamente o documento mais “popular”, o que muitas vezes determina o “melhor” resultado na Web.

Em segundo lugar, as forças sociais por trás da criação de conteúdo nesses dois casos são bem diferentes (Hawking, 2004). A Web reflete o esforço coletivo de muitos autores publicam conteúdo livremente, enquanto que uma intranet geralmente reflete a visão da entidade à qual ela serve. O conteúdo de uma intranet é criado visando a disseminação de informação e não para atrair e manter a atenção de algum grupo de usuários em particular. Além disso, nem todos os usuários têm permissão para publicar conteúdo (Mukherjee e Mao, 2004).

Além disso, o conteúdo em uma intranet é oriundo de repositórios heterogêneos. Por exemplo, servidores de e-mail e sistemas de gerenciamento de conteúdo em geral não fazem referência entre si via *hiperlinks*. Portanto, a estrutura de *links* em uma intranet é diferente da estrutura na Web. Segundo Hawking (2004), cerca de 10% na intranet da IBM, por exemplo, são fortemente conectados entre si, contra cerca de 30% na Web. Nesse sentido, algoritmos que levem em conta essa característica – tal como o *PageRank* (Brin, 1998) – não é tão efetivo em uma intranet como é na Web. Portanto, outras técnicas devem ser empregadas para melhorar os resultados da busca numa intranet.

2.3.1 Requisitos

Os aspectos apresentados mostram que técnicas particulares devem ser aplicadas para a implementação de sistemas de busca empresarial. Esses

aspectos podem ser agrupados conforme as seguintes características encontradas dentro de uma empresa (Abrol et al., 2001; Hawking, 2004; Mukherjee e Mao, 2004; Demartini, 2007):

- **Diversidade de fontes de conteúdo e formato:** empresas produzem e consomem informações estruturadas e não estruturadas oriunda de fontes heterogêneas, tais como: bases de dados, sistemas de gerenciamento de documentos, *e-mail*, páginas HTML, entre outros. Além disso, os documentos são codificados em uma variedade de formatos de arquivos e linguagens de representação.
- **Combinação de busca estruturada e não estruturada:** dada a diversidade de formatos e tipo de repositórios, é esperado que o sistema de busca seja capaz de oferecer uma interface de busca que integre tanto fontes estruturadas como não estruturadas. Um dos desafios aqui é como combinar mecanismos de ordenação (*ranking*) que unifiquem e definam a relevância entre repositórios heterogêneos.
- **Suporte à busca federada:** tal busca permite que um único ponto de acesso visualize o resultado de buscas feitas em múltiplos repositórios, ou feitas por múltiplos motores de busca. O desafio nesse tipo de funcionalidade é o agrupamento eficiente dos resultados e a sua apresentação de forma unificada. A dificuldade aqui está na organização de conjuntos diferentes que tipicamente não têm documentos em comum e empregam diferentes mecanismos de ordenação.
- **Acesso seguro:** o papel desempenhado por um indivíduo em uma empresa indica que tipo de documentos este pode ter acesso. Este requisito exige que os resultados de uma busca sejam filtrados de modo a apresentarem somente os documentos acessíveis pelo usuário. A implementação desse requisito em conjunto com os mecanismos de segurança nativos dos repositórios é um desafio particularmente difícil.
- **Processo de geração de conteúdo:** enquanto a Web tende a crescer democraticamente, intranets são normalmente administradas de maneira burocrática. A criação de conteúdo em uma intranet é normalmente centralizada em um pequeno grupo de pessoas. Além disso, a publicação de conteúdo pode depender de políticas específicas, envolvendo processos de revisão e de aprovação prévias.
- **Contexto de trabalho:** ao contrário dos usuários da Web, os funcionários de uma empresa estão em um contexto de trabalho explícito. Eles desempenham diferentes papéis e estão envolvidos em diversas atividades. Tal contexto deve ser levado em conta pois dá indicações do que pode ser relevante durante o processo de busca.

Hawking (2004) também ressalta como desafio adicional a falta de coleções de teste apropriadas para busca empresarial. Alguns esforços nessa direção vêm sendo feitos, como por exemplo, a coleção desenvolvida por Bailey et al. (2007), no âmbito da conferência TREC.

2.3.2 Abordagens e Tecnologias

Os requisitos para busca empresarial vêm sendo abordados por diversas técnicas, conforme enumerado (Mukherjee e Mao, 2004; Barrows e Traverso, 2006):

1. **Ontologias / Extração de Informação / Web Semântica:** a combinação de técnicas de extração de informações e ontologias para a geração automática de anotações semânticas permite o enriquecimento de documentos de modo a serem mais precisamente processados. Várias linguagens vêm sendo adotadas como padrão para a representação de ontologias, tais como RDF e OWL.
2. **Classificação, clusterização:** consiste na organização dos resultados com base em estruturas hierárquicas (uma ontologia, por exemplo), que podem ser criadas manualmente ou geradas a partir da análise da coleção de documentos.
3. **Utilização de metadados e “tags”:** a associação de metadados (autor, título, data, etc.) permite que a busca possa ser melhor refinada. Além disso, uma abordagem que se popularizou na Web, consiste na utilização de etiquetas (*tags*), que são listas de termos que são usadas para enriquecer o conteúdo. Diversos sites da Web, especialmente os que enfatizam o aspecto social (*weblogs*, sites de compartilhamento de vídeo e redes sociais), permitem que usuários associem etiquetas próprias ao conteúdo criado. Outro tipo de metadado que pode ser adicionado ao conteúdo é a informação referente ao contexto do usuário (processo, tarefa, papel, aplicação, dispositivo), tal como feito por Gross e Klemke (2002).
4. **Sistemas de recomendação:** tal abordagem visa analisar o comportamento dos usuários definindo perfis, de tal forma que o sistema possa recomendar resultados similares para usuários com perfis semelhantes.
5. **Motores de meta-busca:** são sistemas que implementam a *busca federada*, agregando diversos motores de busca e apresentando os resultados de forma unificada.
6. **Interfaces gráficas de usuário ricas e adaptáveis:** tais interfaces devem dar suporte ao usuário tanto para de definição de consultas quanto

para a apresentação e navegação de resultados. Aqui há uma combinação das técnicas acima apresentadas: a interface pode ajudar o usuário na construção de consultas estruturadas (por exemplo, baseadas numa ontologia ou taxonomia); e pode combinar aspectos de categorização, metadados e *tags* visando uma melhor organização dos resultados.

Os próximos capítulos apresentam uma revisão bibliográfica de alguns dos tópicos aqui apresentados: ontologias, Web Semântica e anotações semânticas; e recuperação de informação, incluindo busca semântica e baseada em contexto. Uma análise do estado de arte é feita para cada um dos tópicos apresentados.

Capítulo 3

Ontologias e Web Semântica

Este capítulo apresenta uma revisão bibliográfica sobre algumas áreas de pesquisa relacionadas às *ontologias*. O objetivo é apresentar ao leitor os conceitos da base para o modelo proposto neste trabalho. Além dos conceitos básicos e áreas de aplicação e pesquisa, este capítulo cobre a área da *Web Semântica*, que define padrões e tecnologias utilizados para representação de ontologias.

Adicionalmente, este capítulo também cobre tópicos relacionados a *anotações semânticas*. Uma anotação semântica é um vínculo entre o conteúdo de um documento e sua descrição formal, definida por ontologias. Este tipo de mecanismo enriquece o conteúdo de documentos permitindo que aplicações tratem esse conteúdo de forma mais precisa. Adicionalmente aos conceitos, são apresentadas as principais técnicas as plataformas disponíveis.

3.1 Introdução

O termo “ontologia” é originário da filosofia, tendo sido definido pelos pensadores gregos da Antiguidade como uma maneira de se estudar a natureza do ser e da existência, através da classificação das coisas do mundo. No âmbito da ciência da computação uma ontologia constitui um conjunto de conceitos, propriedades, relações e axiomas que descrevem um dado domínio do conhecimento.

Segundo Studer et al. (1998), “uma ontologia é uma especificação formal e explícita de uma conceitualização compartilhada”. *Conceitualização* se refere a um modelo abstrato de algum fenômeno no mundo. *Explícito* significa que o tipo dos conceitos utilizados e as restrições no seu uso são definidos explicitamente. *Formal* significa que a ontologia deve ser interpretável pelas máquinas. *Compartilhada* reflete a noção de que uma ontologia captura o conhecimento consensual, ou seja, foi aceita por um grupo.

Noy e McGuinness (2001) definem ontologia como sendo uma descrição formal e explícita de conceitos em um domínio de discurso. Esta

descrição contém **classes** (também chamadas de *conceitos*), **propriedades** (também chamadas de *slots* ou *papeis*) que descrevem características e atributos dos conceitos e **restrições nas propriedades** (também chamadas de *facet*as ou *restrições de papeis*). As mesmas autoras definem uma *base de conhecimento* como sendo uma ontologia e suas instâncias.

Por prover as definições acerca de um domínio, ontologias permitem que aplicações usem uma semântica precisa e formal para processarem suas informações. Desse modo, diversas áreas têm tirado vantagem do uso de ontologias, como por exemplo, gestão de conhecimento, sistemas de suporte à aprendizagem (tutores) e comércio eletrônico. Adicionalmente, outra área que vêm impulsionando o uso de ontologias é a *Web Semântica* (Berners-Lee et al., 2001). Tal área tem oferecido padrões e tecnologias para diversos aspectos na implementação e uso de ontologias. Um exemplo dessas tecnologias compreende as linguagens padronizadas para representação de ontologias, que são a base para a implementação da grande maioria das aplicações desse tipo.

3.2 Representação de Ontologias

Uma ontologia pode ser representada de várias maneiras. Grobelnik e Mladenic (2006) se referem a uma ontologia como sendo um grafo direcionado, onde os vértices são os conceitos e os arcos correspondem às relações entre os conceitos. Outros autores definem ontologias como sendo uma tupla contendo vários conjuntos: conceitos, tipos, relações, atributos, instâncias, entre outros (Dill et al., 2003; Ehrig et al., 2005).

Formalmente, ontologias são representadas usando-se principalmente formalismos *orientados a classes e relações*. Este tipo de representação foca as classes de objetos presentes no domínio, buscando relações entre eles, principalmente *herança* ou *especialização*, e criando as taxonomias de classes que ajudam a perceber o domínio como um todo (de Freitas, 2007). Estes formalismos incluem:

- **Frames** (Minsky (1975) *apud* de Freitas (2007)): são considerados os precursores das linguagens orientadas a objetos, pois permitem a definição de classes, atributos, instâncias e herança entre classes. Além disso, permitem a definição de restrições em atributos através de estruturas chamadas *facet*as.
- **Redes Semânticas** (Woods (1975) *apud* de Freitas (2007)): consistem de nós conectados por arcos, onde os nós representam conceitos e os arcos representam relações entre conceitos. Uma rede semântica pode ser considerada como um sistema baseado em *frames*, se os arcos da rede forem atributos do *frame* correspondente a esta rede.

- **Lógica de Descrições** (Brachman e Schmolze (1985) *apud* de Freitas (2007)): baseado em redes semânticas, este formalismo propõe a definição de conceitos por meio de descrições esparsas. Assim, ao contrário dos *frames*, não é necessário que todas as definições relativas a uma classe estejam agrupadas.

Singh e Huhns (2005) enquadram ontologias em relação a outras formas para representação de conhecimento, cada qual com um poder de expressividade maior que a anterior. São elas:

1. **Palavras-chave**: uma lista simples de termos.
2. **Thesauri**: oferecem uma abordagem mais estruturada do que palavras-chave, relacionando os termos conforme as seguintes categorias de classificação: “mais abrangente que”, “menos abrangente que” e “relacionado a”.
3. **Taxonomias**: proveem estruturas de classificação que adicionam o conceito de herança (especialização e generalização).
4. **Ontologias**: oferecem uma variedade de relacionamentos mais rica que as taxonomias, como por exemplo, agregação (relação “parte-todo”) e instanciação (relação “classe-indivíduos”).

Ontologias podem também ser representadas por linguagens que capturam a expressividade de algum formalismo de representação de conhecimento. Dentre as linguagens mais relevantes para representação de ontologias, destacam-se: *Web Ontology Language* (OWL) (McGuinness e van Harmelen, 2004), *Simple HTML Ontology Extensions* (SHOE) (Heflin e Hendler, 2000), e *Unified Modeling Language* (UML) (Booch et al., 1999). Dentre estas linguagens, a OWL se tornou o padrão adotado no contexto da Web Semântica (seção 3.5).

3.3 Tipos de Ontologias

Segundo Studer et al. (1998), embora as ontologias visem capturar o conhecimento sobre algum domínio, estas também podem variar muito. De acordo com o nível de generalidade, as ontologias podem ser agrupadas em diferentes tipos:

- **Ontologias de Domínio** capturam o conhecimento válido para um tipo particular de domínio (por exemplo, eletrônica, medicina e mecânica);
- **Ontologias Genéricas (gerais ou de topo)** são válidas para vários domínios, pois trazem definições mais genéricas necessárias para a

compreensão de aspectos do mundo, como tempo, processos, papéis, espaço, seres e coisas. Exemplos desse tipo de ontologia incluem Cyc e PROTON;

- **Ontologias de Aplicação** contêm todo o conhecimento necessário para modelar uma aplicação em particular;
- **Ontologias de Representação** não estão comprometidas com nenhum domínio em particular. Tais ontologias proveem entidades representacionais sem declarar o que deve ser representado. Um exemplo é uma ontologia que define as primitivas de representação, tais como *frames*, axiomas e atributos.

3.4 Áreas de Pesquisa e Aplicação

Dentre as diversas áreas de pesquisa e aplicação de ontologias, destacam-se as seguintes:

- *Construção e engenharia de ontologias*: esta área está relacionada à definição de metodologias para a construção de ontologias.
- *Editores de ontologias*: seu foco é tanto na usabilidade de ferramentas para edição de ontologias quanto na utilização de linguagens de representação. Destaque para a ferramenta *Protégé*¹.
- *Integração de sistemas*: visa integrar sistemas / bases de dados heterogêneos.
- *Evolução de ontologias*: foca na definição de mecanismos para a captura, representação, propagação, e implementação das mudanças que podem ocorrer em ontologias (Stojanovic et al., 2002).
- *Mediação entre ontologias*: área de intensa pesquisa, visa construir mapeamentos entre partes comuns de diferentes ontologias (Kalfoglou e Schorlemmer, 2003; de Bruijn et al., 2006).
- *Processamento de Linguagem Natural*: neste contexto, duas tarefas distintas podem ser enumeradas:

1. Ontologias podem ser empregadas em aplicações que processam textos. Este aspecto envolve técnicas de Extração de Informação (EI) e a geração de anotações semânticas, conforme apresentado na seção 3.6.

¹<http://protege.stanford.edu/>

2. Ontologias podem ser extraídas a partir de textos, processo chamado aprendizado de ontologias. Embora este aspecto não seja abordado neste trabalho, ele tem alguma relação com a evolução de ontologias, no que diz respeito à captura de possíveis mudanças.

- *Recuperação de informação*: ontologias e anotações semânticas vêm sendo utilizadas para aumentar a precisão das técnicas de RI. Esta área é apresentada na seção 4.1.
- *Web Semântica*: segundo a visão proposta para a Web Semântica, informações estão vinculadas a metadados descritos por ontologias (anotações semânticas). Por oferecer um significado mais explícito para informações, o uso de ontologias permite que aplicações tratem tais informações de forma mais precisa. Esta área vem sendo suportada por uma série de padrões, conforme apresentado na seção 3.5.

A subseção a seguir apresenta a Web Semântica, sua arquitetura e os padrões adotados para a representação de ontologias.

3.5 Web Semântica

Um dos problemas encontrados na Web é que o conteúdo espalhado nos *sites* foi feito para ser consumido por pessoas. Isto significa que aplicações computacionais têm dificuldade em processar de maneira precisa o conteúdo da Web devido à falta do seu enquadramento em um contexto adequado. O exemplo mais claro de tal limitação pode ser visto nas máquinas de busca da Web, onde, por falta de um contexto bem definido, os resultados trazidos aos usuários normalmente contêm informações irrelevantes.

O conceito de Web Semântica surgiu com o intuito de superar esta limitação. A Web Semântica é uma extensão da Web onde informações possuem um significado bem definido (com o suporte de ontologias), permitindo que computadores e pessoas trabalhem em cooperação (Berners-Lee et al., 2001). Portanto, a ideia é que ontologias possam dar o contexto necessário às páginas da Web de modo que não apenas pessoas, mas também sistemas possam acessar o conteúdo da Web de forma mais precisa. O termo “Web Semântica” foi cunhado por Tim Bernes-Lee (Berners-Lee et al., 2001), que é também o idealizador da Web tal como a conhecemos hoje.

3.5.1 Arquitetura da Web Semântica

Para concretizar esta visão, Bernes-Lee idealizou uma arquitetura que estende a Web adicionando uma infraestrutura de suporte à semântica. A arquitetura da Web Semântica vem evoluindo ao longo dos anos e seu estado atual pode ser visto na figura 3.1.

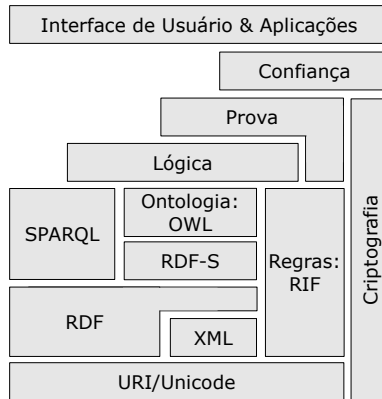


Figura 3.1: Arquitetura da Web Semântica.

Fonte: Bratt (2007).

É importante mencionar que estas camadas têm sido implementadas com base em padrões especificados pelo *World Wide Web Consortium* (W3C)². O W3C é uma entidade que define dezenas de padrões para a Web em geral, e em particular, tem feito esforços também na padronização da Web Semântica. Apenas para enumerar alguns, o W3C está envolvido na especificação da XML, HTML, RDF, OWL, esquema XML, entre outros padrões que serão apresentados na sequência.

Identificadores e Codificação de Caracteres: URI + Unicode

Esta camada forma a base não apenas da Web Semântica, mas também da Web tradicional. A Web utiliza identificadores chamados *Uniform Resource Identifiers* (URIs) como forma de identificar os seus recursos. Por recurso, entende-se qualquer tipo de conteúdo que possa ser identificado na Web: páginas HTML, imagens, vídeos, arquivos em geral, serviços Web, entre outros. Um tipo de URI é o chamado *Uniform Resource Locator* (URL), que além de identificar um recurso, serve também para identificar de que forma este recurso deve ser localizado. Neste caso, um URL indica o tipo de protocolo de comunicação que deve ser utilizado para o acesso ao recurso. Por exemplo, “<http://www.das.ufsc.br>”. Um outro tipo de URI é o *Uniform Resource Name* (URN). Ao contrário do URL, que muda sempre que a localização do recurso muda, um URN é independente da localização do recurso. Para este tipo de identificador, um serviço de nomes se faz necessário para mapear de URNs para URLs.

O outro aspecto tratado nesta camada é a forma de codificação de

²<http://www.w3.org>

caracteres, que neste caso é feita pelo padrão Unicode. Antes da criação deste padrão, centenas de outros padrões eram utilizados, de acordo com a língua e o sistema operacional utilizados. O principal problema desta diversidade é que em um ambiente distribuído como a Web, os sistemas precisam lidar com muitos tipos de codificações, muitas vezes conflitantes: dois tipos de codificação poderiam ter o mesmo código para caracteres diferentes e códigos diferentes para o mesmo caractere.

Assim, no intuito de resolver estes problemas, foi desenvolvido o Unicode. Este padrão visa cobrir os caracteres de todas as línguas escritas do mundo, bem como símbolos em geral. Neste sistema, cada caractere é associado a um número único, independente da língua utilizada. A sua implementação, ou seja, como estes números serão representados na forma de *bytes* é feita por outros padrões, entre eles o UTF-8, UTF-16 e UCS-2.

Sintaxe: XML

Com base na *Extensible Markup Language* (XML), esta camada oferece a base sintática para as demais camadas. XML é um padrão definido pela W3C (Bray et al., 2006b) e que se consolidou como o padrão *de facto* para comunicação entre aplicações. XML é um formato para representação estruturada de dados através da utilização de marcadores (*tags*), tendo a sintaxe semelhante à da *HyperText Markup Language* (HTML). De fato, ambas as linguagens têm a sua origem no *Standard Generalized Markup Language* (SGML), que é uma metalinguagem (padrão ISO 8879) através da qual se pode definir linguagens de marcação para documentos (Sperberg-McQueen e Burnard, 1994).

Entretanto, as semelhanças param na sintaxe. HTML tem a semântica predefinida, que é interpretada pelos navegadores da Web para que as páginas HTML possam ser apresentadas adequadamente. Em outras palavras, o propósito da HTML é definir o aspecto visual de documentos, como o tamanho, cor e tipo das fontes utilizadas. Além disso, o conteúdo de uma página HTML está intrinsecamente vinculado à estrutura do documento. Por outro lado, XML não tem a semântica predefinida, ou seja, é possível definir uma nova estrutura de dados de acordo com o tipo de aplicação. Por esta razão, XML pode ser considerada uma metalinguagem de marcadores. Ao contrário da HTML, documentos XML não definem a maneira pela qual os seus dados serão apresentados. Esta pode ser vista como mais uma vantagem, pois os dados estão desvinculados da forma como são apresentados. A figura 3.2 mostra um exemplo de documento XML contendo informações de um livro.

Dada a flexibilidade da XML, é possível definir o conjunto de marcadores que será utilizado, bem como seus tipos de dados e sua estrutura de aninhamento. Estes aspectos podem ser definidos formalmente através dos *esquemas XML* (*XML schemas*) (Thompson et al., 2006). A vantagem de tal

```
<?xml version="1.0"?>
<livro isbn="123456">
  <titulo>Modern Information Retrieval </titulo>
  <autor>Baeza-Yates </autor>
  <autor>Ribeiro-Neto </autor>
  <editora>Addison Wesley </editora>
  <ano>1999</ano>
</livro>
```

Figura 3.2: Exemplo de Documento XML.

artifício é que, por ser formal, um esquema pode ser usado para validar documentos XML. Esta validação é feita por componentes de *software* chamados *parsers* de XML, que normalmente são implementados como bibliotecas, nas mais diversas linguagens de programação.

Além do XML e seus esquemas, outro aspecto que é definido nesta camada é o uso de *namespaces*. Um *namespace* serve para qualificar os nomes dos marcadores em documentos XML, com o objetivo de evitar conflitos no caso de dois documentos utilizarem os mesmos nomes para os seus marcadores (Bray et al., 2006a). Um *namespace* é identificado por um URI, que normalmente é o URL da entidade que define o esquema. Para fins de legibilidade, este URI é substituído por um prefixo que é colocado antes dos nomes dos marcadores que fazem parte deste *namespace*.

Assim, com a flexibilidade oferecida pelos esquemas e *namespaces*, a XML serve de base para a definição de uma série de outros padrões, incluindo os utilizados nas camadas superiores da arquitetura da Web Semântica.

Intercâmbio de Dados: RDF

O *Resource Description Framework* (RDF) é uma linguagem usada para a representação de recursos da Web e seus metadados (Decker et al., 2000a,b; Manola e Miller, 2004), tais como título, autor, data, e outros mais. RDF é um padrão definido pelo W3C e utiliza a sintaxe da XML. RDF representa os metadados dos recursos da Web usando triplas *sujeito-predicado-objeto*, onde o *sujeito* representa um recurso que é identificado por uma URI, o *predicado* é uma propriedade que representa o metadado em si (por exemplo, o autor de uma página) e o *objeto* se refere ao valor da propriedade, que pode ser um valor literal ou um outro recurso. Por exemplo, para dizer que a página localizada em “http://www.das.ufsc.br/~rui” tem como autor “Rui Tramontin”, a figura 3.3 mostra como seria a declaração em RDF.

Esta tripla também pode ser representada utilizando-se a *Notation3* (N3) (Berners-Lee, 2006), conforme a figura 3.4.

RDF pode ser também representado na forma de uma grafo, como


```
<rdf:Description rdf:about="http://www.das.ufsc.br/~rui">
  <autor>Rui Tramontin</autor>
</rdf:Description>
```

Figura 3.3: Exemplo de declaração RDF.

```
<http://www.das.ufsc.br/~rui <autor> "Rui Tramontin" .
```

Figura 3.4: RDF representado pela *Notation3*.

mostra a figura 3.5. RDF pode ser comparado, em termos formais, a uma rede semântica (de Freitas, 2007).



Figura 3.5: RDF representado na forma de uma grafo.

Taxonomias: Esquemas RDF

Apesar de permitir a associação de metadados a recursos da Web, RDF não permite a definição de classes e propriedades específicas que possam ser reutilizadas por aplicações. Para isto foi definido o *Esquema* RDF ou *RDF Schema* (RDFS), que permite a criação de classes e propriedades personalizadas, da mesma maneira que em uma linguagem de programação orientada a objetos (Brickley e Guha, 2004). Neste sentido, com RDFS é possível criar uma hierarquia de classes (herança), bem como instâncias de classes. Além disso, é possível definir também hierarquias de propriedades, pois, ao contrário das linguagens orientadas a objetos, em RDFS as propriedades são definidas separadamente, sendo depois associadas às classes. A figura 3.6 ilustra um exemplo de classes definidas em RDFS.

Neste exemplo são definidas as classes “Pessoa”, “Veículo” e “Carro”, onde esta última é subclasse de “Veículo”. Adicionalmente é definida a propriedade “registradoPara”, que é uma propriedade da classe “Veículo” (definido por *rdfs:domain*) e seu valor deve ser uma instância da classe “Pessoa” (definido por *rdfs:range*).

Embora ofereça uma certa expressividade para a criação de hierarquias de classes e propriedades, RDFS não é poderosa o suficiente para expressar certas relações necessárias à especificação de ontologias. Manola e Miller (2004) destacam algumas limitações do RDFS:

- Não permite definir restrições de cardinalidade em propriedades.

```

<rdfs:Class rdf:ID="Pessoa"/>
<rdfs:Class rdf:ID="Veiculo"/>
<rdfs:Class rdf:ID="Carro">
  <rdfs:subClassOf rdf:resource="#Veiculo"/>
</rdfs:Class>
<rdf:Property rdf:ID="registradoPara">
  <rdfs:domain rdf:resource="#Veiculo"/>
  <rdfs:range rdf:resource="#Pessoa"/>
</rdf:Property>

```

Figura 3.6: Classes e propriedades definidas em RDFS.

- Não permite especificar que uma propriedade é *transitiva*.
- Não permite especificar que duas classes diferentes (identificadas por URIs diferentes) são equivalentes.
- Não permite especificar que duas instâncias diferentes (identificadas por URIs diferentes) são equivalentes.
- Não consegue descrever novas classes em termos de combinações de outras classes, por exemplo, usando união ou intersecção.

Estas características são providas por linguagens de ontologias, que são definidas na próxima camada da arquitetura.

Ontologias: OWL

A camada responsável pela especificação de ontologias é implementada pela *Web Ontology Language* (OWL). OWL é um padrão da W3C, e representa uma evolução de outra linguagem para ontologias, chamada DAML-OIL. A DAML-OIL, por sua vez, representa um esforço conjunto de pesquisadores americanos, que desenvolveram o *DARPA Agent Markup Language* (DAML), e de europeus, que desenvolveram o *Ontology Inference Layer* (OIL), com o intuito de criarem uma linguagem mais poderosa (Al-Khalifa, 2007). Tanto o DAML-OIL quanto o OWL são baseados em RDF e RDFS e adicionam um maior poder de expressividade para representar ontologias (McGuinness e van Harmelen, 2004).

OWL possui três sublinguagens, cada qual com poder de expressividade diferente: OWL Lite, OWL DL e OWL Full (McGuinness e van Harmelen, 2004). Tais sublinguagens são apresentadas a seguir:

- **OWL Lite:** versão limitada da OWL, que provê uma hierarquia de classificação e restrições simples. Por exemplo, OWL Lite permite apenas cardinalidade 0 ou 1.

- **OWL DL:** Possui toda a expressividade da lógica de descrições (daí o nome DL: *Description Logics*), garantindo também a computabilidade (execução em tempo finito). OWL DL inclui todas as construções da OWL, tais como união, intersecção, complemento e disjunção. Entretanto, tais construções só podem ser usadas sob certas condições, como por exemplo, uma classe não pode ser instância de outra classe.
- **OWL Full:** oferece toda a expressividade da linguagem, mas sem garantias computacionais. OWL Full permite que uma ontologia aumente o significado do vocabulário predefinido da linguagem (tanto da RDF quanto da OWL).

Cada uma destas sublinguagens é uma extensão do seu predecessor mais simples, ou seja, OWL Lite está contido no OWL DL, que por sua vez está contido no OWL Full.

Consultas: SPARQL

SPARQL (Prud'hommeaux e Seaborne, 2008) é uma linguagem para representar consultas para RDF. Ela pode ser usada em diversas fontes de dados, armazenados tanto nativamente em RDF ou visualizados como RDF através de um *middleware*. Os resultados desse tipo de consulta podem ser *result sets* ou grafos RDF.

O exemplo abaixo mostra uma consulta SPARQL para encontrar o autor de uma página Web definido em um grafo RDF. A consulta é organizada em duas partes: a cláusula SELECT identifica as variáveis que devem aparecer nos resultados e a cláusula WHERE provê o padrão básico a ser casado com o grafo. O padrão do grafo neste exemplo consistem em uma única tripla com uma única variável (?autor) na posição do objeto.

Dados:

```
<http://www.das.ufsc.br/~ruirui <autor> "Rui Tramontin" .
```

Consulta:

```
SELECT ?autor
WHERE
{
  <http://www.das.ufsc.br/~ruirui <autor> ?autor .
}
```

Esta consulta, sobre os dados acima, tem uma solução:

| |
|-----------------|
| autor |
| "Rui Tramontin" |

Regras: RIF

O *Rule Interchange Format* (RIF) (Kifer e Boley, 2010) foi criado para o intercâmbio de regras entre sistemas. O foco dado nesta linguagem foi no intercâmbio pois, ao contrário dos demais padrões da Web Semântica, uma única linguagem não satisfaria todas as necessidades dos diversos paradigmas para uso de regras. RIF foi projetado na forma de diversos *dialetos* cobrindo os diversos tipos de linguagens de regras. Os dialetos padrão do RIF são *Core*, *Basic Logic Dialect* (BLD) e *Production Rules Dialect* (PRD):

- **Core:** é a linguagem de base, que fornece o subconjunto comum à maioria dos motores de busca.
- **BLD:** adiciona outros aspectos ao *Core*: funções lógicas, argumentos nomeados, entre outros.
- **PRD:** adiciona a noção de regras de encadeamento para a frente, onde uma regra é disparada e então executa alguma ação, tal como adicionar uma informação à memória de trabalho.

As Camadas de Lógica, Prova e Confiança

As camadas superiores da Web Semântica ainda estão sendo definidas. A camada de Lógica permite que sejam escritas regras que serão validadas na camada de Prova, enquanto que a camada de Confiança decidirá se a prova é confiável ou não (Koivunen e Miller, 2001). É importante reparar que as camadas superiores da Web Semântica estão relacionadas à segurança (criptografia), cujo objetivo é garantir a autenticidade e a integridade das ontologias, regras e provas, e deste modo prover meios para o estabelecimento de confiança.

3.5.2 Ferramentas

É preciso salientar que além do esforço para padronização da Web Semântica, devem ser implementadas aplicações que utilizem estes padrões. De fato, centenas de aplicações foram e vêm sendo desenvolvidas e novas surgem a cada dia. Bergman (2011) mantém uma lista de ferramentas que são classificadas de acordo com o tipo de aplicação. Esta lista é constantemente atualizada e no momento conta com um total de 926 ferramentas (em junho de 2007, essa lista contava com 524 ferramentas!). A seguir é mostrada uma lista contendo os tipos de aplicações mais relevantes para o escopo deste trabalho. Para cada tipo é mostrado também a quantidade de ferramentas (para comparação, os números entre parênteses correspondem a junho de 2007):

- Anotador: 75 (28)

- Aplicação composta / *Framework*: 123 (29)
- Base de Dados: 83 (31)
- Extração de Informação: 118 (45)
- Editor de Ontologia: 59 (10)
- Mapeador/Mediador de Ontologia: 32 (13)
- Linguagem de Consulta: 42 (15)
- Máquina de Inferência / Raciocínio: 39 (23)
- Máquina de Busca: 45 (23)
- *Wrapper* (extrator de dados da Web): 34 (22)

Além disso, a concretização da Web Semântica depende não apenas que as aplicações passem a utilizar ontologias, mas também que o próprio conteúdo da Web esteja vinculado a elas, com a utilização das chamadas *anotações semânticas*. Neste caso, considerando-se o tamanho da Web, a anotação manual do seu conteúdo é uma tarefa inviável. Assim, um tipo particular de aplicação visa justamente efetuar a tarefa de gerar anotações semânticas (semi-)automaticamente nos *sites* da Web.

3.6 Anotações Semânticas

Como dito na seção anterior, a realização da Web Semântica depende da anotação formal do conteúdo espalhado na Web de modo com que sistemas (e não apenas pessoas) possam interpretar esse conteúdo. Em termos simples, uma anotação semântica é um vínculo entre o conteúdo de um texto (expresso em linguagem natural) e sua descrição semântica definida por uma ontologia.

Para criar este vínculo entre texto e ontologia, são necessárias técnicas para análise de linguagem natural. Uma tecnologia que vem tendo papel central nesse contexto é a chamada Extração de Informação (EI) (Bontcheva et al., 2006). Portanto, para entrar em detalhes a respeito de anotações semânticas é preciso apresentar o conceito de EI.

3.6.1 Extração de Informação

EI visa extrair informações contidas em textos e organizá-las na forma de uma tabela (planilha ou base de dados) para que possam ser analisadas futuramente. Existem cinco tipos de informações que podem ser encontradas usando técnicas de EI, cada uma servindo de base para a próxima, de modo a prover informações cada vez mais detalhadas acerca do texto (Cunningham, 2006). São elas: *entidades*, *menções*, *descrições*, *relações* e *eventos*, e são detalhadas a seguir:

- **Entidades:** também chamadas *entidades nomeadas* pois identificam coisas que podem ser referenciadas por nome, tais como, pessoas, organizações e lugares. Entidades podem incluir também valores escalares (números, datas, valores monetários) e endereços.
- **Menções:** todos os lugares onde entidades em particular são referenciadas no texto. Este processo é chamado de *resolução de co-referência* e identifica pronomes que substituem entidades, ou formas ortográficas diferentes para a mesma entidade.
- **Descrições:** com base nas entidades reconhecidas e suas menções, descrições podem ser feitas a estas entidades. Por exemplo, pode-se identificar que um certo nome diz respeito a uma pessoa.
- **Relações:** relações entre entidades podem ser identificadas, como por exemplo uma relação de emprego entre uma pessoa e uma empresa.
- **Eventos:** são inferidos com base nas descrições e relações previamente identificadas. Por exemplo, o processo de reconhecimento de descrições identifica duas pessoas e uma empresa em uma notícia de jornal. O reconhecimento de relações identifica que esta pessoa trabalha para esta empresa. Finalmente, um evento que poderia ser reconhecido seria esta pessoa assinando um contrato (em nome desta empresa) com algum fornecedor.

Enquanto que o reconhecimento automático de entidades tem uma precisão praticamente tão boa quanto uma pessoa (em torno de 95%), o reconhecimento de eventos chega a no máximo 60% (Cunningham, 2006). Isto é devido tanto pela dificuldade de se identificar eventos quanto pelo fato de este processo depender da qualidade das informações extraídas nas etapas anteriores.

Também é interessante mencionar a diferença entre Extração de Informação (EI) e Recuperação de Informação (RI). Na EI, as informações são extraídas e apresentadas ao usuário de forma estruturada, enquanto que na RI, uma lista de documentos é recuperada e cabe ao usuário filtrar as informações desses documentos. Por outro lado, técnicas de RI utilizam modelos matemáticos para representar e ordenar os documentos (conforme é apresentado na seção 4.1) e são, portanto, independentes de domínio, enquanto que técnicas de EI são altamente dependentes do domínio no qual estão vinculadas. EI e RI são técnicas complementares e sua combinação tem o potencial de criar ferramentas mais poderosas para processamento de textos (Gaizauskas e Wilks, 1998).

3.6.2 Anotação Semântica

Uma vez identificadas, as informações extraídas de um documento (entidades, relações, eventos) são então vinculadas a uma descrição formal definida por uma ontologia, formando assim as chamadas anotações semânticas. Um exemplo de anotação semântica pode ser visto na figura 3.7, onde entidades nomeadas são vinculadas às instâncias de uma ontologia.

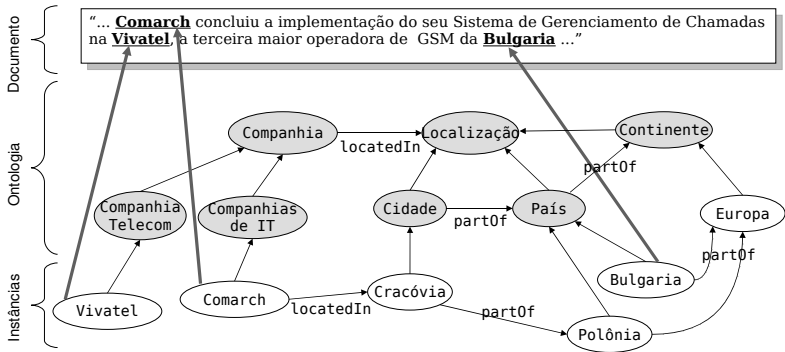


Figura 3.7: Anotações semânticas apontando para algumas instâncias (em branco) de um fragmento de ontologia (em cinza).

Fonte: baseado em Kiryakov et al. (2004).

Apesar do exemplo apresentado acima ser o mais comum, anotações semânticas não estão vinculadas a apenas uma única palavra ou expressão. Uma anotação semântica pode estar vinculada também a uma sentença, parágrafo, seção ou ao documento inteiro (Al-Khalifa, 2007).

Anotações semânticas também podem ser geradas manualmente, sem a necessidade de se extrair informações automaticamente. Embora existam ferramentas que auxiliem de forma amigável a anotação manual de documentos, este processo tende a ter muitos erros em função das características individuais das pessoas que fazem a anotação. Tais características incluem: familiaridade com o domínio, treinamento prévio e interesse e motivação pessoais (Bayerl et al., 2003). Além disso, outro problema com a anotação manual é o volume de documentos disponíveis na Web. Segundo Maedche e Staab (2001), embora as ferramentas para ontologias tenham amadurecido na última década, o processo manual utilizado por tais ferramentas tem levado ao chamado *gargalo da aquisição de conhecimento*³. De modo a solucionar esse "gargalo", técnicas e ferramentas para anotação semântica (semi-) automática vêm sendo propostas.

³Do inglês *knowledge acquisition bottleneck*.

Classificação das Técnicas de Anotação Semântica

Segundo a classificação feita por Reeve e Han (2005), plataformas para anotação semântica podem ser classificadas de acordo com o método usado na EI. Há duas categorias principais, nomeadamente: baseada em **Padrões** e baseada em **Aprendizado de Máquina** (figura 3.8). Adicionalmente, plataformas podem usar métodos provenientes de ambas as categorias, criando-se uma nova classe de plataformas chamada **Multiestratégia**. O objetivo desta última abordagem é ao mesmo tempo se beneficiar dos pontos fortes e compensar as fraquezas dos métodos relacionados a cada categoria.

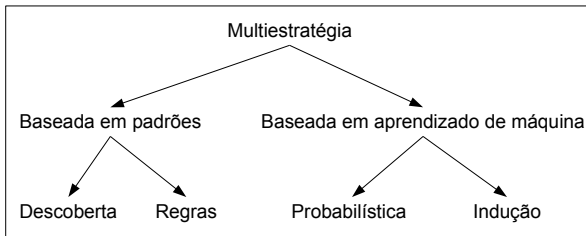


Figura 3.8: Classificação das plataformas para anotação semântica.

Fonte: Reeve e Han (2005).

Plataformas baseadas em Padrões podem efetuar a **descoberta de padrões** ou utilizar padrões definidos manualmente. No primeiro caso, um conjunto inicial de entidades (chamadas “sementes”) é definido e um *corpus*⁴ de treinamento é processado para que se possam identificar os padrões que identifiquem as entidades definidas. Novas entidades são então descobertas, além de novos padrões. Este processo continua recursivamente até que não se descubra nenhuma nova entidade, ou se interrompido pelo usuário. Na outra abordagem, anotações podem ser geradas a partir de **regras** definidas manualmente.

Plataformas baseadas em Aprendizado de Máquina podem utilizar dois métodos: probabilidade e indução. Plataformas **probabilísticas** usam métodos estatísticos para prever a localização de entidades no texto, como por exemplo o *Modelo Oculto de Markov*⁵ (Seymore et al., 1999) ou *Análise de N-gramas* (Downey et al., 2007). No caso de plataformas que usam **indução**, análises linguísticas ou estruturais são feitas em documentos previamente anotados (treinamento) de modo que regras para EI sejam induzidas.

Em resumo, sistemas baseados em regras dependem de regras definidas manualmente ou de algum conjunto inicial de “sementes” (no caso de

⁴Neste contexto, um *corpus* é uma coleção de documentos usados para análise linguística.

⁵Do inglês *Hidden Markov Model*

descoberta de padrões); e sistemas baseados em aprendizado de máquina precisam de um *corpus* de treinamento previamente anotado.

Como dito, plataformas do tipo Multiestratégia são capazes de combinar métodos de ambas as categorias. Entretanto, segundo Reeve e Han (2005), nenhuma plataforma disponível implementa a abordagem de multiestratégia para anotação semântica, embora ela seja adotada em sistemas para extração de ontologia, tal como o *On-To-Knowledge* (Kietz et al., 2000).

3.6.3 Estado da Arte

Com relação ao estado da arte na área de anotação semântica, é preciso destacar a pesquisa feita por Reeve e Han (2005) e Uren et al. (2006). Ambos os trabalhos categorizam as plataformas e técnicas disponíveis, cada um levando em consideração diferentes perspectivas. Reeve e Han (2005) classificam as plataformas de anotação semântica de acordo com a técnica utilizada na anotação, conforme visto na seção anterior.

Uren et al. (2006) classificam as plataformas em **arcabouços** e **ferramentas**. *Arcabouços* definem as diretrizes e especificações relacionadas ao processo de anotação e podem ser implementados de diferentes formas pelas ferramentas. Já as *ferramentas* podem ser classificadas em *manuals*, *automáticas* ou *ambientes integrados para anotação*. A seguir são apresentadas algumas plataformas, de acordo com esta classificação. Foram selecionadas plataformas que ofereçam algum tipo de automatização.

A figura 3.9 apresenta todas as ferramentas para anotação semântica apresentadas, divididas conforme a sua classificação (manual, automática, ambiente integrado) e também são mostrados os arcabouços para anotação. A figura também apresenta as plataformas para EI utilizadas por algumas ferramentas bem como as dependências que existem entre algumas delas.

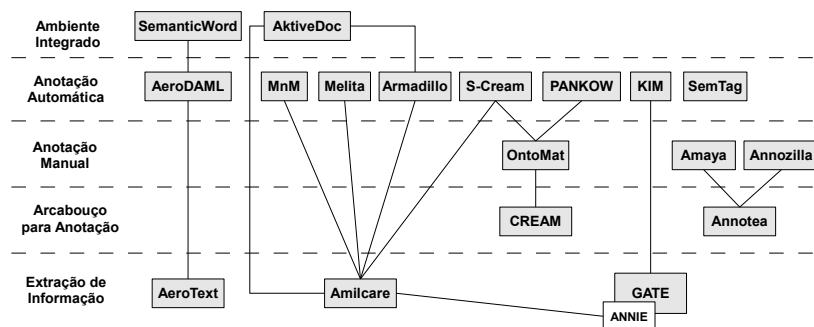


Figura 3.9: Plataformas para anotação semântica e suas dependências.

Medição do Desempenho das Plataformas

Apesar da grande variedade de plataformas disponíveis, poucos autores apresentam dados completos a respeito do desempenho de suas ferramentas. Tais autores adaptaram as medidas de *Precisão*, *Cobertura* e *Medida F*, que são comumente utilizadas na área de RI. Nas equações apresentadas abaixo, os termos “anotações acuradas” e “anotações não acuradas” referem-se às anotações geradas semi-automaticamente por uma plataforma de anotação semântica, e “anotações manuais” refere-se a todas as anotações geradas por um anotador humano.

$$Precisão = \frac{\text{anotações acuradas}}{\text{anotações acuradas} + \text{anotações não acuradas}} \quad (3.1)$$

$$Cobertura = \frac{\text{anotações acuradas}}{\text{anotações manuais}} \quad (3.2)$$

Medida F é calculada como a média harmônica entre precisão e cobertura⁶. A pesquisa feita por Reeve e Han (2005) agrupa algumas poucas plataformas cujos autores publicaram informações relativas a desempenho, conforme pode ser visto na tabela 3.1. Dentre elas, a ferramenta baseada em aprendizado de máquina com maior desempenho é MnM. Das plataformas baseadas em padrões, MUSE (Maynard, 2003) é a melhor. Entretanto, apesar de ter sido classificada por Reeve e Han (2005) como uma ferramenta de anotação semântica, MUSE é na verdade uma ferramenta para EI, que faz parte da plataforma GATE (Cunningham et al., 2002). O pior desempenho dentre todas as implementações analisadas foi o OntoMat usando PANKOW (24,9%). Por outro lado, quando PANKOW é usado em modo semi-automático, ou seja, sugerindo os cinco conceitos mais prováveis para a anotação para que o usuário escolha, o seu desempenho sobe para 49,56% (Cimiano et al., 2004).

| Plataforma | Categoria | Precisão | Cobertura | Medida F |
|-------------------|------------------|----------|-----------|----------|
| Armadillo | Desc. de padrões | 91.0 | 74.0 | 87.0 |
| KIM | Regras | 86.0 | 82.0 | 84.0 |
| MnM | Indução | 95.0 | 90.0 | n/a |
| MUSE | Regras | 93.5 | 92.3 | 92.9 |
| Ont-O-Mat: PANKOW | Desc. de padrões | 65.0 | 28.2 | 24.9 |
| SemTag | Regras | 82.0 | n/a | n/a |

Tabela 3.1: Comparação do desempenho entre algumas plataformas.

Fonte: Reeve e Han (2005).

⁶Para maiores detalhes, ver seção 4.4.4

3.6.4 Comentário sobre Plataformas para Anotação Semântica

Como pôde ser visto, a anotação semântica (e Web Semântica em geral) constitui uma área de intensa pesquisa, com um muitas ferramentas implementadas e que têm alcançado bons resultados. Todavia, apesar de oferecer várias opções para aplicações que precisem utilizar anotações semânticas, a quantidade e a velocidade com que tais contribuições vêm surgindo dificulta a análise e escolha pela ferramenta mais adequada.

Dentre as ferramentas avaliadas, a Plataforma KIM foi escolhida para ser usada neste trabalho. Apesar de outras ferramentas (tais como Armadillo e MnM) terem um desempenho tão bom ou melhor que esta plataforma, KIM apresenta algumas vantagens em relação às demais, tais como:

- Não utiliza aprendizado de máquina e, portanto, não exige que o usuário treine o sistema para que este possa fazer anotações automaticamente.
- Oferece uma ontologia genérica, que pode ser estendida, bem como uma base de conhecimento que contém milhares de instâncias desta ontologia, que por sua vez pode ser também estendida.
- Além da anotação automática, oferece também funcionalidades para indexação e busca de documentos com base nas anotações semânticas geradas.

Portanto, KIM cobre vários aspectos relacionados ao tema deste trabalho, sendo uma das ferramentas de base para a implementação do arcabouço proposto.

Capítulo 4

Recuperação de Informação

Este capítulo apresenta uma revisão bibliográfica da área de *Recuperação de Informação (RI)*. São apresentados os conceitos básicos, os modelos para RI, métodos para avaliação de desempenho, bem como novas abordagens que usam ontologias e anotações semânticas para dar suporte à chamada *busca semântica*.

Este capítulo também discute aspectos relativos ao uso de contexto em sistemas de RI. São abordados aspectos de modelagem e representação, bem como as principais técnicas utilizadas e o estado da arte nesta área.

4.1 Introdução

A disciplina de Recuperação de Informação (RI) trata da representação, armazenamento, organização e acesso a itens de informação (Baeza-Yates e Ribeiro-Neto, 1999). Segundo os mesmos autores, a representação e a organização dos itens de informação (documentos) deve permitir o acesso fácil ao tipo de informação na qual um usuário tem interesse. Entretanto, a definição da *necessidade de informação do usuário* não é uma tarefa simples. A principal razão para isto é que as interfaces dos sistemas de RI atuais não permitem perguntas em linguagem natural e, conseqüentemente, o usuário precisa definir a sua consulta na forma de um conjunto de palavras-chave. Deste modo, o usuário pode não expressar exatamente a sua intenção original, principalmente considerando a ambigüidade inerente às linguagens naturais.

Além disso, é preciso diferenciar o conceito de *Recuperação de Dados* do conceito de *Recuperação de Informação*. Um sistema de recuperação de dados – como um banco de dados – lida com dados que têm estrutura e semântica bem definidas, e sua linguagem de consulta visa recuperar todos os objetos que satisfaçam condições precisamente definidas por uma expressão regular ou álgebra relacional. Assim, para um sistema de recuperação de dados, um simples objeto errôneo dentre milhares de objetos recuperados significa uma falta grave. Por outro lado, sistemas de recuperação de informação

lidam com texto em linguagem natural, que nem sempre é bem estruturado e que pode ser ambíguo. Isto implica que os resultados de um sistema de recuperação de informação podem ser imprecisos e sujeitos a erros.

Portanto, para ser efetivo na satisfação da necessidade de informação do usuário, um sistema de RI deve, de algum modo, “interpretar” o conteúdo dos documentos de uma coleção e ordená-los de acordo com um grau de relevância à consulta do usuário. Esta “interpretação” envolve a extração de informações sintáticas e semânticas do texto de um documento e a comparação de tais informações com a necessidade do usuário. A dificuldade não é apenas saber como extrair estas informações, mas também saber como usá-las para definir a *relevância* do documento (Baeza-Yates e Ribeiro-Neto, 1999). Com base nisso, o principal objetivo de um sistema de RI é recuperar todos os documentos que são relevantes a uma dada consulta e ao mesmo tempo recuperar o menor número possível de documentos não relevantes.

Embora seja uma área intensamente pesquisada no passado (décadas de 70 e 80) e tenha uma certa maturidade, a RI interessava principalmente bibliotecários e cientistas de informação. Entretanto, a introdução da Web nos anos 90 e sua posterior popularização geraram o problema de como seus usuários poderiam facilmente localizar as informações contidas nos documentos espalhados nos milhões de *sites* da Web. Tal problema renovou o interesse nas técnicas de RI como soluções promissoras, e fez surgir os chamados *motores* (ou *máquinas*) de busca da Web (*web search engines*).

4.2 Conceitos

Esta seção apresenta alguns conceitos básicos relacionados à área de RI (Baeza-Yates e Ribeiro-Neto, 1999): índice, visão lógica de documentos, operações de texto, e o processo de recuperação de informação.

4.2.1 O Conceito de Índice

Muito antes dos computadores, a informação precisava ser organizada de modo a poder ser recuperada e usada posteriormente. Foi aí que surgiu o conceito de *índice*, que é usado tanto para organizar as informações dentro de um livro (sumário) como para organizar uma coleção de livros. Um índice é uma coleção de palavras ou conceitos organizados hierarquicamente que são associados a ponteiros para a informação (ou documentos) em questão. Com o advento dos computadores, índices são criados automaticamente e são utilizados pelos sistemas de RI.

4.2.2 Visão Lógica dos Documentos e Operações de Texto

Documentos de uma coleção são representados por conjuntos de termos de índice ou palavras-chave. Tal representação provê a chamada *visão*

lógica do documento (Baeza-Yates e Ribeiro-Neto, 1999). Esta representação pode conter todas as palavras do documento ou este conjunto pode ser reduzido através de operações chamadas *operações de texto*, que podem ser:

- Eliminação das *stopwords* (artigos, preposições e conectivos).
- Redução das palavras, que pode ser feita através de técnicas de *stemming* ou de *lematização* (Gonzalez et al., 2006):
 - *Stemming* é um processo que reduz uma palavra ao seu radical (*stem*). Por exemplo, “livro” e “livros” têm o mesmo *stem*, que é “livr”.
 - A lematização reduz as palavras a uma *forma canônica* correspondente: verbos vão para o infinitivo; substantivos e adjetivos vão para o singular e masculino. Por exemplo, “livro” e “livrinho” têm o mesmo lema: “livro”.

A principal diferença entre os resultados de *stemming* e de lematização é que, no primeiro caso, palavras de diferentes categorias morfológicas podem ter o mesmo *stem*, enquanto que, na lematização, a categoria morfológica é mantida. Por exemplo, “construiu” e “construções” têm o mesmo *stem*: “constru”. Por outro lado, “construiu” tem como lema “construir”, e “construções” tem como lema “construção”. Segundo Braschler e Ripplinger (2004), tais operações têm grande importância nos sistemas de RI, servindo tanto para reduzir o tamanho do arquivo de índice, quanto para tornar a recuperação independente da forma com que o termo ocorre na consulta.

- Identificação de grupos de nomes, que elimina adjetivos, advérbios e verbos.

Sistemas de RI nem sempre utilizam todas estas operações. Portanto, dependendo das operações utilizadas, a visão lógica de um documento pode variar desde uma que contenha todas as palavras até outra que seja simplesmente uma lista de termos. Além disso, sistemas de RI podem considerar também a estrutura interna que um documento pode ter: capítulos, seções, subseções, entre outras.

4.2.3 O Processo de Recuperação de Informação

O processo de recuperação de informação poder ser descrito com base na arquitetura genérica de um sistema de RI, que pode ser vista na figura 4.1. Em primeiro lugar, antes mesmo do processo de recuperação ser iniciado, é preciso definir a coleção de documentos, que é armazenada em uma base de dados, bem como as operações de texto que serão realizadas nos documentos.

Uma vez obtida a representação lógica dos documentos, um índice para tais documentos é gerado. Com os documentos indexados, o processo de recuperação pode ser iniciado. O usuário especifica a sua *necessidade de informação* que é então processada e transformada pelas mesmas operações de texto aplicadas aos documentos (conforme visto na seção anterior). Depois, *operações de consulta* são aplicadas de modo a se obter uma representação do sistema para a consulta. A consulta é então processada para obter os *documentos recuperados*. Esta busca é feita rapidamente graças ao índice gerado previamente.

Antes de serem enviados ao usuário, os documentos recuperados são ordenados de acordo com uma *probabilidade* de relevância. Tal ordenação é definida pelo modelo adotado pelo sistema de RI, conforme será visto na seção 4.3. O usuário então examina o conjunto de documentos recuperados e pode selecionar os que são de maior interesse e iniciar um ciclo de realimentação de relevância. Nesse ciclo, o sistema utiliza os documentos selecionados pelo usuário para modificar a formulação da consulta, em um processo chamado *expansão da consulta*. É esperado que esta consulta modificada represente melhor a real necessidade do usuário (mais detalhes sobre esse processo são apresentados na seção 4.7.5).

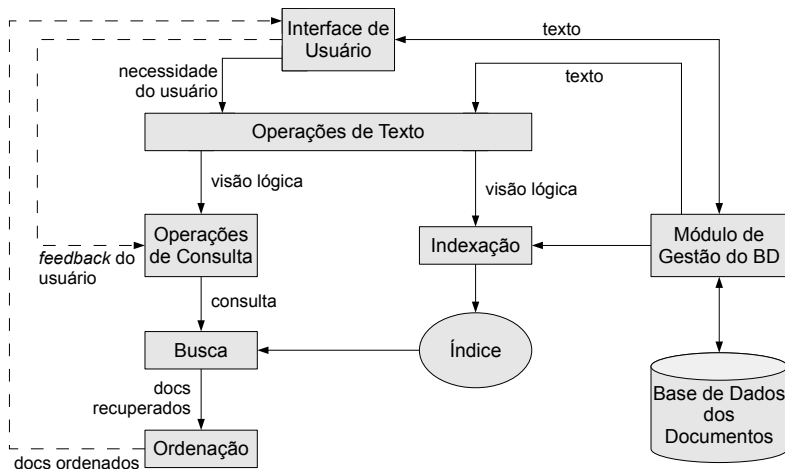


Figura 4.1: Arquitetura conceitual de um sistema de RI.

Fonte: Baeza-Yates e Ribeiro-Neto (1999).

É preciso ressaltar também que no caso dos motores de busca da Web, os documentos não estão centralizados em uma base de dados, mas sim espalhados pela Web. O motor de busca usa então o chamado *Web crawler* (também conhecido como *spider*) que é um navegador automático responsá-

vel por vasculhar os *links* que encontrar e devolver o conteúdo das páginas que são processadas para serem indexadas.

4.3 Modelos para RI

Os modelos de RI são usados para representar os documentos de alguma forma, bem como definir um algoritmo para a recuperação e ordenação de documentos de acordo com a sua relevância. Os modelos de RI são classificados em três grupos:

1. **Baseados em Teoria de Conjuntos:** representam os documentos como conjuntos de termos. A recuperação é feita com base em operações sobre conjuntos. Alguns modelos desta categoria são: o *modelo booleano padrão* e o *modelo booleano estendido*.
 - *Modelo booleano:* modelo de recuperação simples, baseado na teoria de conjuntos e álgebra booleana. A consulta é uma expressão booleana de termos ligados pelos conectivos “não”, “e” e “ou”. Aplicando a teoria de conjuntos, o resultado da consulta contém os documentos que satisfaçam a expressão definida na consulta. A desvantagem deste modelo é que a relevância dos documentos é binária, isto é, o documento é relevante ou não. Assim, não há uma ordem de relevância nos resultados. Por esta razão, o modelo booleano é considerado mais um modelo para recuperação de dados do que um modelo para recuperação de informação (Baeza-Yates e Ribeiro-Neto, 1999).
2. **Algébricos:** representam os documentos e as consultas normalmente como vetores, matrizes ou tuplas e operações algébricas são usadas para calcular a ordem de relevância dos documentos. Esta categoria inclui: o *modelo vetorial*, o *modelo vetorial generalizado*, o *modelo booleano estendido*¹ e a *indexação da semântica latente*².
 - *Modelo vetorial:* modelo definido por Salton et al. (1975), onde tanto documentos quanto consultas são definidos como vetores. Estes vetores contêm pesos que representam os termos indexados. Cada peso é calculado em função da frequência com que ocorre no documento e do número de documentos onde aparecem. O grau de similaridade de um documento em relação a uma consulta é definido por uma distância algébrica entre os seus vetores

¹O *modelo booleano estendido* é um modelo híbrido que relaxa a álgebra booleana, que é interpretada em termos de distância algébrica, tal com é feito no modelo vetorial (Salton et al., 1983).

²Do inglês “*latent semantic indexing*”.

representantes, como por exemplo, o *cosse*no do ângulo entre os vetores. Isso permite, ao contrário do modelo booleano, que documentos possam ser ordenados de acordo com esta distância, ou seja, de acordo com a sua relevância.

3. **Probabilísticos:** tratam o processo de RI como experimento aleatório de vários estágios e similaridades são representadas por probabilidades. Exemplos deste tipo são: o *modelo da independência binária*, o *modelo de relevância probabilística* e a *rede de inferência*.

Além dos modelos clássicos, é importante mencionar o algoritmo *PageRank* (Brin e Page, 1998), utilizado pela máquina de busca *Google*. Este algoritmo leva em consideração a estrutura hipertextual da Web e calcula a relevância de documentos com base no número de documentos que apontam (contêm um *link*) para ele.

4.4 Avaliação de Sistemas de RI

A avaliação de um sistema depende do seu tipo e seus objetivos. De uma maneira geral, espera-se que qualquer sistema de *software* proveja a funcionalidade para a qual ele foi concebido. Outro tipo de análise consiste na utilização errônea do sistema, visando avaliar o seu comportamento em situações excepcionais e detectar possíveis falhas (Baeza-Yates e Ribeiro-Neto, 1999).

Outras medidas comumente utilizadas na avaliação de sistemas são *tempo* e *espaço*. Em outras palavras, quanto menor o tempo de resposta e menor o espaço ocupado, melhor o sistema. Em geral, há uma relação de custo-benefício entre essas duas medidas, onde um sempre pode ser melhorado em detrimento do outro. Em se tratando de sistemas para RI, outras medidas são de particular interesse. Ao contrário de sistemas de recuperação de dados, os documentos recuperados por sistemas de RI não são respostas exatas e devem ser classificados (ordenados) de acordo com sua *relevância* à consulta. Tal classificação, inexistente em sistemas de recuperação de dados, desempenha um papel central no processo de RI. Em suma, a avaliação de um sistema de RI consiste na avaliação do quão *preciso* é o conjunto de respostas desse sistema.

Tal avaliação é normalmente baseada numa coleção de referência e na obtenção de algumas medidas. A coleção de referência (ou de teste) consiste de uma coleção de documentos, um conjunto de consultas e um conjunto de documentos relevantes para cada consulta. Esse último elemento é definido por especialistas. Assim, para cada sistema *S* a ser avaliado, essa medida de avaliação quantifica a *similaridade* entre o conjunto de documentos recuperados por *S* e o conjunto de documentos relevantes definidos por especialistas.

Desse modo, é possível estimar a qualidade da estratégia de busca adotada por S .

Outro aspecto que deve ser considerado nesse tipo de avaliação é como a *tarefa de busca* será executada. Por exemplo, a tarefa pode simplesmente consistir de uma série de (ou uma única) consultas processadas em lote, ou seja, o usuário submete uma ou mais consultas e recebe os resultados. Por outro lado, a tarefa de busca pode ser uma sessão interativa completa, onde o usuário constrói sua consulta através de uma série de passos interagindo com o sistema. Outra alternativa consiste na combinação desses dois tipos.

Tarefas em lote e interativas são distintas e, portanto, a forma de avaliação é também distinta. Numa sessão interativa, o esforço do usuário em termos de interação com o sistema (eventos de interface) e tempo de uso são aspectos críticos que devem ser observados e mensurados. Numa sessão em lote, por outro lado, nenhum desses aspectos é tão importante quanto a *qualidade* dos resultados obtidos.

Neste trabalho, o tipo de atividade a ser avaliada é uma *sessão em lote*. A razão para tal escolha está na maior facilidade na condução dos experimentos, pois não é preciso mensurar os eventos de interface do usuário. Além disso, no protótipo desenvolvido não foi priorizada a interface gráfica do sistema, não sendo portanto adequado a esse tipo de análise. O foco, portanto, será nas medidas de qualidade dos resultados oferecidos pelo sistema.

4.4.1 Medidas Básicas: Precisão e Cobertura

Considere uma consulta C (para uma dada coleção de referência) e o seu conjunto R de documentos relevantes, sendo $|R|$ o número de documentos desse conjunto. Assuma que um dado sistema de RI que está sendo avaliado processa a consulta C e gera o conjunto A de documentos resultantes, sendo $|A|$ o número de documentos desse conjunto. Finalmente, definimos $|Ra|$ como o número de documentos na intersecção entre R e A . A figura 4.2 ilustra esses conjuntos.

Com base nas definições apresentadas, *precisão* e *cobertura* para uma dada consulta C são definidas como:

- **Precisão:** é a fração dos documentos recuperados (conjunto A) que é relevante. Por exemplo, se o sistema recupera 8 documentos para uma dada consulta, onde 4 deles são relevantes, a precisão do sistema nessa consulta é 0.5 ou 50%.

$$\text{Precisão} = \frac{|Ra|}{|A|} \quad (4.1)$$

- **Cobertura:** fração dos documentos relevantes da coleção (ou seja, em R) que foi recuperada.

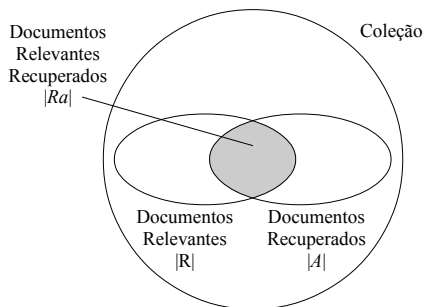


Figura 4.2: Documentos relevantes para uma dada consulta.

Fonte: adaptado de Baeza-Yates e Ribeiro-Neto (1999).

$$Cobertura = \frac{|Ra|}{|R|} \quad (4.2)$$

Em outras palavras, essas medidas são calculadas com base no número de documentos relevantes recuperados em relação ao número total de documentos recuperados (precisão) ou ao número total de documentos relevantes na coleção (cobertura).

É importante salientar que tais medidas devem ser analisadas em conjunto, pois um sistema com uma boa precisão não necessariamente terá uma boa cobertura, e vice-versa. De fato, a observação isolada dessas duas medidas pode levar a uma interpretação incorreta a respeito do sistema sendo avaliado. Por exemplo, para uma dada consulta, basta o sistema recuperar apenas um único documento que é relevante para obter 100% de precisão. Por outro lado, a recuperação de todos os documentos levará a uma cobertura de 100%. Logo, um bom sistema de RI busca o equilíbrio entre essas duas medidas. Uma alternativa para esse tipo de análise consiste na utilização de outras medidas, que combinam precisão e cobertura, conforme será mostrado mais adiante (seção 4.4.4).

Outro aspecto a ser destacado é o fato de que a simples observação da precisão e cobertura nem sempre será o suficiente para indicar a qualidade de um dado sistema, ou mesmo para a comparação entre dois sistemas. A razão para isso é que tais medidas representam um valor baseado em conjuntos, ou seja, assume-se que todos os documentos recuperados foram examinados pelo usuário. No entanto, sistemas de RI apresentam os documentos recuperados na forma de uma lista ordenada pelo grau de relevância à consulta. O usuário então examina essa lista seguindo a ordem estabelecida. A análise desse aspecto é avaliada conforme uma *curva de precisão x cobertura*, cuja definição é explicada na seção a seguir.

4.4.2 Curva de Precisão x Cobertura

A definição de tal gráfico parte do princípio que as medidas de cobertura e precisão variam à medida que o usuário percorre a lista de documentos recuperados. Desse modo, para cada documento recuperado, é feito o cálculo dessas medidas de acordo com a situação parcial do percurso do usuário. A tendência desse tipo de curva é a precisão decrescer à medida que a cobertura aumenta de valor.

A construção dessa curva é explicada com base num exemplo genérico, adaptado de Baeza-Yates e Ribeiro-Neto (1999), conforme segue. Dada uma consulta C , o conjunto R_c , contendo os documentos relevantes para C é definido abaixo. Esse conjunto contém 10 documentos, selecionados por especialistas.

$$R_c = \{d_3, d_5, d_9, d_{25}, d_{39}, d_{44}, d_{56}, d_{71}, d_{89}, d_{123}\}$$

Considerando um dado sistema a ser avaliado, o processamento de consulta C retorna o seguinte conjunto de documentos:

- | | | |
|----------------|----------------|---------------|
| 1. d_{123} • | 6. d_9 • | 11. d_{38} |
| 2. d_{84} | 7. d_{511} | 12. d_{48} |
| 3. d_{56} • | 8. d_{129} | 13. d_{250} |
| 4. d_6 | 9. d_{187} | 14. d_{113} |
| 5. d_3 | 10. d_{25} • | 15. d_3 • |

Os documentos relevantes à consulta C estão destacados com um marcador à direita do seu número. Examinando essa lista a partir do primeiro documento (tal como o usuário faria), é possível observar os seguintes pontos. Primeiramente, o documento d_{123} , classificado como número um, é relevante. Além disso, ele corresponde a 10% de todos os documentos relevantes (do conjunto R_c). Portanto, diz-se que a precisão é de 100% a uma cobertura de 10%. Em seguida, o documento d_{56} , classificado em terceiro lugar, é o próximo documento relevante. Nesse ponto, é dito que a precisão é de aproximadamente 66% (dois de três documentos são relevantes) a uma cobertura de 20% (dois de dez documentos relevantes foram examinados). Continuando essa análise nos demais documentos, é possível desenhar uma *curva de precisão x cobertura*, tal como ilustrado na figura 4.3. A precisão em níveis de cobertura acima de 50% cai para zero, pois nem todos os documentos relevantes foram recuperados.

Esse tipo de curva é normalmente baseada em 11 níveis de cobertura padrão, que são 0%, 10%, 20%, ..., 100%. Como os níveis de cobertura para cada consulta podem ser diferentes dos 11 níveis padrão, a utilização de um procedimento de interpolação se faz necessária. Tal procedimento funciona

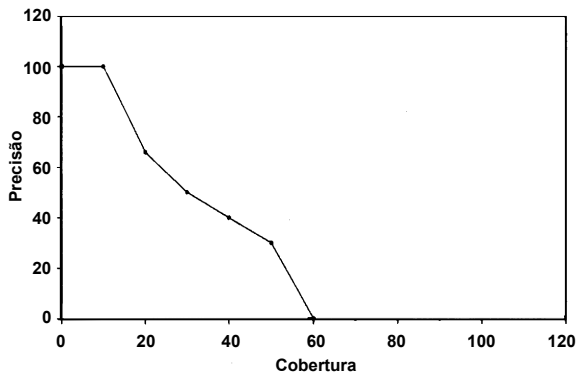


Figura 4.3: Precisão nos 11 níveis de cobertura padrão.
 Fonte: adaptado de Baeza-Yates e Ribeiro-Neto (1999).

da seguinte maneira. Seja $r_j, j \in \{0, 1, 2, \dots, 10\}$, uma referência ao j -ésimo nível de cobertura padrão (por exemplo, r_5 representa o nível a 50%), e $P(r_j)$ a precisão nesse nível, temos:

$$P(r_j) = \max_{r_j \leq r \leq r_{j+1}} P(r) \quad (4.3)$$

que diz que a precisão interpolada no j -ésimo nível padrão é a maior precisão conhecida em qualquer nível de cobertura entre o j -ésimo nível e o $(j+1)$ -ésimo nível.

O procedimento apresentado leva em consideração uma única consulta. Porém, sistemas de RI são normalmente avaliados através da análise de várias consultas distintas. Desse modo, para cada consulta uma curva distinta é gerada. Para avaliar o desempenho de um sistema dentre todas as consultas, é preciso calcular a média dos pontos de precisão conforme a equação 4.4.

$$\bar{P}(r) = \sum_{i=1}^{N_c} \frac{P_i(r)}{N_c} \quad (4.4)$$

onde $\bar{P}(r)$ é a precisão média no nível de cobertura r , N_c é o número de consultas e $P_i(r)$ é a precisão no nível de cobertura r para a i -ésima consulta.

A curva média resultante é normalmente utilizada para comparar o desempenho entre dois ou mais sistemas, conforme mostra a figura 4.4. Nesse exemplo, um sistema tem precisão mais alta nos níveis mais baixos de cobertura, enquanto o segundo sistema é superior em níveis mais altos de cobertura.

Segundo Baeza-Yates e Ribeiro-Neto (1999), tais gráficos representam a estratégia padrão para a avaliação de sistemas de RI e são amplamente usados na literatura da área. Esse tipo de gráfico permite avaliar quantita-

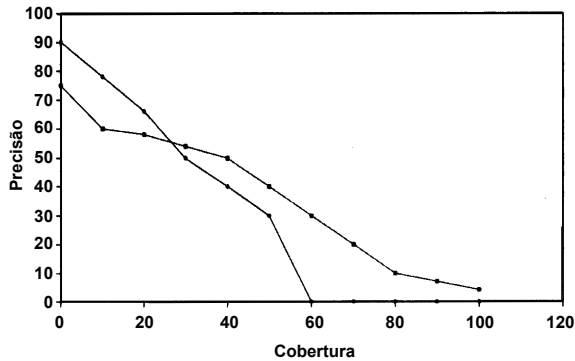


Figura 4.4: Curvas médias de precisão x cobertura entre dois sistemas.

Fonte: adaptado de Baeza-Yates e Ribeiro-Neto (1999).

tivamente o desempenho de um sistema de maneira simples e intuitiva, com seus resultados combinados numa única curva.

4.4.3 Valores Detalhados

De maneira complementar ao gráfico de precisão x cobertura, é possível fazer uma análise mais detalhada levando-se em consideração as consultas individualmente. A razão principal para isso é que a média da precisão de muitas consultas pode ocultar anomalias importantes no sistema sob análise. Além disso, quando da comparação entre dois sistemas, pode-se estar interessado em investigar se um sistema é superior ao outro em *cada* consulta da coleção de teste. Esta segunda situação também pode ser facilmente ocultada por um valor médio.

Nesses casos, um valor individual de precisão para cada consulta pode ser usado, sendo interpretados como um detalhamento da curva de precisão x cobertura. Algumas dessas medidas são apresentadas a seguir.

Precisão Média para os Documentos Relevantes Observados: a ideia é gerar um valor único através da média dos pontos de precisão obtidos quando cada novo documento relevante é observado na lista de resultados. No exemplo da figura 4.3, os pontos de precisão calculados quando cada novo documento relevante é observado são 1, 0.66, 0.5, 0.4 e 0.3. Portanto, a *precisão média para os documentos relevantes observados* é dada por $(1 + 0.66 + 0.5 + 0.4 + 0.3)/5 = 0.57$. Essa medida favorece sistemas que posicionam documentos relevantes no topo da lista de resultados. Todavia, mesmo tendo um bom valor para esta medida, um sistema ainda pode apresentar um desempenho fraco em termos de cobertura geral.

Precisão-R: este valor é definido como a precisão da R -ésima posição

na lista de resultados, onde R é o número total de documentos relevantes para a consulta (ou seja, o número de documentos no conjunto R_c). Para o exemplo da figura 4.3, o valor da precisão-R é 0.4, pois $R=10$ e há quatro documentos relevantes entre os primeiros 10 documentos recuperados. Esta medida é um parâmetro útil para observar o comportamento do sistema para cada consulta individual. Adicionalmente, é possível também calcular a precisão-R média de todas as consultas. Porém, um único valor que resume o comportamento de um sistema em relação a várias consultas pode ser bem impreciso.

Histogramas de Precisão: as medidas de precisão-R para cada consulta podem ser usadas para comparar o desempenho entre dois sistemas da seguinte forma. Sejam $RP_A(i)$ e $RP_B(i)$ os valores de precisão-R para os sistemas A e B para a i -ésima consulta, Seja a diferença entre esses valores definida na equação 4.5:

$$RP_{A/B}(i) = RP_A(i) - RP_B(i) \quad (4.5)$$

Um valor para $RP_{A/B}(i)$ igual a zero indica que ambos os sistemas tem desempenho equivalente (em termos de precisão-R) para a i -ésima consulta. Um valor positivo indica que A tem desempenho melhor e, no caso de um valor negativo, B é superior.

Tabela de Estatísticas Gerais: outras medidas podem ser armazenadas em uma tabela que oferece um resumo estatístico referente ao conjunto de todas as consultas executadas. Esses valores incluem, por exemplo, o número de consultas, o número total de documentos recuperados por todas as consultas, o número total de documentos relevantes que foram efetivamente recuperados em todas as consultas, o número total de documentos relevantes que deveriam ter sido recuperados em todas as consultas, entre outros.

4.4.4 Outras Medidas

Apesar da popularidade das medidas de precisão e cobertura, em alguns casos outras medidas são mais apropriadas para a avaliação de sistemas de RI. Nesse sentido, medidas alternativas têm sido propostas ao longo dos anos, com destaque para a *média harmônica* e a *medida E*.

- **Média Harmônica:** também conhecida como medida F (*F measure*), é uma medida que combina precisão e cobertura. É calculada conforme a equação 4.6:

$$F(j) = \frac{2}{\frac{1}{r(j)} + \frac{1}{P(j)}} \quad (4.6)$$

onde $r(j)$ é a cobertura para o j -ésimo documento na lista de resultados, $P(j)$ é a sua precisão, e $F(j)$ é a média harmônica de $r(j)$ e $P(j)$.

A função F assume valores no intervalo $[0,1]$, onde 0 significa que nenhum documento relevante foi recuperado e 1 indica que todos os documentos recuperados são relevantes. Por combinar as duas medidas básicas, a maximização do valor de F pode ser interpretada como uma tentativa para se encontrar o melhor compromisso entre cobertura e precisão.

- **Medida E:** proposta por Van Rijsbergen (1979), esta medida permite que o usuário possa especificar se está mais interessado na precisão ou na cobertura. Tal medida é definida segundo a equação 4.7:

$$E(j) = 1 - \frac{1 + b^2}{\frac{b^2}{r(j)} + \frac{1}{P(j)}} \quad (4.7)$$

onde $r(j)$ e $P(j)$ são, respectivamente, a cobertura e a precisão do j -ésimo documento recuperado, $E(j)$ é a medida E relativa a $r(j)$ e $P(j)$, e b é um parâmetro especificado pelo usuário que reflete a importância relativa entre cobertura e precisão. Caso b tenha valores maiores que 1, o usuário está mais interessado na precisão; valores menores que 1 indicam o interesse pela cobertura; para $b = 1$, a medida E funciona como o complemento da medida F.

4.4.5 Coleções de Referência

Conforme mencionado, a avaliação de sistemas de RI é feita com base numa coleção de referência. Existem algumas coleções que vêm sendo usadas ao longo dos anos por pesquisadores da área. A maior e mais largamente utilizada é a coleção da conferência TREC (Baeza-Yates e Ribeiro-Neto, 1999).

A *Text REtrieval Conference* (TREC)³ é uma conferência anual (a primeira edição ocorreu em 1992) dedicada à experimentação de sistemas de RI usando uma grande coleção contendo mais de um milhão de documentos. Seu propósito é dar suporte à pesquisa feita pela comunidade de RI provendo a infraestrutura necessária para a avaliação de metodologias para recuperação de textos (TREC, 2004). A conferência conta com a participação de dezenas de grupos de pesquisa de diversos países, que podem comparar seus sistemas com base em um conjunto padronizado de experimentos.

A coleção da TREC é composta de três partes: documentos (escritos em inglês e codificados em SGML); requisições de informação (definidas em linguagem natural) para a definição das consultas; e um conjunto de documentos relevantes para cada requisição (Baeza-Yates e Ribeiro-Neto, 1999). Além disso, há uma série de análises padronizadas que levam em

³<http://trec.nist.gov/>

consideração as medidas para RI, muitas das quais foram apresentadas nesta seção.

A maior limitação para a utilização da coleção da TREC é que o seu tamanho faz com o processo de experimentação seja extenso, além de necessitar de uma etapa de preparação também longa. Além disso, é preciso pagar uma taxa para a sua aquisição.

Além da TREC, outras coleções de testes foram definidas. Nos estudos experimentais feitos por (Fox, 1983) *apud* (Baeza-Yates e Ribeiro-Neto, 1999), cinco pequenas coleções de teste foram desenvolvidas: ADI (documentos relacionados à ciência da informação), CACM, INSPEC (resumos relacionados à eletrônica, computação e física), ISI, e *Medlars* (artigos médicos).

Uma vantagem dessas coleções menores é que elas são gratuitas e estão disponíveis publicamente. A tabela 4.1 apresenta as coleções disponíveis no *site* do Grupo de Pesquisa em RI da Universidade de Glasgow⁴:

| Coleção | Documentos | Consultas | Tamanho (Mb) |
|-----------|------------|-----------|--------------|
| LISA | 5872 | 35 | 3.4 |
| NPL | 11429 | 93 | 3.1 |
| CACM | 3204 | 64 | 2.2 |
| CISI | 1460 | 112 | 2.2 |
| Cranfield | 1400 | 225 | 1.6 |
| Time | 423 | 83 | 1.5 |
| Medline | 1033 | 30 | 1.1 |
| ADI | 82 | 35 | 0.04 |

Tabela 4.1: Coleções de teste disponíveis publicamente (Universidade de Glasgow).

Fonte: Glasgow (2007).

4.4.6 Comentários a Respeito das Coleções de Teste

Apesar da área de RCs desenvolver-se rapidamente, a coleta de dados de uma RC operacional ainda é uma tarefa muito difícil, o que requer o desenvolvimento de novas abordagens e infraestruturas de suporte (Camarinha-Matos e Afsarmanesh, 2006). Por outro lado, uma ideia que se teve ao longo do trabalho foi a utilização e/ou adaptação de alguma das coleções de teste aqui apresentadas.

No entanto, alguns problemas foram detectados. Em primeiro lugar, os assuntos cobertos por essas coleções não estão relacionados tanto à ontologia quanto ao modelo de contexto definido neste trabalho. Além disso, e mais importante, por não levarem em consideração o conceito de contexto adotado neste trabalho, o conjunto de documentos relevantes para cada consulta tende

⁴<http://ir.dcs.gla.ac.uk/>

a ser diferente, pois as consultas são livres de contexto. De fato, a falta de coleções de teste é um problema comum para a área de *busca empresarial*, e particularmente nos trabalhos envolvidos com busca usando contexto (Belkin, 2008). Portanto, houve a necessidade de se definir uma coleção de teste específica para este tipo de trabalho.

4.5 Problemas relacionados à RI

Por serem baseados em palavras-chave, sistemas tradicionais de RI apresentam uma série de limitações, tal com enumerado por Riloff e Lehnert (1994):

- **Sinonímia:** palavras diferentes podem expressar o mesmo conceito. Por exemplo, “empresa” e “companhia” representam o mesmo conceito. A sinonímia pode reduzir a cobertura do sistema pois, documentos relevantes referindo-se somente a sinônimos de uma palavra usada na consulta podem não ser recuperados.
- **Polissemia:** palavras podem ter vários significados. Por exemplo, o termo “banco” pode significar o móvel que serve de assento ou uma agência bancária. Polissemia pode levar a baixas taxas de precisão, pois documentos irrelevantes podem ser recuperados.
- **Frases:** algumas palavras são boas como termos de indexação somente em frases ou expressões específicas. Por exemplo, na expressão “banco de horas”, as palavras “banco” e “horas” podem ser interpretadas de maneiras diferentes se forem analisadas separadamente.
- **Contexto Local:** algumas palavras e frases são bons termos de indexação apenas em contextos locais específicos. Por exemplo, em uma busca por documentos relacionados a roubo de bancos, a palavra “roubo” não é suficiente; o objeto do roubo deve ser um banco.
- **Contexto Global:** alguns documentos podem não conter qualquer palavra ou expressão que sejam bons termos de indexação. Neste caso, a relevância de um documento pode depender do contexto inteiro de uma sentença, parágrafo, ou mesmo de todo o documento. Por exemplo, a sentença “um homem armado pegou o dinheiro e fugiu” refere-se a um roubo, embora nenhuma das palavras consiga expressar isso individualmente.

Esses problemas estão incluídos na chamada *barreira da palavra-chave*⁵ (Mauldin, 1991) *apud* (Paz-Trillo et al., 2005). Para superar tais limitações, sistemas de RI necessitam de outras estruturas, baseadas em conceitos

⁵*keyword barrier*

ao invés de apenas palavras-chave. O problema da sinonímia vem sendo tratado com o uso de taxonomias e ontologias (Mahesh et al., 1999; Mihalcea e Moldovan, 2000). Técnicas para eliminação da ambiguidade de palavras⁶ são aplicadas para lidar com o problema da polissemia (Sanderson, 1994, 2000). Já os demais problemas são mais complexos, envolvendo uma análise linguística mais elaborada, bem como novos métodos para indexação e busca (Croft et al., 1991; Jones e Staveley, 1999; Bahle et al., 2002; Williams et al., 2004).

Além do problema da sinonímia e da polissemia, Kalina Bontcheva e Thurlow (2006) argumentam que máquinas de busca convencionais não levam em consideração os vínculos semânticos entre termos. Estes vínculos podem ser obtidos através de formas de representação de conhecimento, tais como ontologias. A próxima seção aborda as novas tendências, onde sistemas de RI, com o auxílio de ontologias e anotações semânticas, tentam lidar com estes problemas.

4.6 RI Baseada de Ontologias

A busca com base em conceitos, ou seja, baseada também no significado ao invés de apenas em sequências de caracteres, foi foco de muita pesquisa na área de RI muito antes de surgir a visão proposta para a Web Semântica (Järvelin et al., 2001). Exemplos dessas iniciativas incluem a Indexação da Semântica Latente (Deerwester et al., 1990) e abordagens para identificação de conceitos nas palavras (Gonzalo et al., 1998).

Com a visão proposta pela Web Semântica, as ontologias passaram a ser a base para o melhor processamento e busca por informações. A RI baseada em ontologias envolve a anotação semântica de documentos e a indexação baseada nestas anotações. Em outras palavras, não é feita apenas a indexação tradicional (com palavras-chave), mas também a indexação semântica (usando conceitos e instâncias de ontologias).

Além disso, a ontologia serve também para a definição da consulta, ou seja, a consulta passa a ser uma expressão formal com base nos conceitos da ontologia. Em paralelo às linguagens para a representação de ontologias, há algumas linguagens para a definição de consultas semânticas: por exemplo RDQL (Seaborne, 2004), RQL (Karvounarakis et al., 2002) e SPARQL (Prud'hommeaux e Seaborne, 2008). Por exemplo, considerando uma ontologia que tenha as classes *Organização*, *Lugar*, e uma relação entre essas classes chamada *localizadaEm*, que relaciona uma organização a um lugar. Uma consulta semântica que leva em conta esta relação pode ser escrita, em linguagem natural como: “*Companhias de TI localizadas na Europa*”. Em RDQL, a consulta ficaria codificada da seguinte forma:

⁶do inglês “*word sense disambiguation*”.

```
SELECT ?company
WHERE (?company <rdf:type> <kb:IT_Company>)
      (?company <kb:locatedIn> <kb:Europe>)
```

É importante destacar também que, do ponto de vista do usuário, tal consulta pode ser construída de diversas maneiras. Ela pode ser gerada a partir de uma consulta baseada em palavras-chave, tal como em (Guha et al., 2003), por uma consulta em linguagem natural (Contreras et al., 2004), ou com base numa interface rica que permite que o usuário selecione as classes da ontologia e defina restrições e valores às suas propriedades (Maedche et al., 2003b; Kiryakov et al., 2004).

A execução do motor de busca semântico é feita da seguinte forma. A consulta semântica é processada e suas restrições são aplicadas sobre uma base de conhecimento (instâncias da ontologia). As instâncias que satisfazem a consulta são então retornadas. Os documentos que foram anotados com as instâncias retornadas no passo anterior são selecionados, ordenados e apresentados ao usuário. A visão geral desse processo é ilustrada na figura 4.5.

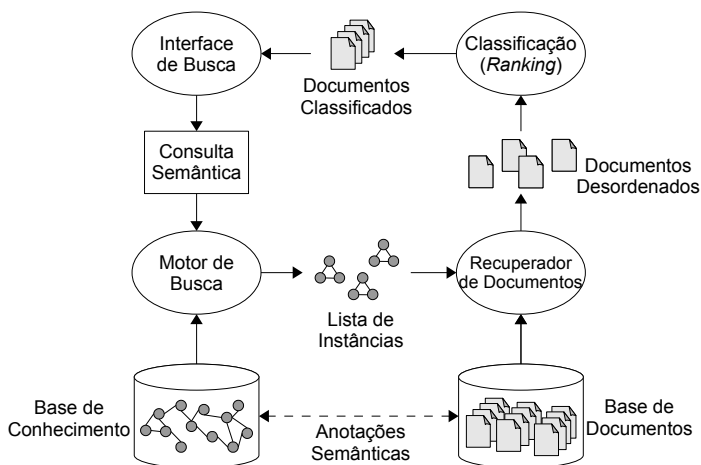


Figura 4.5: Busca baseada em ontologia.
Fonte: baseado em Castells et al. (2007).

Um aspecto importante de ser ressaltado é a necessidade de uma base de conhecimento com uma quantidade razoável de instâncias, aspecto que tem impacto direto na qualidade dos resultados da busca. Castells et al. (2007) ilustra casos onde a busca semântica chega a ser pior do que a busca com palavras-chave pelo fato de haver poucas instâncias quando da execução de determinada consulta. Outro problema tem a ver com a qualidade das

anotações geradas, que em certas situações não conseguem lidar com ambiguidades.

4.6.1 Estado da Arte

TAP (Guha et al., 2003) apresentam um modelo onde documentos e conceitos são nós tratados da mesma forma dentro de uma rede semântica. Nesse sentido, não há a separação explícita entre conteúdo e metadados, tal como feito em outros trabalhos correlatos (Kiryakov et al., 2004; Castells et al., 2007). Um dos focos desse trabalho é a apresentação dos resultados, que são expandidos a partir de dados dos nós vizinhos.

Mayfield e Finin (2003) oferecem outro sistema de busca usando ontologia. Ele combina técnicas baseadas em ontologia e em recuperação de texto num mecanismo de inferência visando a criação de novas anotações semânticas.

Bonino et al. (2004) apresentam um modelo para busca semântica com base no modelo vetorial, que foi modificado para lidar com conceitos de uma ontologia, ao invés de palavras. É proposto um método automático para expansão da consulta, onde a busca é refinada com base na navegação na hierarquia de conceitos da ontologia. O método proposto realiza refinamentos sucessivos na consulta original, onde conceitos mais gerais ou mais específicos substituem os utilizados na iteração anterior, de modo a ajustar os parâmetros da consulta. Este algoritmo faz parte de uma plataforma para anotação e busca semântica chamada **DOSE** (*Distributed Open Semantic Platform*) (Bonino et al., 2003; DOSE, 2007).

Kiryakov et al. (2003, 2004); Popov et al. (2003) apresentam a Plataforma **KIM** que provê *Application Programming Interfaces* (APIs) e ferramentas para anotação semântica automática, indexação e busca semântica de documentos. Esta plataforma foi apresentada anteriormente na seção sobre anotações semânticas. KIM utiliza como base ferramentas de código aberto que são bem sucedidas nas suas respectivas áreas: **Sesame** (Broekstra et al., 2002; Sesame, 2007), como repositório de ontologias; **GATE** (Cunningham et al., 2002; GATE, 2007), plataforma para processamento de texto utilizada para a geração das anotações semânticas; e **Lucene** (Lucene, 2007), máquina de busca baseada em palavras-chave que foi modificada para fazer buscas baseadas em conceitos de ontologias. KIM também oferece uma ontologia genérica e extensível (chamada *PROTON*) e uma base de conhecimento que conta com cerca de 80 mil entidades. KIM está sendo utilizado por outras plataformas para busca semântica, tal como **Squirrel** (Duke et al., 2007).

Squiggle (Squiggle, 2007) é um arcabouço cujo propósito é servir de base para o desenvolvimento de máquinas de buscas semânticas. Tal como a plataforma KIM, Squiggle utiliza o **Sesame** como repositório de ontologias e o **Lucene** como máquina de busca. Squiggle não é uma máquina de busca

pronta, mas permite que usuários personalizem sua própria máquina de busca utilizando uma ontologia particular.

Castells et al. (2007) propõem uma adaptação do modelo vetorial para lidar com a busca baseada em ontologia. A abordagem é bastante similar à plataforma KIM, mas seu foco é no algoritmo para a ordenação dos documentos usando um esquema de indexação semântica baseado em pesos associados às anotações semânticas. Além disso, a abordagem procura combinar a busca semântica com a busca textual para a melhoria dos resultados.

Outro tipo de aplicação que realiza busca semântica com base em ontologia é conhecido como *portal semântico* (Maedche et al., 2003a; Contreras et al., 2004). Tais aplicações oferecem funcionalidades simples de busca, mas retornando apenas instâncias de ontologia ao invés de documentos.

Dentre os trabalhos que usam ontologias para busca em aplicações particulares, podemos citar Paz-Trillo et al. (2005), que descrevem o **OnAIR**, um sistema de RI com suporte de ontologias utilizado para busca de vídeos. Müller et al. (2006) apresentam uma abordagem para anotação semântica, indexação e busca em *wikis* corporativos. Köhler et al. (2006) apresentam um protótipo para a definição de mapeamentos entre ontologias, bem como implementa métodos para indexação e busca baseada em ontologias.

Conforme foi dito na seção 3.6.4, a Plataforma KIM foi escolhida para ser usada como base para a perspectiva semântica do arcabouço proposto. Esta plataforma oferece, de forma integrada, funcionalidades (na forma de APIs) tanto para anotação semântica automática quanto para indexação e busca. Além disso, oferece uma ontologia a uma base de conhecimento que podem ser adaptadas para um domínio em particular.

4.7 RI Baseada em Contexto

Segundo Mylonas et al. (2008), a noção de contexto já foi reconhecida há tempos como tendo importância fundamental em uma variedade de áreas, tais como computação móvel e pervasiva, linguística computacional, análise automática de imagens e recuperação de informação. Apesar da diversidade de áreas, todas estão relacionadas ao tópico *sistemas sensíveis ao contexto*, onde um modelo representa o contexto do usuário e o contexto então é utilizado para fazer adaptações no sistema (Baldauf et al., 2007).

O foco deste trabalho é no uso do contexto para a RI, ou seja, como ele pode ser usado para melhorar os resultados da busca. São cobertos aspectos de representação, captura, bem como são apresentadas as principais abordagens utilizadas para a customização da busca.

4.7.1 Definição de Contexto

Conforme mencionado, há várias definições encontradas na literatura para o termo contexto, tal como definido por Dey (2001): qualquer infor-

mação que pode ser usada para caracterizar a situação de uma entidade que é considerada relevante para a interação entre um usuário e uma aplicação. Além disso, o contexto exerce algum impacto, alguma transformação no ambiente ou aplicação sendo utilizada, tal como mencionado por Bolchini et al. (2007): contexto normalmente tem um impacto significativo na forma como humanos (ou máquinas) atuam e como interpretam coisas; além disso, uma mudança no contexto causa uma transformação na experiência que está sendo vivenciada.

No entanto, tais definições apenas dão uma ideia geral do que é contexto e sua utilidade. Quando avaliado dentro de um domínio ou aplicação mais específica, o termo “contexto” pode ser interpretado de diferentes formas, pode modelar diferentes aspectos e pode ter um impacto igualmente diverso. De acordo com a análise da literatura, o contexto pode representar diferentes aspectos, dependendo dos requisitos da aplicação ou problema:

1. **Condições Ambientais / Situação Dinâmica:**

- Aspectos físicos, como localização, tempo e dispositivo (Fahy e Clarke, 2004).
- Aspectos organizacionais tais como processos, tarefas e papéis (Hawking, 2004).

2. **Perfil Individual:** são os tópicos de interesse do usuário (Tamine-Lechani et al., 2007). Segundo Teevan et al. (2005) uma vasta gama de atividades implícitas do usuário têm sido propostas como fontes de informação para a melhoria da busca na Web, incluindo o histórico de buscas, histórico de navegação, comunidades na Web, e interações em aplicações com interfaces de usuário ricas.

3. **Linguagem Natural / Semântica e Estrutura do Documento:** esta noção é particularmente usada na área de RI, onde contexto não é associado ao usuário - como nas definições anteriores - mas a termos de documentos.

- “Distância” semântica entre palavras (contexto pode ser visto como o significado de uma palavra, portanto palavras mais próximas em termos de semântica estão em contextos similares).
- O contexto de um termo também pode ser um conjunto de palavras na sua vizinhança (Bhogal et al., 2007).
- O contexto pode ser os *hyperlinks* entre páginas da Web (Yu et al., 2001).

Conforme apresentado na introdução, este trabalho considera a primeira perspectiva. Portanto, contexto significa a situação dinâmica de um usuário no ambiente da RC, particularmente aspectos organizacionais.

Há três aspectos principais que devem ser considerados quando se trata de aplicações que usam contexto: (i) Como capturar o contexto?; (ii) Como representar o contexto?; e (iii) Como usar o contexto na busca de informação? Tais aspectos são discutidos nas seções a seguir.

4.7.2 Captura do Contexto

A captura do contexto depende da sua natureza, como por exemplo:

- Aspectos físicos: são capturados por sensores (por exemplo, a temperatura e posicionamento).
- Aspectos Organizacionais / Funcionais: mecanismos de autorização de sistemas de segurança e ferramentas de *Business Process Management* (BPM) podem ser usadas para capturar estes aspectos.
- Perfil de interesses: pode ser obtidos a partir de informações históricas, tais como o histórico de buscas ou de uso de aplicações.

Como o contexto para este trabalho refere-se ao segundo tipo, ele pode ser capturado através de metadados oriundos de aplicações de gestão de processos de negócios (BPM). Todavia, este aspecto não é o foco deste trabalho: assume-se que há alguma ferramenta de gestão de processos no ambiente de trabalho do usuário e que a informação contextual será obtida de tal aplicação.

4.7.3 Representação do Contexto

No que diz respeito à representação do contexto, Bolchini et al. (2007) apresenta três grupos de características que devem ser consideradas:

1. **Aspectos modelados**: o conjunto de dimensões gerenciadas pelo modelo, tais como *espaço, tempo, histórico do contexto, assunto e perfil do usuário*. Em outras palavras, tem a ver com os elementos de contexto modelados, tal como visto na seção 4.7.1.
2. **Características de representação**: são as características gerais do modelo. Incluem:
 - Tipo de formalismo: ou seja, a linguagem de representação do modelo. Por exemplo, pares nome-valor, linguagens de marcadores, linguagens baseadas em lógica, grafos, ontologias.

- **Flexibilidade:** habilidade do modelo de se adaptar a diferentes contextos. Por exemplo, um modelo pode ser específico a um domínio ou aplicação em particular, ou pode ser genérico.
 - **Granularidade variável:** diz respeito à habilidade do modelo para representar o contexto em diferentes níveis de detalhe.
3. **Gestão e uso do contexto:** tem a ver com a forma como o contexto é construído, gerenciado e explorado.

Em geral, tais aspectos estão fortemente relacionados com a abordagem adotada para o seu uso, conforme será visto a seguir.

4.7.4 Uso do Contexto

Várias técnicas para uso do contexto em sistemas de RI vêm sendo utilizadas. De acordo com os trabalhos pesquisados, as seguintes abordagens foram observadas: refinamento da consulta (expansão da consulta e sistemas de recomendação), reordenação e categorização.

Refinamento da consulta: segundo Stojanovic (2005) o objetivo do refinamento da consulta é modificar consultas ou seus resultados de acordo com as necessidades dos usuários. Isso pode ser feito de duas maneiras: *expansão da consulta* ou com *sistemas de recomendação*.

- **Expansão da Consulta:** visa ajudar usuários a fazer uma consulta melhor, isto é, tenta melhorar a qualidade da busca através de substituição ou adição de novos termos à consulta inicial (Stojanovic, 2005). Expansão da consulta é necessária devido à ambiguidade das linguagens naturais e também pela dificuldade em se utilizar um único termo para representar um conceito. Com a expansão da consulta, o usuário é guiado a formular consultas que permitem a obtenção de resultados mais úteis.
- **Sistemas de Recomendação:** sistemas de recomendação podem ser de dois tipos. O primeiro tipo é chamado de *recomendação baseada em conteúdo* e visa recomendar documentos similares àqueles que um dado usuário se interessou no passado (Belkin e Croft, 1992). O segundo tipo, chamado *recomendação colaborativa*, identifica usuários cujos gostos são similares aos de um certo usuário e recomenda documentos que tenham interesse (Balabanović e Shoham, 1997; Renda e Straccia, 2005). Sistemas de recomendação são também conhecidos como sistemas de filtragem, pois determinam quais documentos na lista de resultados são relevantes e quais não são. Isto pode ser feito através de comparação dos documentos com uma lista de palavras-chave que descrevem um usuário ou um conjunto de documentos que o

usuário julgou previamente como relevantes e irrelevantes. É esperado de um bom filtro remover muitos documentos não relevantes e manter os relevantes na lista de resultados (Pretschner e Gauch, 1999).

Reordenação (*Re-ranking*): visa mudar a ordem dos resultados da busca de acordo com o contexto. Uma preparação prévia pode se fazer necessária como, por exemplo, a adição de metadados sobre o contexto aos documentos durante a indexação. Este tipo de algoritmo aplica uma função aos números do *ranking* que foram retornados pelo motor de busca. O objetivo dessa função é trazer mais documentos relevantes para o topo de lista. Uma plataforma de busca que faz re-ranking usando contexto é apresentada por Gross e Klemke (2002).

Categorização: visa organizar os documentos em categorias, no intuito de melhorar os resultados da busca. De acordo com Kāki (2005), duas abordagens principais são tradicionalmente utilizadas em categorização: *clusterização* e *classificação*. *Clusterização* de documentos define categorias – também chamadas de *contextos* por Ratprasartporn et al. (2009) – através do agrupamento de documentos similares. Já na *classificação*, documentos são associados a um conjunto predefinido de categorias. Técnicas de categorização podem ser utilizadas tanto na apresentação (organização) de resultados quanto em conjunto com técnicas de ordenação (*ranking*) de documentos.

Como a abordagem adotada neste trabalho foca na *expansão da consulta*, a seção a seguir apresenta em mais detalhes as principais técnicas utilizadas nessa abordagem.

4.7.5 Expansão da Consulta

O principal objetivo da expansão da consulta (também conhecida como aumento da consulta) é adicionar novos termos significativos à consulta original (Bhogal et al., 2007). Segundo Pretschner e Gauch (1999), em geral consultas tendem a ser muito amplas. Portanto, se uma consulta puder ser expandida de acordo com os interesses do usuário, é mais provável que os resultados sejam mais focados. No entanto, isso nem sempre é fácil pois a consulta deve ser expandida com termos relevantes. Caso isso não aconteça, mais documentos não relevantes serão retornados.

Segundo Rieh e Xie (2001), outra motivação por trás da expansão da consulta é baseada na observação de que nem sempre o usuário sabe como expressar exatamente o que está procurando e, portanto, há a necessidade de se fazer um *refinamento* ou *reformulação* na consulta, com base nos resultados inicialmente obtidos. Efthimiadis (1996) divide o processo de busca em duas partes: a formulação da consulta inicial, onde é definida a estratégia de busca; e o estágio de reformulação, onde a consulta é ajustada manualmente ou com a ajuda do sistema.

Com base nos estudos feitos em diversos trabalhos (Rieh e Xie, 2001; Jones et al., 2006; Xiang et al., 2010), há várias motivações que levam usuários a reformularem suas consultas:

- **Substituição por sinônimos:** termos originais são substituídos por palavras que tenham o mesmo significado.
- **Correção de erros:** a mudança ocorre por causa de erros ortográficos ou de digitação.
- **Especificação:** o significado da consulta é especificado através da adição de mais termos ou da substituição de alguns termos por palavras com significado mais específico.
- **Generalização:** feita através da remoção de alguns termos ou da substituição por outros que tenham um significado mais geral.
- **Substituição por termos relacionados:** ao contrário da substituição por sinônimos, aqui há uma mudança de significado. No entanto, a substituição é feita por termos que tenham alguma relação com os termos originais.

Abordagens para Expansão da Consulta

Segundo Efthimiadis (1996); Bhogal et al. (2007), a expansão da consulta pode ser manual, automática ou assistida pelo usuário:

- A expansão manual depende do conhecimento do usuário, que decide quais novos termos serão incluídos na consulta.
- No caso da expansão automática, pesos são calculados para todos os termos e os que tiverem maiores pesos são adicionados à consulta inicial. A qualidade da expansão depende de como os pesos são calculados, o que pode variar de acordo com a função utilizada no cálculo.
- Com a expansão assistida pelo usuário, o sistema gera os possíveis termos para a expansão e o usuário seleciona quais deles serão incluídos. Os novos termos selecionados devem prover o contexto necessário para a consulta original e conseqüentemente permitem a melhoria nos resultados obtidos.

A expansão pode ser feita de maneira interativa, sendo normalmente baseada em hipertexto ou hierarquias de conceitos construídas automaticamente a partir do repositório de documentos (Stojanovic, 2005).

De um modo geral, há três tipos básicos de abordagens para a realização da expansão da consulta (Efthimiadis, 1996; Bhogal et al., 2007): baseada

nos resultados (ou por realimentação de relevância⁷), baseada em modelos de conhecimento dependentes da coleção e baseada em modelos de conhecimento independentes da coleção. Nessa última abordagem, normalmente é utilizada uma ontologia, onde seus conceitos são usados para a desambiguação dos termos e a subsequente expansão da consulta. A figura 4.6 posiciona tais abordagens.

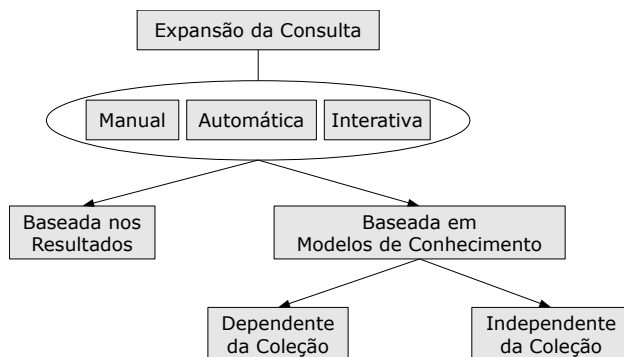


Figura 4.6: Tipos e abordagens para expansão da consulta.

Fonte: (Efthimiadis, 1996)

É importante salientar que as técnicas para a expansão da consulta se baseiam na noção linguística de contexto, ou seja, contexto é tido como o significado dos termos, que algumas vezes é obtido através das palavras que se encontram próximas aos termos em questão.

Realimentação de Relevância

Realimentação de relevância é uma técnica bem estabelecida para a modificação da consulta original usando palavras dos documentos mais bem classificados ou identificados como relevantes. É considerada uma forma mais fácil de se melhorar a busca ao invés do usuário ter que construir uma nova consulta.

O ciclo de realimentação de relevância pede que o usuário entre com uma consulta inicial com os resultados apresentados na forma de uma lista ordenada de documentos (normalmente os títulos / resumos). A partir desses resultados, o usuário seleciona os documentos que julgar relevantes. Os termos relevantes desses documentos são então adicionados à consulta inicial. Uma alternativa a essa abordagem é o pseudo-realimentação de relevância, onde os primeiros n documentos são assumidos como relevantes.

⁷Do inglês *relevance feedback*.

Independente da abordagem usada (tradicional ou pseudo), a seleção dos termos é um fator chave para o desempenho das consultas expandidas. É preciso considerar como associar pesos aos novos termos, seja para excluir os termos originais da consulta, seja para incluir todos os novos termos, ou apenas alguns deles (e nesse caso quantos devem ser incluídos).

Expansão da Consulta usando modelos de conhecimento dependentes da coleção

Este tipo de abordagem obtém o contexto a partir da coleção de documentos. As técnicas mais comuns são *stemming*, clusterização e co-ocorrência de termos.

- *Stemming* é um processo onde todas as variações de termos são geradas pela adição e/ou remoção de prefixos e sufixos, conforme o caso.
- *Clusterização de termos* consiste em agrupar documentos similares em grupos chamados *clusters*. Assume-se que documentos similares são relevantes às mesmas consultas. Tendo isso como base, se os termos da consulta forem mapeados para um ou mais *clusters*, os termos desses *clusters* seriam usados na expansão da consulta.
- *Co-ocorrência de termos* é outra forma de se inferir o contexto de uma consulta. Esta técnica consiste na detecção de dois ou mais termos que se situam uma ao lado do outro (ou próximos entre si) no documento.

Expansão da consulta usando modelos de conhecimento independentes da coleção

O problema com as técnicas tradicionais de realimentação de relevância e das técnicas dependentes da coleção é que elas são dirigidas ao conteúdo. O conteúdo da coleção é analisado para a extração dos termos candidatos à expansão da consulta. Assim sendo, tais técnicas funcionam quando há documentos relevantes suficientes para serem usados e também que tais documentos contenham um conjunto razoável de termos que representem a área relacionada à consulta. Modelos independentes da coleção não sofrem dessa limitação.

Em linguagem natural, uma palavra tem múltiplos significados dependendo do contexto aplicável. O propósito do modelo de conhecimento é prover o contexto para o vocabulário dessa linguagem, visando a eliminação das ambiguidades das palavras contidas em um texto. Modelos de conhecimento independentes da coleção podem ser definidos na forma de *thesauri* ou de ontologias.

As vantagens do uso de ontologias incluem o fato de já existir diversas ontologias disponíveis (Bateman, 2005) e que estão codificadas em formatos

padronizados (XML e outros padrões da Web Semântica). Já problemas relativos ao uso de ontologias incluem a incompatibilidade de vocabulário entre os termos da consulta e os conceitos da ontologia. Um mapeamento se faz necessário para resolver esse problema. Além disso, se uma ontologia para um domínio particular não existe, seria necessário construí-la. É importante salientar que essas tarefas não são triviais, exigindo tempo e esforço consideráveis. Nesses casos, o uso de metodologias e ferramentas são indicados, tanto para a definição de mapeamentos quanto para o suporte à construção de ontologias.

4.7.6 Estado da arte na expansão de consultas usando ontologias

Embora a noção de contexto utilizada no estado da arte seja diferente da adotada neste trabalho, foram selecionados os trabalhos que utilizam ontologias para este fim. É interessante destacar o papel da ontologia WordNet, que foi utilizada em diversos trabalhos sobre expansão da consulta.

Gonzalo et al. (1998) usaram uma coleção de teste envolvendo consultas cuja ambiguidade foi removida manualmente. Os experimentos cobriram três tipos de indexação, que foi baseada nos: termos originais; nos sentidos dos termos, definidos manualmente; e nos *synsets*⁸ da WordNet. Os autores observaram que se não for removida a ambiguidade das consultas, a indexação por *synsets* tem o mesmo desempenho da indexação tradicional de termos. Já a indexação através dos sentidos das palavras leva a uma melhoria na busca de mais de 29%.

Finkelstein et al. (2002) definem um sistema de busca que combina aspectos linguísticos oferecidos pela ontologia WordNet com uma rede semântica e faz uma análise estatística para avaliar o relacionamento entre palavras. Novos termos relacionados são então selecionados para a expansão da consulta. Os resultados mostraram que o uso desse contexto para guiar o processo de busca oferece melhorias na busca.

Navigli e Velardi (2003) argumentam que a expansão com sinônimos e hiperônimos tem um efeito limitado no desempenho da busca na Web. Eles sugerem que outros tipos de relações semânticas derivadas de uma ontologia são mais efetivos. Foi usada a ontologia WordNet para extrair o domínio semântico de uma palavra e a consulta foi então expandida usando-se termos com significados mais próximos. Os resultados dos experimentos mostraram uma melhoria em relação à consulta inicial.

Bonino et al. (2004) desenvolveram um sistema onde a consulta é refinada com base nas relações hierárquicas entre os conceitos da ontologia. Em outras palavras esse trabalho foca na *especificação* e na *generalização* (conforme apresentado na seção 4.7.5). Tal processo é feito automaticamente

⁸*Synsets* são conjuntos de sinônimos agrupados de acordo com a sua equivalência semântica.

com base em um limiar predefinido. Esse limiar é baseado no número de documentos relevantes em relação a um dado conceito.

Baziz et al. (2005) usaram redes semânticas para representar o conteúdo de documentos e conceitos da ontologia WordNet foram extraídos. Os conceitos foram então indexados, e combinados com a indexação tradicional por palavras-chave, houve uma melhoria significativa no desempenho da busca.

Fu et al. (2005) apresentam uma técnica de expansão da consulta baseada em duas ontologias: uma para dados geográficos e outra no domínio de turismo. Nesse trabalho, a consulta foi expandida com base em informações geográficas, sendo que tal abordagem melhorou os resultados das buscas.

Nilsson et al. (2005) usaram uma ontologia de domínio baseada no sistema SUIs (*Stockholm University Information System*) para a expansão da consulta. SUIs é diferente de outros sistemas de busca pois não permite a entrada de qualquer tipo de consulta. Os tipos de questões são restritos a *quem, o quê, quando e onde*. Ao invés de expandir consultas com base em todos os relacionamentos oferecidos por uma ontologia, apenas sinônimos e hipônimos são usados para aumentar a precisão. Os experimentos mostraram uma melhoria nos resultados.

Stojanovic (2005) apresenta uma abordagem para determinar os possíveis refinamentos em uma consulta baseada em ontologia. O primeiro passo consiste na detecção do tipo de ambiguidade que a consulta pode ter. Em seguida, uma lista ordenada dos refinamentos potencialmente úteis é apresentada ao usuário.

4.7.7 Outras abordagens e Comentários sobre o Estado da Arte

Devido à diversidade dos tipos de contexto, tipos de modelo, abordagens para seu uso e áreas de aplicação, há uma dificuldade natural para o posicionamento e comparação dos trabalhos correlatos, tanto entre si, quanto com a presente tese. Considerando esses pontos, foi observado que esta tese tem relação em alguns aspectos com os trabalhos correlatos, mas não foi detectado nenhum trabalho que tenha uma relação direta entre todos os aspectos ao mesmo tempo.

Bhokal et al. (2007) observam que as técnicas de filtragem colaborativa têm alguma similaridade com técnicas de expansão da consulta baseadas em ontologias. Enquanto a primeira se baseia na obtenção de informações de uma comunidade de usuários que compartilham os mesmos interesses com o usuário que faz a busca, a segunda usa uma ontologia, que normalmente é uma representação coletiva de um domínio que foi definida por uma comunidade de especialistas. No entanto, foi possível observar uma dificuldade de posicionamento, onde a maioria dos trabalhos não apresenta uma comparação quantitativa em relação aos demais trabalhos correlatos. Todavia, é possível

se fazer uma análise qualitativa a respeito das técnicas utilizadas, conforme é apresentado a seguir.

A tabela 4.2 apresenta o posicionamento desta tese em relação ao estado da arte. A tabela é organizada em função do tipo de contexto (seção 4.7.1) em relação com a técnica utilizada (seção 4.7.4).

| Contexto ⇒ Abordagem ↓ | Perfil Individual | Linguístico | Situação Dinâmica |
|--------------------------------------|--|--|---|
| Categoriz. | Käki (2005) | Chanana et al. (2004); Ratprasartporn et al. (2009) | – |
| Ranking | D'Agostini et al. (2008); Finkelstein et al. (2001); Pitkow et al. (2002); Kraft et al. (2006); Navrat e Taraba (2007); Mylonas et al. (2008); Bianchini et al. (2010); Xiang et al. (2010) | D'Agostini et al. (2008); Mylonas et al. (2008) | Gross e Klemke (2002); Bianchini et al. (2010) Redon et al. (2007) |
| Sistemas de Recomend. | Renda e Straccia (2005); Wang et al. (2008) | Bai e Nie (2008) | Kirsch-Pinheiro et al. (2005); Bolchini et al. (2007) |
| Expansão Consulta | Finkelstein et al. (2001); Pitkow et al. (2002); Teevan et al. (2005); Kraft et al. (2006) | Bonino et al. (2004); Stojanovic (2005); Bai et al. (2007) | Esta Tese |

Tabela 4.2: Estado da arte na busca usando contexto.

Há alguns trabalhos que adotam abordagens semelhantes, mas o tipo de contexto é outro. Por exemplo, Bonino et al. (2004); Stojanovic (2005) fazem expansão de consultas semânticas, mas o contexto que lidam é do tipo linguístico. Bonino et al. (2004) focam nos refinamentos para *especificação* e *generalização* da consulta, esta tese lida também com outras relações entre os conceitos da ontologia. Além disso, o refinamento nesse trabalho é feito com base no número de documentos relevantes, enquanto que aqui a motivação é o contexto do usuário, conforme será detalhado no próximo capítulo. Já Stojanovic (2005) lida com o refinamento da consultas semânticas, sugerido com base na ambiguidade da consulta original.

Por outro lado, há trabalhos que seguem a mesma definição de contexto, mas usam abordagens distintas. Por exemplo, o projeto **SaiMotion**

(Gross e Klemke, 2002) define o contexto com elementos organizacionais (processos e tarefas do usuário) e adota uma abordagem que associa o contexto às fontes de informação e indexa a informação contextual (fazendo *re-ranking*). **Context-ADDICT** (Bolchini et al., 2007) também toma uma definição de contexto similar, mas a abordagem adotada para o uso do contexto é baseada na filtragem de informação. Outro trabalho correlato é o sistema **ESTEEM**, cujo foco é em redes *peer-to-peer* (P2P) como suporte à troca de conhecimento e de serviços entre profissionais na área médica (Bianchini et al., 2010). O sistema proposto dá suporte à descoberta de conhecimento e serviços com base em semântica (definida por uma ontologia) e também permite que tal busca seja feita com base no contexto. No caso desse trabalho o contexto é definido como a situação atual do usuário, suas coordenadas geográficas, seus interesses tanto em termos de conteúdo quanto em termos de formato (multimídia, por exemplo).

É importante observar que alguns trabalhos aplicam uma combinação de técnicas. Finkelstein et al. (2001); Pitkow et al. (2002); Kraft et al. (2006) propõem uma combinação entre expansão da consulta a *re-ranking*. Além disso, outra combinação que é feita é no tipo de contexto. Bianchini et al. (2010) modelam o contexto tanto como a situação dinâmica como os interesses do usuário. Além disso, foi observado que a maioria dos trabalhos foca na busca na Web.

Do ponto de vista da modelagem do contexto, o modelo baseado em ontologia aqui proposto foi adaptado (para os aspectos relativos a RCs) do modelo orientado a objetos definido por Kirsch-Pinheiro et al. (2004) e Gensel et al. (2008).

Capítulo 5

Modelo Proposto

Este capítulo descreve o modelo proposto, que é organizado em duas perspectivas: a semântica e a de contexto. Conforme mencionado nos procedimentos metodológicos, a perspectiva semântica foca na busca com suporte de ontologias e a perspectiva de contexto é definida a partir da semântica. Esta perspectiva tem como foco um modelo de contexto e regras para expansão da consulta. Adicionalmente, o modelo é posicionado numa arquitetura de um sistema, definida com base em serviços de *software*. O modelo de contexto compreende consiste na ligação dos tópicos (conceitos da ontologia) com os elementos de contexto e então faz-se uma filtragem com base nessa ligação.

5.1 Visão Geral da Abordagem

A abordagem adotada baseia-se em três elementos presentes em RCs: a fontes de informação, o contexto bem definido do usuário e a ontologia geral. Os dois primeiros são elementos comuns mesmo em um ambiente de uma única empresa, pois os seus funcionários produzem e/ou consomem informações e trabalham dentro de um contexto, ou seja, desempenhando um certo papel em um processo de negócio. O terceiro elemento, todavia, é tomado como um pressuposto, pois não é algo que se encontra naturalmente nesse tipo de ambiente.

De fato, a ontologia está mais para parte da solução, dando o suporte para alguma facilidade ou funcionalidade dentro dos sistemas da empresa ou RC. A ontologia pode ser, por exemplo, usada como um guia pelos funcionários / parceiros como uma fonte de informação acerca da terminologia, normas, processos e quaisquer outros aspectos relevantes à organização. Além disso, como foi observado na revisão bibliográfica, ontologias vêm sendo usadas para permitir o processamento e busca de informação de maneira mais precisa. É preciso enfatizar também o seu uso em aplicações de suporte a RCs (Loss, 2007; Ermilova e Afsarmanesh, 2008; Baldo et al., 2009), sendo que algumas ontologias de referência na área foram desenvolvidas (Plisson e

Ljubic, 2006; Ermilova e Afsarmanesh, 2008).

Portanto, embora não seja uma realidade para empresas e as diversas manifestações de RCs, a abordagem aqui apresentada tomará como pressuposto básico a existência de uma ontologia comum adotada pela RC. A figura 5.1 dá uma visão geral da abordagem. As três características são a base para as perspectivas semântica e de contexto, que por sua vez fazem parte de uma arquitetura de serviços para busca, bem como a sua implementação.

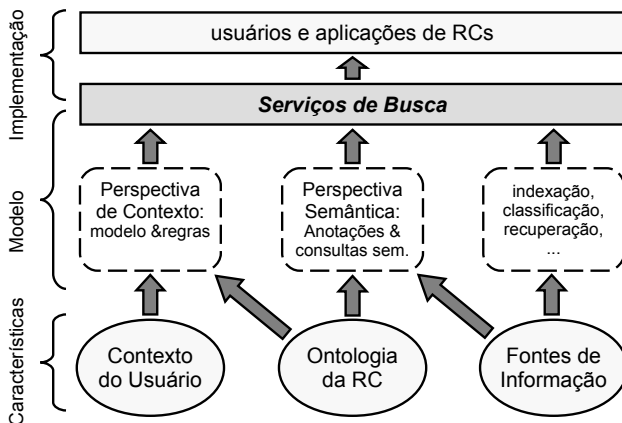


Figura 5.1: Posicionamento da abordagem em relação aos elementos de busca em RCs.

A *perspectiva semântica* leva em consideração as fontes de informação e a ontologia. Ela trata da geração e uso de anotações semânticas bem como de consultas semânticas para o suporte à busca. A *perspectiva de contexto* complementa a semântica, onde elementos de contexto são adicionados à ontologia e regras são usadas para a expansão da consulta.

5.2 Perspectiva Semântica

Esta perspectiva corresponde à visão da Web Semântica, onde uma ontologia é usada para enriquecer documentos permitindo um processamento mais preciso. Neste caso em particular, a ontologia também dá suporte à busca, sendo usada para a definição de consultas semânticas. A visão aqui apresentada é, portanto, a mesma mostrada na revisão bibliográfica, sendo o ponto de partida para o modelo de busca usando contexto proposto nesta tese.

Esta perspectiva pode ser dividida em dois passos distintos: anotação semântica automática e indexação de documentos, e busca semântica.

No primeiro passo, a ontologia é usada para dar suporte à geração de anotações semânticas aos documentos. Esse processo é feito de forma (semi-

)automática, de acordo com as técnicas apresentadas da seção 3.6. É preciso destacar aqui a importância crucial de que a ontologia possua uma quantidade de instâncias suficientes – também chamada de base de conhecimento – para a geração de anotações com uma boa qualidade.

Na busca semântica, o motor de busca deve ser capaz de processar consultas semânticas, ou seja, consultas definidas em função dos conceitos da ontologia adotada. A busca é dividida em dois passos. No primeiro, as entidades da base de conhecimento que casem com as restrições da consulta são selecionadas. Esse processo é similar à recuperação de dados em um banco de dados, onde os valores são selecionados sem qualquer tipo de ordenação entre si. No segundo passo, são selecionados os documentos cujas anotações semânticas estão vinculadas às entidades selecionadas no primeiro passo. Um algoritmo de classificação (*ranking*) é utilizado para a ordenação dos resultados. A figura 5.2 abaixo resume a perspectiva semântica.

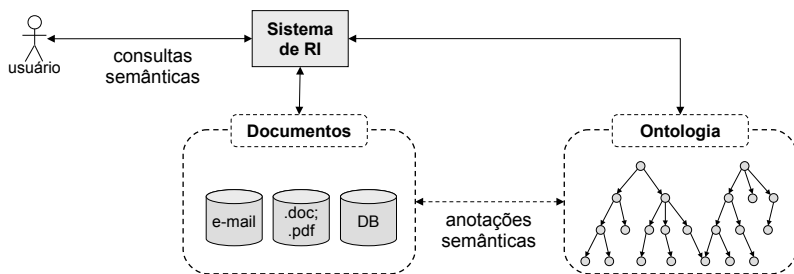


Figura 5.2: Perspectiva Semântica.

Uma consulta semântica aplica restrições a classes e relações presentes na ontologia, de modo a buscar documentos que contenham anotações que casem com estas restrições. Tomando o mesmo exemplo apresentado na seção 4.6, a consulta semântica “*Companhias de TI localizadas na Europa*” retorna documentos que contenham anotações semânticas relacionadas à instâncias da classe *Organização* cuja propriedade *localizadaEm* esteja relacionada a uma instância em particular da classe *Lugar*, cujo nome é “Europa”.

No modelo conceitual, uma consulta semântica é definida na forma de predicados. Nesse sentido, o exemplo apresentado acima é representado da seguinte forma: $x | IT\ company(x) \wedge locatedIn('Europe')$.

Esta notação é usada ao longo deste capítulo, permitindo o entendimento de como as regras da perspectiva de contexto são aplicadas.

5.2.1 Ontologia para RCs

Embora a perspectiva apresentada não defina o tipo de ontologia, o foco deste trabalho é a busca em RCs. Nesse sentido, foi definida uma ontologia que define os elementos dessa área. Ela consiste em uma versão unificada e simplificada das ontologias definidas por Ermilova e Afsarmanesh (2008) e Plisson e Ljubic (2006), no âmbito do Projeto ECOLEAD. A figura 5.3 mostra os principais conceitos e relações dessa ontologia, que são descritos a seguir.

- **Organização:** classe que pode ser tanto uma RC quanto um participante. Ela pode ter processos, recursos, competências e entrega algum resultado (produto ou serviço).
- **RC:** representa uma rede colaborativa de organizações. Pode ser um Ambiente de Criação de Organizações Virtuais (ACV) ou uma Organização Virtual (OV).
- **Participante:** organização que faz parte de uma RC.
- **Competência:** está associada a uma capacidade, que por sua vez, está ligada a um recurso.
- **Processo:** faz uso de recursos, é composta por tarefas e gera um resultado.
- **Recurso:** pode ser humano, físico ou tecnológico.
- **Papel:** é desempenhado por um participante e está associado a uma tarefa.
- **Oportunidade de Colaboração:** é detectada e mantida pelo ACV e dispara uma OV.
- **Capital / Bem:** representa os recursos compartilhados pelos parceiros dentro de um ACV. Pode ser documentos, *software* ou ontologias.

A ontologia também define a classe **Indicador de Desempenho**. Esse aspecto da ontologia foi adaptado da ontologia desenvolvida por Baldo (2008). Um indicador está relacionado a outras classes que definem seus atributos. Segundo a ontologia, um indicador:

- mede um **Objeto** (organização, processo, recurso);
- é medido com um **Objetivo de Medição** (planejamento, gestão, etc.);
- considera uma **Perspectiva de Desempenho** específica (custo, confiabilidade, qualidade, etc.);

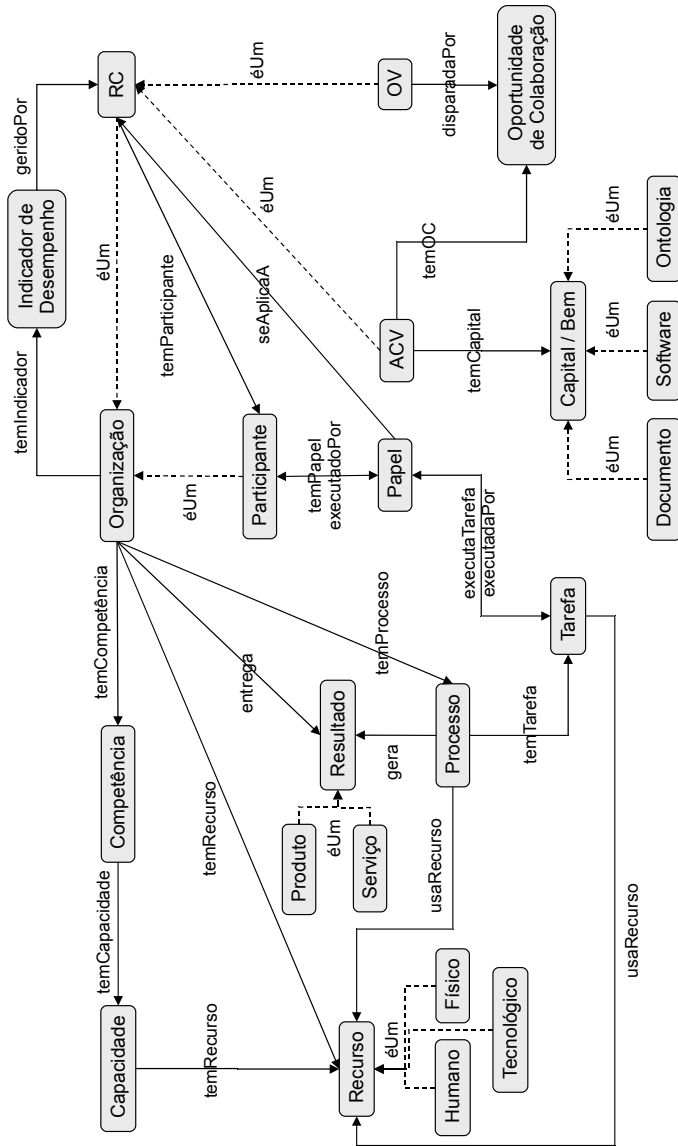


Figura 5.3: Ontologia para RCs.

- é aplicado a um **Domínio**;
- está relacionado a um **Contexto Organizacional**. Esta classe define o escopo do indicador, que pode ser intra ou interorganizacional;
- é calculado usando uma **Regra de Cálculo** e;
- provê resultados em uma certa **Unidade de Medida**.

A figura 5.4 apresenta parte da ontologia referente a indicadores de desempenho.

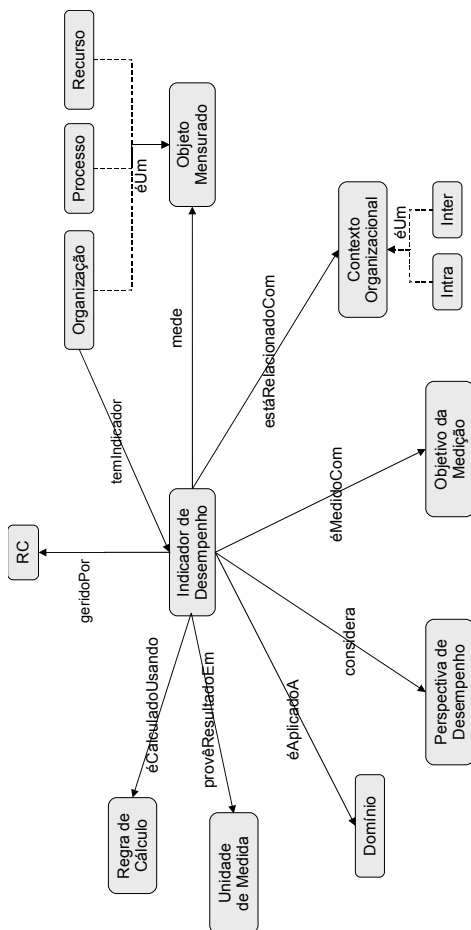


Figura 5.4: Fragmento da ontologia que define indicadores de desempenho.

dor, conselheiro). Já numa OV, uma organização pode ser um *parceiro* ou um *provedor de suporte* (*broker*, coordenador, *panejador*).

A razão para a definição de uma nova ontologia ao invés da reutilização de alguma já existente foi basicamente a simplificação de algumas relações, permitindo assim que consultas possam expressar mais facilmente o que está sendo buscado. Por exemplo, na ontologia definida por Ermilova e Afsarmanesh (2008), a o elementos associados a uma organização (recursos, processos, competências) são agrupados dentro de um *perfil*. Assim, uma consulta por “organizações que tenha um certo recurso” deveriam ser expressas como “organização que tenha um certo perfil que contenham um certo recurso”. Além disso, outra simplificação feita foi a remoção de alguns conceitos e relações que não teriam tanta utilidade tanto para a anotação automática quanto para a busca. Exemplos desse tipo conceito incluem descrições dos sistemas de gestão, confiança e de valor.

Um aspecto que foi herdado das ontologias originais é o fato de que ela foi definida de tal forma que extensões possam ser mais facilmente adicionadas. A ontologia foi dividida em três níveis, do mais geral ao mais específico: *topo*, *domínio* e *aplicação* (figura 5.6). Essa divisão permite uma fácil construção de cima para baixo (*top-down*), ou seja, a partir dos conceitos mais gerais. Tais níveis são enumerados a seguir:

1. **Topo:** especifica os principais conceitos referentes a RCs: Participante, Perfil, Competência, Processo e Recurso.
2. **Domínio:** inclui subclasses dos conceitos definidos no nível *topo*, relativas a algum domínio. Por exemplo, manufatura, turismo, comércio.
3. **Aplicação:** consiste em especializações do nível de domínio que são específicas a uma aplicação em particular.

A ontologia aqui apresentada compreende o primeiro nível e contém os conceitos presentes em todos os tipos de RCs. Espera-se, portanto, que especializações possam ser feitas de acordo com o domínio e com a aplicação. Tais extensões podem ser associadas à ontologia comum na forma de subclasses dos conceitos mais gerais.

Para a validação, alguns conceitos vinculados ao domínio / aplicação foram definidos (p. ex. processos / competências). Além disso, foram definidas instância na base de conhecimento, de modo a viabilizar o processo de anotação semântica automática.

5.3 Perspectiva de Contexto

A perspectiva de contexto toma como base a perspectiva semântica adicionando a noção de *contexto*. Conforme foi apresentado na revisão bibliográfica, o contexto pode ser interpretado de diversas formas. Neste trabalho,



Figura 5.6: Níveis da ontologia de RCs.

contexto é interpretado como a situação atual e dinâmica do usuário, representado pelos seguintes elementos: a rede na qual participa, a organização para a qual trabalha, o papel por ele desempenhado, e o processo e atividade em execução no momento.

A abordagem proposta consiste em capturar este contexto e, com base em um modelo que o representa e de um conjunto de regras, aplicar um refinamento na consulta original definida pelo usuário. A captura é feita por um serviço externo, conforme mostra a seção 5.6. Em linhas gerais, o modelo de contexto é vinculado à ontologia de tal forma que os elementos de contexto estejam associados aos conceitos ali definidos. Um conjunto de regras é então usado para selecionar – de acordo com o contexto atual do usuário – quais conceitos são candidatos à expansão da consulta. A figura 5.7 resume a perspectiva de contexto, mostrando como esses componentes são associados entre si.

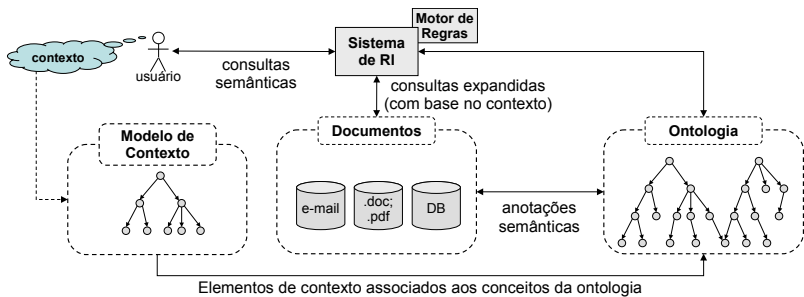


Figura 5.7: Perspectiva de Contexto.

A justificativa para o uso de tal abordagem é que esse tipo de contexto pode ser capturado dinamicamente através de plataformas de segurança e de gestão de processos. Além disso, tais elementos de contexto são conceitos presentes na ontologia de RCs e, portanto, estão vinculados aos tópicos buscados. Em outras palavras, a motivação da abordagem é aproveitar a ontologia

(que já é utilizada para processar buscas semânticas) e ligá-la ao contexto do usuário.

É importante ressaltar também que a ontologia continua sendo um elemento central da abordagem, estando agora vinculada ao modelo de contexto. Consequentemente, a abordagem trata de consultas semânticas, ao invés de consultas baseadas em palavras-chave.

A expansão da consulta baseia-se em algumas das motivações apresentadas na seção 4.7.5 que levam ao refinamento da consulta. Tal refinamento pode ser feito por:

- **Substituição, através de generalização ou especialização:** onde são selecionados conceitos que são sub ou superclasses do conceito usado originalmente na consulta. Por exemplo, um usuário em um contexto de “manufatura” (um processo) e fazendo uma busca por “serviços de *software*”, pode ter sua consulta refinada para “serviços de *software* para manufatura”. É importante salientar aqui que o refinamento é feito com base nas relações ontológicas entre os conceitos. Neste caso a consulta $x|\text{Servico}(x)$ é substituída por $x|\text{ServicoDeManufatura}(x)$, onde *ServicoDeManufatura* é subclasse de *Servico*.
- **Relaxamento ou restrição das variáveis da busca:** novas condições são adicionadas à consulta, visando torná-la mais restrita ou mais ampla. Novamente são usadas as relações ontológicas entre os conceitos, neste caso selecionando conceitos que são relacionados ao conceito em questão. Por exemplo, busca por indicadores. A consulta $x|\text{Indicator}(x) \wedge \text{considers}(x, \text{“quality”})$ pode adicionar mais uma restrição, supondo-se que o usuário esteja em um contexto de “produção”: $x|\text{Indicator}(x) \wedge \text{considers}(x, \text{“quality”}) \wedge \text{measures}(x, \text{“production”})$.

O funcionamento dessa abordagem é detalhado a seguir.

5.4 Modelo de Contexto

Para que o sistema de busca possa capturar o contexto do usuário e usá-lo para a expansão das consultas, tal contexto deve ser modelado. O contexto em uma RC é formado por uma série de elementos, que foram determinados com base em observações feitas na análise da literatura.

Gross e Klemke (2002) definem o contexto como uma tupla (*pessoa, atividade, tempo, lugar*), onde *pessoa* indica os papéis desempenhados, interesses e experiência; *atividade* indica a tarefa a ser cumprida (e que está dentro de um processo, incluindo as ferramentas utilizadas); *tempo* pode ser absoluto ou relativo ao progresso de uma atividade; *lugar* indica as coordenadas do usuário, que podem ser absolutas ou definidas de forma relativa, tal como o número da sala ou departamento.

Hawking et al. (2005) ressaltam que os elementos contextuais típicos para a busca em uma empresa são os *usuários* (e seus papéis), a *tarefa* sendo desempenhada (pois as buscas são normalmente relacionadas ao trabalho) e as *fontes de informação*, que são geradas e consumidas no âmbito do negócio em questão, e cujos metadados (tal como o autor) podem prover o contexto necessário para a busca.

Demartini (2007) enfatiza que a buscas dentro de uma empresa são orientadas pelo trabalho em andamento e seus resultados podem ser personalizados de acordo com o papel desempenhado. Portanto, elementos como a *tarefa* (atividade, processo) e *papel* desempenhado podem ser capturados e utilizados para a melhoria do desempenho da busca.

Belkin (2008) destaca a necessidade de integração do sistema de busca com o ambiente de trabalho. Tal afirmação é feita com base na observação de que a busca por informação é quase sempre a consequência dos objetivos de uma pessoa visando realizar uma tarefa.

Com base nessas evidências, os seguintes elementos de contexto foram estabelecidos:

- A *RC* e a *organização*, que oferecem a noção de lugar;
- O *processo* e *tarefa* atuais, que indicam os objetivos e o estado atual do trabalho;
- O *papel* desempenhado pelo usuário.

Tais elementos são agrupados e associados à entidade *contexto*, que por sua vez é associada a outras entidades que compõem o modelo proposto: *usuário*, *consulta*, *restrição* e *conceito*. Tais entidades são organizadas na forma de uma ontologia, conforme mostra a figura 5.8.

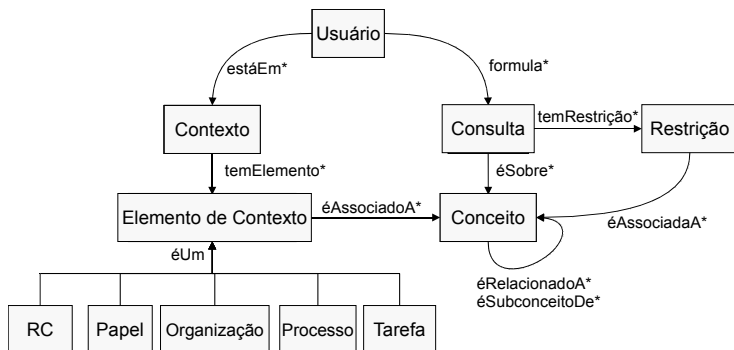


Figura 5.8: Modelo de contexto.

De acordo como esse modelo, um *usuário* está em um certo *contexto* e formula *consultas*. Consultas são compostas por um ou mais *conceitos*, bem como *restrições* sobre as suas propriedades. Um contexto é composto por *elementos de contexto*, que representam aspectos em particular que compõem tal contexto: a RC na qual participa, sua *organização*, seu *papel* e a *atividade* e *processo* nos quais está envolvido. Os elementos do modelo de contexto e seus relacionamentos são definidos na forma de predicados, que então podem ser usados pelas regras de customização. Eles podem ser vistos na tabela 5.1.

| Predicados | Descrição |
|-----------------------------|--|
| $formula(U, Q)$ | Usuário U formula consulta Q . |
| $ehSobre(Q, C)$ | Consulta Q é sobre o conceito C . |
| $estahEm(U, X)$ | Usuário U está no contexto X . |
| $ehRelacionadoA(C_1, C_2)$ | Conceito C_1 tem relação com C_2 . |
| $ehSubconceitoDe(C_1, C_2)$ | Conceito C_1 é subconceito de C_2 . |
| $temElemento(X, E)$ | Contexto X tem o elemento de contexto E . |
| $ehAssociadoA(E, C)$ | Elemento de contexto E é associado ao conceito C . |

Tabela 5.1: Modelo de contexto definido como predicados.

Embora muitos desses predicados sejam auto-explicativos, é importante explicar em maiores detalhes o significado de $ehRelacionadoA$, pois ele é a base para várias regras. Ele significa, em termos de notação RDF, que C_1 tem uma propriedade do tipo objeto (ou seja, uma relação) cujo *range* é C_2 . Portanto, tendo-se a notação (*sujeito propriedade valor*) para as triplas em RDF, o predicado $ehRelacionadoA(C_1, C_2)$ é verdadeiro quando:

$$(c_1 \text{ rdf:type rdfs:Class}) \wedge (c_2 \text{ rdf:type rdfs:Class}) \\ \wedge (p \text{ rdf:type rdf:Property}) \wedge (p \text{ rdfs:domain } c_1) \wedge (p \text{ rdfs:range } c_2)$$

A primeira linha significa que tanto C_1 quanto C_2 são classes, isto é, conceitos da ontologia. Os outros elementos da expressão significam que há uma propriedade da classe C_1 (definido por *domain*) que aponta para C_2 (através de *range*). Em outras palavras, isso quer dizer que $ehRelacionadoA$ não é simétrica, ou seja, se C_1 tem uma relação com C_2 , o contrário não é necessariamente verdadeiro. É importante fazer essa distinção, pois tal relação tem um impacto na forma como certas regras selecionam os conceitos. A seção a seguir detalha esses aspectos.

5.4.1 Ligação do Modelo de Contexto à Ontologia

Esta seção apresenta a ligação entre o modelo de contexto e a ontologia. A figura 5.9 apresenta as classes da ontologia que estão associadas aos elementos de contexto. A ideia inicial que motivou tal associação consistiu em vincular os elementos de contexto diretamente ao seu equivalente na

ontologia. Desse modo, o elemento “Processo” do modelo de contexto está vinculado à classe “Processo” da ontologia e o mesmo ocorre com os demais elementos: tarefa, organização, papel e RC.

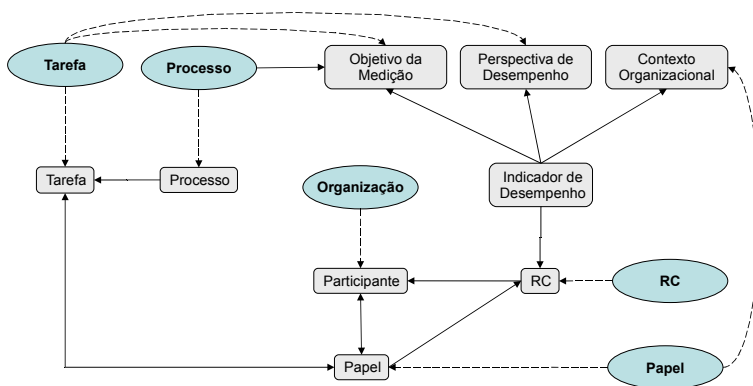


Figura 5.9: Ligação entre os elementos de contexto com a ontologia.

Além dessas ligações mais óbvias, alguns elementos de contexto estão também vinculados a outras classes, que servem para qualificar indicadores de desempenho:

- *Papel*: este elemento pode definir, entre outras coisas, o escopo de visibilidade do usuário. Assim, ele foi associado à classe *Contexto Organizacional*, que representa o escopo atuação do indicador de desempenho (intra ou interorganizacional). Dessa maneira, dependendo do papel do usuário (e do seu nível de visibilidade), este elemento pode filtrar determinados indicadores que estão sendo buscados.
- *Processo*: foi vinculado à classe *Objetivo da Medição* pois representa o propósito da medição a ser realizada: gestão, produção, planejamento, entrega, entre outros. Assim, indicadores podem também ser selecionados dependendo do contexto (processo) do usuário.
- *Tarefa*: assim como o processo, a tarefa pode indicar qual o propósito da busca. Na ontologia, tal propósito pode ser relacionado às classes *Objetivo da Medição* e *Perspectiva de Desempenho*.

Com base nessas associações, as regras de expansão da consulta podem, com base em instâncias do modelo de contexto (ou seja, de um contexto capturado dinamicamente) e em instâncias da ontologia (base de conhecimento), selecionar novas restrições a serem adicionadas ou substituídas na consulta original do usuário.

5.5 As Regras de Customização

Conforme apresentado no exemplo da seção 5.3, a ideia básica desta abordagem é selecionar novos conceitos a serem adicionados na consulta, sendo que tais conceitos têm alguma relação com os conceitos definidos na consulta original e, ao mesmo tempo, estão associados ao contexto do usuário.

Nesse sentido, de modo a fazer tal seleção, 12 regras de customização foram definidas. Tais regras selecionam os conceitos de acordo com o tipo de relação que pode haver entre eles. Considerando o conceito C_q que pertence à consulta Q , e o conceito C_x associado ao contexto do usuário que formula Q , as regras são agrupadas de acordo com três perspectivas:

1. Selecionando conceitos **que são relacionados pelos** conceitos definidos na consulta. Em termos do modelo de contexto, isto é representado como $ehRelacionadoA(C_q, C_x)$. Esse tipo de regras adiciona restrições à consulta original, ou seja, seus resultados são focados em menos documentos.
2. Selecionando conceitos **que têm relação com** os conceitos mencionados na consulta. Esta é a situação inversa ao item anterior, sendo representada como $ehRelacionadoA(C_x, C_q)$. Esse tipo de regra adiciona conceitos adicionais a serem buscados.
3. Selecionando conceitos **que fazem parte da hierarquia** dos conceitos definidos na consulta. É representado tanto por $ehSubconceitoDe(C_q, C_x)$ quanto por $ehSubconceitoDe(C_x, C_q)$. O propósito desse tipo de regra é sugerir a substituição dos conceitos originais pelos seus sub ou super-conceitos.

As duas primeiras perspectivas visam expandir a consulta através do relaxamento ou restrição das variáveis da consulta. Isso é feito com base nas relações entre os conceitos que se encontram no mesmo contexto. A terceira perspectiva visa expandir a consulta através da substituição, com base na generalização ou especialização.

Tais regras são apresentadas e explicadas a seguir. De modo a manter as regras o mais simples possível, algumas delas lidam com aspectos particulares e então regras mais complexas são construídas com base nas mais simples. Seguindo essa abordagem, a primeira regra é usada para deduzir o contexto da consulta.

$$formula(U, Q) \wedge estahEm(U, X) \Rightarrow ehFormuladaEm(Q, X) \quad (5.1)$$

A regra 5.1 deduz o contexto no qual a consulta é formulada, considerando o contexto do usuário que a formula. Dado que esse aspecto é fundamental, essa regra serve como base para as demais.

5.5.1 Selecionando conceitos relacionados pelos conceitos da consulta

O propósito da próxima regra é selecionar conceitos que possam adicionar novas restrições à consulta original:

$$\begin{aligned} & ehSobre(Q, C_1) \wedge ehRelacionadoA(C_1, C_2) \wedge not(ehSobre(Q, C_2)) \\ & \Rightarrow ehSobreConceitoRelacionado(Q, C_2) \quad (5.2) \end{aligned}$$

De acordo com a regra 5.2, se uma consulta Q se refere a um certo conceito C_1 que por sua vez está relacionado ao conceito C_2 , e C_2 não é mencionado em Q , então Q também tem relação ao conceito C_2 . O propósito dessa regra é deduzir novos conceitos que possam ser adicionados à consulta, na forma de restrições. Para ilustrar esta situação, toma-se como exemplo os conceitos e relações mostrados na figura 5.10.

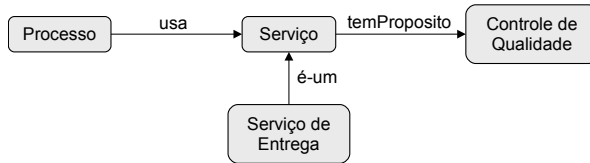


Figura 5.10: Fragmento de uma ontologia.

Considerando que o usuário define a seguinte consulta: $x|Servico(x)$, e aplicando-se a regra 5.2, o único conceito relacionado a “*Serviço*” é “*Controle de Qualidade*”. Portanto, se o usuário decidir adicionar este novo conceito, isto será feito de acordo com a relação definida entre eles. A nova consulta seria: $x|Servico(x) \wedge temProposito(x, “ControleDeQualidade”)$. Novamente, é importante salientar que tal adição representa uma restrição (em termos de escopo de busca) à consulta original.

Seguindo a estratégia de manter as regras o mais simples possível, essa regra não considera o contexto. Isto é feito pela próxima regra:

$$\begin{aligned} & ehFormuladaEm(Q, X) \wedge temElemento(X, E) \\ & \wedge ehAssociadoA(E, C) \wedge ehSobreConceitoRelacionado(Q, C) \\ & \Rightarrow ehSobreConceitoRelacionadoNoContexto(Q, C) \quad (5.3) \end{aligned}$$

A regra 5.3 deduz conceitos que são relacionados pelos conceitos da consulta original (Q) e que também são associados ao contexto (X) do usuário. Portanto, ela usa os fatos deduzidos pelas regras 5.1 (selecionando o contexto do usuário) e 5.2 (selecionando os conceitos relacionados). O exemplo anterior pode ser considerado aqui também, assumindo-se que “*Controle de Qualidade*” é associado ao contexto do usuário.

5.5.2 Selecionando conceitos que têm relação com os conceitos da consulta

A regra 5.4 adiciona conceitos que têm relação com os conceitos definidos pela consulta original, ou seja:

$$\begin{aligned} & ehSobre(Q, C_1) \wedge ehRelacionadoA(C_2, C_1) \wedge not(ehSobre(Q, C_2)) \\ & \Rightarrow ehSobreConceitoQueSeRelaciona(Q, C_2) \quad (5.4) \end{aligned}$$

De acordo com a regra 5.4, se uma consulta Q se refere a um certo contexto C_1 e há um conceito C_2 que tem relação com C_1 , e também que C_2 não é mencionada em Q , então Q também tem relação ao conceito C_2 . A diferença entre esta regra e a regra 5.2 é que os parâmetros estão invertidos no segunda condição.

Como exemplo, vamos considerar os mesmo conceitos apresentados na figura 5.10 e a mesma consulta: $x|Servico(x)$. O conceito selecionado neste caso é “*Processo*”. Ao invés de adicionar restrições à consulta original, esta regra sugere novos conceitos a serem buscados. Novamente, esta regra não considera o contexto, que é complementado pela regra 5.5:

$$\begin{aligned} & ehFormuladaEm(Q, X) \wedge temElemento(X, E) \\ & \wedge ehAssociadoA(E, C) \wedge ehSobreConceitoQueSeRelaciona(Q, C) \\ & \Rightarrow ehSobreConceitoQueSeRelacionaNoContexto(Q, C) \quad (5.5) \end{aligned}$$

A regra 5.5 deduz conceitos que têm relação com aqueles presentes na consulta (Q) e que são também associados ao contexto (X) do usuário. Para tal, ela usa os fatos deduzidos pelas regras 5.1 e 5.4.

5.5.3 Selecionando conceitos que pertencem à hierarquia dos conceitos da consulta

A regra 5.6 simplesmente deduz que o conceito C , declarado em Q , está na sua própria hierarquia (tanto superior quanto inferior).

$$\begin{aligned} ehSobre(Q, C) \Rightarrow estahAbaixoNaHierarquia(C, C) \\ \wedge estahAcimaNaHierarquia(C, C) \end{aligned} \quad (5.6)$$

Uma vez disparada, esta regra faz com que outras também sejam executadas, de modo a se deduzir os demais conceitos da hierarquia: regras 5.7 e 5.8 deduzem os conceitos superiores a C e, de maneira similar, as regra 5.9 e 5.10 deduzem os conceitos inferiores.

$$\begin{aligned} estahAcimaNaHierarquia(C_1, C_1) \wedge ehSubconceitoDe(C_1, C_2) \\ \wedge not(estahAcimaNaHierarquia(C_1, C_2)) \\ \Rightarrow estahAcimaNaHierarquia(C_1, C_2) \end{aligned} \quad (5.7)$$

$$\begin{aligned} estahAcimaNaHierarquia(C_1, C_2) \wedge ehSubconceitoDe(C_2, C_3) \\ \wedge not(estahAcimaNaHierarquia(C_1, C_3)) \\ \Rightarrow estahAcimaNaHierarquia(C_1, C_3) \end{aligned} \quad (5.8)$$

$$\begin{aligned} estahAbaixoNaHierarquia(C_1, C_1) \wedge ehSubconceitoDe(C_2, C_1) \\ \wedge not(estahAbaixoNaHierarquia(C_1, C_2)) \\ \Rightarrow estahAbaixoNaHierarquia(C_1, C_2) \end{aligned} \quad (5.9)$$

$$\begin{aligned} estahAbaixoNaHierarquia(C_1, C_2) \wedge ehSubconceitoDe(C_3, C_2) \\ \wedge not(estahAbaixoNaHierarquia(C_1, C_3)) \\ \Rightarrow estahAbaixoNaHierarquia(C_1, C_3) \end{aligned} \quad (5.10)$$

A regra 5.11 complementa as regras anteriores associando à consulta Q novos conceitos, pertencentes à hierarquia dos conceitos já associados a Q :

$$\begin{aligned} ehSobre(Q, C_1) \wedge not(ehSobre(Q, C_2)) \\ \wedge (estahAbaixoNaHierarquia(C_1, C_2) \\ \vee estahAcimaNaHierarquia(C_1, C_2)) \\ \Rightarrow ehSobreConceitoDaHierarquia(Q, C_2) \end{aligned} \quad (5.11)$$

Tal como nos casos anteriores, é preciso então filtrar os conceitos selecionados através do contexto. Nesse sentido, a regra 5.12 é definida:

$$\begin{aligned} & ehFormuladaEm(Q, X) \wedge temElemento(X, E) \\ \wedge & ehAssociadoA(E, C) \wedge ehSobreConceitoDaHierarquia(Q, C) \\ \Rightarrow & ehSobreConceitoDaHierarquiaNoContexto(Q, C) \quad (5.12) \end{aligned}$$

De modo similar aos outros casos, as primeiras três condições selecionam o contexto X da consulta Q , um elemento de contexto E e um dado conceito C associado a E . Em seguida, se C faz parte da hierarquia dos conceitos pertencentes à consulta Q , então pode-se deduzir que Q também tem relação com o conceito C .

Ao contrário das outras duas perspectivas, espera-se neste caso que os conceitos selecionados substituam o conceito original. Por exemplo, considerando os mesmos conceitos apresentado na figura 5.10 e a mesma consulta: $x|Servico(x)$. Assumindo-se que “*Serviço de Entrega*” está associado ao contexto do usuário, este será selecionado pela regra 5.11 pois é um subconceito de “*Serviço*”. Neste caso, a nova consulta seria: $x|ServicoDeEntrega(x)$ (busca por algo relacionado ao contexto), ao invés da consulta mais geral: $x|Servico(x)$.

Considerando três perspectivas cobertas pelas regras, não é esperado que todos os grupos de regras sejam úteis em todos os casos. Isso dependerá da ontologia e de como os elementos de contexto são associados aos seus contextos. Além disso, é esperado que o usuário possa ter controle sobre a expansão da consulta, aceitando ou não as modificações sugeridas. Assim, uma vez expandida e aceita pelo usuário, a nova consulta pode então ser enviada ao motor de busca.

5.6 Posicionamento numa Arquitetura de Sistema

As perspectivas semântica e de contexto apresentadas foram enquadradas em uma arquitetura de sistema, visando a sua implementação num segundo passo. Tal arquitetura é dita conceitual pois ela é independente de tecnologia e vários de seus aspectos não foram especificados.

A arquitetura conceitual é apresentada na figura 5.11. Ela é composta pelos seguintes elementos: serviços¹, aplicativos e elementos externos.

- **Serviços:** implementam as funcionalidades oferecidas pelo sistema. São eles:

¹No nível conceitual, o termo serviço significa um módulo de *software* que provê uma certa funcionalidade, independentemente da linguagem e da tecnologia envolvida na sua implementação.

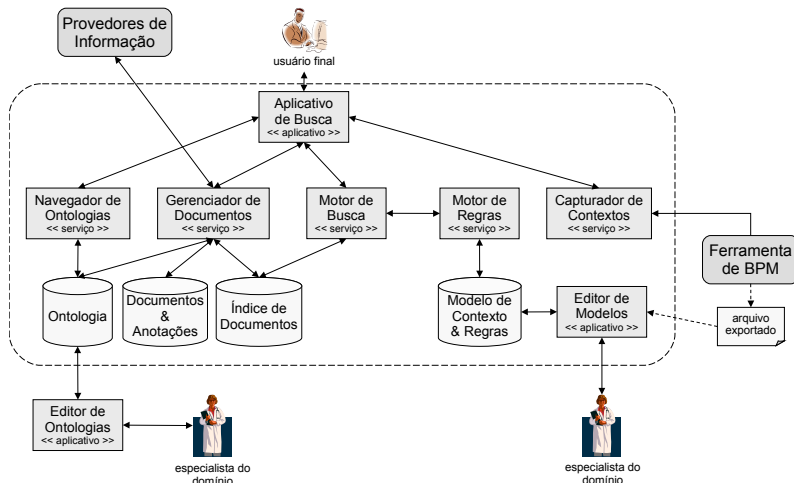


Figura 5.11: Arquitetura Conceitual.

- *Navegação de ontologias*: oferece operações para leitura de classes, propriedades, instâncias. Provê suporte ao aplicativo de busca durante a construção de consultas semânticas.
- *Gerenciador de documentos*: oferece operações para anotação semântica automática, indexação e armazenamento de documentos. Utilizado pelos provedores de informação para a publicação e pelo aplicativo de busca para a recuperação de documentos.
- *Motor de busca*: usado para a busca de documentos.
- *Motor de regras*: motor de inferência que aplica as regras para expansão das consultas.
- *Capturador de contextos*: serviço integrado a uma ferramenta de gestão de processos de negócio (BPM) em particular. Responsável por recuperar o contexto atual do usuário.

- **Aplicativos**: permitem interação com os usuários. São eles:

- *Aplicativo de busca*: usado para definição de consultas, busca e recuperação de documentos.
- *Editor de modelos*: ferramenta também usada por especialistas numa etapa preparatória. Usada para a vinculação do modelo de contexto à ontologia, bem como para o refinamento do modelo/regras, caso necessário.

- **Elementos externos:** não fazem parte da arquitetura em si, mas são essenciais para o seu funcionamento:
 - *Provedores de informação:* pode ser qualquer sistema que gere conteúdo a ser buscado. Neste caso, ele deve ter acesso ao serviço *gerenciador de documentos*. Exemplos típicos de provedores de informação são ferramentas de suporte a CSCW (e-mail, grupos de discussão e sistemas de mensagens instantâneas).
 - *Editor de ontologias:* usado por especialistas para a manipulação de gestão da ontologia e suas instâncias. Em geral, é uma atividade preparatória para o uso do sistema.
 - *Ferramenta de BPM:* integrada ao capturador de contextos, oferece a informação a respeito do contexto do usuário.

É importante salientar que tal arquitetura não foca na forma de implementação das funcionalidades. Aspectos como anotação semântica automática, busca, usabilidade, segurança, entre outros, não foram definidos aqui mas devem ser especificados em nível de implementação. Por exemplo, no que diz respeito à anotação semântica automática, há diversas técnicas para tal, conforme mencionado na seção 3.6.

5.6.1 Casos de Uso

Os seguintes casos de uso foram definidos: (1) Gestão da ontologia e (2) Gestão do modelo de contexto e regras, ambos executados por um Especialista no Domínio; (3) Publicação de documentos, feita por um Provedor de Informações; e (4) Busca de Informação, realizada por um usuário final (membro de uma RC). O diagrama de casos de uso pode ser visto na figura 5.12.

Atenção especial deve ser dada ao caso de uso “Busca de Informação”, pois nele é possível posicionar o funcionamento do processo de expansão da consulta. A sequência de ações apresentadas abaixo explica como tal abordagem funciona, de acordo com a arquitetura conceitual:

1. O usuário define a consulta usando o *aplicativo de busca*;
2. O *aplicativo de busca* solicita ao serviço de *captura de contexto* qual o contexto atual do usuário;
3. A consulta e o contexto do usuário são enviados ao serviço *motor de busca*;
4. O *motor de busca* repassa a consulta e o contexto para o *motor de regras*;

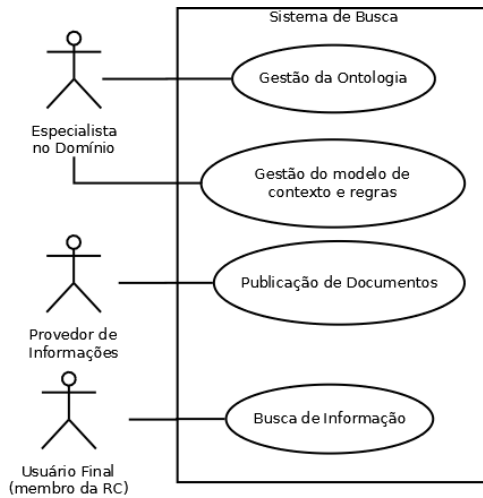


Figura 5.12: Casos de uso do sistema de busca.

5. O *motor de regras* aplica as regras de acordo com o contexto do usuário e envia a nova consulta ao *motor de busca*;
6. O *motor de busca* envia a nova consulta ao *aplicativo de busca* para que o usuário possa fazer modificações, caso necessário;
7. A nova consulta é então enviada de volta ao *motor de busca*;
8. O *motor de busca* processa a consulta e retorna a lista de resultados;
9. O usuário pode navegar nos resultados ou definir uma nova consulta.

5.7 Considerações

Este capítulo apresentou uma abordagem para expansão de consultas com base em um modelo para representar o contexto e de regras para seleção de novas restrições à consulta original. Uma vantagem dessa abordagem é que a expansão da consulta pode, além de facilitar o processo de definição de consultas semânticas, também pode auxiliar os usuários a descobrirem novos tópicos a serem buscados.

A abordagem também se baseou em uma ontologia de topo para RCs cujas classes estão associadas aos elementos do modelo de contexto. Desse modo, as regras são muito genéricas, mas que podem ser substituídas ou utilizadas em conjunto com novas regras, de acordo com requisitos particulares de domínio e/ou aplicação, conforme ilustrado no esquema da figura 5.13.



Figura 5.13: Níveis da ontologia em relação às regras.

Um detalhe a respeito do modelo é que, embora na prática o usuário possa estar em múltiplos contextos simultaneamente, é assumido que este está num único contexto.

Capítulo 6

Protótipo Computacional

Este capítulo apresenta a implementação do protótipo computacional, com base na arquitetura conceitual apresentada no capítulo anterior. Conforme definido na metodologia, a abordagem adotada para o desenvolvimento do trabalho foi dividida em duas etapas. A primeira etapa consiste na instanciamento da arquitetura utilizando-se ferramentas e soluções disponíveis. Esta etapa cobre a perspectiva semântica apresentada no modelo conceitual. Finalmente, a segunda etapa envolve a implementação da perspectiva de contexto.

6.1 Arquitetura de Implementação

A implementação da arquitetura conceitual para busca de informação em RCs foi feita com base numa arquitetura *Service Oriented Architecture* (SOA), usando serviços Web, e os serviços em si foram escritos em linguagem Java. Esta arquitetura foi denominada *Knowledge Search Services*, ou simplesmente *K-search*, termo que será utilizado a partir de agora para se referir a esta implementação. A figura 6.1 mostra os elementos que compõem a arquitetura de implementação do *K-search*.

As interfaces de implementação dos serviços (codificadas em *Web Service Definition Language* (WSDL)) encontram-se no apêndice A.

6.1.1 Tecnologias e Ferramentas Utilizadas

A arquitetura conceitual foi instanciada com base em uma série de tecnologias e ferramentas, a saber:

- Linguagem Java, usando o ambiente de desenvolvimento Eclipse¹.
- Padrões e ferramentas para serviços Web, em particular o Apache Tomcat². Conforme identificado no Capítulo 2, uma arquitetura orientada a

¹<http://www.eclipse.org/>

²<http://tomcat.apache.org/>

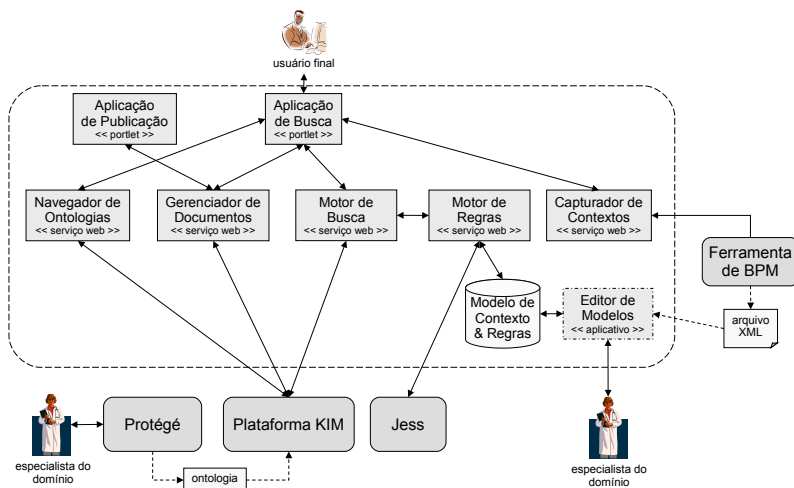


Figura 6.1: Arquitetura de Implementação do *K-search*.

serviços (SOA) é um requisito tecnológico essencial para a implementação de infraestruturas para RCs.

- O portal Liferay³, bem como a implementação de *portlets* em Java.
- Linguagem para representação de ontologias, em particular os padrões RDF e OWL.
- Editor de ontologias Protégé⁴.

Adicionalmente, duas ferramentas de suporte foram utilizadas na implementação: a Plataforma KIM, que serviu de base para a perspectiva semântica; e o Jess, que permitiu a implementação do motor de regras para implementação da perspectiva de contexto. Essas ferramentas são detalhadas nas subseções a seguir.

Plataforma KIM

Conforme mencionado na revisão bibliográfica, a Plataforma KIM foi escolhida para implementar parte da arquitetura, visto que esta ferramenta oferece suporte para a *perspectiva semântica*: gestão de ontologias, anotação semântica automática, indexação e busca de documentos (Popov et al., 2003). Todas essas funcionalidades são oferecidas pela Plataforma KIM através de

³<http://www.liferay.com/>

⁴<http://protege.stanford.edu/>

um servidor que oferece um conjunto de APIs (acessíveis via Java *Remote Method Invocation* (RMI)). Estas APIs serão detalhadas mais adiante, quando a implementação dos serviços for descrita. KIM também oferece uma ontologia de alto nível e uma base de conhecimento que proporciona a cobertura de entidades de importância geral.

Outro detalhe interessante a favor da Plataforma KIM é que algumas de suas funcionalidades são suportadas por outras plataformas. Estas plataformas são de código aberto e representam implementações largamente utilizadas, cada qual com um propósito específico. São elas:

- **GATE**⁵: plataforma para processamento de texto, que oferece funcionalidades personalizáveis e escaláveis para o processo de Extração de Informação (EI), bem como anotação e gerenciamento de documentos (Cunningham et al., 2002; GATE, 2007). Esta plataforma foi mencionada na seção 3.6, pois tem sido utilizada por várias plataformas de anotação semântica para dar suporte à EI e à anotação em si. Para ser utilizada pela KIM, GATE teve vários de seus módulos modificados para que trabalhassem com base em ontologias.
- **Sesame**: plataforma que pode ser usada para armazenamento de ontologias ou como uma biblioteca Java para aplicações que precisem trabalhar com ontologias internamente (Broekstra et al., 2002; Sesame, 2007). É usada pela Plataforma KIM para armazenar a ontologia e suas instâncias. Sesame trabalha com padrões como RDF, RDFS e *Notation3*.
- **Apache Lucene**: máquina de busca escrita em Java (Lucene, 2007). Como sua busca é baseada em palavras-chave, foi adaptada e utilizada pela KIM para implementar a funcionalidade de busca semântica.

Outro aspecto interessante da Plataforma KIM é que ela utiliza uma ontologia chamada *PROTo ONtology* (PROTON)⁶. Desenvolvida no escopo do projeto SEKT (SEKT, 2006), PROTON é uma ontologia de propósito geral, contendo mais de 300 classes e 100 propriedades que cobrem a maioria dos conceitos de mais alto nível, necessários para anotação semântica, indexação e busca. PROTON também proporciona independência de domínios, definições lógicas, alinhamento com padrões populares e boa cobertura de nomes e entidades de domínios concretos (por exemplo, pessoas, organizações, localidades, números, datas e endereços). Esta ontologia é codificada em *OWL Lite* e dividida em quatro módulos: i) *System*; ii) *Upper*; iii) *Top*; e iv) *KM* (Terziev et al., 2005). O propósito desta divisão é permitir que se

⁵General Architecture for Text Engineering

⁶<http://proton.semanticweb.org/>

possa adicionar facilmente extensões específicas a um domínio em particular (aspectos de integração são apresentados na seção 6.3).

A Plataforma KIM também oferece uma base de conhecimento, que corresponde às instâncias da ontologia PROTON. Esta base de conhecimento cobre várias entidades de importância geral e é utilizada no processo de anotação semântica para auxiliar no reconhecimento de entidades nomeadas. Tais entidades incluem:

- Lugares: cidades, países, continentes, etc.
- Organizações: comerciais, acadêmicas, esportivas, governamentais, etc.
- Pessoas específicas, governantes, gestores de empresas, etc.

Jess

Jess é um motor de regras e ambiente de escrita de *scripts* desenvolvidos na linguagem Java por Ernest Friedman-Hill (Jess, 2008). Usando Jess, é possível desenvolver aplicações em Java que sejam capazes de processar o conhecimento expresso na forma declarativa (regras). Jess implementa uma versão melhorada do algoritmo Rete, que é eficiente na solução de problemas de associação N-para-N de padrões (Forgy, 1982). Jess é composto por:

- Um motor de regras e APIs para a sua utilização em Java; ou seja, é possível fazer o Jess processar objetos em Java diretamente.
- Uma linguagem de *script* que permite o acesso às APIs do Java; de maneira alternativa, é possível manipular objetos sem a necessidade de se escrever código Java.
- Ambiente de desenvolvimento, disponível na forma de *plugin* para o ambiente Eclipse. Tal ambiente oferece as seguintes facilidades: colorização com base na sintaxe; assistente de preenchimento automático; verificação de erros em tempo real; formatação automática de código; execução e *debug* de código.

Com base nessas características, a sua escolha para utilização neste trabalho foi natural, visto que possui implementação eficiente, é integrado ao Java e possui licença gratuita para uso acadêmico.

6.2 Implementação dos Serviços do *K-search*

Esta seção descreve os aspectos de implementação de cada um dos serviços do *K-search*. A figura 6.2 apresenta um diagrama de classes contendo as classes principais e os pacotes envolvidos. A implementação propriamente dita do *k-search* está no pacote `org.ecolead.icti.ks`, que contém uma

classe para cada um dos serviços oferecidos. Este pacote contém também sub-pacotes que contêm classes específicas para cada serviço. Tais pacotes serão detalhados a seguir. Este diagrama também mostra outros pacotes, que correspondem às APIs das ferramentas de suporte: os pacotes `org.openrdf` e `com.ontotext.kim.client`, da plataforma KIM, e o pacote `jess`, do Jess. Esse pacotes serão detalhados na medida que a implementação dos serviços correspondentes forem detalhadas.

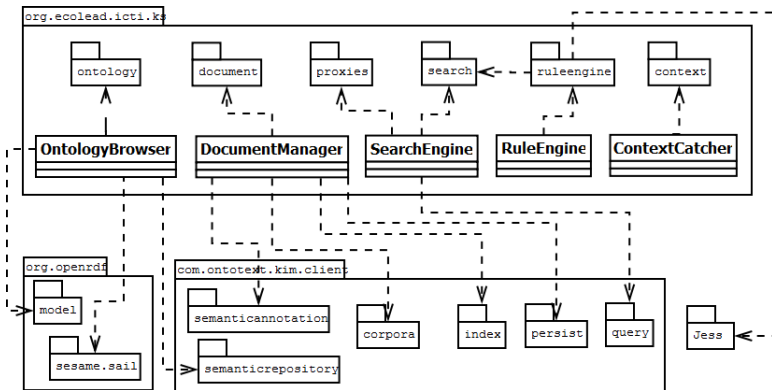


Figura 6.2: Diagrama de classes do *k-search*, mostrando as principais classes e pacotes.

Conforme pode ser visto pelas dependências das classes, os serviços de navegação de ontologias (*ontology browser*), gestão de documentos (*document manager*) e motor de busca (*search engine*) são suportados pelas APIs da plataforma KIM, e o motor de regras (*rule engine*) depende da API do Jess. O capturador de contextos (*context catcher*) não possui dependências, visto que não foi integrado a um sistema de gestão de processos. Os diagramas UML detalhando cada um dos serviços são apresentados no apêndice B.

6.2.1 Serviço para Navegação de Ontologias

Os serviço de navegação de ontologias é implementado pela classe `OntologyBrowser` e suas operações são listadas a seguir:

- **Carregar todas as classes da ontologia:** retorna uma lista contendo as URIs de todas as classes da ontologia.
- **Obter as superclasses:** dada uma URI (que identifica uma classe), esta operação retorna uma lista contendo as URIs das superclasses.
- **Obter subclasses:** de maneira análoga à operação anterior, retorna as subclasses de uma classe.

- **Obter Propriedades:** retorna uma lista contendo as propriedades de uma classe.
- **Obter instâncias:** retorna as instâncias de uma dada classe.
- **Obter valores de uma propriedade:** dada uma instância e uma propriedade, esta operação retorna os valores que esta propriedade tem para a instância dada.
- **Obter classes:** dada uma instância, esta operação retorna as classes desta instância.

6.2.2 Serviço para Gestão de Documentos

Este serviço é implementado pela classe `DocumentManager`. As seguintes operações são oferecidas pelo serviço Gerenciador de Documentos:

- **Publicação de documentos:** usada por provedores de informação para indexarem seus documentos. O documento é passado como parâmetro (de um URL ou texto) e esta operação retorna um identificador único gerado dinamicamente. É interessante notar que várias funcionalidades da KIM são agrupadas para formar esta operação: primeiramente o documento é criado (`CorporaAPI`); em seguida ele é semanticamente anotado (`SemanticAnnotationAPI`); e finalmente ele é indexado e armazenado (`IndexAndPersistAPI`).
- **Carregamento de um documento:** dado o seu identificador, esta operação retorna o documento. Esta funcionalidade faz uso da capacidade de armazenamento oferecida pela KIM. Entretanto, ela só é útil em casos onde o provedor de informação não armazena localmente seus documentos, ou pode servir como uma cópia de segurança de documentos.
- **Atualização de um documento:** atualiza o documento armazenado na KIM, incluindo seu conteúdo, suas anotações semânticas, e seus índices.
- **Remoção de um documento:** remove o documento, suas anotações, e seus índices.
- **Carregamento das anotações semânticas de um documento:** dado o seu identificador, esta operação retorna uma lista contendo as suas anotações semânticas (objetos da classe `Annotation`).

6.2.3 Serviço Motor de Busca

O serviço motor de busca é implementado pela classe `SearchEngine` e oferece as seguintes operações:

- **Busca de documentos baseada em palavras-chave:** esta é uma busca semelhante à tradicional, onde uma consulta formada por palavras-chave é passada de parâmetro. Um diferencial nesta operação é que a consulta pode ser filtrada de acordo com as *features* (metadados) dos documentos. Por exemplo, pode-se procurar documentos cuja *feature* “Autor” contenha o termo “Rui”.
- **Busca de entidades:** dada uma palavra-chave e um operador de comparação, uma lista de entidades (instâncias da ontologia) é retornada. Quando passadas de parâmetro em outras operações, estas instâncias podem ser utilizadas para buscar documentos.
- **Busca de documentos baseada em uma entidade:** uma entidade da base de conhecimento (retornada pela operação anterior) é passada de parâmetro e os documentos que contenham anotações semânticas referentes a esta entidade são retornados.
- **Busca de documentos baseada em uma consulta semântica:** uma consulta semântica aplica restrições a classes e relações presentes na ontologia, de modo a buscar documentos que contenham anotações que casem com estas restrições. O modelo de implementação da consulta semântica representa os predicados do modelo conceitual na forma de objetos que definem *variáveis* e *restrições de relações*:
 - *Variáveis* são identificadores que associam nomes a classes e podem definir restrições nas propriedades das mesmas. No exemplo $x|Organization(x) \wedge locatedIn('Europe')$, a consulta contém duas variáveis, uma associada à Organização, chamada **ORG**, e outra à classe Lugar, chamada **LUG**. A variável **LUG** tem uma restrição na propriedade *nome*, que deve conter o termo “Europa”.
 - *Restrições de relações* vinculam duas variáveis com base em alguma relação definida entre as respectivas classes na ontologia. Seguindo o exemplo, a consulta terá uma restrição de relação entre as duas variáveis definidas, ou seja, **ORG** localizada em **LUG**. A figura 6.3 apresenta o diagrama de objetos dessa consulta semântica.

Com esta estrutura, é possível então definir consultas que levem em consideração a ontologia adotada, ou seja, consultas semânticas. Adicionalmente, é possível associar as variáveis às *features* dos documentos. Assim, assumindo que os documentos foram indexados contendo

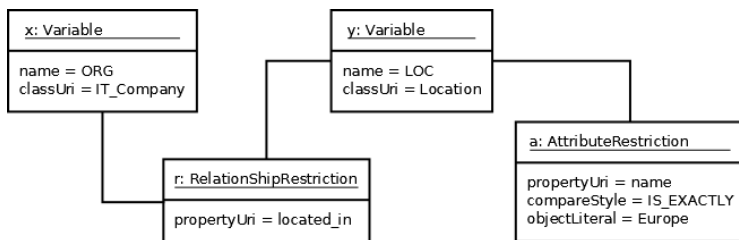


Figura 6.3: Diagrama de objetos para uma consulta semântica.

uma *feature* chamada “Título”, o exemplo acima apresentado poderia associar a variável **ORG** a esta *feature*, e assim buscar documentos que contenham organizações localizadas na Europa no seu título. Apesar de mais precisa, a desvantagem de uma consulta semântica é que a sua construção é trabalhosa e depende do conhecimento da ontologia por parte do usuário. A interface gráfica do cliente que utiliza este serviço deve, portanto, dar o suporte necessário para tornar este processo o mais fácil possível.

- **Busca com consulta semântica e contexto:** funciona quase da mesma maneira que a consulta semântica simples, com a diferença de que a consulta original do usuário é modificada de acordo com as regras de expansão e o contexto do usuário.

6.2.4 Serviço Capturador de Contextos

A classe `ContextCatcher` implementa o serviço de captura de contextos. Ele oferece uma única operação, responsável por recuperar o contexto do usuário, dado o seu identificador.

O contexto é modelado pela classe `GenericContext`, que armazena, além do identificador do usuário do seu contexto, um `Map`⁷ contendo os elementos de contexto e seus valores. Este tipo de estrutura de dados foi escolhida para tornar o modelo mais flexível, ou seja, qualquer elemento de contexto pode ser definido. Tal definição é feita num arquivo de configuração, chamado “*context.properties*”. A figura 6.4 mostra o conteúdo deste arquivo.

A primeira linha contém o número de elementos que serão considerados e as demais linhas definem os respectivos elementos. Neste caso, quatro elementos de contexto são definidos: *cno*, *process*, *task* e *role*. As configurações desse arquivo são carregadas pelo construtor da classe `GenericContext`. A lista de elementos de contexto pode ser recuperada através do método

⁷A interface `Map`, oferecida pela biblioteca padrão do Java, permite o armazenamento de coleções de pares nome-valor.


```
num_elements = 4
element_1 = cno
element_2 = process
element_3 = task
element_4 = role
```

Figura 6.4: Arquivo “*context.properties*”.

`getContextElements`. A leitura e escrita dos valores contidos nesses elementos é feita pelos métodos `getContextElement` e `setContextElement`, respectivamente.

Integração com ferramentas externas

Embora este trabalho não seja particularmente focado na captura do contexto, esta seção discute alguns aspectos referentes a esse tópico.

Conforme já mencionado, como o tipo de contexto utilizado neste trabalho é relativo a aspectos organizacionais e funcionais / de processo, a alternativa para a sua captura seria através da integração com ferramentas de suporte à gestão de processos de negócio (BPM). Em geral, esse tipo de plataforma provê funcionalidades para o monitoramento do estado dos processos em execução, de modo que dados como o processo, tarefa e papel do usuário podem ser recuperados.

De maneira complementar, plataformas de segurança podem também servir para a obtenção de alguns dos elementos de contexto em uma RC. Como parte dos requisitos do Projeto *European Collaborative Networked Organization LEADership Initiative* (ECOLEAD), o protótipo implementado neste trabalho integrou-se a uma plataforma de segurança especialmente desenvolvida para RCs, chamada DRACO (Sowa e Sniezynski, 2007). Essa plataforma serviu tanto para questões puramente de segurança (autenticação e autorização), mas também para a captura de alguns elementos de contexto: particularmente a organização e o papel do usuário.

Apesar de não serem utilizados no processo de avaliação deste trabalho, o sucesso na obtenção dos elementos a partir da plataforma de segurança serviu para mostrar a viabilidade técnica dessa abordagem. De maneira similar, há várias alternativas *open source* de ferramentas BPM que podem ser integradas ao *k-search*. Um exemplo é desse tipo de ferramenta é o Apache ODE (ODE, 2011), que é um motor de execução de processos de negócios que permite a execução e monitoramento de processos escritos na linguagem WS-BPEL.

6.2.5 Serviço Motor de Regras

O serviço motor de regras é implementado pela classe `RuleEngine`. Através do método `customizeQuery`, esta classe aplica as regras e expande a consulta de acordo com o contexto do usuário. Esta classe faz uso do motor de regras oferecido pela plataforma Jess e utiliza algumas de suas API.

O Arquivo de Regras

Conforme mencionado, a classe `Rete` implementa o motor de regras do Jess. Tal classe, quando instanciada, carrega os fatos e regras de um arquivo, para que então possa executar as regras ali definidas. O arquivo regras é dividido em várias seções⁸, a saber: *templates*, *metamodelo RDF*, *ontologia* e *regras*.

Templates: definem a estrutura dos fatos, tal como classes ou tabelas de bancos de dados. Um *template* é formado por *slots* que definem os campos que um dado fato deve possuir. A figura 6.5 apresenta alguns templates que definem a estrutura do modelo de contexto.

```

::: Declaring the templates for the Context Model
-----
(deftemplate formulatesQuery (slot user) (slot query))
(deftemplate isAbout (slot query) (slot class))
(deftemplate hasConstraint (slot query) (multislot constraint))
(deftemplate isIn (slot user) (slot context))
(deftemplate hasContextElement (slot context) (slot contextElement))
(deftemplate isAssociatedTo (slot contextElement) (slot class))

```

Figura 6.5: *Templates* para estrutura do modelo de contexto.

Metamodelo RDF/OWL: Como a ontologia é associada ao modelo de contexto, esta também é, armazenada na forma de fatos, dentro do arquivo de regras. A inclusão da ontologia no arquivo foi feita na forma de triplas, e teve como base o metamodelo RDF definido por Gandon (2007). Tal metamodelo pode então ser usado para a especificação de qualquer ontologia definida em RDF/OWL. A figura 6.6 abaixo mostra um trecho contendo o *template* que define o formato das triplas, e dois fatos relacionados ao metamodelo RDF.

Ontologia: além do metamodelo RDF, Gandon (2007) definiu uma folha de estilos XSL que permitiu a geração da ontologia do formato RDF para o formato de triplas. Embora seja uma transformação automática, uma limitação desta abordagem é que o arquivo de regras deve ser atualizado cada vez que a ontologia evoluir. Este aspecto pode ser melhorado com o uso

⁸Por questões de simplicidade, aqui são apresentadas apenas um trecho dessas seções. O arquivo de regras completo pode ser visto no apêndice D.

```

;;; RDFS and OWL Meta model
;;; Declaring the triple template
(deftemplate triple "Template representing a triple"
  (slot predicate (default ""))
  (slot subject (default ""))
  (slot object (default ""))
)

;;; Declaring facts of the meta-model of OWL
(deffacts OWLModel "Facts of the meta-model of RDFS and OWL"

;;; Resource is a Class
(triple
  (predicate "http://www.w3.org/1999/02/22-rdf-syntax-ns#type")
  (subject "http://www.w3.org/2000/01/rdf-schema#Resource")
  (object "http://www.w3.org/2002/07/owl#Class")

;;; Class is a subtype of Resource
(triple
  (predicate "http://www.w3.org/2000/01/rdf-schema#subClassOf")
  (subject "http://www.w3.org/2002/07/owl#Class")
  (object "http://www.w3.org/2000/01/rdf-schema#Resource")
)

```

Figura 6.6: Trecho do metamodelo RDF/OWL, definido na forma de triplas.

do mecanismo de integração apresentado na subseção a seguir. No entanto, devido à limitações de tempo, não foi possível implementar este aspecto no protótipo. A figura 6.7 apresenta um trecho da Ontologia de Indicadores de Desempenho, codificada na forma de triplas.

Regras: finalmente, as regras são definidas na sintaxe do Jess. A figura 6.8 apresenta a regra 5.1, definida na seção 5.5 (página 94), codificada em Jess. As demais regras encontram-se no apêndice D.

Integração do k-search com o motor de regras do Jess

Apesar do arquivo de regras possuir vários fatos pré-definidos, é preciso fazer uma troca de dados entre o *k-search* e o Jess. Isto porque as consultas e contextos são definidos de forma dinâmica e portanto não há razão para estarem pré-definidos no arquivo de regras.

A troca de dados é feita através de objetos cujas classes servem para a definição dos *templates*. No caso particular deste trabalho, tais objetos estão relacionados aos predicados que definem o modelo de contexto. A figura 6.9 apresenta a listagem dos *templates* que refletem esse mecanismo.

Assim, ao invés de definir os *templates* através de *slots* (tal como na figura 6.5), a sua definição é feita através da classe Java que é importada dentro do arquivo de regras. Essas classes estão definidas no pacote `ruleengine.model`, conforme o diagrama da figura 6.10.

```

;;; Performance Indicators Ontology
(assert

(triple
 (predicate "http://www.w3.org/1999/02/22-rdf-syntax-ns#type")
 (subject "http://www.gsigma.ufsc.br/pi-co#CollaborativeNetwork")
 (object "http://www.w3.org/2002/07/owl#Class")
)

(triple
 (predicate "http://www.w3.org/1999/02/22-rdf-syntax-ns#type")
 (subject "http://www.gsigma.ufsc.br/pi-co#PerformanceIndicator")
 (object "http://www.w3.org/2002/07/owl#Class")
)

(triple
 (predicate "http://www.w3.org/1999/02/22-rdf-syntax-ns#type")
 (subject "http://www.gsigma.ufsc.br/pi-co#Domain")
 (object "http://www.w3.org/2002/07/owl#Class")
)

```

Figura 6.7: Trecho da Ontologia de Indicadores de Desempenho, codificada na forma de triplas.

```

;;; Declaring the rules for query expansion -----
(defrule R1
  "associates queries to contexts"
  (FormulatesQuery (user ?u) (query ?q))
  (IsIn (user ?u) (context ?cx))
  =>
  (assert (IsFormulatedIn (query ?q) (context ?cx)))
  (printout t ?q " is formulated in Context " ?cx "(R1)" crlf)
)

```

Figura 6.8: Regra 1 (associação da consulta ao contexto) codificada na sintaxe do Jess.

Através de objetos dessas classes, é possível então inserir novos fatos no motor de busca do Jess. Desse modo, é possível converter os objetos que representam a consulta e o contexto na forma de predicados, que são depois adicionados ao motor de regras. De maneira similar, novos fatos gerados pelas regras podem ser recuperados – também na forma de objetos – e então convertidos para restrições da consulta.

6.3 Integração com a Ontologia PROTON

Conforme mencionado, a Plataforma KIM é baseada na ontologia PROTON. Tal ontologia é dividida em módulos, permitindo assim uma maior flexibilidade de modificação, seja por substituição ou por extensão, através de ontologias de um domínio em particular. Ela é separada em três módulos

```

(import org.ecolead.icti.ks.ruleengine.model.*)

;;; Declaring the templates for the Context Model

(deftemplate FormulatesQuery      (declare (from-class FormulatesQuery))
 (deftemplate IsAbout             (declare (from-class IsAbout)))
 (deftemplate HasConstraint       (declare (from-class HasConstraint)))
 (deftemplate IsIn               (declare (from-class IsIn)))
 (deftemplate HasContextElement  (declare (from-class HasContextElement))
 (deftemplate IsAssociatedTo     (declare (from-class IsAssociatedTo)))

;;; Declaring the templates for the rules

(deftemplate IsFormulatedIn       (declare (from-class IsFormulatedIn))
 (deftemplate IsAboutPointedConcept
   (declare (from-class IsAboutPointedConcept)))
 (deftemplate IsAboutPointedConceptFromContext
   (declare (from-class IsAboutPointedConceptFromContext)))
 (deftemplate IsInSubClassHierarchy
   (declare (from-class IsInSubClassHierarchy)))
 (deftemplate IsInSuperClassHierarchy
   (declare (from-class IsInSuperClassHierarchy)))

```

Figura 6.9: *Templates* para estrutura do modelo de contexto, importadas de classes Java.

(KimDocs, 2010):

- **Módulo “System”**: contém algumas primitivas no nível “meta”, que são normalmente elementos específicos dentro do código de aplicações baseadas em ontologias. Este nível introduz a noção de “entidade” que pode ter vários apelidos (*aliases*).
- **Módulo “Top”**: nível conceitual mais alto e mais geral. Consiste de cerca de 20 classes, que asseguram um bom equilíbrio entre utilidade, independência de domínio e facilidade de entendimento e uso. Este nível é normalmente o melhor para o estabelecimento de um alinhamento com outras ontologias.
- **Módulo “Upper”**: contém mais de 200 classes de entidade, que normalmente aparecem em múltiplos domínios (por exemplo, vários tipos de organizações, uma ampla gama de lugares, etc.)

Com base nesses módulos, a extensão da ontologia PROTON pode ser feita de duas maneiras. A primeira consiste na substituição (parcial ou total) dos módulos “Top” e “Upper” por módulos customizados para um domínio.

Outra maneira é a integração de uma ontologia de domínio de tal forma que as suas classes herdem pelo menos da classe “Entity”. Tal herança pode ser feita direta ou indiretamente, e recomenda-se, se aplicável, que as

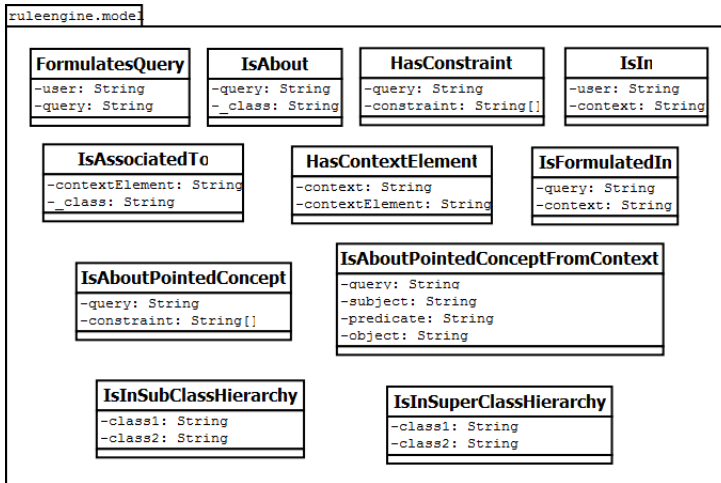


Figura 6.10: Classes utilizadas para a integração do *k-search* como o Jess.

novas classes herdem uma dessas classes do módulo “*Top*”: Pessoa (*Person*), Organização (*Organization*) ou Lugar (*Location*).

Para a integração com a ontologias de RCs, a integração foi feita utilizando a segunda abordagem. Foi criada uma ontologia usando a ferramenta Protégé, onde os módulos “*Top*” e “*Upper*” foram importados. As relações de herança entre as classes são apresentadas na tabela 6.1, onde as classes na coluna à esquerda herdam das classes na coluna à direita.

| Classe na Ontologia de RCs | Classe na KIMO |
|-----------------------------|--------------------------------------|
| Organização | Organization (<i>Top</i>) |
| Produto | Product (<i>Top</i>) |
| Serviço | Service (<i>Top</i>) |
| Indicador de Desempenho | Abstract (<i>Top</i>) |
| Oportunidade de Colaboração | Abstract (<i>Top</i>) |
| Objeto Mensurado | Object (<i>Top</i>) |
| Recurso | Object (<i>Top</i>) |
| Recurso Humano | Person (<i>Top</i>) |
| Domínio | BusinessAbstraction (<i>Upper</i>) |

Tabela 6.1: Pontos de integração entre a ontologia de RCs e a KIMO.

6.4 Demais Elementos da Arquitetura

6.4.1 Aplicativos

Há dois clientes especialmente desenvolvidos para os serviços do *K-search*, que servem tanto para serem aplicações de utilidade geral, como também para servirem de base para a integração de aplicações com os serviços do *K-search*. Eles foram implementados também em Java e na forma de *portlets*, o que permite a fácil instalação e utilização em portais. São eles:

- **Cliente para publicação de documentos:** permite a publicação de documentos a partir de um arquivo, um URL, ou de um texto, que é digitado pelo usuário. Depois de entrar com o documento, o usuário pode incluir uma lista de *features* com seus respectivos valores. A figura 6.11 mostra uma tela do *portlet* de publicação.

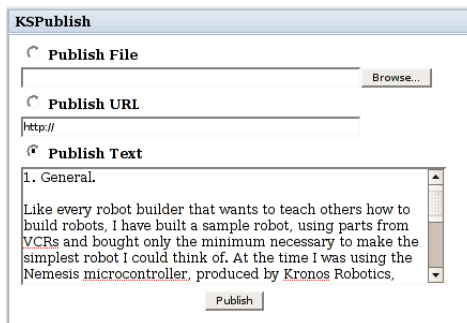


Figura 6.11: *Portlet* para publicação de documentos.

- **Cliente para buscas semânticas:** este cliente faz uso dos serviços de navegação na ontologia para auxiliar o usuário na construção de consultas semânticas, da máquina de busca para executar a consulta, e no indexador de documentos, para recuperar documentos (lembrando que nesta implementação, o indexador de documentos também os armazena). Os três tipos de consultas oferecidas pelo serviço *Máquina de Busca* podem ser feitas pelo usuário: busca baseada em palavra-chave, busca baseada em entidades e busca baseada em consulta semântica. A figura 6.12 mostra três telas do *portlet* de publicação, ilustrando um exemplo de busca semântica.

6.4.2 Elementos Externos

- **Provedores de informação:** pode ser qualquer tipo de aplicação que lide com documentos: aplicações que lidam com e-mail, grupos de dis-



Figura 6.12: *Portlet* para busca semântica: (a) definição da consulta; (b) resultados; (c) visualização de um documento.

cussão, gerenciadores de documento em geral, ferramentas de CSCW, entre outras. Particularmente no âmbito do projeto ECOLEAD, há um outro conjunto de serviços da ICT-I que foi desenvolvido (por outro parceiro do projeto) para dar suporte à interação entre pessoas, que consiste basicamente de serviços para CSCW (e-mails, *instant messaging*, etc.) (Ratti e Rabelo, 2007; Rodrigo e Ratti, 2007). Este pacote de serviços foi integrado ao *K-search*, e todos os documentos publicados pelos serviços de CSCW são indexados (utilizando *features* para diferenciar o tipo de documento) e podem ser posteriormente buscados com base na máquina de busca do *K-search*.

- **Aplicações de RCs:** pode ser qualquer tipo de aplicação que necessite buscar e consumir informações.
- **Editor de ontologias:** esta aplicação pode ser qualquer editor que trabalhe com os padrões RDF/OWL. Para este trabalho, o editor de ontologias escolhido foi o Protégé.
- **Ferramenta de BPM:** como mencionado na seção 6.2.4, não foi feita a integração com uma ferramenta de BPM. No entanto, uma alternativa seria a integração com o Apache ODE (ODE, 2011)

6.5 Considerações sobre o Protótipo

Este capítulo apresentou a implementação do arcabouço para busca de informação em RCs, que leva em consideração aspectos de semântica e contexto. A plataforma foi desenvolvida na forma de uma arquitetura SOA, implementada como serviços Web. Ela oferece:

- Serviços para navegação na ontologia, publicação e busca de documentos. Além disso, há serviços para captura de contextos e execução das regras.
- Aplicativos cliente, na forma de *portlets*, para publicação de documentos e para busca.

É importante salientar que tanto os serviços quanto as aplicações são completamente funcionais, tendo sido validadas no âmbito do ECOLEAD. Além disso, as aplicações podem servir de base para implementações customizadas. Dentre as dificuldades na implementação, pode-se destacar:

- Integração com as API do KIM, pois há detalhes não triviais. Por exemplo, leitura dos dados da ontologia, manipulação de documentos (leitura e manipulação dos seus identificadores), utilização da estrutura de dados da consulta semântica.
- Integração com a ontologia do KIM, apesar das facilidades oferecidas.
- Integração do motor de regras do Jess com o modelo de dados interno do *k-search*, particularmente ao modelo da consulta semântica.

Além disso, apesar de ser um esforço de caráter puramente técnico, foi preciso o conhecimento de várias tecnologias, tais como, serviços Web, representação de ontologias (RDF / triplas), a ferramenta Protégé, implementação das regras em Jess e uso das APIs do Jess e do KIM.

Além disso, foi agregado um certo valor sobre as APIs dos KIM, no sentido de tornar a sua utilização mais abstrata / simples. Em outras palavras, várias das suas funcionalidades foram agregadas num único serviço, ou mesmo numa única operação de um serviço (como no caso do serviço de gestão de documentos). Embora possa-se questionar a perda de poder na utilização do *k-search* em relação ao uso da plataforma KIM, esse aspecto era desejável para a integração com os demais sistemas no âmbito do projeto ECOLEAD. Da forma como foi feito, vários aspectos (anotação semântica e indexação, por exemplo) se tornaram mais abstratos, o que facilitou a integração com outros serviços.

Dentre as limitações do protótipo, destacam-se o fato da ontologia ser replicada no arquivo de regras, o que torna a propagação de mudanças na

ontologia uma tarefa mais trabalhosa. Além disso, outra limitação diz respeito à captura do contexto. Para fins de execução e avaliação, o serviço de captura de contextos já terá definida *a priori* a lista de contextos a serem utilizados nos experimentos. É importante ressaltar, todavia, que do ponto de arquitetura, os demais serviços acessam de maneira transparente o serviço de captura. Este desacoplamento permitirá que esta limitação seja mais facilmente resolvida no futuro, pois apenas este serviço terá que ser reimplementado. Além, disso, tal integração mostrou-se tecnicamente viável. Outro aspecto que merece ser melhorado é a parte referente a interfaces com o usuário, tanto na definição de consultas semânticas, quando na apresentação dos resultados (incluindo também aspectos de contexto).

É importante salientar que se trata de uma pesquisa experimental e que as várias simplificações feitas não impactaram no funcionamento correto do sistema. Ou seja, a despeito das limitações, o protótipo foi suficiente para avaliar a abordagem adotada para lidar com o problema de pesquisa.

Capítulo 7

Avaliação do Modelo

Este capítulo apresenta a avaliação do modelo de contexto proposto, com base em experimentos no protótipo computacional. De acordo com as etapas descritas na seção 1.7.2, a avaliação deu-se em duas etapas, conforme a metodologia incremental definida:

1. Avaliação da perspectiva semântica, ou seja, do modelo e protótipos parciais através da sua integração com outros serviços e aplicações (projeto ECOLEAD).
2. Avaliação da perspectiva de contexto, em termos de medição do desempenho (usando-se medidas de RI) do protótipo.

7.1 Avaliação da Perspectiva Semântica

A primeira avaliação deste trabalho foi feita na versão parcial do arcabouço proposto (cobrindo a perspectiva semântica), no âmbito do projeto ECOLEAD¹. O ECOLEAD foi um dos projetos integrados do Sexto Programa Quadro da Comissão Europeia (FP6 IP 506958). O projeto iniciou em abril de 2004, contou com a participação de 26 parceiros de 14 países diferentes e teve seu término em junho de 2008.

O objetivo do projeto ECOLEAD foi criar fundamentos e os mecanismos necessários para estabelecer a mais avançada sociedade industrial europeia baseada em redes e colaboração. O pressuposto fundamental deste projeto é que um impacto substancial na materialização de redes colaborativas de organizações requer uma abordagem holística ampla. Dada a complexidade da área e as múltiplas interdependências entre as entidades de negócios além dos atores sociais e as abordagens tecnológicas envolvidos, inovações substanciais não podem ser atingidas através de inovações incrementais em

¹ Acrônimo para *European Collaborative Networked Organization LEADership Initiative* (Iniciativa Europeia para a Liderança em Organizações em Redes Colaborativas), <http://www.ecolead.org>

áreas isoladas. Assim sendo, esse projeto atuou em três áreas fundamentais e inter-relacionadas, que são a base para organizações em rede: Ambientes de Criação de Organizações Virtuais, Organizações Virtuais e Comunidades Virtuais Profissionais.

Nesse projeto, uma infraestrutura de TIC para o suporte de RCs foi desenvolvida e uma das funcionalidades oferecidas foi a busca de informação em RCs. Essa funcionalidade, por sua vez, teve como base o arcabouço proposto neste trabalho. Para o projeto, o desenvolvimento do arcabouço de busca cobriu a perspectiva semântica (conforme apresentado na metodologia), e foi validado no âmbito deste projeto, dando suporte a algumas aplicações e serviços de mais alto nível para RCs.

Tal validação foi feita através do uso por outros serviços e aplicações no ECOLEAD, sendo avaliado tanto em termos de **arquitetura** quanto em termos de **funcionalidade**. No primeiro aspecto, a abordagem baseada em uma *arquitetura orientada a serviços* foi testada através de integrações bem sucedidas com outros serviços e aplicações. Já no ponto de vista de funcionalidade, a *perspectiva semântica* da plataforma serviu, em certos casos, para agregar valor a outros serviços aplicações. Os seguintes módulos foram integrados ao arcabouço proposto:

- Plataforma de serviços para CSCW: utilizou o serviço de maneira a usufruir dos benefícios desse tipo de busca. Por um lado, serviu como provedor de informações e, por outro, como cliente para buscas. A interface padrão de busca dessa plataforma foi adaptada para realizar as buscas semânticas.
- Serviços e aplicações para busca de indicadores de desempenho: esta plataforma adaptou os serviços do *k-search* para uso em um domínio em particular, onde uma ontologia e base de conhecimento específicas serviram para dar suporte à busca por indicadores de desempenho armazenados em um banco de dados.
- Plataforma de suporte ao aprendizado organizacional: o *k-search* foi integrado como parte de uma das funcionalidades dessa plataforma.

Outro ponto que também contribuiu para a avaliação positiva deste trabalho foi a publicação do mesmo em conferências reconhecidas internacionalmente bem como em uma revista internacional indexada. Isto demonstra que o trabalho foi avaliado pela comunidade de pesquisadores da área de RCs e afins, creditando os resultados aqui obtidos. A lista de publicações produzidas no escopo de realização deste trabalho pode ser vista no capítulo 8.

7.2 Avaliação da Perspectiva de Contexto

a versão final do modelo foi avaliada através de experimentos e determinados cenários, com base em medidas de desempenho. Como foi mencionado no capítulo introdutório, o problema de pesquisa foi abordado de maneira essencialmente quantitativa, conforme a metodologia tradicional para avaliação de sistemas de RI, tal como descrito na seção 4.4. No entanto, há também um aspecto qualitativo nesta análise, pois os critérios de relevância – que é uma das bases para a medição do desempenho – foram obtidos através de questionários a usuários, especialistas na área de RCs.

Para avaliação de um sistema baseado em contexto, o que talvez seria a abordagem ideal consiste na implantação da plataforma num ambiente operacional, onde os usuários, além de usarem o sistema, dariam o *feedback* necessário de alguma maneira. Isso poderia ser feito implicitamente através de *logs* de execução, ou explicitamente, com um sistema de avaliação respondido pelo usuário.

No entanto, essa abordagem é mais trabalhosa de se implementar, demanda mais tempo e, principalmente, exige a existência de tal ambiente operacional. Isso se tornou inviável, dadas as limitações tanto do protótipo (a captura do contexto, por exemplo) bem como da falta de RCs reais que ofereçam tal ambiente. Além disso, o protótipo tem muitas limitações em nível de interface com o usuário, o que tornaria difícil esse tipo de análise, mesmo que em um ambiente reduzido.

Portanto, a abordagem que se mostrou viável para a avaliação foi a execução de uma série de consultas em lote e a expansão que é sugerida não é controlada pelo usuário. É interessante ressaltar que, embora se perca na flexibilidade, por outro lado tem-se um critério uniforme de relevância. Desse modo, é possível interpretar de maneira uniforme as medidas de desempenho.

Mesmo assim, o critério de relevância será definido pelo usuário, mas será feito *a priori*. Ou seja, ao invés de usar o sistema e dar opiniões sobre seu uso, ou indicar se tal documento retornado é ou não relevante, o usuário será questionado com várias consultas, e indicará quais documentos ele espera em retorno. Tal questionário explica o cenário no qual o usuário está inserido, bem como o seu contexto de trabalho e o tipo de consulta que está sendo executada. Assim, uma vez entrando nesse contexto, o usuário escolhe quais documentos são relevantes para a busca. O critério de relevância definido pelo usuário é então usado para se fazer análises quantitativas (baseadas em precisão e cobertura).

Os experimentos cobrem dois cenários básicos de busca: *busca por indicadores* e *busca por documentos*. A escolha de tais cenários foi feita visando aproveitar material existente produzido ao longo do ECOLEAD. O cenário de indicadores foi obtido do trabalho de Baldo (2008), e o cenário de documentos foi obtido a partir de material extraído de relatórios de projeto.

7.2.1 Metodologia de Avaliação

Como o objetivo é medir o ganho que o contexto oferece ao sistema em termos de desempenho, a qualidade dos resultados será avaliada através de medidas utilizadas na literatura para sistemas de RI: precisão e cobertura, bem como as demais medidas apresentadas na seção 4.4. Portanto, a metodologia de avaliação consiste na preparação de uma coleção de teste, na execução de consultas, e a posterior análise dos resultados.

Apesar desse tipo de avaliação ser dita quantitativa, pois é feita com base nas medidas típicas usadas em sistemas de RI, a preparação da coleção de teste requer a opinião de especialistas, que vão definir quais documentos são relevantes para cada consulta (no seu respectivo contexto). Portanto, a avaliação feita neste trabalho é dita *qualitativa* e *quantitativa*.

As etapas da avaliação compreendem a *preparação*, a *execução*, e a *análise dos resultados*, conforme detalhado a seguir.

Preparação: esta etapa consiste na preparação dos experimentos, ou seja, as coleções de teste, o arcabouço de semântica e contexto, e a indexação dos documentos.

1. População da base de conhecimento. elemento fundamental que permite o processo de anotação semântica automática (na etapa de identificação de instâncias da ontologia), bem como nas busca semânticas. A seção 7.3.1 apresenta a base de conhecimento utilizada.
2. Preparação do modelo de contexto. Vinculação do modelo de contexto à ontologias de RCs, conforme detalhado na seção 7.3.4.
3. Definição das coleções de teste. São duas coleções focadas em RCs, conforme mostra a seção 7.3.2. Cada coleção é utilizado em cenário diferente:
 - (a) *Indicadores de Desempenho:* a ideia aqui é aproveitar uma coleção de documentos existente, bem como ontologia e base de conhecimento (Baldo, 2008). tal coleção também é focada na área de RCs, mais especificamente, em OVs.
 - (b) *Deliverables do projeto ECOLEAD:* formada por seções de alguns relatórios selecionados. A justificativa para o seu uso é devido à ausência de uma coleção adequada para avaliar este trabalho, bem como o fato de seu conteúdo estar relacionado à área de RCs e poder ser facilmente anotado semanticamente e melhor contextualizado.
4. Definição das consultas: para cada cenário, uma série de consultas foi definida (ver seção 7.3.3).

5. Definição dos critérios de relevância: feita por especialistas na área, com base em um questionário, conforme mostrado na seção 7.3.5.
6. Indexação dos documentos. Esta tarefa é executada em lote, usando um cliente do *K-search*.

Execução: consiste na execução das consultas e captura dos resultados. As consultas são executadas em lote: aspectos de interface de usuário não interessam. O processo é executado da seguinte forma:

- Para cada configuração do *K-search* (*palavra-chave, semântica, semântica + contexto*):
 - Processamento de todas as consultas e retorno dos resultados;
 - Resultados ficam todos num arquivo, para posterior análise.

Justificativa para esta abordagem: dados os requisitos específicos deste trabalho, não é possível fazer uma comparação de desempenho com outros sistemas. No entanto, como as duas primeiras configurações de busca (*palavra-chave, semântica*) são oferecidas pela plataforma KIM, a abordagem proposta pode ser comparada com relação a essa plataforma.

Análise dos Resultados: nesta etapa, os resultados das consultas são processados por uma ferramenta, e os relatórios contendo as medidas são interpretados. Esta etapa compreende as seguintes atividades:

- Para cada coleção de teste e para cada config. do *K-search*:
 - Geração dos relatórios de desempenho (medidas).
- Geração dos relatórios comparativos entre as configurações do *K-search*;
- Interpretação dos resultados.

A figura 7.1 apresenta a visão geral dessa metodologia, incluindo as ferramentas envolvidas em cada etapa, o fluxo de dados entre elas, e a ordem de execução (indicada por números).

7.3 Preparação dos Experimentos

A preparação dos experimentos foi dividida nas seguintes etapas:

1. Definição de instâncias da ontologia, para alimentar a base de conhecimento: ACVs, OVs, empresas, processos, recursos, indicadores, entre outros.

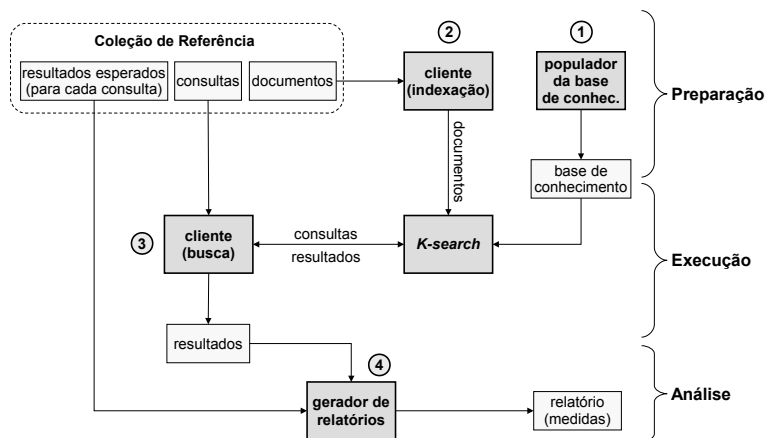


Figura 7.1: Metodologia de Avaliação.

2. Seleção de documentos para a coleção de teste.
3. Definição das instâncias de contextos.
4. Definição das consultas.
5. Definição dos critérios de relevância.

7.3.1 Instâncias para a Base de Conhecimento

A base de conhecimento foi alimentada com instâncias da ontologia de RCs, de modo que os documentos a serem buscados possam ser anotados. Essas instâncias foram extraídas de relatórios do Projeto ECOLEAD, particularmente com dados referentes às empresas piloto do projeto. Os indicadores de desempenho foram obtidos do trabalho de Baldo (2008).

As instâncias são enumeradas a seguir, e são organizadas de acordo com o seu tipo.

Redes Colaborativas:

- ACVs: IECOS, Helice Foundation, CeBeNetwork.
- OVs: Virtuelle Fabrik, OIN (Orona Innovation Network).

Organizações Participantes:

- IECOS: IPACSA, Especialidad y Repuestos Petroquimicos, Alta Tecnologia en Maquinados, Maquinados AML, Polimaquinados Moreno, Orona, EIC, ELECTRA Vitoria Group, MONDRAGON CORPORACION COOPERATIVA.

- Helice: ISOIN, MEUPE, INESPASA, AEROSUR, GHESA, Regional Development Agency.
- CeBeNetwork: EuroEngineering aerospace, P+Z, Aerocon, AIDA Development.
- OIN: Orona, EIC, IKERLAN, Mondragón Faculty Of Engineering, ELECTRA Vitoria Group, MONDRAGON CORPORACION COOPERATIVA, Herreros and Associates, ZABALA.

Processos:

- IECOS: Registration of new members, VBE Performance management, VO Creation.
- Helice: VBE Characterization, VO Creation, VO Launch.
- Virtuelle Fabrik: VO Performance Measurement, Integrated VO Management.
- OIN: OIN Roadmapping, VO Management.

Recursos Humanos:

- ISOIN: Carmen M. Aguilera, Jose A. Parejo, Ricardo Galan.
- HELICE: Simón Vázquez.
- CeBeNetwork: Berthold Tiefensee, Daniel Lässig.

Recursos Tecnológicos (*software*):

- Membership and Structure Management System (MSMS)
- Profiling and Competency Management System (PCMS)
- Ontology and Discovery Management System (ODMS)
- Trust Management System (TrustMan)
- Virtual Organization Information Management System (VIMS)
- Collaboration Opportunity Characterization and Rough Planning (COC-Plan)
- Partners Search and Suggestion (PSS)
- Decision Support System (DSS)

- Agreement Negotiation Wizard (WizAN)
- VO-Model (VO-Mod, VOMod)
- Decision Support System (DSS)
- Supporting Indicator Definitions (SID)
- Distributed Indicator Information Integrator (DI3)
- Monitor and Finance (MAF)

Relações entre RCs e Recursos Tecnológicos:

- IECOS: MSMS, PCMS, ODMS, TrustMan, DSS, VIMS, COC-Plan, PSS.
- Helice: PCMS, MSMS, WizAN, PSS.
- Virtuelle Fabrik: DI3, SID, VO-Mod Wizard.
- OIN: VO-Mod, SID, DI3, MAF, DSS.

Oportunidades de Colaboração (IECOS):

- Centrifuge Machine.
- Punch for biopsies.

Competências:

- IECOS: Manufacturing.
- Helice: Aerospace Engineering.

Indicadores de Desempenho:

- Percentage of Qualified Suppliers which Meet Defined Requirements
- Cost per Invoice
- Demand/Supply Planning Costs
- Incoming Production Material Quality
- Order Management Costs
- Order Management Cycle Time
- Package Cycle Time

- Packaging Cost
- Cost of storage space
- Production Plan Adherence
- Production Process Validation Frequency
- Percentage of Faultless production
- Production Material Cycle Time
- Deviation from planned start date
- Number of re-planning
- Problem Compensation performance of a VO member
- Reliability of VO member's statements / predictions
- Percentage of available of resources in relation to overall assigned resources
- Satisfaction with collaboration performance of a VO member
- Number of call backs as Percentage of total inquiries

A extração dessas entidades foi feita manualmente, sendo uma tarefa que consumiu um certo tempo. A consequência é que a base não ficou muito grande, assim como nem todas as entidades ficaram completamente consistentes no que diz respeito aos relacionamentos entre si. Por exemplo, a OV “Virtuelle Fabrik” não possui participantes e a única RC que possui oportunidades de colaboração é a “IECOS”. Algumas RC não se relacionam com recursos tecnológicos.

Algumas instâncias e relações não foram apresentadas aqui, como por exemplo, as tarefas, a relação entre cada processo (e suas tarefas) com os recursos tecnológicos, bem como sua relação com os papéis na RC. Esses detalhes são encontrados no apêndice C, que mostra a base completa.

Visão Detalhada da Base de Conhecimento

A base de conhecimento pode ser vista como um grafo, onde os vértices são as instâncias e as relações são os arcos. No entanto, a quantidade de instâncias e relações pode levar a um grafo muito grande, o que dificulta a sua visualização. Para dar um exemplo simples, a figura 7.2 apresenta um fragmento da base de conhecimento, contendo as instâncias referentes ao ACV “Helice”.

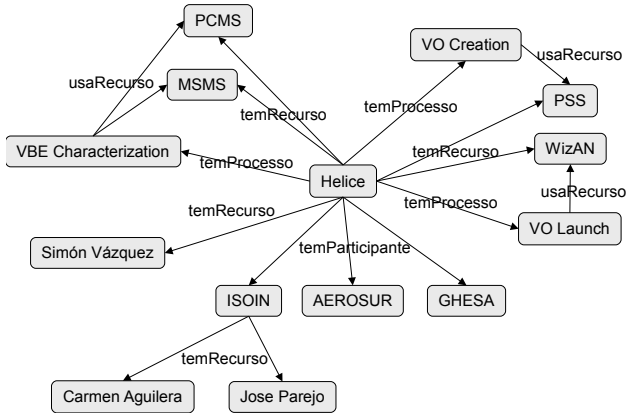


Figura 7.2: Fragmento da base de conhecimento, representado na forma de grafo.

A figura apresenta a entidade “Helice” circundada por outras instâncias, referentes a processos, recursos e participantes. É importante reparar que este grafo compreende somente uma pequena parte da base de conhecimento. Para facilitar a visualização, vários elementos foram omitidos. Por exemplo, as tarefas que compõem cada processo não foram incluídas. Além disso, outro aspecto implícito nessa figura é a relação entre cada instância e sua respectiva classe.

Outra maneira de representar a base de conhecimento é na forma de predicados. Nesse caso, um predicado representa cada arco ligando dois vértices. Por exemplo, a relação “Helice” tem o participante “ISOIN” pode ser expressa na forma: `temParticipante(Helice, ISOIN)`. A figura 7.3 apresenta o mesmo exemplo representado na forma de predicados.

```

temParticipante(Helice, ISOIN)
temParticipante(Helice, AEROSUR)
temParticipante(Helice, GHESA)
temRecurso(Helice, SimonVazquez)
temRecurso(ISOIN, CarmenAguilera)
temRecurso(ISOIN, JoseParejo)
temRecurso(Helice, MSMS)
temRecurso(Helice, WizAN)
temRecurso(Helice, PSS)
temProcesso(Helice, VBE_Characterization)
temProcesso(Helice, VO_Creation)
temProcesso(Helice, VO_Launch)

```

Figura 7.3: Fragmento da base de conhecimento, representado na forma de predicados.

A base de conhecimento é composta por mais de 150 instâncias e cerca de 280 relações, que podem ser consultadas no apêndice C.

Base de Conhecimento na Sintaxe usada pelo KIM

A forma de representação apresentada se aproxima de maneira como a base de conhecimento é de fato codificada e utilizada pela Plataforma KIM. Ela fica armazenada em um arquivo texto, onde cada predicado é codificado na forma de triplas (sujeito, predicado, objeto). Cada tripla fica em uma linha.

Além dessa diferença sintática, é importante destacar que cada instância, além de ser associada à sua classe, deve ser também associada a um ou mais “apelidos”². Cada apelido é uma instância da classe `Alias`, que por sua vez é subclasse de `Lexical Resource`. Um apelido é, portanto, a forma como uma instância é representada de maneira textual, servindo para o processo de EI e posterior anotação semântica. A figura 7.4 apresenta a representação da instância do ACV “Helice”, contendo dois apelidos: “*Helice Foundation*” e “*HELICE.NET*”.

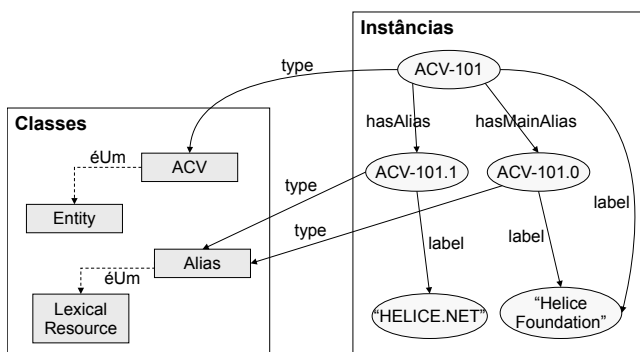


Figura 7.4: Representação de uma instância e seus dois apelidos.

Fonte: adaptado de Kiryakov et al. (2004)

A definição de mais de um apelido permite que uma entidade seja melhor reconhecida em um texto. A figura 7.5 apresenta esse mesmo exemplo na sintaxe de triplas.

Cada linha do arquivo contém as URIs dos elementos das triplas, terminando com um ponto final. É importante reparar que para a criação de uma instância com apenas dois apelidos, oito triplas foram escritas. Além disso, a nomenclatura adotada (URIs) faz com que os nomes de classes e relações sejam muito longos. Portanto, a geração manual de cada instância acaba sendo uma tarefa lenta e passível de erros. Para auxiliar nesse processo,

²Do inglês *alias*.

```

<http://www.ontotext.com/kim/2005/04/wkb#ACV-101>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
  <http://www.gsigma.ufsc.br/pi-co#ACV> .
<http://www.ontotext.com/kim/2005/04/wkb#ACV-101>
  <http://www.w3.org/2000/01/rdf-schema#label> "Helice Foundation" .
<http://www.ontotext.com/kim/2005/04/wkb#ACV-101.0>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
  <http://proton.semanticweb.org/2005/04/protons#Alias> .
<http://www.ontotext.com/kim/2005/04/wkb#ACV-101.0>
  <http://www.w3.org/2000/01/rdf-schema#label> "Helice Foundation" .
<http://www.ontotext.com/kim/2005/04/wkb#ACV-101>
  <http://proton.semanticweb.org/2005/04/protons#hasMainAlias>
  <http://www.ontotext.com/kim/2005/04/wkb#ACV-101.0> .
<http://www.ontotext.com/kim/2005/04/wkb#ACV-101.1>
  <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
  <http://proton.semanticweb.org/2005/04/protons#Alias> .
<http://www.ontotext.com/kim/2005/04/wkb#ACV-101.1>
  <http://www.w3.org/2000/01/rdf-schema#label> "HELICE.NET" .
<http://www.ontotext.com/kim/2005/04/wkb#ACV-101>
  <http://proton.semanticweb.org/2005/04/protons#hasAlias>
  <http://www.ontotext.com/kim/2005/04/wkb#ACV-101.1> .

```

Figura 7.5: Instâncias representando o ACV “Helice” e seus dois apelidos.

uma ferramenta foi implementada, permitindo a definição das instâncias e seus apelidos, e a posterior geração de código. Uma vez definida a base de conhecimento, pode-se passar para a próxima etapa, que é a anotação semântica e indexação dos documentos.

7.3.2 Documentos para a Coleção de Teste

Assim como a base de conhecimento, os documentos foram gerados a partir de relatórios do Projeto ECOLEAD. Esta tarefa foi feita em paralelo à extração das instâncias para a base de conhecimento, de tal forma que anotações semânticas possam ser geradas para os documentos.

Os documentos consistem em fragmentos de seções, e cobrem vários aspectos, em geral técnicos, referentes a: descrições de processos, ferramentas de *software*, empresas, RCs, informações sobre funcionários, entre outros. Os documentos são curtos, visando uma melhor análise da sua relevância. A figura 7.6 ilustra um desses documentos.

Os documentos estão no formato XML, segundo a estrutura adotada pela ferramenta que faz a indexação em lote. O documento contém metadados (features) que indicam o seu identificador, o título e de onde foi extraído. Foram definidos 50 documentos que podem ser vistos no apêndice E. Além deles, há 20 documentos contendo descrições de indicadores. Esses últimos foram extraídos da base de testes de Baldo (2008).

```
<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1038" />
    <feature name="Title" value="ISOIN / HELICE.NET" />
    <feature name="ExtractedFrom" value="D81.2c - section 6.1.3" />
  </features>
  <content>ISOIN / HELICE.NET

Three business processes will be implemented, from the formation of a VBE from the
existing cluster to the creation and launch of resulting VOs. The pilots will be
run in the real life environment of four members of HELICE.NET: MEUPE, INESPASA,
AEROSUR and GHESA. The required implementation plan is assisted by ISOIN, the core
technological partner of HELICE.

The engagement of key industrial players of the cluster in the ECOLEAD
demonstration activities is ensured by the foundation own organisation structure,
since it is participated by the main industrial stakeholders, thus providing the
suitable industrial scenarios and relevant commitment for a successful adoption and
implementation of ECOLEAD result suite.

  </content>
</ksDocument>
```

Figura 7.6: Documento usado na coleção de testes.

7.3.3 Consultas

A avaliação foi feita com base em buscas em dois cenários: *busca por indicadores* e *busca por documentos*. Cada cenário apresenta características que destacam as vantagens e desvantagens da abordagem proposta. Além disso, algumas consultas e respectivos contextos foram definidos de tal forma enfatizar quando a abordagens tem um bom ou mal desempenho. O cenário de indicadores, por exemplo, é o mais significativo pois é baseado num dos cenários de avaliação para o arcabouço proposto por Baldo (2008), que também foi validado no âmbito do projeto ECOLEAD. Já o cenário de busca documentos envolve buscas onde o contexto tem pouca ou nenhuma influência, seja porque os tópicos buscados não estão associados ao contexto atual do usuário, seja porque o usuário busca justamente o que seria selecionado pelo contexto.

Cenário 1: Busca por Indicadores de Desempenho

A busca por indicadores de desempenho é uma atividade essencial em uma RC, especialmente durante a criação de uma OV, onde parceiros devem ser selecionados e seus desempenhos em negócios passados devem ser analisados como critério para tal seleção. Outra situação que pode utilizar desse tipo de busca é quanto há a necessidade de se tomar alguma decisão (pelo gestor da organização, por exemplo), e algum indicador pode ser usado nessa avaliação. Este cenário de teste foi adaptado do trabalho de Baldo et al. (2009), e consistiu de uma coleção de indicadores de desempenho, cujas

descrições foram armazenadas em documentos XML³

Conforme a ontologia de RCs, um indicador pode ser qualificado por vários atributos. Pode-se dizer, dentre outras coisas, que um indicador:

- mede *algo*: recurso, processo, organização;
- é medido com um certo *objetivo*: planejamento, gestão, produção, distribuição;
- considera uma *perspectiva de desempenho* específica: tempo, custo, confiabilidade, qualidade, flexibilidade.

Foram definidas quatro situações onde consultas foram executadas pelo sistema. Cada uma delas é caracterizada pelas instâncias de contextos definidas na seção anterior.

É interessante reparar que essa variedade de atributos pode ser usada como parâmetro para uma consulta, permitindo a seleção de indicadores de maneira bem detalhada. Pode-se buscar, por exemplo, indicadores que meçam certo recurso, que tenham um certo objetivo, e que são associados a um certo domínio.

Além disso, tal como foi definido no modelo conceitual, vários dos conceitos associados a um indicador estão vinculados ao modelo de contexto. Portanto, além do fato da busca ser bem parametrizada, tais parâmetros podem ser filtrados de acordo com o contexto do usuário. Na mesma consulta hipotética apresentada no parágrafo anterior, a abordagem proposta filtraria apenas os indicadores cujos atributos (recurso, objetivo e domínio) que estão vinculados ao contexto atual do usuário.

Outro aspecto que merece destaque é o fato de que cada instância de indicador na base de conhecimento está associada a um único documento (que descreve de maneira textual o indicador) e que tal documento está associado a somente essa instância. Ou seja, há uma relação de 1-para-1 entre as instâncias de indicadores e os documentos que as descrevem.

Tal fato torna a busca por indicadores diferente da busca por documentos em geral, pois não há a associação de múltiplas instâncias em um único documento (na forma de anotações) ou a existência de várias anotações referentes a uma mesma instância. Assim, apesar dos resultados da busca por indicadores serem também documentos, tais documentos “representam” as instâncias da base de conhecimento. Já numa busca de documentos em geral, não há o controle desse tipo de relação, sendo que é possível haver casos onde a busca que retorne apenas uma única instância, por exemplo, acabe retornando mais de um documento, ou mesmo nenhum, caso não haja documentos com anotações para essa instância.

³Tais indicadores estavam originalmente armazenados em um banco de dados.

Para este cenário foram definidas quatro consultas, conforme apresentado abaixo. Para cada consulta o contexto do usuário é caracterizado. As consultas e respectivos contextos aqui apresentados são expressos em linguagem natural. Algumas palavras chave das consultas e contextos estão destacados e são também apresentados em inglês, pois os documentos da coleção foram escritos nesse idioma.

- **Consulta 1:** “Indicadores que meçam **entrega** (*delivery*) e considerem **custo** (*cost*)”
 - Contexto: O usuário é **membro de uma OV** e está num processo de **gestão** (*management*) executando a atividade de **análise de custos** (*cost analysis*).
- **Consulta 2:** “indicadores que meçam **produção** (*production*)”
 - Contexto: O usuário é **membro de uma OV** e está num processo de **gestão** (*management*) executando uma atividade de **avaliação de qualidade** (*quality evaluation*).
- **Consulta 3:** “indicadores que considerem **confiabilidade** (*reliability*)”
 - Contexto: O usuário é o **gestor de uma OV** e está num processo de **gestão da OV** (*VO management*) executando a atividade de **reescalonamento** (*rescheduling*).
- **Consulta 4:** “indicadores que considerem **qualidade** (*quality*)”
 - Contexto: O usuário é um **broker** e participa de um **ACV**. O processo atual é a **criação de uma nova OV** (*VO creation*), sendo a tarefa em execução o **planejamento** (*planning*).

Cenário 2: Busca por Documentos

Ao contrário da busca por indicadores, este cenário apresenta tipo de busca mais geral. Neste cenário, a busca pode ser estruturada com base em diversos tipos de entidades e as relações entre si: Organizações, redes de organizações, competências, processos, produtos, recursos e tarefas.

Conforme mencionado na descrição do cenário anterior, a busca por documentos em geral difere pelo fato da possibilidade de uma instância estar associada a mais de um documento (ou mesmo nenhum), e que um documento pode ter múltiplas anotações para várias (ou a mesma) instâncias. Assim, diferente da busca por indicadores, mesmo que uma busca não retorne nenhum documento, é possível que alguma instância tenha sido recuperada, mas não há documentos que se refiram a ela.

Foram definidas três consultas para este cenário, conforme apresentado a seguir.

- **Consulta 5: “recursos pertencentes ao ACV Helice”**
 - Contexto: O usuário é o **administrador de um ACV** chamado “Helice”. O processo atual é a **criação de uma nova OV** (*VO creation*).
- **Consulta 6: “organizações participantes da OV OIN”**
 - Contexto: O usuário é o **administrador de um ACV** chamado “IECOS”, executando um processo de **gestão** (*VBE management*).
- **Consulta 7: “ACVs que tenham competência na área aeroespacial”**
 - Contexto: O usuário é o **administrador de um ACV** chamado “CeBeNetwork”, executando um processo de **gestão** (*VBE management*).

Nas próximas seções, os contextos e consultas são caracterizados mais formalmente, de acordo com o modelo de contexto e com a linguagem de representação de consulta.

7.3.4 Instâncias do Modelo de Contexto

Esta seção apresenta os contextos utilizados nos experimentos, na forma de instâncias do modelo de contexto. Tais instâncias podem ser representadas como uma tupla na forma:

$$InstanciaContexto = \{RC, Papel, Processo, Tarefa\}$$

De acordo com essa estrutura, as instâncias utilizadas para as sete consultas de teste são apresentadas abaixo:

1. $VoMember1 = \{VoOIN, member, management, costanalysis\}$
2. $VoMember2 = \{VoOIN, member, management, qualityevaluation\}$
3. $VoManager = \{VoOIN, manager, VManagement, rescheduling\}$
4. $VbeBroker = \{VbeHelice, Broker, VOcreation, planning\}$
5. $VbeAdmin1 = \{VbeHelice, Admin, VOcreation, partnerssearch\}$
6. $VbeAdmin2 = \{VbeIECOS, Admin, Management, -\}$

7. $VbeAdmin3 = \{VbeCeBeNetwork, Admin, Management, -\}$

Conceitualmente, as instâncias dos elementos de contexto representam os metadados capturados de ferramentas externas (de BPM / segurança). Como o protótipo não foi integrado a esse tipo de ferramenta, as instâncias dos contextos são fixas, preparadas para estes experimentos.

Para que as regras possam funcionar apropriadamente, as instâncias dos elementos de contexto devem ser associadas às instâncias da base de conhecimento. Este trabalho deve ser feito previamente por um especialista no domínio. Tal integração pode ser vista como uma limitação dessa abordagem. No entanto, essa dependência também ocorre na perspectiva semântica, onde é preciso haver uma base de conhecimento consistente para que a busca possa trazer bons resultados. De fato, da mesma maneira que o modelo de contexto pode ser visto como parte da ontologia, as suas instâncias também fazem parte da base de conhecimento.

A tabela 7.1 a seguir apresenta as instâncias dos contextos utilizados nesses experimentos e o seu vínculo com a base de conhecimento. A coluna da esquerda mostra as instâncias dos elementos de contexto e na coluna da direita estão as respectivas instâncias da base de conhecimento.

Na implementação, tais relações são expressas na forma de fatos que podem então ser processados pelo motor de regras. O apêndice D mostra essas relações na sintaxe utilizada pelo Jess.

7.3.5 Critérios de Relevância

Os critérios de relevância são definidos por usuários (especialistas na área de RCs) com base em um questionário. Nesse questionário, para cada consulta o contexto do usuário é apresentado, bem como a situação que levou à necessidade da busca. O especialista então assinala os documentos que julgar relevantes. O questionário pode ser visto no apêndice G.

Foram entrevistados cinco especialistas, cujas respostas foram usadas para a definição dos critérios de relevância, essenciais para a definição das medidas e gráficos.

A ideia aqui é mover a parte qualitativa (subjativa) da utilização do sistema para a definição dos resultados esperados nas buscas. Ou seja, ao invés de se perguntar se o sistema é “bom” depois de usá-lo, é perguntado o que o usuário quer como resultados das buscas.

Neste caso, o sistema não é usado pelo usuário, que apenas indica o que ele deve retornar. O sistema então é testado quantitativamente em relação à coleção de teste definida. Naturalmente, os usuários desse tipo de pesquisa devem ter conhecimento da área e devem ser instruídos a “entrarem no contexto” definido para cada consulta. Assim, eles podem dar a devida noção de relevância de acordo com esses contextos.

| | Instâncias do Modelo de Contexto | Base de Conhecimento |
|---|--|---|
| | RC: OIN | OV: OIN |
| 1 | Papel: OIN Partner | Papel: ParceiroDaOIN Contexto Org.: Intra |
| | Processo: Management | Processo: VO Management Obj. da Medição: Management |
| | Tarefa: Cost Analysis | Persp. Desempenho: Cost |
| | RC: OIN | OV: OIN |
| 2 | Papel: OIN Partner | Papel: ParceiroDaOIN Contexto Org.: Intra |
| | Processo: Management | Processo: VO Management Obj. da Medição: Management |
| | Tarefa: Quality Evaluation | Persp. Desempenho: Quality |
| | RC: OIN | OV: OIN |
| 3 | Papel: OIN Manager | Papel: CoordenadorDaOIN Contexto Org.: Inter |
| | Processo: OIN VO Management | Processo: VO Management Obj. da Medição: Management |
| | Tarefa: VO1 rescheduling | – |
| | RC: VBE Helice | ACV: Helice |
| 4 | Papel: Broker Helice | Contexto Org.: Inter |
| | Processo: Helice VO Creation | Processo: VO Creation in Helice |
| | Tarefa: Helice Planning | Obj. da Medição: Planning |
| | RC: VBE Helice | ACV: Helice |
| 5 | Papel: Broker Helice | Contexto Org.: Inter |
| | Processo: Helice VO Creation | Processo: VO Creation in Helice |
| | Tarefa: Helice partners selection | Tarefa: VO partners selection |
| | RC: IECOS | ACV: IECOS |
| 6 | Papel: IECOS Admin | Papel: AdministradorDoIECOS Contexto Org.: Inter |
| | Processo: IECOS Management | Processo: VBE Performance management Obj. da Medição: Management |
| | Tarefa: – | – |
| | RC: CeBeNetwork | ACV: CeBeNetwork |
| 7 | Papel: CeBeNetwork Admin | Papel: AdministradorDoCeBeNetwork Contexto Org.: Inter |
| | Processo: CBN Management | Processo: CBN Management Obj. da Medição: Management |
| | Tarefa: – | – |

Tabela 7.1: Ligação entre as instâncias do contexto e a base de conhecimento.

Outra coisa importante aqui é que essa abordagem tira o pesquisador da parte subjetiva (exceto na definição dos documentos e consultas). Os usuários são quem definem os resultados esperados.

Outro fator a favor dessa abordagem é que é mais complicado “simular” o ganho qualitativo da busca contextual. Não é preciso criar nenhum cenário fictício que simule o contexto, para daí o usuário dar sua opinião durante o processo de busca. Por outro lado, perde-se no fator “interatividade”,

pois o usuário não tem controle sobre a expansão das consultas.

O critério de relevância foi definido através da unificação das respostas do questionário. Os documentos que foram assinalados pela maioria absoluta dos entrevistados – neste experimento, por pelo menos três especialistas – foram selecionados para a coleção de referência. O apêndice H apresenta as respostas do questionário tabuladas, bem como os documentos selecionados. Outras opções de implementação desta abordagem foram:

1. Seleção de documentos votados por **todos** os especialistas: isto corresponde à intersecção entre os documentos votados. Além esta situação não ocorrer para todas as consultas, ela gera, naturalmente, poucos documentos. Ela é, portanto, uma abordagem muito restritiva, e que acabaria favorecendo o sistema que retornasse poucos resultados.
2. Seleção de documentos com pelo menos **um** voto: esta é a situação oposta, e corresponde à operação de união. Neste caso, seriam selecionados muitos documentos, o que favoreceria o sistema que retornasse mais resultados.

Essas duas estratégias foram descartadas pois descaracterizam a opinião geral dos entrevistados. A opção pela maioria acabou sendo a mais fiel, excluindo documentos que tenham apenas um ou dois votos.

7.4 Análise dos Resultados

Os resultados, medidas e discussões para cada consulta são apresentados a seguir. Para cada situação, os três tipos de consultas são apresentados (*palavra-chave*, *semântica* e *semântica+contexto*). A consulta baseada em semântica e contexto é gerada pela aplicação de todas as regras. Conforme dito, a execução das consultas é feita em lote e assim o usuário não tem controle sobre a expansão da consulta.

As medidas utilizadas na avaliação são:

- **Precisão**: relação entre o número de documentos relevantes recuperados e o total de documentos recuperados.
- **Cobertura**: relação entre o número de documentos relevantes recuperados e o total de documentos relevantes na coleção.
- **Medida F**: indica o equilíbrio entre precisão e cobertura.
- **Precisão R**: dado um valor R que é o número de documentos relevantes na coleção, a precisão R é calculada como a relação entre o número de documentos relevantes presentes na R primeiras posições na lista de resultados e R .

- **Curva de precisão e cobertura:** gráfico gerado com base nos cálculos de precisão a cada documento relevante observado na lista de resultados. É desenhado com base em 11 pontos de cobertura padrão (0%, 10%, ..., 100%).
- **Precisão média para os documentos relevantes observados:** valor único calculado através da média dos pontos de precisão obtidos quando cada novo documento relevante é observado na lista de resultados. Pode ser interpretada como um resumo da curva de precisão e cobertura.

A razão para o uso de diversas medias é que, individualmente, elas são insuficientes para uma avaliação mais abrangente do sistema de busca. Além disso, várias dessas medidas são complementares entre si.

7.4.1 Busca por Indicadores de Desempenho

Consulta 1

A tabela 7.2 apresenta a consulta 1 e seu contexto, expressos em linguagem natural.

| | |
|-----------------|---|
| Consulta | “Indicadores que meçam entrega (<i>delivery</i>) e considerem custo (<i>cost</i>)” |
| Contexto | O usuário é membro de uma OV e está num processo de gestão (<i>management</i>) executando a atividade de análise de custos (<i>cost analysis</i>). |

Tabela 7.2: Consulta 1 e seu contexto.

A tabela 7.3 apresenta a consulta 1 expressa em três formas: palavras-chave, expressão lógica baseada na ontologia e expressão lógica expandida com base no contexto. Para cada tipo de consulta, os respectivos resultados (identificadores de documentos) são apresentados.

| Tipo | Consulta | Resultados |
|---------------|---|-------------------|
| Palavra-chave | “delivery AND cost” | – |
| Semântica | $x \mid \text{Indicator}(x) \wedge \text{measures}(x, \text{“delivery”}) \wedge \text{considers}(x, \text{“cost”})$ | 8, 5, 2 |
| Contexto | $x \mid \text{Indicator}(x) \wedge \text{measures}(x, \text{“delivery”}) \wedge \text{considers}(x, \text{“cost”}) \wedge \text{isRelatedTo}(x, \text{“intraorganizational”}) \wedge \text{hasObjective}(x, \text{“management”})$ | 5, 2 |

Tabela 7.3: Diferentes expressões para a consulta 1.

A consulta expandida tomou como base aspectos do contexto do usuário: o custo (que já estava na consulta semântica original), o escopo intraorganizacional (pois é um membro a OV) e gestão (pois é o processo

em execução no momento). Uma consequência natural dessa expansão é que os resultados da busca por contexto são sempre subconjuntos da busca puramente semântica. Para esta consulta, a busca semântica retornou os indicadores 8, 5 e 2, e a busca por contexto recuperou o 5 e 2. A busca por palavras chave não retornou documento algum.

Medidas

As medidas obtidas para os três tipos de busca podem ser vistas na tabela 7.4.

| Medida | palavra-chave | semântico | contexto |
|--|---------------|-----------|------------|
| Número de docs. relev. esperados | 1 | 1 | 1 |
| Número de docs. recuperados | 0 | 3 | 2 |
| Número de docs. relev. recuperados | 0 | 1 | 1 |
| Precisão | 0 | 0,34 | 0,5 |
| Cobertura | 0 | 1 | 1 |
| Medida F | 0 | 0,5 | 0,67 |
| Precisão-R | 0 | 0 | 0 |
| Precisão média para docs. relev. observ. | 0 | 0,34 | 0,5 |

Tabela 7.4: Medidas de desempenho para a consulta 1.

A figura 7.7 mostra as curvas de precisão x cobertura da consulta 1, para cada tipo de busca.

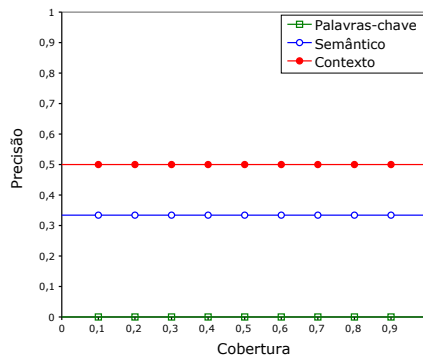


Figura 7.7: Curvas de Precisão x Cobertura para a consulta 1.

Análise: segundo o critério definido, o indicador 2 é o único resultado relevante para esta consulta. Assim, tanto a abordagem semântica como a por contexto apresentaram uma cobertura de 100%. Já a abordagem por contexto possui uma precisão maior (50% contra 33%) pois recupera menos

indicadores.

Consulta 2

A tabela 7.5 apresenta a consulta 2 e seu contexto, expressos em linguagem natural.

| | |
|-----------------|---|
| Consulta | “Indicadores que meçam produção (<i>production</i>)” |
| Contexto | O usuário é membro de uma OV e está num processo de gestão (<i>management</i>) executando a atividade de avaliação de qualidade (<i>quality evaluation</i>). |

Tabela 7.5: Consulta 2 e seu contexto.

A tabela 7.6 apresenta a consulta 2 expressa nas três formas, bem como os seus respectivos resultados.

| Tipo | Consulta | Resultados |
|---------------|--|-------------------|
| Palavra-chave | “production” | 9, 10, 11, 12, 4 |
| Semântica | $x \mid \text{Indicator}(x) \wedge \text{measures}(x, \text{“production”})$ | 13, 12, 11, 10 |
| Contexto | $x \mid \text{Indicator}(x) \wedge \text{measures}(x, \text{“production”}) \wedge \text{isRelatedTo}(x, \text{“intraorganizational”}) \wedge \text{hasObjective}(x, \text{“management”}) \wedge \text{considers}(x, \text{“quality”})$ | 12 |

Tabela 7.6: Diferentes expressões para a consulta 2.

Neste caso, a consulta baseada em contexto também leva em consideração o objetivo de gestão e a perspectiva de qualidade, ambos elementos associados ao contexto do usuário. Assim sendo, os resultados reduziram de quatro (13, 12, 11, 10) da busca semântica para apenas um (12) para a busca por contexto.

Medidas

As medidas obtidas para os três tipos de busca podem ser vistas na tabela 7.7. A figura 7.8 mostra as curvas de precisão x cobertura.

Análise: os documentos relevantes para esta consulta são o 12 e o 15. Assim, todos os tipos de buscas tiveram 100% de cobertura. Já a precisão da busca por contexto foi novamente maior (50% contra 25%) pois recuperou somente um documento (contra quatro da busca semântica). A busca por contexto teve um desempenho superior, conforme pode ser observado pelas demais medidas e pela curva.

| Medida | palavra-chave | semântico | contexto |
|--|---------------|------------|------------|
| Número de docs. relev. esperados | 2 | 2 | 2 |
| Número de docs. recuperados | 5 | 4 | 1 |
| Número de docs. relev. recuperados | 1 | 1 | 1 |
| Precisão | 0,2 | 0,25 | 1 |
| Cobertura | 0,5 | 0,5 | 0,5 |
| Medida F | 0,29 | 0,34 | 0,67 |
| Precisão-R | 0 | 0,5 | 0,5 |
| Precisão média para docs. relev. observ. | 0,25 | 0,5 | 1 |

Tabela 7.7: Medidas de desempenho para a consulta 2.

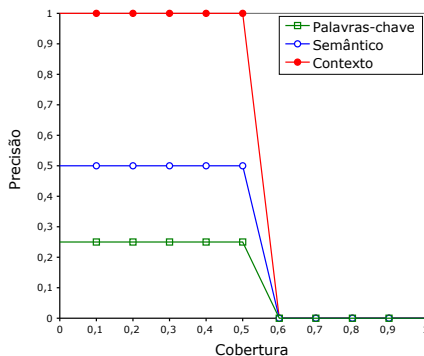


Figura 7.8: Curvas de Precisão x Cobertura para a consulta 2.

Consulta 3

A tabela 7.8 apresenta a consulta 3 e seu contexto, expressos em linguagem natural.

| | |
|-----------------|---|
| Consulta | “Indicadores que considerem confiabilidade (<i>reliability</i>)” |
| Contexto | O usuário é o gestor de uma OV e está num processo de gestão da OV (<i>VO management</i>) executando a atividade de reescalonamento (<i>rescheduling</i>). |

Tabela 7.8: Consulta 3 e seu contexto.

Nesta busca, o contexto do usuário faz com que a consulta também leve em consideração o objetivo de gestão (relativo ao processo do qual o usuário participa) e do escopo interorganizacional (pois o usuário é o gestor da OV). Desse modo, os resultados caíram de seis (da busca semântica) para dois indicadores (da busca por contexto), conforme pode ser visto na tabela 7.9.

| Tipo | Consulta | Resultados |
|---------------|---|-----------------------|
| Palavra-chave | “reliability” | 17, 15 |
| Semântica | $x \mid \text{Indicator}(x) \wedge \text{considers}(x, \text{“reliability”})$ | 17, 15, 14, 11, 10, 1 |
| Contexto | $x \mid \text{Indicator}(x) \wedge \text{considers}(x, \text{“reliability”}) \wedge \text{hasObjective}(x, \text{“management”}) \wedge \text{isRelatedTo}(x, \text{“interorganizational”})$ | 15, 14 |

Tabela 7.9: Diferentes expressões para a consulta 3.

Medidas

As medidas obtidas para os três tipos de busca podem ser vistas na tabela 7.10.

| Medida | palavra-chave | semântico | contexto |
|--|---------------|-------------|----------|
| Número de docs. relev. esperados | 4 | 4 | 4 |
| Número de docs. recuperados | 2 | 6 | 2 |
| Número de docs. relev. recuperados | 1 | 3 | 2 |
| Precisão | 0,5 | 0,5 | 1 |
| Cobertura | 0,25 | 0,75 | 0,5 |
| Medida F | 0,34 | 0,59 | 0,67 |
| Precisão-R | 0,25 | 0,5 | 0,5 |
| Precisão média para docs. relev. observ. | 0,5 | 0,56 | 1 |

Tabela 7.10: Medidas de desempenho para a consulta 3.

A figura 7.9 mostra as curvas de precisão x cobertura da consulta 3.

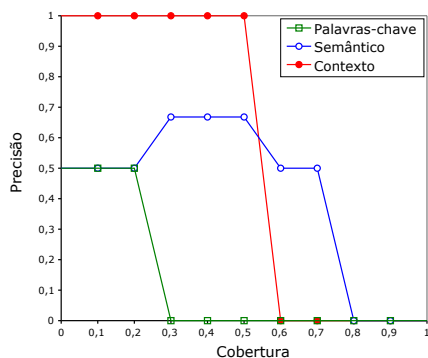


Figura 7.9: Curvas de Precisão x Cobertura para a consulta 3.

Análise: novamente, a busca por contexto apresentou um desempenho melhor em termos de precisão (100% contra 50%), mas com uma cobertura inferior (50% contra 75%). Todavia, a busca por contexto apresentou valores superiores nas demais medidas, bem como uma curva com melhor desempenho (exceto na área de cobertura entre cerca de 55% e 80%). É interessante notar que houve um aumento incomum de precisão na busca semântica nos pontos de cobertura entre 20% e 30%. Isso ocorreu devido ao fato desse tipo de busca ter recuperado dois documentos relevantes (15 e 14) em sequência. Isso fez com que a precisão subisse de 50% (pois o 17 não é relevante) para 67%.

Consulta 4

A tabela 7.11 apresenta a consulta 4 e seu contexto, expressos em linguagem natural.

| | |
|-----------------|---|
| Consulta | “Indicadores que considerem qualidade (<i>quality</i>)” |
| Contexto | O usuário é um <i>broker</i> e participa de um ACV . O processo atual é a criação de uma nova OV (<i>VO creation</i>), sendo a tarefa em execução o planejamento . |

Tabela 7.11: Consulta 4 e seu contexto.

A tabela 7.12 apresenta a consulta 4 expressa nas três formas, bem como os seus respectivos resultados.

| Tipo | Consulta | Resultados |
|---------------|---|-------------------|
| Palavra-chave | “quality” | 12, 4 |
| Semântica | $x \mid \text{Indicator}(x) \wedge \text{considers}(x, \text{“quality”})$ | 19, 12, 4 |
| Contexto | $x \mid \text{Indicator}(x) \wedge \text{considers}(x, \text{“quality”}) \wedge \text{isRelatedTo}(x, \text{“interorganizational”}) \wedge \text{hasObjective}(x, \text{“planning”})$ | 19 |

Tabela 7.12: Diferentes expressões para a consulta 4.

Aqui a expansão fez a consulta se considerar também o planejamento (que é a tarefa executada pelo usuário) bem como o escopo interorganizacional (pois o usuário é um *broker*). Como esperado, os resultados foram mais restritos.

Medidas

As medidas obtidas para os três tipos de busca podem ser vistas na tabela 7.13.

| Medida | palavra-chave | semântico | contexto |
|--|---------------|-----------|----------|
| Número de docs. relev. esperados | 3 | 3 | 3 |
| Número de docs. recuperados | 2 | 3 | 1 |
| Número de docs. relev. recuperados | 1 | 2 | 1 |
| Precisão | 0,5 | 0,67 | 1 |
| Cobertura | 0,34 | 0,67 | 0,34 |
| Medida F | 0,4 | 0,67 | 0,5 |
| Precisão-R | 0,34 | 0,67 | 0,34 |
| Precisão média para docs. relev. observ. | 1 | 1 | 1 |

Tabela 7.13: Medidas de desempenho para a consulta 4.

A figura 7.10 mostra as curvas de precisão x cobertura da consulta 4.

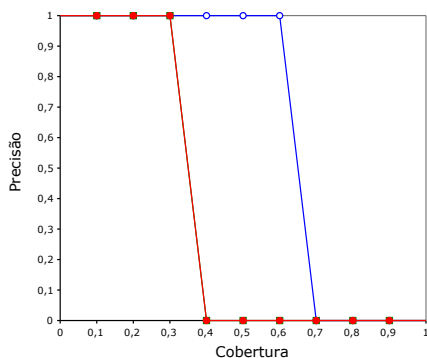


Figura 7.10: Curvas de Precisão x Cobertura para a consulta 4.

Análise: embora a busca por contexto apresente precisão superior em relação aos outros tipos de busca, as demais medidas tiveram valores muito baixos. Nestas, a busca por contexto foi praticamente igual à busca por palavras chave. O gráfico ilustra essa situação.

Desempenho Médio da Busca por IDs

As médias das medidas obtidas nas quatro consultas do cenário de busca por IDs são apresentadas na tabela 7.14. As curvas médias de precisão x cobertura podem ser vistas na figura 7.11.

| Medida | palavra-chave | semântico | contexto |
|--|---------------|-------------|-------------|
| Número de docs. relev. esperados | 10 | 10 | 10 |
| Número de docs. recuperados | 9 | 16 | 6 |
| Número de docs. relev. recuperados | 3 | 7 | 5 |
| Precisão | 0,3 | 0,44 | 0,88 |
| Cobertura | 0,27 | 0,73 | 0,58 |
| Medida F | 0,25 | 0,53 | 0,63 |
| Precisão-R | 0,15 | 0,42 | 0,33 |
| Precisão média para docs. relev. observ. | 0,44 | 0,6 | 0,88 |

Tabela 7.14: Medidas de desempenho médio para o busca de indicadores.

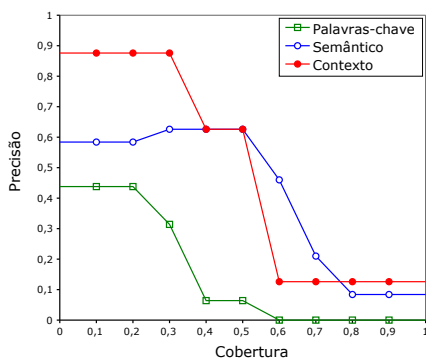


Figura 7.11: Curvas de Precisão x Cobertura médias para a busca por indicadores.

Comentário: assim como nas consultas individuais, o desempenho médio da busca por contexto é superior, em geral sendo melhor em termos de precisão, pois restringe os resultados. Todavia, tal restrição comprometeu a cobertura, que em geral é melhor na busca semântica. É importante ressaltar aqui que isto ocorre porque todas as restrições inferidas pelo contexto são adicionadas à consulta. Tal excesso não aconteceria se o usuário tivesse o controle da expansão da consulta.

7.4.2 Busca por Documentos

Este cenário mostra as limitações da abordagem. A ideia é mostrar situações onde o contexto pode não se aplicar muito bem à busca ou simplesmente não faz diferença.

Consulta 5

A tabela 7.15 apresenta a consulta 5 e seu contexto, expressos em linguagem natural. A tabela 7.16 apresenta a consulta expressa nas três formas, bem como os seus respectivos resultados.

| | |
|-----------------|--|
| Consulta | “recursos pertencentes ao ACV Helice” |
| Contexto | O usuário é o administrador de um ACV chamado “Helice”. O processo atual é a criação de uma nova OV (<i>VO creation</i>). |

Tabela 7.15: Consulta 5 e seu contexto.

| Tipo | Consulta | Resultados |
|---------------|---|------------------------|
| Palavra-chave | “resource AND Helice” | – |
| Semântica | $x \mid \text{Recurso}(x) \wedge \text{temRecurso}(\text{“Helice”}, x)$ | 1050, 1049, 1048, 1003 |
| Contexto | $x \mid \text{Recurso}(x) \wedge \text{temRecurso}(\text{“Helice”}, x) \wedge \text{usaRecurso}(\text{“VO Creation in Helice”}, x)$ | 1050, 1049, 1003 |

Tabela 7.16: Diferentes expressões para a consulta 5.

Neste caso, o contexto adicionou uma nova restrição à consulta. Como o usuário está em um processo de criação de uma OV, a consulta foi modificada de modo a selecionar recursos utilizados nesse processo.

Medidas

As medidas obtidas para os três tipos de busca podem ser vistas na tabela 7.17.

| Medida | palavra-chave | semântico | contexto |
|--|---------------|-------------|----------|
| Número de docs. relev. esperados | 3 | 3 | 3 |
| Número de docs. recuperados | 0 | 4 | 3 |
| Número de docs. relev. recuperados | 0 | 3 | 2 |
| Precisão | 0 | 0,75 | 0,67 |
| Cobertura | 0 | 1 | 0,67 |
| Medida F | 0 | 0,86 | 0,67 |
| Precisão-R | 0 | 1 | 0,67 |
| Precisão média para docs. relev. observ. | 0 | 1 | 1 |

Tabela 7.17: Medidas de desempenho para a consulta 5.

A figura 7.12 mostra as curvas de precisão x cobertura da consulta 5.

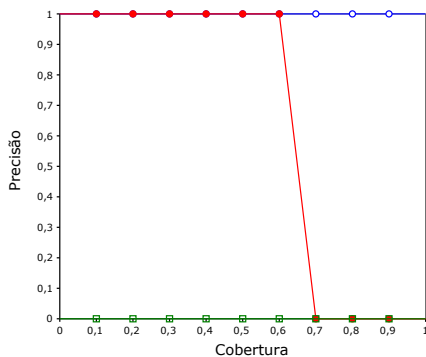


Figura 7.12: Curvas de Precisão x Cobertura para a consulta 5.

Análise: a consulta semântica teve desempenho igual ou superior à consulta baseada em contexto. Esse é um caso onde o contexto não ajudou na busca, levando a resultados um pouco abaixo da busca semântica.

Consulta 6

A tabela 7.18 apresenta a consulta 6 e seu contexto, expressos em linguagem natural.

| | |
|-----------------|---|
| Consulta | “organizações participantes da OV OIN” |
| Contexto | O usuário é o administrador de um ACV chamado “IECOS”, executando um processo de gestão (<i>VBE management</i>). |

Tabela 7.18: Consulta 6 e seu contexto.

A tabela 7.19 apresenta a consulta 6 expressa nas três formas, bem como os seus respectivos resultados.

| Tipo | Consulta | Resultados |
|---------------|---|------------------------------|
| Palavra-chave | “partner AND OIN” | 1030, 1032 |
| Semântica | $x \mid \text{Parceiro}(x) \wedge \text{temParticipante}(\text{“OIN”}, x)$ | 1030, 1032, 1031, 1034, 1035 |
| Contexto | $x \mid \text{Parceiro}(x) \wedge \text{temParticipante}(\text{“OIN”}, x) \wedge \text{temParticipante}(\text{“IECOS”}, x)$ | 1030, 1032, 1031, 1034, 1035 |

Tabela 7.19: Diferentes expressões para a consulta 6.

Medidas

As medidas obtidas para os três tipos de busca podem ser vistas na tabela 7.20.

| Medida | palavra-chave | semântico | contexto |
|--|---------------|-----------|----------|
| Número de docs. relev. esperados | 3 | 3 | 3 |
| Número de docs. recuperados | 2 | 5 | 5 |
| Número de docs. relev. recuperados | 2 | 3 | 3 |
| Precisão | 1 | 0,6 | 0,6 |
| Cobertura | 0,67 | 1 | 1 |
| Medida F | 0,8 | 0,75 | 0,75 |
| Precisão-R | 0,67 | 0,67 | 0,67 |
| Precisão média para docs. relev. observ. | 1 | 0,92 | 0,92 |

Tabela 7.20: Medidas de desempenho para a consulta 6.

A figura 7.13 mostra as curvas de precisão x cobertura da consulta 6.

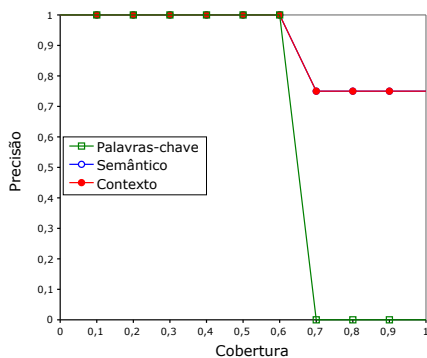


Figura 7.13: Curvas de Precisão x Cobertura para a consulta 6.

Análise: este é um caso onde a busca não tem relação com o contexto do usuário. Neste caso, o contexto acabou não fazendo diferença em relação à busca semântica. No entanto, os resultados foram os mesmos porque os participantes da OV “OIN” também fazem parte do ACV “IECOS”. Se não houvesse tal intersecção, o contexto atrapalharia a busca, que retornaria um conjunto vazio. Esse tipo de problema seria amenizado em um processo interativo, pois o usuário poderia refinar a consulta de modo a remover a restrição adicionada automaticamente.

Consulta 7

A tabela 7.21 apresenta a consulta 7 e seu contexto, expressos em linguagem natural.

| | |
|-----------------|---|
| Consulta | “ACVs que tenham competência na área aeroespacial ” |
| Contexto | O usuário é o administrador de um ACV chamado “CeBeNetwork”, executando um processo de gestão (<i>VBE management</i>). |

Tabela 7.21: Consulta 7 e seu contexto.

A tabela 7.22 apresenta a consulta 7 expressa nas três formas, bem como os seus respectivos resultados.

| Tipo | Consulta | Resultados |
|---------------|---|-------------------|
| Palavra-chave | “ACV AND aerospace” | 1041 |
| Semântica | $x \mid \text{ACV}(x) \wedge \text{temCompetencia}(x, \text{“aeroespacial”})$ | 1038, 1040 |
| Contexto | $x \mid \text{ACV}(x) \wedge \text{temCompetencia}(x, \text{“aeroespacial”})$ | 1038, 1040 |

Tabela 7.22: Diferentes expressões para a consulta 7.

Medidas

As medidas obtidas para os três tipos de busca podem ser vistas na tabela 7.23.

| Medida | palavra-chave | semântico | contexto |
|--|----------------------|------------------|-----------------|
| Número de docs. relev. esperados | 2 | 2 | 2 |
| Número de docs. recuperados | 1 | 2 | 2 |
| Número de docs. relev. recuperados | 0 | 2 | 2 |
| Precisão | 0 | 1 | 1 |
| Cobertura | 0 | 1 | 1 |
| Medida F | 0 | 1 | 1 |
| Precisão-R | 0 | 1 | 1 |
| Precisão média para docs. relev. observ. | 0 | 1 | 1 |

Tabela 7.23: Medidas de desempenho para a consulta 7.

A figura 7.14 mostra as curvas de precisão x cobertura da consulta 7.

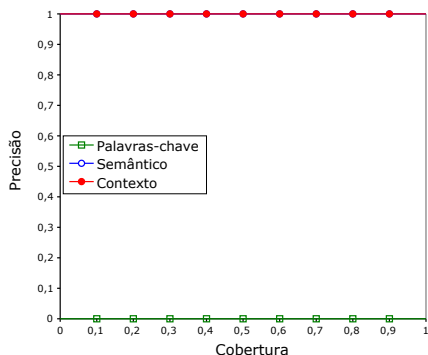


Figura 7.14: Curvas de Precisão x Cobertura para a consulta 7.

Análise: outro caso onde o contexto não exerce influência. Aqui, a consulta original não sofre modificações, e as buscas retornaram os mesmos resultados.

Desempenho Médio da Busca por Documentos

As médias das medidas obtidas nas três consultas do cenário de busca por documentos são apresentadas na tabela 7.24.

| Medida | palavra-chave | semântico | contexto |
|--|---------------|-------------|----------|
| Número de docs. relev. esperados | 8 | 8 | 8 |
| Número de docs. recuperados | 3 | 11 | 10 |
| Número de docs. relev. recuperados | 2 | 8 | 7 |
| Precisão | 0,34 | 0,78 | 0,76 |
| Cobertura | 0,22 | 1 | 0,89 |
| Medida F | 0,27 | 0,87 | 0,81 |
| Precisão-R | 0,22 | 0,89 | 0,78 |
| Precisão média para docs. relev. observ. | 0,34 | 0,97 | 0,98 |

Tabela 7.24: Medidas de desempenho médio para a busca por documentos.

A figura 7.15 mostra as curvas médias de precisão x cobertura para o cenário de busca por documentos.

Comentário: este cenário mostrou casos onde a busca baseada em contexto não se aplicou muito bem. Novamente, caso o usuário tivesse controle durante a expansão, os resultados poderiam ser diferentes. Mesmo assim, ficou claro que o contexto não teve um efeito esperado nesses casos.

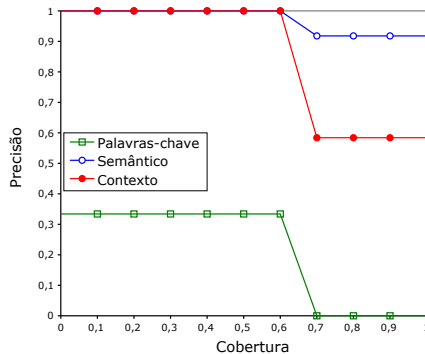


Figura 7.15: Curvas de Precisão x Cobertura médias para a busca por documentos.

7.5 Considerações

Com base no cenário de busca de indicadores, o modelo para busca em contexto apresentou um ganho médio de desempenho, especialmente em termos de precisão. Todavia, a cobertura média foi comprometida, sendo levemente inferior à abordagem puramente semântica. Por outro lado, o cenário de busca em documentos mostrou que quando a busca é por entidades que não tenham associação com o contexto atual do usuário, temos um desempenho semelhante à busca semântica. Isso era esperado, visto que as consultas foram preparadas fora do contexto do usuário. De qualquer modo, existem algumas razões intrínsecas que permitem explicar tais resultados, conforme detalhado a seguir.

Características da busca baseada em ontologias: como a primeira etapa da busca é feita sobre a base de conhecimento (instâncias da ontologia), pode haver situações onde uma entidade resultante não esteja associada a nenhum documento. Desse modo, é possível haver situações onde, mesmo havendo tal entidade na base de conhecimento, a busca não retorne resultado algum. Tal situação atrapalharia a medição de desempenho, visto que esta é feita com base nos documentos retornados. Além disso, há cenários onde cada entidade está associada a um único documento (como no caso dos indicadores), e outros onde as entidades fazem parte de múltiplas anotações no conteúdo de documentos em geral (por exemplo, referências a pessoas e organizações em um contrato).

O modelo acaba sendo muito restritivo: como a mudança é feita na expressão lógica da consulta, não há uma variação suave entre os resultados, pois não há flexibilidade em termos de granularidade. Em outras palavras, a adição de uma certa restrição pode fazer com que uma entidade, que antes era

relevante, passe a ser irrelevante para a busca em contexto.

Características dos procedimentos de avaliação: devido à dificuldades na aplicação em um ambiente real, as consultas foram executadas em lote em um cenário de testes reduzido. Isso fez com que todas as restrições fossem aplicadas automaticamente, tornando a busca muito restritiva. Caso a busca fosse interativa, o usuário poderia escolher quais restrições aplicar, ou mesmo quais elementos de contexto são usados na expansão da consulta. Isso poderia gerar buscas com resultados mais abrangentes. No entanto, outras métricas seriam necessárias. Como cada usuário teria um resultado diferente, medidas baseadas em precisão e cobertura não seriam adequadas. Neste caso, medidas orientadas ao uso da interface (número de telas, “cliques” do *mouse*, indicando o esforço do usuário para encontrar o resultado esperado) poderiam ser utilizadas. Outra opção seria uma análise puramente qualitativa, com base em questionários para se obter a opinião dos usuários.

É importante ressaltar, no entanto, que mesmo com a execução completamente automática (mais restritiva), o modelo mostrou uma melhora no desempenho geral: isso pode ser observado graficamente ou em termos de medida F, mostrando um melhor equilíbrio entre precisão e cobertura.

Em relação à complexidade computacional, a plataforma Jess utiliza um algoritmo eficiente chamado *Rete* (Friedman-Hill, 2007), que tem complexidade linear em relação ao número de regras e fatos.

Para dar um exemplo do impacto no tempo de execução, pode-se usar alguns valores capturados em experimentos preliminares. A aplicação das regras nos 20 indicadores levou cerca de 48 milissegundos⁴. Já base de indicadores completa (cerca de 230) custou levou cerca de 172 milissegundos. É interessante notar que, embora a base tenha aumentado mais de dez vezes, o tempo de execução não chegou a quadruplicar. Portanto, embora este aspecto não tenha sido o foco do trabalho, os números mostram que a abordagem se mostrou viável em um cenário maior.

⁴Executando em uma máquina equipada com processador Intel®Core™2 Duo 3.0 GHz.

Capítulo 8

Conclusões

A busca por informação em RCs é um requisito fundamental para os parceiros de tais redes. O grande volume de informações produzido e intercambiado nessas redes é um bem valioso que precisa ser recuperado, podendo ser usado para as mais diversas aplicações: gestão de conhecimento, seleção de indicadores e de parceiros, suporte à decisão, herança de organizações virtuais, entre outros. Por outro lado, é preciso destacar que a área de RI vem crescendo e o foco com buscas usando ontologias e contexto vem ganhado cada vez maior atenção pela comunidades de pesquisa.

Este trabalho apresentou um modelo baseado em ontologia para representar o contexto de um usuário de uma RC e regras para dar suporte à expansão da consulta. O modelo de contexto é composto por diversos elementos que caracterizam a situação dinâmica dentro de uma RC. Tanto o modelo quanto as regras foram definidos de tal forma que novos elementos possam ser adicionados substituindo parcial ou totalmente os elementos aqui utilizados.

O modelo foi construído sobre um arcabouço de busca semântica, conforme o estado da arte na área de RI. Ele é baseado no uso de anotações semânticas que enriquecem as fontes de informação com base no casamento entre palavras em linguagem natural e conceitos de uma ontologia. A ontologia é também usada como base para a construção das consultas e, em conjunto com as anotações, permite um melhor desempenho na busca.

A abordagem proposta mostrou que o contexto pode melhorar os resultados da busca – em termos de precisão – apesar de perdas no que diz respeito à cobertura dos resultados obtidos. Outro aspecto que ficou evidente foi o papel do usuário durante o processo, papel esse que foi focado somente na definição dos critérios de relevância. Todavia, apesar de não atuar durante a expansão da consulta, os resultados apresentados foram superiores à busca por palavras-chave e baseada em ontologias. Isso mostra que os resultados podem ser ainda melhores com o usuário interagindo com o sistema.

A maior dificuldade detectada foi talvez a preparação dos cenários

para os experimentos. Por um lado, há um esforço considerável para a definição da ontologia, base de conhecimento, documentos, consultas e critérios de relevância, além da implantação e execução de diversas ferramentas. Por outro lado, a dificuldade encontra-se também na quase inexistência de ambientes reais que permitam testes com resultados efetivos. Mesmo assim, pode-se afirmar que o arcabouço contribuiu para o melhoramento da busca nesse tipo de ambiente.

8.1 Contribuições

A principal contribuição deste trabalho consiste na convergência entre diversos elementos: RI, ontologias e contexto e sua aplicação à área de RCs. Foi desenvolvido uma arcabouço para busca de informação em RCs, baseado numa arquitetura orientada a serviços. Tal arcabouço foi validado no âmbito do projeto ECOLEAD e serviu também de base, em diferentes estágios de seu desenvolvimento, para outros trabalhos de pesquisa, a saber: Baldo (2008); Loss (2007). O arcabouço é formado por:

- Modelo de contexto e regras para expansão de consultas: explora as relações entre os conceitos da ontologia / base de conhecimento e as instâncias do modelo de contexto, visando selecionar novos tópicos a serem buscados.
- Ontologia de topo para RCs, que é ligada ao modelo de contexto: serve de base para o funcionamento do modelo.
- Serviços de *software* para funcionalidades relacionadas à RI: anotação semântica, indexação e busca de documentos, navegação na ontologia, entre outros.

É preciso destacar que o modelo, as regras e a ontologia podem ser estendidos para um domínio ou aplicação mais específicos. Além disso, o método de expansão da consulta proposto pode ser adaptado para ser usado em outras áreas. Adicionalmente, o arcabouço foi implementado em um protótipo computacional que foi integrado a uma infraestrutura de TIC de suporte a RCs. Outros resultados foram alcançados, e que podem ser reutilizados em trabalhos futuros:

- Coleções de teste para buscas em RCs:
 - Documentos (indicadores de desempenho e trechos de relatórios);
 - Consultas e critérios de relevância;
 - Base de conhecimento (para anotações semânticas).

- Ferramentas de suporte: para medição de sistemas de RI e para população da base de conhecimento.

8.2 Produção Bibliográfica

Segue abaixo a lista de publicações científicas produzidas como resultado do trabalho apresentado nesta tese:

Artigos publicados em congressos internacionais

1. Tramontin Jr., R. J. e Rabelo, R. J.; 2007. A knowledge search framework for collaborative networks. Em: *Establishing The Foundation of Collaborative Networks, IFIP TC 5 Working Group 5.5 Eighth IFIP Working Conference on Virtual Enterprises*, September 10-12, 2007, Guimarães, Portugal, IFIP, páginas 573–582. Springer.
2. Tramontin Jr., R. J., Hanachi, C., e Rabelo, R. J.; 2008. A rule-based approach for customizing knowledge search in cnos. Em: *Pervasive Collaborative Networks, IFIP TC 5 WG 5.5 Ninth Working Conference on Virtual Enterprises*, September 8-10, 2008, Poznan, Poland, IFIP, páginas 243–252. Springer.
3. Tramontin Jr., R. J., Rabelo, R. J., e Hanachi, C.; 2009. Using semantics and context for searching in networked organizations. Em: *Proceedings of I2TS'2009 - 8th International Information and Telecommunication Technologies Symposium*, December 9-11, 2009, Florianópolis, Brasil, páginas 17-24. Fundação Barddal de Educação e Cultura.

Artigos publicados em revistas internacionais

1. Tramontin Jr., R. J., Rabelo, R. J., e Hanachi, C.; 2010. Customizing knowledge search in collaborative networked organisations through context-based query expansion. *Production Planning and Control*, 21:229–246.

8.3 Limitações da Abordagem Proposta

Apesar das contribuições para a área em questão, algumas considerações devem ser feitas a respeito das limitações do trabalho.

Em relação à abordagem proposta, algumas limitações ficaram evidentes nos experimentos. Uma delas é o fato da abordagem fazer a filtragem por entidades, o que por vezes faz a consulta muito restritiva. Além disso, percebe-se a necessidade de um especialista no domínio para a vinculação da base de conhecimento às instâncias dos contextos. No entanto esta acaba

sendo uma limitação inerente à busca semântica, pois há a necessidade de uma base de conhecimento bem consistente.

Finalmente, embora extensível, o modelo definido é muito geral, pois está associado ao nível *topo* da ontologia (conceitos mais gerais). Novos elementos de contexto e novas regras podem ser definidos para um dado domínio ou aplicação. Esses níveis mais específicos poderiam ser obtidos através do estudo de manifestações reais de RCs, também com o auxílio de técnicas de mineração de dados para se extrair conhecimento específico que possa servir para a especialização das regras.

8.4 Sugestões para Trabalhos Futuros

Com base nas limitações do trabalho e na amplitude de abordagens que a área oferece, vários trabalhos podem ser realizados no futuro:

1. Melhoramentos no modelo e no sistema. Melhorar a interface de usuário, de modo a permitir a escolha na utilização de elementos de contexto individualmente. Além disso, os resultados podem ser exibidos de maneira contextualizada, organizados pelos elementos de contexto. Outro aspecto é o tratamento de múltiplos contextos simultaneamente;
2. Execução de experimentos, com foco no usuário (interface gráfica) e com mais consultas (outros cenários);
3. A ontologia e a base de conhecimento utilizadas são pequenas e simples. É preciso um cenário mais complexo para que se possa testar as regras que lidam com superclasses e subclasses;
4. Outras abordagens para uso do contexto durante a busca podem ser vistas como complementares. Assim, a combinação com outras técnicas pode tratar as limitações existentes ou adicionar novas funcionalidades. Exemplos:
 - Combinação com técnicas de classificação (*ranking*), associando-se, por exemplo, o contexto aos documentos;
 - Ajuste automático de restrições, de acordo com a quantidade de entidades que são retornadas. Ou seja, aplicam-se todas as restrições, vê-se os resultados, e faz-se o ajuste, afrouxando ou restringindo esses critérios.

Apêndice A

Interfaces dos Serviços do *K-Search*

Este apêndice apresenta as interfaces WSDL dos serviços do *K-Search*.

A.1 Serviço para Navegação de Ontologias

```
<?xml version="1.0" encoding="UTF-8"?>
<wsdl:definitions
  targetNamespace="http://www.gsigma.ufsc.br/services/ks"
  xmlns:apachesoap="http://xml.apache.org/xml-soap"
  xmlns:impl="http://www.gsigma.ufsc.br/services/ks"
  xmlns:intf="http://www.gsigma.ufsc.br/services/ks"
  xmlns:wSDL="http://schemas.xmlsoap.org/wsdl/"
  xmlns:wsoap="http://schemas.xmlsoap.org/wsdl/soap/"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <wsdl:documentation>
    This Web Service is built on top of KIM Platform, using the
    SemanticRepositoryAPI. It is used to browse the ontology,
    allowing read-only operations like loading its classes,
    properties and instances.
  </wsdl:documentation>
  <wsdl:types>
    <schema elementFormDefault="qualified"
      targetNamespace="http://www.gsigma.ufsc.br/services/ks"
      xmlns="http://www.w3.org/2001/XMLSchema">
      <element name="getClasses">
        <complexType>
          <sequence>
            <element name="instanceUri" type="xsd:QName" />
          </sequence>
        </complexType>
      </element>
      <element name="getClassesResponse">
        <complexType>
          <sequence>
            <element maxOccurs="unbounded"
              name="getClassesReturn" type="xsd:QName" />
          </sequence>
        </complexType>
      </element>
      <element name="getOntologyClassStructure">
        <complexType />
      </element>
      <element name="getOntologyClassStructureResponse">
        <complexType>

```

```

    <sequence>
      <element maxOccurs="unbounded" minOccurs="0"
        name="getOntologyClassStructureReturn" type="impl:class" />
    </sequence>
  </complexType>
</element>
<complexType name="property">
  <sequence>
    <element name="name" type="xsd:QName" />
    <element name="range" minOccurs="0">
      <complexType>
        <sequence>
          <element maxOccurs="unbounded"
            minOccurs="0" name="item" type="xsd:QName" />
        </sequence>
      </complexType>
    </element>
  </sequence>
</complexType>
<complexType name="property-list">
  <sequence>
    <element maxOccurs="unbounded" minOccurs="0"
      name="item" type="impl:property" />
  </sequence>
</complexType>
<complexType name="class">
  <sequence>
    <element name="name" type="xsd:QName" />
    <element name="superClasses" minOccurs="0">
      <complexType>
        <sequence>
          <element maxOccurs="unbounded"
            minOccurs="0" name="item" type="xsd:QName" />
        </sequence>
      </complexType>
    </element>
    <element name="properties" minOccurs="0"
      type="impl:property-list" />
  </sequence>
</complexType>
<complexType name="ontology-fault">
  <sequence />
</complexType>
<element name="fault" type="impl:ontology-fault" />
<element name="getProperties">
  <complexType>
    <sequence>
      <element name="classUri" type="xsd:QName" />
      <element name="includeSuperClasses"
        minOccurs="0" type="xsd:boolean" />
    </sequence>
  </complexType>
</element>
<element name="getPropertiesResponse">
  <complexType>
    <sequence>
      <element maxOccurs="unbounded"
        name="getPropertiesReturn" type="xsd:QName" />
    </sequence>
  </complexType>
</element>
<element name="getAllClasses">
  <complexType />

```

```
</element>
<element name="getAllClassesResponse">
  <complexType>
    <sequence>
      <element maxOccurs="unbounded"
        name="getAllClassesReturn" type="xsd:QName" />
    </sequence>
  </complexType>
</element>

<element name="getRootClasses">
  <complexType />
</element>
<element name="getRootClassesResponse">
  <complexType>
    <sequence>
      <element maxOccurs="unbounded"
        name="getRootClassesReturn" type="xsd:QName" />
    </sequence>
  </complexType>
</element>

<element name="getInstances">
  <complexType>
    <sequence>
      <element name="classUri" type="xsd:QName" />
    </sequence>
  </complexType>
</element>
<element name="getInstancesResponse">
  <complexType>
    <sequence>
      <element maxOccurs="unbounded"
        name="getInstancesReturn" type="xsd:QName" />
    </sequence>
  </complexType>
</element>
<element name="getSubClasses">
  <complexType>
    <sequence>
      <element name="classUri" type="xsd:QName" />
    </sequence>
  </complexType>
</element>
<element name="getSubClassesResponse">
  <complexType>
    <sequence>
      <element maxOccurs="unbounded"
        name="getSubClassesReturn" type="xsd:QName" />
    </sequence>
  </complexType>
</element>
<element name="getSuperClasses">
  <complexType>
    <sequence>
      <element name="classUri" type="xsd:QName" />
    </sequence>
  </complexType>
</element>
<element name="getSuperClassesResponse">
  <complexType>
    <sequence>
      <element maxOccurs="unbounded"
```

```

        name="getSuperClassesReturn" type="xsd:QName" />
    </sequence>
</complexType>
</element>
<element name="getPropertyValues">
    <complexType>
        <sequence>
            <element name="instanceUri" type="xsd:QName" />
            <element name="propertyUri" type="xsd:QName" />
        </sequence>
    </complexType>
</element>
<element name="getPropertyValuesResponse">
    <complexType>
        <sequence>
            <element maxOccurs="unbounded"
                name="getPropertyValuesReturn" type="xsd:anyType" />
        </sequence>
    </complexType>
</element>
<element name="getPropertyRange">
    <complexType>
        <sequence>
            <element name="propertyUri" type="xsd:QName" />
        </sequence>
    </complexType>
</element>
<element name="getPropertyRangeResponse">
    <complexType>
        <sequence>
            <element maxOccurs="unbounded"
                name="getPropertyRangeReturn" type="xsd:QName" />
        </sequence>
    </complexType>
</element>
<element name="getPropertiesValues">
    <complexType>
        <sequence>
            <element name="instanceUri" type="xsd:QName" />
        </sequence>
    </complexType>
</element>
<element name="getPropertiesValuesResponse">
    <complexType>
        <sequence>
            <element maxOccurs="unbounded"
                name="getPropertiesValuesReturn" type="impl:property-values"
                />
        </sequence>
    </complexType>
</element>
<element name="getReversePropertiesValues">
    <complexType>
        <sequence>
            <element name="instanceUri" type="xsd:QName" />
        </sequence>
    </complexType>
</element>
<element name="getReversePropertiesValuesResponse">
    <complexType>
        <sequence>
            <element maxOccurs="unbounded"

```

```

        name="getReversePropertiesValuesReturn" type="impl:property
        -values" />
    </sequence>
</complexType>
</element>

<complexType name="property-values">
    <sequence>
        <element name="property" type="xsd:QName" />
        <element name="values" minOccurs="0">
            <complexType>
                <sequence>
                    <element maxOccurs="unbounded"
                    name="item" type="xsd:anyType" />
                </sequence>
            </complexType>
        </element>
    </sequence>
</complexType>
</schema>
</wsdl:types>
<wsdl:message name="getPropertiesValuesRequest">
    <wsdl:part element="impl:getPropertiesValues" name="parameters" />
</wsdl:message>
<wsdl:message name="getPropertiesValuesResponse">
    <wsdl:part element="impl:getPropertiesValuesResponse"
    name="parameters" />
</wsdl:message>
<wsdl:message name="getReversePropertiesValuesRequest">
    <wsdl:part element="impl:getReversePropertiesValues" name="parameters"
    />
</wsdl:message>
<wsdl:message name="getReversePropertiesValuesResponse">
    <wsdl:part element="impl:getReversePropertiesValuesResponse"
    name="parameters" />
</wsdl:message>
<wsdl:message name="getOntologyClassStructureRequest">
    <wsdl:part element="impl:getOntologyClassStructure"
    name="parameters" />
</wsdl:message>
<wsdl:message name="getOntologyClassStructureResponse">
    <wsdl:part element="impl:getOntologyClassStructureResponse"
    name="parameters" />
</wsdl:message>
<wsdl:message name="getSubClassesRequest">
    <wsdl:part element="impl:getSubClasses" name="parameters" />
</wsdl:message>
<wsdl:message name="getPropertiesRequest">
    <wsdl:part element="impl:getProperties" name="parameters" />
</wsdl:message>
<wsdl:message name="getSubClassesResponse">
    <wsdl:part element="impl:getSubClassesResponse"
    name="parameters" />
</wsdl:message>
<wsdl:message name="OntologyException">
    <wsdl:part element="impl:fault" name="fault" />
</wsdl:message>
<wsdl:message name="getPropertyValuesRequest">
    <wsdl:part element="impl:getPropertyValues" name="parameters" />
</wsdl:message>
<wsdl:message name="getSuperClassesResponse">
    <wsdl:part element="impl:getSuperClassesResponse"
    name="parameters" />

```

```

</wsdl:message>
<wsdl:message name="getPropertiesResponse">
  <wsdl:part element="impl:getPropertiesResponse"
    name="parameters" />
</wsdl:message>
<wsdl:message name="getPropertyRangeRequest">
  <wsdl:part element="impl:getPropertyRange" name="parameters" />
</wsdl:message>
<wsdl:message name="getAllClassesResponse">
  <wsdl:part element="impl:getAllClassesResponse"
    name="parameters" />
</wsdl:message>

<wsdl:message name="getRootClassesResponse">
  <wsdl:part element="impl:getRootClassesResponse"
    name="parameters" />
</wsdl:message>
<wsdl:message name="getPropertyRangeResponse">
  <wsdl:part element="impl:getPropertyRangeResponse"
    name="parameters" />
</wsdl:message>
<wsdl:message name="getClassesResponse">
  <wsdl:part element="impl:getClassesResponse" name="parameters" />
</wsdl:message>
<wsdl:message name="getInstancesResponse">
  <wsdl:part element="impl:getInstancesResponse"
    name="parameters" />
</wsdl:message>
<wsdl:message name="getAllClassesRequest">
  <wsdl:part element="impl:getAllClasses" name="parameters" />
</wsdl:message>

<wsdl:message name="getRootClassesRequest">
  <wsdl:part element="impl:getRootClasses" name="parameters" />
</wsdl:message>
<wsdl:message name="getInstancesRequest">
  <wsdl:part element="impl:getInstances" name="parameters" />
</wsdl:message>
<wsdl:message name="getPropertyValuesResponse">
  <wsdl:part element="impl:getPropertyValuesResponse"
    name="parameters" />
</wsdl:message>
<wsdl:message name="getClassesRequest">
  <wsdl:part element="impl:getClasses" name="parameters" />
</wsdl:message>
<wsdl:message name="getSuperClassesRequest">
  <wsdl:part element="impl:getSuperClasses" name="parameters" />
</wsdl:message>
<wsdl:portType name="OntologyBrowser">
  <wsdl:operation name="getPropertiesValues">
    <wsdl:input message="impl:getPropertiesValuesRequest"
      name="getPropertiesValuesRequest" />
    <wsdl:output message="impl:getPropertiesValuesResponse"
      name="getPropertiesValuesResponse" />
    <wsdl:fault message="impl:OntologyException"
      name="OntologyException" />
  </wsdl:operation>
  <wsdl:operation name="getReversePropertiesValues">
    <wsdl:input message="impl:getReversePropertiesValuesRequest"

```

```

        name="getReversePropertiesValuesRequest" />
<wsdl:output message="impl:getReversePropertiesValuesResponse"
    name="getReversePropertiesValuesResponse" />
<wsdl:fault message="impl:OntologyException"
    name="OntologyException" />
</wsdl:operation>
<wsdl:operation name="getOntologyClassStructure">
<wsdl:input message="impl:getOntologyClassStructureRequest"
    name="getOntologyClassStructureRequest" />
<wsdl:output
    message="impl:getOntologyClassStructureResponse"
    name="getOntologyClassStructureResponse" />
<wsdl:fault message="impl:OntologyException"
    name="OntologyException" />
</wsdl:operation>
<wsdl:operation name="getClasses">
<wsdl:documentation>
    Returns the classes of an instance.
</wsdl:documentation>
<wsdl:input message="impl:getClassesRequest"
    name="getClassesRequest" />
<wsdl:output message="impl:getClassesResponse"
    name="getClassesResponse" />
<wsdl:fault message="impl:OntologyException"
    name="OntologyException" />
</wsdl:operation>
<wsdl:operation name="getProperties">
<wsdl:documentation>
    Returns the properties of a class. It is possible to
    specify if the super classes properties should be
    included in the result.
</wsdl:documentation>
<wsdl:input message="impl:getPropertiesRequest"
    name="getPropertiesRequest" />
<wsdl:output message="impl:getPropertiesResponse"
    name="getPropertiesResponse" />
<wsdl:fault message="impl:OntologyException"
    name="OntologyException" />
</wsdl:operation>
<wsdl:operation name="getAllClasses">
<wsdl:documentation>
    Returns all classes defined in the ontology. It returns
    an array containing QNames for all defined classes.
</wsdl:documentation>
<wsdl:input message="impl:getAllClassesRequest"
    name="getAllClassesRequest" />
<wsdl:output message="impl:getAllClassesResponse"
    name="getAllClassesResponse" />
</wsdl:operation>

<wsdl:operation name="getRootClasses">
<wsdl:documentation>
    Returns all classes that do not have super classes
    defined in the ontology. It returns an array containing
    QNames for all root classes.
</wsdl:documentation>

<wsdl:input message="impl:getRootClassesRequest"
    name="getRootClassesRequest" />

<wsdl:output message="impl:getRootClassesResponse"
    name="getRootClassesResponse" />

```

```

</wsdl:operation>
<wsdl:operation name="getInstances">
  <wsdl:documentation>
    Returns the instances of a class.
  </wsdl:documentation>
  <wsdl:input message="impl:getInstancesRequest"
    name="getInstancesRequest" />
  <wsdl:output message="impl:getInstancesResponse"
    name="getInstancesResponse" />
  <wsdl:fault message="impl:OntologyException"
    name="OntologyException" />
</wsdl:operation>
<wsdl:operation name="getSubClasses">
  <wsdl:documentation>
    Returns the sub classes of a given class.
  </wsdl:documentation>
  <wsdl:input message="impl:getSubClassesRequest"
    name="getSubClassesRequest" />
  <wsdl:output message="impl:getSubClassesResponse"
    name="getSubClassesResponse" />
  <wsdl:fault message="impl:OntologyException"
    name="OntologyException" />
</wsdl:operation>
<wsdl:operation name="getSuperClasses">
  <wsdl:documentation>
    Returns the super classes of a given class.
  </wsdl:documentation>
  <wsdl:input message="impl:getSuperClassesRequest"
    name="getSuperClassesRequest" />
  <wsdl:output message="impl:getSuperClassesResponse"
    name="getSuperClassesResponse" />
  <wsdl:fault message="impl:OntologyException"
    name="OntologyException" />
</wsdl:operation>
<wsdl:operation name="getPropertyValues">
  <wsdl:documentation>
    Returns the values of a property for a given instance.
    Properties values may be QNames or simple Strings.
  </wsdl:documentation>
  <wsdl:input message="impl:getPropertyValuesRequest"
    name="getPropertyValuesRequest" />
  <wsdl:output message="impl:getPropertyValuesResponse"
    name="getPropertyValuesResponse" />
  <wsdl:fault message="impl:OntologyException"
    name="OntologyException" />
</wsdl:operation>
<wsdl:operation name="getPropertyRange">
  <wsdl:documentation>
    Returns the range of a property. The range states that
    the values of a property are instances of one or more
    classes.
  </wsdl:documentation>
  <wsdl:input message="impl:getPropertyRangeRequest"
    name="getPropertyRangeRequest" />
  <wsdl:output message="impl:getPropertyRangeResponse"
    name="getPropertyRangeResponse" />
  <wsdl:fault message="impl:OntologyException"
    name="OntologyException" />
</wsdl:operation>
</wsdl:portType>
<wsdl:binding name="OntologyBrowserServiceSoapBinding"
  type="impl:OntologyBrowser">
  <wsdlsoap:binding style="document"

```



```
    transport="http://schemas.xmlsoap.org/soap/http" />
<wsdl:operation name="getPropertiesValues">
  <wsdlsoap:operation soapAction="" />
  <wsdl:input name="getPropertiesValuesRequest">
    <wsdlsoap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="getPropertiesValuesResponse">
    <wsdlsoap:body use="literal" />
  </wsdl:output>
  <wsdl:fault name="OntologyException">
    <wsdlsoap:fault name="OntologyException" use="literal" />
  </wsdl:fault>
</wsdl:operation>

<wsdl:operation name="getReversePropertiesValues">
  <wsdlsoap:operation soapAction="" />
  <wsdl:input name="getReversePropertiesValuesRequest">
    <wsdlsoap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="getReversePropertiesValuesResponse">
    <wsdlsoap:body use="literal" />
  </wsdl:output>
  <wsdl:fault name="OntologyException">
    <wsdlsoap:fault name="OntologyException" use="literal" />
  </wsdl:fault>
</wsdl:operation>

<wsdl:operation name="getOntologyClassStructure">
  <wsdlsoap:operation soapAction="" />
  <wsdl:input name="getOntologyClassStructureRequest">
    <wsdlsoap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="getOntologyClassStructureResponse">
    <wsdlsoap:body use="literal" />
  </wsdl:output>
  <wsdl:fault name="OntologyException">
    <wsdlsoap:fault name="OntologyException" use="literal" />
  </wsdl:fault>
</wsdl:operation>

<wsdl:operation name="getClasses">
  <wsdlsoap:operation soapAction="" />
  <wsdl:input name="getClassesRequest">
    <wsdlsoap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="getClassesResponse">
    <wsdlsoap:body use="literal" />
  </wsdl:output>
  <wsdl:fault name="OntologyException">
    <wsdlsoap:fault name="OntologyException" use="literal" />
  </wsdl:fault>
</wsdl:operation>

<wsdl:operation name="getProperties">
  <wsdlsoap:operation soapAction="" />
  <wsdl:input name="getPropertiesRequest">
    <wsdlsoap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="getPropertiesResponse">
    <wsdlsoap:body use="literal" />
  </wsdl:output>
  <wsdl:fault name="OntologyException">
    <wsdlsoap:fault name="OntologyException" use="literal" />
  </wsdl:fault>
</wsdl:operation>

<wsdl:operation name="getAllClasses">
```

```

<wsdlsoap:operation soapAction="" />
<wsdl:input name="getAllClassesRequest">
  <wsdlsoap:body use="literal" />
</wsdl:input>
<wsdl:output name="getAllClassesResponse">
  <wsdlsoap:body use="literal" />
</wsdl:output>
</wsdl:operation>

<wsdl:operation name="getRootClasses">
  <wsdlsoap:operation soapAction="" />
  <wsdl:input name="getRootClassesRequest">
    <wsdlsoap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="getRootClassesResponse">
    <wsdlsoap:body use="literal" />
  </wsdl:output>
</wsdl:operation>

<wsdl:operation name="getInstances">
  <wsdlsoap:operation soapAction="" />
  <wsdl:input name="getInstancesRequest">
    <wsdlsoap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="getInstancesResponse">
    <wsdlsoap:body use="literal" />
  </wsdl:output>
  <wsdl:fault name="OntologyException">
    <wsdlsoap:fault name="OntologyException" use="literal" />
  </wsdl:fault>
</wsdl:operation>

<wsdl:operation name="getSubClasses">
  <wsdlsoap:operation soapAction="" />
  <wsdl:input name="getSubClassesRequest">
    <wsdlsoap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="getSubClassesResponse">
    <wsdlsoap:body use="literal" />
  </wsdl:output>
  <wsdl:fault name="OntologyException">
    <wsdlsoap:fault name="OntologyException" use="literal" />
  </wsdl:fault>
</wsdl:operation>

<wsdl:operation name="getSuperClasses">
  <wsdlsoap:operation soapAction="" />
  <wsdl:input name="getSuperClassesRequest">
    <wsdlsoap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="getSuperClassesResponse">
    <wsdlsoap:body use="literal" />
  </wsdl:output>
  <wsdl:fault name="OntologyException">
    <wsdlsoap:fault name="OntologyException" use="literal" />
  </wsdl:fault>
</wsdl:operation>

<wsdl:operation name="getPropertyValues">

```

```

    <wsdlsoap:operation soapAction="" />
    <wsdl:input name="getPropertyValuesRequest">
      <wsdlsoap:body use="literal" />
    </wsdl:input>
    <wsdl:output name="getPropertyValuesResponse">
      <wsdlsoap:body use="literal" />
    </wsdl:output>
    <wsdl:fault name="OntologyException">
      <wsdlsoap:fault name="OntologyException" use="literal" />
    </wsdl:fault>
  </wsdl:operation>
  <wsdl:operation name="getPropertyRange">
    <wsdlsoap:operation soapAction="" />
    <wsdl:input name="getPropertyRangeRequest">
      <wsdlsoap:body use="literal" />
    </wsdl:input>
    <wsdl:output name="getPropertyRangeResponse">
      <wsdlsoap:body use="literal" />
    </wsdl:output>
    <wsdl:fault name="OntologyException">
      <wsdlsoap:fault name="OntologyException" use="literal" />
    </wsdl:fault>
  </wsdl:operation>
</wsdl:binding>
<wsdl:service name="OntologyBrowserService">
  <wsdl:port binding="impl:OntologyBrowserServiceSoapBinding"
    name="OntologyBrowserService">
    <wsdlsoap:address
      location="http://localhost:8080/ks/services/
        OntologyBrowserService" />
  </wsdl:port>
</wsdl:service>
</wsdl:definitions>

```

A.2 Serviço para Gestão de Documentos

```

<?xml version="1.0" encoding="UTF-8"?>
<wsdl:definitions targetNamespace="http://www.gsigma.ufsc.br/services/ks"
  xmlns:apacheSOAP="http://xml.apache.org/xml-soap" xmlns:impl="http
    ://www.gsigma.ufsc.br/services/ks"
  xmlns:intf="http://www.gsigma.ufsc.br/services/ks" xmlns:wsdl="http://
    schemas.xmlsoap.org/wsdl/" xmlns:wsdlsoap="http://schemas.xmlsoap.
    org/wsdl/soap/" xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <wsdl:documentation>
    This Web Service is built on top of KIM Platform, using the following
    APIs: (a) CorporaAPI, for document creation. (b)
    IndexAndPersistAPI, for document indexing and storage. (c)
    SemanticAnnotationAPI, for semantic annotation of documents.
  </wsdl:documentation>
  <wsdl:types>
    <schema elementFormDefault="qualified" targetNamespace="http://www.
      gsigma.ufsc.br/services/ks" xmlns="http://www.w3.org/2001/
      XMLSchema">
      <element name="publishStringDocument">
        <complexType>
          <sequence>
            <element name="content" type="xsd:string" />
            <element name="encoding" minOccurs="0" type="xsd:string" />
            <element maxOccurs="unbounded" minOccurs="0" name="
              customFeatures" type="impl:feature" />
          </sequence>

```

```

    </complexType>
  </element>
  <element name="publishStringDocumentResponse">
    <complexType>
      <sequence>
        <element name="publishStringDocumentReturn" type="xsd:string"
          />
      </sequence>
    </complexType>
  </element>
  <element name="publishUriDocument">
    <complexType>
      <sequence>
        <element name="url" type="xsd:string" />
        <element name="encoding" minOccurs="0" type="xsd:string" />
        <element maxOccurs="unbounded" name="customFeatures"
          minOccurs="0" type="impl:feature" />
      </sequence>
    </complexType>
  </element>
  <element name="publishUriDocumentResponse">
    <complexType>
      <sequence>
        <element name="publishUriDocumentReturn" type="xsd:string" />
      </sequence>
    </complexType>
  </element>
  <element name="loadDocument">
    <complexType>
      <sequence>
        <element name="id" type="xsd:string" />
      </sequence>
    </complexType>
  </element>
  <element name="loadDocumentResponse">
    <complexType>
      <sequence>
        <element name="loadDocumentReturn" type="impl:document" />
      </sequence>
    </complexType>
  </element>
  <element name="loadDocuments">
    <complexType>
      <sequence>
        <element maxOccurs="unbounded" name="idList" type="xsd:string"
          />
      </sequence>
    </complexType>
  </element>
  <element name="loadDocumentsResponse">
    <complexType>
      <sequence>
        <element maxOccurs="unbounded" name="loadDocumentsReturn"
          type="impl:document" />
      </sequence>
    </complexType>
  </element>
  <element name="updateDocument">
    <complexType>
      <sequence>
        <element name="doc" type="impl:document" />
      </sequence>
    </complexType>
  </element>

```

```

</element>
<element name="updateDocumentResponse">
  <complexType />
</element>
<element name="getAnnotations">
  <complexType>
    <sequence>
      <element name="docId" type="xsd:string" />
    </sequence>
  </complexType>
</element>
<element name="getAnnotationsResponse">
  <complexType>
    <sequence>
      <element maxOccurs="unbounded" minOccurs="0" name="
        getAnnotationsReturn" type="impl:annotation" />
    </sequence>
  </complexType>
</element>
<element name="deleteDocument">
  <complexType>
    <sequence>
      <element name="id" type="xsd:string" />
    </sequence>
  </complexType>
</element>
<element name="deleteDocumentResponse">
  <complexType />
</element>
<element name="fault" type="impl:persistence-fault" />
<complexType name="persistence-fault">
  <sequence />
</complexType>
<complexType name="feature">
  <annotation>
    <documentation>
      A feature is a name-value pair used to attach metadata, such
      as title, author, version, date, etc., to documents.
      Features are also semantically annotated and indexed,
      allowing
      searches upon them.
    </documentation>
  </annotation>
  <sequence>
    <element name="name" nillable="true" type="xsd:string" />
    <element name="value" nillable="true" type="xsd:string" />
  </sequence>
</complexType>
<complexType name="feature-list">
  <sequence>
    <element maxOccurs="unbounded" name="item" type="impl:feature"
      />
  </sequence>
</complexType>
<complexType name="document">
  <annotation>
    <documentation>Representation of a document used by the
      Document Manager Service.</documentation>
  </annotation>
  <sequence>
    <element name="id" type="xsd:string">
      <annotation>
        <documentation>

```

```

        An unique id is generated by the Document Manager Service
        when a new document is published. It can be used to
        load, update or delete the document.
    </documentation>
</annotation>
</element>
<element name="content" type="xsd:string">
    <annotation>
        <documentation>Document content.</documentation>
    </annotation>
</element>
<element name="features" minOccurs="0" type="impl:feature-list"
/>
</sequence>
</complexType>
<complexType name="annotation">
    <annotation>
        <documentation>
            An Annotation used to associate a given instance of the
            ontology to a piece of text from the document content.
            The location of the annotated text is defined by
            startOffset and
            endOffset.
        </documentation>
    </annotation>
    <sequence>
        <element name="startOffset" type="xsd:int" />
        <element name="endOffset" type="xsd:int" />
        <element name="instance" type="xsd:QName" />
    </sequence>
</complexType>
<element name="addCustomId">
    <complexType>
        <sequence>
            <element name="docId" type="xsd:string" />
            <element name="customIdName" type="xsd:string" />
            <element name="customIdValue" type="xsd:string" />
        </sequence>
    </complexType>
</element>
<element name="addCustomIdResponse" />
<element name="convertCustomIdToDocId">
    <complexType>
        <sequence>
            <element name="customIdName" type="xsd:string" />
            <element minOccurs="unbounded" name="customId" type="xsd:
                string" />
        </sequence>
    </complexType>
</element>
<element name="convertCustomIdToDocIdResponse">
    <complexType>
        <sequence>
            <element minOccurs="unbounded" name="
                convertCustomIdToDocIdReturn" type="xsd:string" />
        </sequence>
    </complexType>
</element>
<element name="convertDocIdToCustomId">
    <complexType>
        <sequence>
            <element name="customIdName" type="xsd:string" />

```

```

        <element maxOccurs="unbounded" name="docId" type="xsd:string"
            />
    </sequence>
</complexType>
</element>
<element name="convertDocIdToCustomIdResponse">
    <complexType>
        <sequence>
            <element maxOccurs="unbounded" name="
                convertDocIdToCustomIdReturn" type="xsd:string" />
        </sequence>
    </complexType>
</element>
</schema>
</wsdl:types>

<wsdl:message name="deleteDocumentResponse">
    <wsdl:part element="impl:deleteDocumentResponse" name="parameters" />
</wsdl:message>
<wsdl:message name="updateDocumentResponse">
    <wsdl:part element="impl:updateDocumentResponse" name="parameters" />
</wsdl:message>
<wsdl:message name="publishUrlDocumentResponse">
    <wsdl:part element="impl:publishUrlDocumentResponse" name="parameters"
        />
</wsdl:message>
<wsdl:message name="deleteDocumentRequest">
    <wsdl:part element="impl:deleteDocument" name="parameters" />
</wsdl:message>
<wsdl:message name="updateDocumentRequest">
    <wsdl:part element="impl:updateDocument" name="parameters" />
</wsdl:message>
<wsdl:message name="publishStringDocumentRequest">
    <wsdl:part element="impl:publishStringDocument" name="parameters" />
</wsdl:message>
<wsdl:message name="publishUrlDocumentRequest">
    <wsdl:part element="impl:publishUrlDocument" name="parameters" />
</wsdl:message>
<wsdl:message name="getAnnotationsRequest">
    <wsdl:part element="impl:getAnnotations" name="parameters" />
</wsdl:message>
<wsdl:message name="publishStringDocumentResponse">
    <wsdl:part element="impl:publishStringDocumentResponse" name="
        parameters" />
</wsdl:message>
<wsdl:message name="loadDocumentRequest">
    <wsdl:part element="impl:loadDocument" name="parameters" />
</wsdl:message>
<wsdl:message name="PersistenceException">
    <wsdl:part element="impl:fault" name="fault" />
</wsdl:message>
<wsdl:message name="getAnnotationsResponse">
    <wsdl:part element="impl:getAnnotationsResponse" name="parameters" />
</wsdl:message>
<wsdl:message name="addCustomIdRequest">
    <wsdl:part element="impl:addCustomId" name="parameters" />
</wsdl:message>
<wsdl:message name="addCustomIdResponse">
    <wsdl:part element="impl:addCustomIdResponse" name="parameters" />
</wsdl:message>
<wsdl:message name="convertDocIdToCustomIdRequest">
    <wsdl:part element="impl:convertDocIdToCustomId" name="parameters" />
</wsdl:message>

```

```

<wsdl:message name="convertCustomIdToDocIdResponse">
  <wsdl:part element="impl:convertCustomIdToDocIdResponse" name="
    parameters" />
</wsdl:message>
<wsdl:message name="convertDocIdToCustomIdResponse">
  <wsdl:part element="impl:convertDocIdToCustomIdResponse" name="
    parameters" />
</wsdl:message>
<wsdl:message name="convertCustomIdToDocIdRequest">
  <wsdl:part element="impl:convertCustomIdToDocId" name="parameters" />
</wsdl:message>
<wsdl:message name="loadDocumentResponse">
  <wsdl:part name="parameters" element="impl:loadDocumentResponse"></
    wsdl:part>
</wsdl:message>
<wsdl:message name="loadDocumentsRequest">
  <wsdl:part element="impl:loadDocuments" name="parameters" />
</wsdl:message>
<wsdl:message name="loadDocumentsResponse">
  <wsdl:part element="impl:loadDocumentsResponse" name="parameters" />
</wsdl:message>

<wsdl:portType name="DocumentManager">
  <wsdl:operation name="addCustomId">
    <wsdl:documentation>Adds a custom Id to a Document.</wsdl:
      documentation>
    <wsdl:input message="impl:addCustomIdRequest" name="
      addCustomIdRequest" />
    <wsdl:output message="impl:addCustomIdResponse" name="
      addCustomIdResponse" />
    <wsdl:fault message="impl:PersistenceException" name="
      PersistenceException" />
  </wsdl:operation>
  <wsdl:operation name="convertCustomIdToDocId">
    <wsdl:documentation>Converts custom document ids to internal
      document ids.</wsdl:documentation>
    <wsdl:input message="impl:convertCustomIdToDocIdRequest" name="
      convertCustomIdToDocIdRequest" />
    <wsdl:output message="impl:convertCustomIdToDocIdResponse" name="
      convertCustomIdToDocIdResponse" />
    <wsdl:fault message="impl:PersistenceException" name="
      PersistenceException" />
  </wsdl:operation>
  <wsdl:operation name="convertDocIdToCustomId">
    <wsdl:documentation>Converts internal document ids to custom
      document ids.</wsdl:documentation>
    <wsdl:input message="impl:convertDocIdToCustomIdRequest" name="
      convertDocIdToCustomIdRequest" />
    <wsdl:output message="impl:convertDocIdToCustomIdResponse" name="
      convertDocIdToCustomIdResponse" />
    <wsdl:fault message="impl:PersistenceException" name="
      PersistenceException" />
  </wsdl:operation>
  <wsdl:operation name="deleteDocument">
    <wsdl:documentation>Deletes a document from the index and storage
      .</wsdl:documentation>
    <wsdl:input message="impl:deleteDocumentRequest" name="
      deleteDocumentRequest" />
    <wsdl:output message="impl:deleteDocumentResponse" name="
      deleteDocumentResponse" />
    <wsdl:fault message="impl:PersistenceException" name="
      PersistenceException" />
  </wsdl:operation>

```



```

<wsdl:operation name="publishStringDocument">
  <wsdl:documentation>Stores a document and returns its unique
    generated id.</wsdl:documentation>
  <wsdl:input message="impl:publishStringDocumentRequest" name="
    publishStringDocumentRequest" />
  <wsdl:output message="impl:publishStringDocumentResponse" name="
    publishStringDocumentResponse" />
</wsdl:operation>
<wsdl:operation name="publishUrlDocument">
  <wsdl:documentation>Stores a document and returns its unique
    generated id.</wsdl:documentation>
  <wsdl:input message="impl:publishUrlDocumentRequest" name="
    publishUrlDocumentRequest" />
  <wsdl:output message="impl:publishUrlDocumentResponse" name="
    publishUrlDocumentResponse" />
  <wsdl:fault message="impl:PersistenceException" name="
    PersistenceException" />
</wsdl:operation>
<wsdl:operation name="loadDocument">
  <wsdl:documentation>Returns a document.</wsdl:documentation>
  <wsdl:input message="impl:loadDocumentRequest" name="
    loadDocumentRequest" />
  <wsdl:output message="impl:loadDocumentResponse" name="
    loadDocumentResponse" />
  <wsdl:fault message="impl:PersistenceException" name="
    PersistenceException" />
</wsdl:operation>
<wsdl:operation name="loadDocuments">
  <wsdl:documentation>Returns a list of documents.</wsdl:
    documentation>
  <wsdl:input message="impl:loadDocumentsRequest" name="
    loadDocumentsRequest" />
  <wsdl:output message="impl:loadDocumentsResponse" name="
    loadDocumentsResponse" />
  <wsdl:fault message="impl:PersistenceException" name="
    PersistenceException" />
</wsdl:operation>
<wsdl:operation name="updateDocument">
  <wsdl:documentation>Updates a document in the index and storage.</
    wsdl:documentation>
  <wsdl:input message="impl:updateDocumentRequest" name="
    updateDocumentRequest" />
  <wsdl:output message="impl:updateDocumentResponse" name="
    updateDocumentResponse" />
  <wsdl:fault message="impl:PersistenceException" name="
    PersistenceException" />
</wsdl:operation>
<wsdl:operation name="getAnnotations">
  <wsdl:documentation>Returns a list of annotations of a given
    document.</wsdl:documentation>
  <wsdl:input message="impl:getAnnotationsRequest" name="
    getAnnotationsRequest" />
  <wsdl:output message="impl:getAnnotationsResponse" name="
    getAnnotationsResponse" />
  <wsdl:fault message="impl:PersistenceException" name="
    PersistenceException" />
</wsdl:operation>
</wsdl:portType>

<wsdl:binding name="DocumentManagerServiceSoapBinding" type="impl:
  DocumentManager">
  <wsdl:soap:binding style="document" transport="http://schemas.xmlsoap.
    org/soap/http" />

```

```

<wsdl:operation name="addCustomId">
  <wsdlsoap:operation soapAction="" />
  <wsdl:input name="addCustomIdRequest">
    <wsdlsoap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="addCustomIdResponse">
    <wsdlsoap:body use="literal" />
  </wsdl:output>
  <wsdl:fault name="PersistenceException">
    <wsdlsoap:fault name="PersistenceException" use="literal" />
  </wsdl:fault>
</wsdl:operation>
<wsdl:operation name="convertCustomIdToDocId">
  <wsdlsoap:operation soapAction="" />
  <wsdl:input name="convertCustomIdToDocIdRequest">
    <wsdlsoap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="convertCustomIdToDocIdResponse">
    <wsdlsoap:body use="literal" />
  </wsdl:output>
  <wsdl:fault name="PersistenceException">
    <wsdlsoap:fault name="PersistenceException" use="literal" />
  </wsdl:fault>
</wsdl:operation>
<wsdl:operation name="convertDocIdToCustomId">
  <wsdlsoap:operation soapAction="" />
  <wsdl:input name="convertDocIdToCustomIdRequest">
    <wsdlsoap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="convertDocIdToCustomIdResponse">
    <wsdlsoap:body use="literal" />
  </wsdl:output>
  <wsdl:fault name="PersistenceException">
    <wsdlsoap:fault name="PersistenceException" use="literal" />
  </wsdl:fault>
</wsdl:operation>
<wsdl:operation name="deleteDocument">
  <wsdlsoap:operation soapAction="" />
  <wsdl:input name="deleteDocumentRequest">
    <wsdlsoap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="deleteDocumentResponse">
    <wsdlsoap:body use="literal" />
  </wsdl:output>
  <wsdl:fault name="PersistenceException">
    <wsdlsoap:fault name="PersistenceException" use="literal" />
  </wsdl:fault>
</wsdl:operation>
<wsdl:operation name="publishStringDocument">
  <wsdlsoap:operation soapAction="" />
  <wsdl:input name="publishStringDocumentRequest">
    <wsdlsoap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="publishStringDocumentResponse">
    <wsdlsoap:body use="literal" />
  </wsdl:output>
</wsdl:operation>
<wsdl:operation name="publishUrlDocument">
  <wsdlsoap:operation soapAction="" />
  <wsdl:input name="publishUrlDocumentRequest">
    <wsdlsoap:body use="literal" />
  </wsdl:input>
  <wsdl:output name="publishUrlDocumentResponse">

```

```

        <wsdlsoap:body use="literal" />
    </wsdl:output>
    <wsdl:fault name="PersistenceException">
        <wsdlsoap:fault name="PersistenceException" use="literal" />
    </wsdl:fault>
</wsdl:operation>
<wsdl:operation name="loadDocument">
    <wsdlsoap:operation soapAction="" />
    <wsdl:input name="loadDocumentRequest">
        <wsdlsoap:body use="literal" />
    </wsdl:input>
    <wsdl:output name="loadDocumentResponse">
        <wsdlsoap:body use="literal" />
    </wsdl:output>
    <wsdl:fault name="PersistenceException">
        <wsdlsoap:fault name="PersistenceException" use="literal" />
    </wsdl:fault>
</wsdl:operation>
<wsdl:operation name="loadDocuments">
    <wsdlsoap:operation soapAction="" />
    <wsdl:input name="loadDocumentsRequest">
        <wsdlsoap:body use="literal" />
    </wsdl:input>
    <wsdl:output name="loadDocumentsResponse">
        <wsdlsoap:body use="literal" />
    </wsdl:output>
    <wsdl:fault name="PersistenceException">
        <wsdlsoap:fault name="PersistenceException" use="literal" />
    </wsdl:fault>
</wsdl:operation>
<wsdl:operation name="updateDocument">
    <wsdlsoap:operation soapAction="" />
    <wsdl:input name="updateDocumentRequest">
        <wsdlsoap:body use="literal" />
    </wsdl:input>
    <wsdl:output name="updateDocumentResponse">
        <wsdlsoap:body use="literal" />
    </wsdl:output>
    <wsdl:fault name="PersistenceException">
        <wsdlsoap:fault name="PersistenceException" use="literal" />
    </wsdl:fault>
</wsdl:operation>
<wsdl:operation name="getAnnotations">
    <wsdlsoap:operation soapAction="" />
    <wsdl:input name="getAnnotationsRequest">
        <wsdlsoap:body use="literal" />
    </wsdl:input>
    <wsdl:output name="getAnnotationsResponse">
        <wsdlsoap:body use="literal" />
    </wsdl:output>
    <wsdl:fault name="PersistenceException">
        <wsdlsoap:fault name="PersistenceException" use="literal" />
    </wsdl:fault>
</wsdl:operation>
</wsdl:binding>

<wsdl:service name="DocumentManagerService">
    <wsdl:port binding="impl:DocumentManagerServiceSoapBinding" name="
        DocumentManagerService">
        <wsdlsoap:address location="http://localhost:8080/ks/services /
            DocumentManagerService" />
    </wsdl:port>
</wsdl:service>

```

```
</wsdl:definitions >
```

A.3 Serviço Motor de Busca

```
<?xml version="1.0" encoding="UTF-8"?>
<wsdl:definitions targetNamespace="http://www.gsigma.ufsc.br/services/ks"
  xmlns:apachesoap="http://xml.apache.org/xml-soap" xmlns:impl="http
  ://www.gsigma.ufsc.br/services/ks"
  xmlns:intf="http://www.gsigma.ufsc.br/services/ks" xmlns:wsdl="http://
  schemas.xmlsoap.org/wsdl/" xmlns:wdsloap="http://schemas.xmlsoap.
  org/wsdl/soap/" xmlns:xsd="http://www.w3.org/2001/XMLSchema">
  <wsdl:documentation>This Web Service is built on top of KIM Platform ,
  using the QueryAPI.</wsdl:documentation>
  <wsdl:types>
    <schema elementFormDefault="qualified" targetNamespace="http://www.
      gsigma.ufsc.br/services/ks" xmlns="http://www.w3.org/2001/
     /XMLSchema">
      <element name="searchDocumentsWithKeyWords">
        <complexType>
          <sequence>
            <element maxOccurs="unbounded" minOccurs="0" name="
              keywordRestrictions" type="impl:keyword-restriction" />
          </sequence>
        </complexType>
      </element>
      <element name="searchDocumentsWithKeyWordsResponse">
        <complexType>
          <sequence>
            <element maxOccurs="unbounded" minOccurs="0" name="
              searchDocumentsWithEntityReturn" type="xsd:string" />
          </sequence>
        </complexType>
      </element>
      <element name="searchDocuments">
        <complexType>
          <sequence>
            <element maxOccurs="unbounded" name="variables" type="impl:variable
              " />
            <element maxOccurs="unbounded" minOccurs="0" name="
              relationshipRestrictions" type="impl:relationship-restriction"
              />
            <element maxOccurs="unbounded" minOccurs="0" name="
              keywordRestrictions" type="impl:keyword-restriction" />
          </sequence>
        </complexType>
      </element>
      <element name="searchEntitiesAdvanced">
        <complexType>
          <sequence>
            <element maxOccurs="unbounded" name="variables" type="impl:variable
              " />
            <element maxOccurs="unbounded" minOccurs="0" name="
              relationshipRestrictions" type="impl:relationship-restriction"
              />
            <element maxOccurs="unbounded" minOccurs="0" name="
              keywordRestrictions" type="impl:keyword-restriction" />
          </sequence>
        </complexType>
      </element>
      <simpleType name="compare-style">
        <restriction base="xsd:string">
```

```

    <enumeration value="is-unknown" />
    <enumeration value="is-exactly" />
    <enumeration value="starts-with" />
    <enumeration value="ends-with" />
    <enumeration value="contains" />
    <enumeration value="is-greater-before" />
    <enumeration value="is-less-after" />
  </restriction>
</simpleType>
<complexType name="attribute-restriction">
  <annotation>
    <documentation>Attribute restrictions narrow the search by
      restricting some of the properties of the class bound to the
      variable.</documentation>
  </annotation>
  <sequence>
    <element name="propertyUri" minOccurs="0" type="xsd:QName">
      <annotation>
        <documentation>URI of the property to be restricted. It should be
          retrieved by the Ontology Browser Service.</documentation>
      </annotation>
    </element>
    <element name="compareStyle" minOccurs="0" type="impl:compare-style
      ">
      <annotation>
        <documentation>Style of comparison. If absent, it is assumed the
          value "is-exactly".</documentation>
      </annotation>
    </element>
    <element name="objectLiteral" type="xsd:string">
      <annotation>
        <documentation>Value to be compared.</documentation>
      </annotation>
    </element>
  </sequence>
</complexType>
<complexType name="attribute-restriction-list">
  <sequence>
    <element maxOccurs="unbounded" minOccurs="1" name="item" type="impl:
      attribute-restriction" />
  </sequence>
</complexType>
<complexType name="variable">
  <annotation>
    <documentation>
      Variables are part of a semantic query. A variable is bound to a
      class of the ontology, and can have a set of attribute
      restrictions. Variables can also be bounded to document's
      features.
    </documentation>
  </annotation>
  <sequence>
    <element name="name" type="xsd:string" />
    <element name="classUri" type="xsd:QName">
      <annotation>
        <documentation>URI of the class the variable will be bound. It
          should be retrieved by the Ontology Browser Service.</
          documentation>
      </annotation>
    </element>
    <element name="attributeRestrictions" minOccurs="0" type="impl:
      attribute-restriction-list">
      <annotation>

```

```

    <documentation>A set of attribute restrictions.</documentation>
  </annotation>
</element>
<element name="featureBinding" minOccurs="0" type="xsd:string">
  <annotation>
    <documentation>Binds the variable to a document's feature.</
      documentation>
  </annotation>
</element>
</sequence>
</complexType>
<complexType name="relationship-restriction">
  <annotation>
    <documentation>Relationship restrictions are part of a semantic
      query, they narrow the search by restricting relations between
      classes.</documentation>
  </annotation>
</sequence>
  <element name="subjectVar" type="xsd:string">
    <annotation>
      <documentation>Variable name of the subject of the relation.</
        documentation>
    </annotation>
  </element>
  <element name="propertyUri" type="xsd:QName">
    <annotation>
      <documentation>URI of a property from subjectVar that is a
        relation to objectVar. This URI should be retrieved by the
        Ontology Browser Service.</documentation>
    </annotation>
  </element>
  <element name="objectVar" type="xsd:string">
    <annotation>
      <documentation>Variable name of the object of the relation.</
        documentation>
    </annotation>
  </element>
</sequence>
</complexType>
<complexType name="keyword-restriction">
  <annotation>
    <documentation>Keyword restrictions are the non-semantic part of a
      semantic query, they narrow the search by restricting some
      document's features.</documentation>
  </annotation>
</sequence>
  <element name="featureName" type="xsd:string">
    <annotation>
      <documentation>Name of the feature to be restricted.</
        documentation>
    </annotation>
  </element>
  <element name="keyword" type="xsd:string">
    <annotation>
      <documentation>A String containing some keyword-based pattern.</
        documentation>
    </annotation>
  </element>
</sequence>
</complexType>
<element name="searchDocumentsResponse">
  <complexType>
    <sequence>

```

```

        <element maxOccurs="unbounded" minOccurs="0" name="
            searchDocumentsReturn" type="xsd:string" />
    </sequence>
</complexType>
</element>
<complexType name="query-fault">
    <sequence />
</complexType>
<element name="fault" type="impl:query-fault" />
<element name="searchEntities">
    <complexType>
        <sequence>
            <element name="name" type="xsd:string" />
            <element name="compareStyle" type="impl:compare-style" minOccurs
                ="0" />
        </sequence>
    </complexType>
</element>
<element name="searchEntitiesResponse">
    <complexType>
        <sequence>
            <element maxOccurs="unbounded" minOccurs="0" name="
                searchEntitiesReturn" type="xsd:QName" />
        </sequence>
    </complexType>
</element>

<element name="searchEntitiesAdvancedResponse">
    <complexType>
        <sequence>
            <element maxOccurs="unbounded" minOccurs="0" name="
                searchEntitiesReturn" type="xsd:QName" />
        </sequence>
    </complexType>
</element>

<element name="searchDocumentsWithEntity">
    <complexType>
        <sequence>
            <element name="entity" type="xsd:QName" />
        </sequence>
    </complexType>
</element>
<element name="searchDocumentsWithEntityResponse">
    <complexType>
        <sequence>
            <element maxOccurs="unbounded" minOccurs="0" name="
                searchDocumentsWithEntityReturn" type="xsd:string" />
        </sequence>
    </complexType>
</element>
</schema>
</wsdl:types>
<wsdl:message name="searchDocumentsResponse">
    <wsdl:part element="impl:searchDocumentsResponse" name="parameters" />
</wsdl:message>
<wsdl:message name="searchDocumentsRequest">
    <wsdl:part element="impl:searchDocuments" name="parameters" />
</wsdl:message>
<wsdl:message name="searchDocumentsWithKeyWordsResponse">
    <wsdl:part element="impl:searchDocumentsWithKeyWordsResponse" name="
        parameters" />
</wsdl:message>

```

```

<wsdl:message name="searchDocumentsWithKeyWordsRequest">
  <wsdl:part element="impl:searchDocumentsWithKeyWords" name="parameters"
  />
</wsdl:message>
<wsdl:message name="QueryException">
  <wsdl:part element="impl:fault" name="fault" />
</wsdl:message>
<wsdl:message name="searchEntitiesRequest">
  <wsdl:part element="impl:searchEntities" name="parameters" />
</wsdl:message>

<wsdl:message name="searchEntitiesAdvancedRequest">
  <wsdl:part element="impl:searchEntitiesAdvanced" name="parameters" />
</wsdl:message>
<wsdl:message name="searchEntitiesResponse">
  <wsdl:part element="impl:searchEntitiesResponse" name="parameters" />
</wsdl:message>

<wsdl:message name="searchEntitiesAdvancedResponse">
  <wsdl:part element="impl:searchEntitiesAdvancedResponse" name="
  parameters" />
</wsdl:message>
<wsdl:message name="searchDocumentsWithEntityRequest">
  <wsdl:part element="impl:searchDocumentsWithEntity" name="parameters"
  />
</wsdl:message>
<wsdl:message name="searchDocumentsWithEntityResponse">
  <wsdl:part element="impl:searchDocumentsWithEntityResponse" name="
  parameters" />
</wsdl:message>
<wsdl:portType name="SearchEngine">
  <wsdl:operation name="searchDocuments">
    <wsdl:documentation>
      Search for documents indexed and stored by the Document Manager
      Service. The parameters comprise the so-called &quot;semantic
      query&quot;. It returns a list of document ids.
    </wsdl:documentation>
    <wsdl:input message="impl:searchDocumentsRequest" name="
    searchDocumentsRequest" />
    <wsdl:output message="impl:searchDocumentsResponse" name="
    searchDocumentsResponse" />
    <wsdl:fault message="impl:QueryException" name="QueryException" />
  </wsdl:operation>
  <wsdl:operation name="searchEntities">
    <wsdl:input message="impl:searchEntitiesRequest" name="
    searchEntitiesRequest" />
    <wsdl:output message="impl:searchEntitiesResponse" name="
    searchEntitiesResponse" />
    <wsdl:fault message="impl:QueryException" name="QueryException" />
  </wsdl:operation>

  <wsdl:operation name="searchEntitiesAdvanced">
    <wsdl:input message="impl:searchEntitiesAdvancedRequest" name="
    searchEntitiesAdvancedRequest" />
    <wsdl:output message="impl:searchEntitiesAdvancedResponse" name="
    searchEntitiesAdvancedResponse" />
    <wsdl:fault message="impl:QueryException" name="QueryException" />
  </wsdl:operation>
  <wsdl:operation name="searchDocumentsWithEntity">
    <wsdl:input message="impl:searchDocumentsWithEntityRequest" name="
    searchDocumentsWithEntityRequest" />
    <wsdl:output message="impl:searchDocumentsWithEntityResponse" name="
    searchDocumentsWithEntityResponse" />
  </wsdl:operation>

```



```
</wsdl:operation>
<wsdl:operation name="searchDocumentsWithKeyWords">
  <wsdl:input message="impl:searchDocumentsWithKeyWordsRequest" name="
    searchDocumentsWithKeyWordsRequest" />
  <wsdl:output message="impl:searchDocumentsWithKeyWordsResponse" name="
    searchDocumentsWithKeyWordsResponse" />
</wsdl:operation>
</wsdl:portType>
<wsdl:binding name="SearchEngineServiceSoapBinding" type="impl:
  SearchEngine">
  <wsdlsoap:binding style="document" transport="http://schemas.xmlsoap.
    org/soap/http" />
  <wsdl:operation name="searchDocuments">
    <wsdlsoap:operation soapAction="" />
    <wsdl:input name="searchDocumentsRequest">
      <wsdlsoap:body use="literal" />
    </wsdl:input>
    <wsdl:output name="searchDocumentsResponse">
      <wsdlsoap:body use="literal" />
    </wsdl:output>
    <wsdl:fault name="QueryException">
      <wsdlsoap:fault name="QueryException" use="literal" />
    </wsdl:fault>
  </wsdl:operation>
  <wsdl:operation name="searchEntities">
    <wsdlsoap:operation soapAction="" />
    <wsdl:input name="searchEntitiesRequest">
      <wsdlsoap:body use="literal" />
    </wsdl:input>
    <wsdl:output name="searchEntitiesResponse">
      <wsdlsoap:body use="literal" />
    </wsdl:output>
    <wsdl:fault name="QueryException">
      <wsdlsoap:fault name="QueryException" use="literal" />
    </wsdl:fault>
  </wsdl:operation>
  <wsdl:operation name="searchEntitiesAdvanced">
    <wsdlsoap:operation soapAction="" />
    <wsdl:input name="searchEntitiesAdvancedRequest">
      <wsdlsoap:body use="literal" />
    </wsdl:input>
    <wsdl:output name="searchEntitiesAdvancedResponse">
      <wsdlsoap:body use="literal" />
    </wsdl:output>
    <wsdl:fault name="QueryException">
      <wsdlsoap:fault name="QueryException" use="literal" />
    </wsdl:fault>
  </wsdl:operation>
  <wsdl:operation name="searchDocumentsWithEntity">
    <wsdlsoap:operation soapAction="" />
    <wsdl:input name="searchDocumentsWithEntityRequest">
      <wsdlsoap:body use="literal" />
    </wsdl:input>
    <wsdl:output name="searchDocumentsWithEntityResponse">
      <wsdlsoap:body use="literal" />
    </wsdl:output>
  </wsdl:operation>
  <wsdl:operation name="searchDocumentsWithKeyWords">
    <wsdlsoap:operation soapAction="" />
    <wsdl:input name="searchDocumentsWithKeyWordsRequest">
      <wsdlsoap:body use="literal" />
    </wsdl:input>
```

```
<wsdl:output name="searchDocumentsWithKeyWordsResponse">
  <wsdlsoap:body use="literal" />
</wsdl:output>
</wsdl:operation>
</wsdl:binding>
<wsdl:service name="SearchEngineService">
  <wsdl:port binding="impl:SearchEngineServiceSoapBinding" name="
    SearchEngineService">
    <wsdlsoap:address location="ks/services/SearchEngineService" />
  </wsdl:port>
</wsdl:service>
</wsdl:definitions>
```

Apêndice B

Diagramas UML do Protótipo

Este apêndice apresenta os diagramas UML referentes à implementação do protótipo.

B.1 Serviço para Navegação de Ontologias

Os serviço de navegação de ontologias é implementado pela classe `OntologyBrowser`, conforme mostra o diagrama da figura B.1. Tal classe depende dos seguintes pacotes:

- `ontology`: contém o modelo de dados utilizado por esse serviço, ou seja, classes e propriedades da ontologia, exceções, e uma classe contendo constantes para URIs. Como a ontologia e suas instâncias estão representadas segundo a linguagem OWL, as classes, propriedades e instâncias são identificadas por URIs. Portanto, o serviço de navegação de ontologias basicamente lida com parâmetros e valores de retorno do tipo URI.
- `org.openrdf`: contém classes e interfaces utilizadas internamente pelo KIM / Sesame para manipular a ontologia.
- `com.ontotext.kim.client`: contém as APIs do KIM que são utilizadas por este serviço. As principais classes são:
 - `SemanticRepositoryAPI`: encapsula o repositório de ontologias do Sesame e provê métodos para gestão da ontologia e suas instâncias.
 - `GetService` e `KIMService`: a primeira é uma classe usada para acessar o servidor RMI do KIM, e a segunda é uma interface que representa o serviço (ponto de acesso) propriamente dito do KIM. É também utilizada pelos serviços de *gestão de documentos e motor busca*.

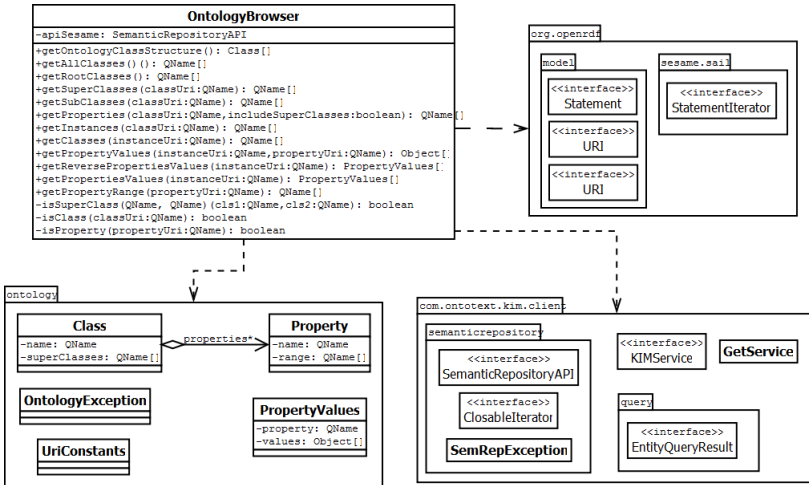


Figura B.1: Diagrama de classes do serviço de navegação na ontologia.

B.2 Serviço para Gestão de Documentos

Este serviço é implementado pela classe `DocumentManager`, que depende dos seguintes pacotes, conforme o diagrama da figura B.2:

- `document`: este pacote contém as classes de suporte às operações desse serviço:
 - `Document`: classe que representa um documento, armazenando o seu conteúdo, um identificador, uma lista de *features*, e uma lista de anotações semânticas.
 - `Feature`: uma *feature* consiste em um par “nome-valor” usado para adicionar metadados personalizados aos documentos, como por exemplo, título, autor, versão, data, entre outros. As features também são anotadas semanticamente e indexadas, oferecendo um filtro adicional durante o processo de busca.
 - `Annotation`: representa uma anotação semântica. Ela contém a URI da instância usada na anotação bem como a posição de início e fim (localização) dentro do documento.
 - `PersistenceException`: representa situações de erro como, por exemplo, o carregamento de um documento inexistente.
- `com.ontotext.kim.client`: contém as APIs do KIM que são utilizadas por este serviço. Este pacote é organizado em vários sub-pacotes,

sendo algumas classes de especial interesse:

- CorporaAPI, que provê métodos para manipulação de documentos.
- SemanticAnnotationAPI, que provê anotação semântica automática de documentos.
- IndexAndPersistAPI, que provê, além da indexação, o armazenamento e a recuperação de documentos.
- Classes do pacote `corpora`, usadas pelo KIM para representar internamente os documentos, suas *features* e anotações.
- `GetService` e `KIMService`: usadas para acessar utilizar o servidor RMI do KIM.

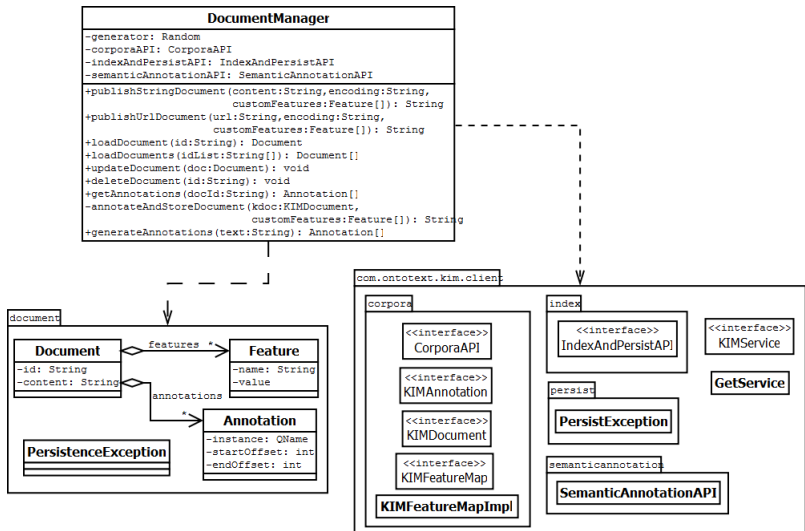


Figura B.2: Diagrama de classes do serviço de gestão de documentos.

B.3 Serviço Motor de Busca

O serviço motor de busca é implementado pela classe `SearchEngine`, que tem relação com os seguintes pacotes (figura B.3):

- `proxies`: contém as classes para acesso aos serviços `ContextCatcher` e `RuleEngine`. Tais serviços são invocados durante a busca baseada em contexto para a captura e expansão da consulta, respectivamente.

- `search`: contém algumas classes que implementam o modelo de dados utilizados pelo serviço. A classe `QueryException` representa exceções relacionadas a consultas incorretas. A classe `SemanticQuery` representa uma consulta semântica. Esta classe é formada por uma agregação de outras classes (variáveis e restrições), conforme é explicado a seguir. Já a classe `SemanticQueryDaoXml` é usada para carregar e salvar consultas semânticas em formato XML. Esta última classe é muito útil para execuções de consultas em lote, pois estas podem ficar armazenadas em arquivos XML e carregadas automaticamente.
- `com.ontotext.kim.client`: contém as APIs do KIM que são utilizadas por este serviço. Este pacote é organizado em vários sub-pacotes, com destaque para o pacote `query`, que contém as classes utilizadas nas operações de busca e para a representação interna de consultas e resultados usadas pela plataforma KIM. Além disso, há também a classe `GetService` e a interface `KIMService`, usadas para acessar o KIM via RMI.

O diagrama de sequência da figura B.4 ilustra o funcionamento da busca usando semântica e contexto. Pode-se reparar no diagrama que o identificador do usuário é usado para se recuperar o seu contexto, que em seguida é usado para a expansão da consulta.

B.4 Serviço Capturador de Contextos

A classe `ContextCatcher` implementa o serviço de captura de contextos, conforme ilustra o diagrama da figura B.5. Ele oferece uma única operação, responsável por recuperar o contexto do usuário, dado o seu identificador.

B.5 Serviço Motor de Regras

O serviço motor de regras é implementado pela classe `RuleEngine`. Através do método `customizeQuery`, esta classe aplica as regras e expande a consulta de acordo com o contexto do usuário. Esta classe faz uso do motor de regras oferecido pela plataforma Jess e utiliza algumas de suas API. A figura B.6 apresenta o diagrama de classes deste serviço.

A classe `RuleEngine` depende dos seguintes pacotes:

- `context`: contém a classe `GenericContext`, utilizada para representar o contexto.
- `search`: este pacote contém as classes que definem uma consulta semântica.

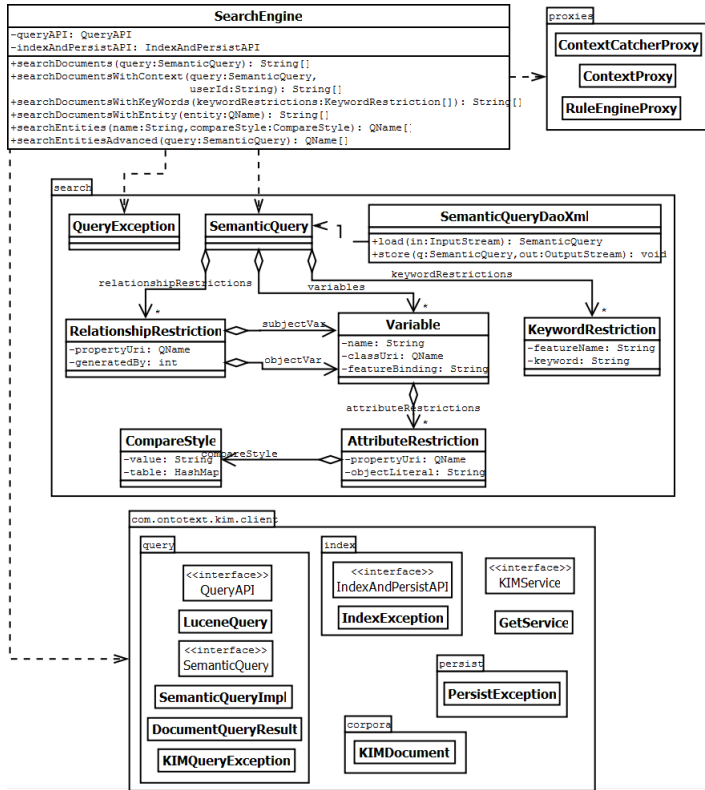


Figura B.3: Diagrama de classes do serviço motor de busca.

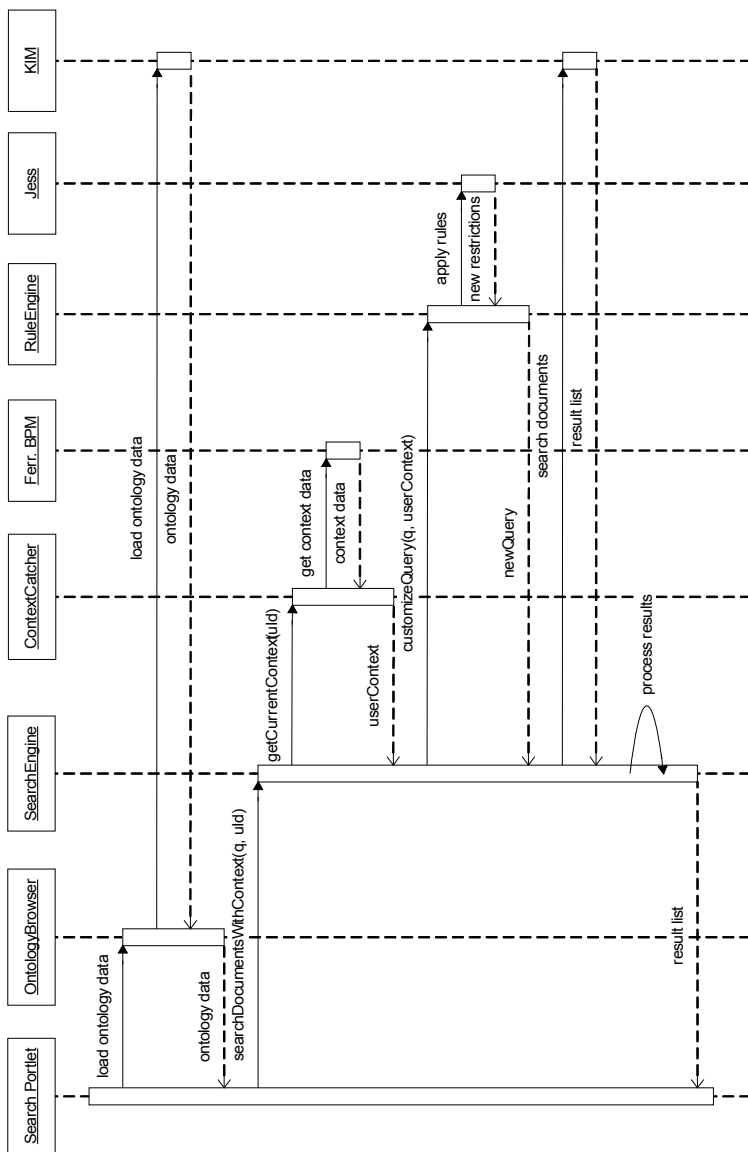


Figura B.4: Diagrama de sequência da busca usando o contexto.

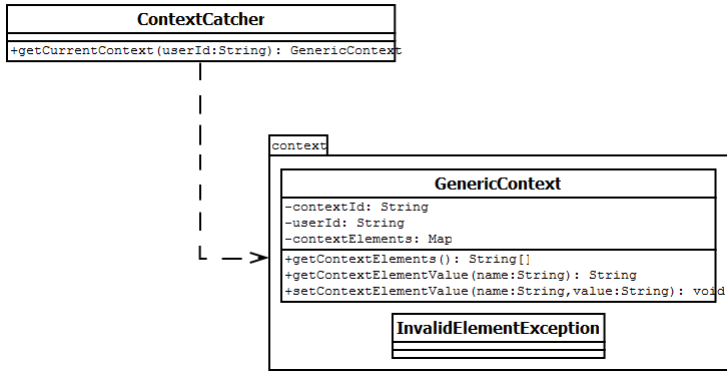


Figura B.5: Diagrama de classes do serviço de captura de contextos.

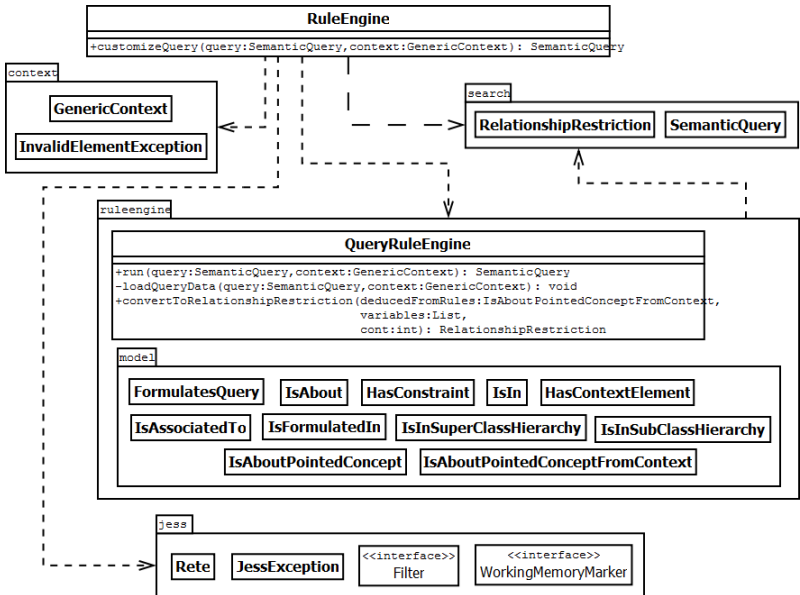


Figura B.6: Diagrama de classes do serviço motor de regras.

- `ruleengine`: contém a implementação propriamente dita do motor de busca. A razão para tal separação é tornar o código do serviço mais claro. Neste caso, a classe `RuleEngine` apenas delega a implementação para a classe `QueryRuleEngine` através do método `run()`. Este pacote também contém o sub-pacote `model`, onde estão definidas classes que permitem a troca de dados entre o motor de regras e o modelo de dados do *k-search*. Este aspecto é detalhado mais adiante, na subseção 6.2.5.
- `jess`: contém as classes que permitem a utilização do Jess. A classe `Rete` implementa o motor de regras do Jess.

B.6 Implantação do Protótipo

Como as APIs da plataforma KIM são acessíveis via protocolo Java RMI, a implantação pode ser feita de maneira distribuída. Além disso, os serviços podem ser implantados em um servidor diferente do servidor do portal (para as aplicações cliente). Esta seria a configuração com a carga de trabalho melhor distribuída, pois tanto a plataforma KIM quanto a implementação do portal consomem muitos recursos. A plataforma KIM, por exemplo, exige pelo menos 1 GB de memória para ter uma execução satisfatória. Tal configuração de implantação é apresentada na figura B.7.

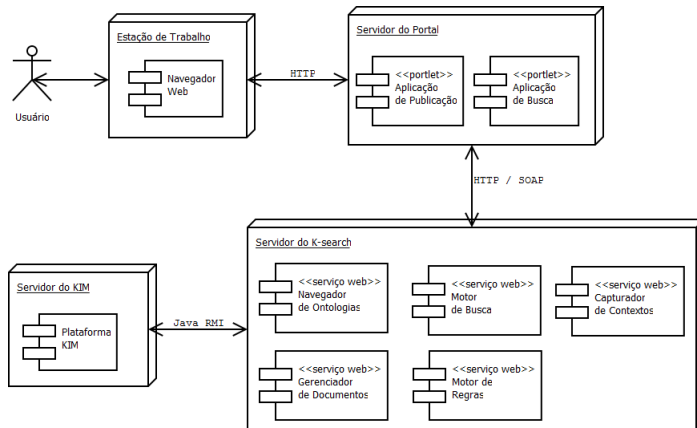


Figura B.7: Diagrama de implantação do *k-search*.

Apêndice C

Base de conhecimento

A base de conhecimento usada nos experimentos é apresentada em detalhes a seguir. Alguns números dessa base podem ser vistos nas tabelas C.1 e C.2.

| | |
|-------------------------------|------------|
| Total de instâncias: | 159 |
| RCs: | 5 |
| Participantes: | 23 |
| Papéis: | 9 |
| Recursos: | 19 |
| Oportunidades de Colaboração: | 2 |
| Competências: | 2 |
| Indicadores de Desempenho: | 20 |
| Atributos dos Indicadores: | 27 |
| Processos: | 11 |
| Tarefas: | 41 |

Tabela C.1: Instâncias da base de conhecimento.

| | |
|----------------------------|------------|
| Total de relações: | 281 |
| RCs e Participantes: | 23 |
| Participantes e Papéis: | 23 |
| Participantes e Recursos: | 28 |
| Participantes e Processos: | 11 |
| Processos e Recursos: | 22 |
| Processos e Tarefas: | 42 |
| Papéis e Tarefas: | 24 |
| Tarefas e Recursos: | 23 |

Tabela C.2: Relações da base de conhecimento.

Instâncias

ACV(IECOS)
 ACV(Helice)
 ACV(CeBeNetwork)
 OV(VirtuelleFabrik)
 OV(OIN)

Participante(IPACSA)
 Participante(EspecialidadyRepuestosPetroquimicos)
 Participante(AltaTecnologiaEnMaquinados)
 Participante(MaquinadosAML)
 Participante(PolimaquinadosMoreno)
 Participante(ISOIN)
 Participante(MEUPE)
 Participante(INESPASA)
 Participante(AEROSUR)
 Participante(GHESA)
 Participante(RegionalDevelopmentAgency)
 Participante(EuroEngineeringAerospace)
 Participante(P+Z)
 Participante(Aerocon)
 Participante(AIDADevelopment)
 Participante(Orona)
 Participante(EIC)
 Participante(IKERLAN)
 Participante(MFE)
 Participante(ELECTRA)
 Participante(MONDRAGON)
 Participante(Herreros)
 Participante(ZABALA)

MembroDoACV(MembroDoIECOS)
 MembroDoACV(MembroDoHelice)
 MembroDoACV(MembroDoCeBeNetwork)
 AdministradorDoACV(AdministradorDoIECOS)
 AdministradorDoACV(AdministradorDoHelice)
 AdministradorDoACV(AdministradorDoCeBeNetwork)
 ProvedorDeSuporteAoACV(ProvedorDeSuporteAoHelice)
 ParceiroDaOV(ParceiroDaOIN)
 CoordenadorDaOV(CoordenadorDaOIN)

Humano(CarmenAguilera)
 Humano(JoseParejo)
 Humano(RicardoGalan)
 Humano(SimonVazquez)
 Humano(BertholdTiefensee)
 Humano(DanielLassig)

Tecnologico(MSMS)
 Tecnologico(PCMS)
 Tecnologico(ODMS)
 Tecnologico(TrustMan)
 Tecnologico(VIMS)
 Tecnologico(COC-Plan)
 Tecnologico(PSS)
 Tecnologico(WizAN)
 Tecnologico(VO-Mod)
 Tecnologico(DSS)
 Tecnologico(SID)
 Tecnologico(DI3)
 Tecnologico(MAF)

OportunidadeDeColaboracao(Centrifuge_Machine)
 OportunidadeDeColaboracao(Punch_for_biopsies)

Competencia (Manufacturing)
 Competencia (Aerospace_Engineering)
 Processo (Registration_of_new_members)
 Processo (VBE_Performance_management)
 Processo (VO_Creation_in_IECOS)
 Processo (VBE_Characterization)
 Processo (VO_Creation_in_Helice)
 Processo (VO_Launch)
 Processo (VO_Performance_Measurement)
 Processo (Integrated_VO_Management)
 Processo (OIN_Roadmapping)
 Processo (VO_Management)
 Processo (CBN_Management)
 Tarefa (New_Applicant)
 Tarefa (Registration_Process)
 Tarefa (Profiling_and_Compentency_Management)
 Tarefa (Update_Information)
 Tarefa (Assigment_of_Roles_Rights_and_Responsabilities)
 Tarefa (New_Compentencies)
 Tarefa (Indicators_Definition)
 Tarefa (Collect_and_Store_Information)
 Tarefa (Performance_Assessment)
 Tarefa (Collect_and_Store_Trust_Information)
 Tarefa (Trust_Assessment)
 Tarefa (Performance_Trust_Monitoring)
 Tarefa (Activate_Alarming_Mechanism)
 Tarefa (Analysis_and_Decision_Making)
 Tarefa (Collaboration_Oportunity_Detection)
 Tarefa (CO_Characterization_Process)
 Tarefa (Material_Selection_Process)
 Tarefa (VO_Rough_Plan_Process)
 Tarefa (Partners_Search_and_Selection)
 Tarefa (Negotiation_Process)
 Tarefa (Contracting : _Send_Purchase_Order)
 Tarefa (Receive_Purchase_Order)
 Tarefa (VO_Registration)
 Tarefa (Requirement_identification)
 Tarefa (Characterization_of_the_actors)
 Tarefa (Definition_of_the_organizational_framework)
 Tarefa (Member_profiling)
 Tarefa (Definition_of_the_business_strategy)
 Tarefa (Proactiveness_measurement_of_members)
 Tarefa (Information_forwarding)
 Tarefa (Identification_of_business_opportunities)
 Tarefa (VO_partners_selection)
 Tarefa (VO_rough_planning)
 Tarefa (Identification_of_contract_objects)
 Tarefa (Collaborative_agreement)
 Tarefa (Contract_negotiation)
 Tarefa (VO_launch)
 Tarefa (Performance_tests)
 Tarefa (VO_Modelling)
 Tarefa (VO_Monitoring)
 Tarefa (VO_Analysis)
 IndicadorDeDesempenho (Percentage_of_Qualified_Suppliers)
 IndicadorDeDesempenho (Cost_per_Invoice)
 IndicadorDeDesempenho (Demand/Supply_Planning_Costs)
 IndicadorDeDesempenho (Incoming_Production_Material_Quality)
 IndicadorDeDesempenho (Order_Management_Costs)

IndicadorDeDesempenho (Order_Management_Cycle_Time)
 IndicadorDeDesempenho (Package_Cycle_Time)
 IndicadorDeDesempenho (Packaging_Cost)
 IndicadorDeDesempenho (Cost_of_storage_space)
 IndicadorDeDesempenho (Production_Plan_Adherence)
 IndicadorDeDesempenho (Production_Process_Validation_Frequency)
 IndicadorDeDesempenho (Percentage_of_Faultless_production)
 IndicadorDeDesempenho (Production_Material_Cycle_Time)
 IndicadorDeDesempenho (Deviation_from_planned_start_date)
 IndicadorDeDesempenho (Number_of_re-planning)
 IndicadorDeDesempenho (Problem_Compensation_performance_of_a_VO_member)
 IndicadorDeDesempenho (Reliability_of_VO_member's_statements/_predictions)
 IndicadorDeDesempenho (Percentage_of_available_of_resources)
 IndicadorDeDesempenho (Satisfaction_with_collaboration_performance)
 IndicadorDeDesempenho (Number_of_call_backs_as_Percentage_of_total_inquiries)

ObjetoMensurado (Acquisition)
 ObjetoMensurado (Packaging)
 ObjetoMensurado (Invoice)
 ObjetoMensurado (Delivery)
 ObjetoMensurado (Planning)
 ObjetoMensurado (Management)
 ObjetoMensurado (Orders)
 ObjetoMensurado (Product)
 ObjetoMensurado (Inventory)
 ObjetoMensurado (Production)
 ObjetoMensurado (Material)

ObjetivoDaMedição (Planning)
 ObjetivoDaMedição (Delivery)
 ObjetivoDaMedição (Management)
 ObjetivoDaMedição (Production)
 ObjetivoDaMedição (Delivery)
 ObjetivoDaMedição (Inventory)

PerspectivaDeDesempenho (Reliability)
 PerspectivaDeDesempenho (Cost)
 PerspectivaDeDesempenho (Quality)
 PerspectivaDeDesempenho (Time)
 PerspectivaDeDesempenho (Flexibility)
 PerspectivaDeDesempenho (Productivity)
 PerspectivaDeDesempenho (Collaboration)

ContextoOrganizacional (Intra)
 ContextoOrganizacional (Inter)
 ContextoOrganizacional (IntraInter)

Relações

temParticipante (IECOS , IPACSA)
 temParticipante (IECOS , EspecialidadyRepuestosPetroquimicos)
 temParticipante (IECOS , AltaTecnologiaEnMaquinados)
 temParticipante (IECOS , MaquinadosAML)
 temParticipante (IECOS , PolimaquinadosMoreno)
 temParticipante (IECOS , Orona)
 temParticipante (IECOS , EIC)
 temParticipante (IECOS , ELECTRA)
 temParticipante (IECOS , MONDRAGON)
 temParticipante (Helice , ISOIN)
 temParticipante (Helice , MEUPE)
 temParticipante (Helice , INESPASA)

```

temParticipante (Helice , AEROSUR)
temParticipante (Helice , GHESA)
temParticipante (Helice , RegionalDevelopmentAgency)
temParticipante (CeBeNetwork , EuroEngineeringAerospace)
temParticipante (CeBeNetwork , P+Z)
temParticipante (CeBeNetwork , Aerocon)
temParticipante (CeBeNetwork , AIDADevelopment)
temParticipante (OIN , Orona)
temParticipante (OIN , EIC)
temParticipante (OIN , IKERLAN)
temParticipante (OIN , MFE)
temParticipante (OIN , ELECTRA)
temParticipante (OIN , MONDRAGON)
temParticipante (OIN , Herreros)
temParticipante (OIN , ZABALA)

temPapel (IPACSA , AdministradorDoIECOS)
temPapel (EspecialidadyRepuestosPetroquimicos , MembroDoIECOS)
temPapel (AltaTecnologiaEnMaquinados , MembroDoIECOS)
temPapel (MaquinadosAML , MembroDoIECOS)
temPapel (PolimaquinadosMoreno , MembroDoIECOS)
temPapel (ISOIN , AdministradorDoHelice)
temPapel (MEUPE , MembroDoHelice)
temPapel (INESPASA , MembroDoHelice)
temPapel (AEROSUR , MembroDoHelice)
temPapel (GHESA , MembroDoHelice)
temPapel (RegionalDevelopmentAgency , ProvedorDeSuporteAoHelice)
temPapel (EuroEngineeringAerospace , AdministradorDoCeBeNetwork)
temPapel (P+Z , MembroDoCeBeNetwork)
temPapel (Aerocon , MembroDoCeBeNetwork)
temPapel (AIDADevelopment , MembroDoCeBeNetwork)
temPapel (Orona , CoordenadorDaOIN)
temPapel (EIC , ParceiroDaOIN)
temPapel (IKERLAN , ParceiroDaOIN)
temPapel (MFE , ParceiroDaOIN)
temPapel (ELECTRA , ParceiroDaOIN)
temPapel (MONDRAGON , ParceiroDaOIN)
temPapel (Herreros , ParceiroDaOIN)
temPapel (ZABALA , ParceiroDaOIN)

temRecurso (ISOIN , CarmenAguilera)
temRecurso (ISOIN , JoseParejo)
temRecurso (ISOIN , RicardoGalan)
temRecurso (Helice , SimonVazquez)
temRecurso (CeBeNetwork , BertholdTiefensee)
temRecurso (CeBeNetwork , DanielLassig)
temRecurso (IECOS , MSMS)
temRecurso (IECOS , PCMS)
temRecurso (IECOS , ODMS)
temRecurso (IECOS , TrustMan)
temRecurso (IECOS , DSS)
temRecurso (IECOS , VIMS)
temRecurso (IECOS , COC-Plan)
temRecurso (IECOS , PSS)
temRecurso (Helice , PCMS)
temRecurso (Helice , MSMS)
temRecurso (Helice , TrustMan)
temRecurso (Helice , WizAN)
temRecurso (Helice , COC-Plan)
temRecurso (Helice , PSS)
temRecurso (VirtuelleFabrik , DI3)
temRecurso (VirtuelleFabrik , SID)
temRecurso (VirtuelleFabrik , VO-Mod)

```

```

temRecurso(OIN, VO-Mod)
temRecurso(OIN, SID)
temRecurso(OIN, DI3)
temRecurso(OIN, MAF)
temRecurso(OIN, DSS)

temProcesso(IECOS, Registration_of_new_members)
temProcesso(IECOS, VBE_Performance_management)
temProcesso(IECOS, VO_Creation_in_IECOS)
temProcesso(Helice, VBE_Characterization)
temProcesso(Helice, VO_Creation_in_Helice)
temProcesso(Helice, VO_Launch)
temProcesso(VirtuelleFabrik, VO_Performance_Measurement)
temProcesso(VirtuelleFabrik, Integrated_VO_Management)
temProcesso(OIN, OIN_Roadmapping)
temProcesso(OIN, VO_Management)
temProcesso(CeBeNetwork, CBN_Management)

usaRecurso(Registration_of_new_members, MSMS)
usaRecurso(Registration_of_new_members, PCMS)
usaRecurso(Registration_of_new_members, ODMS)
usaRecurso(VBE_Performance_management, TrustMan)
usaRecurso(VBE_Performance_management, DSS)
usaRecurso(VO_Creation_in_IECOS, VIMS)
usaRecurso(VO_Creation_in_IECOS, COC-Plan)
usaRecurso(VO_Creation_in_IECOS, PSS)
usaRecurso(VBE_Characterization, MSMS)
usaRecurso(VBE_Characterization, PCMS)
usaRecurso(VBE_Characterization, TrustMan)
usaRecurso(VO_Creation_in_Helice, COC-Plan)
usaRecurso(VO_Creation_in_Helice, PSS)
usaRecurso(VO_Launch, WizAN)
usaRecurso(VO_Performance_Measurement, DI3)
usaRecurso(VO_Performance_Measurement, SID)
usaRecurso(Integrated_VO_Management, VO-Mod)
usaRecurso(VO_Management, VO-Mod)
usaRecurso(VO_Management, SID)
usaRecurso(VO_Management, DI3)
usaRecurso(VO_Management, MAF)
usaRecurso(VO_Management, DSS)

temTarefa(Registration_of_new_members, New_Applicant)
temTarefa(Registration_of_new_members, Registration_Process)
temTarefa(Registration_of_new_members,
  Profiling_and_Compentency_Management)
temTarefa(Registration_of_new_members, Update_Information)
temTarefa(Registration_of_new_members,
  Assigment_of_Roles_Rights_and_Responsibilities)
temTarefa(Registration_of_new_members, New_Compentencies)
temTarefa(VBE_Performance_management, Indicators_Definition)
temTarefa(VBE_Performance_management, Collect_and_Store_Information)
temTarefa(VBE_Performance_management, Performance_Assessment)
temTarefa(VBE_Performance_management, Collect_and_Store_Trust_Information
)
temTarefa(VBE_Performance_management, Trust_Assessment)
temTarefa(VBE_Performance_management, Performance_Trust_Monitoring)
temTarefa(VBE_Performance_management, Activate_Alarming_Mechanism)
temTarefa(VBE_Performance_management, Analysis_and_Decision_Making)
temTarefa(VO_Creation_in_IECOS, Collaboration_Oportunity_Detection)
temTarefa(VO_Creation_in_IECOS, CO_Characterization_Process)
temTarefa(VO_Creation_in_IECOS, Material_Selection_Process)
temTarefa(VO_Creation_in_IECOS, VO_Rough_Plan_Process)

```



```

temTarefa(VO_Creation_in_IECOS , Partners_Search_and_Selection)
temTarefa(VO_Creation_in_IECOS , Negotiation_Process)
temTarefa(VO_Creation_in_IECOS , Contracting:_Send_Purchase_Order)
temTarefa(VO_Creation_in_IECOS , Receive_Purchase_Order)
temTarefa(VO_Creation_in_IECOS , VO_Registration)
temTarefa(VBE_Characterization , Requirement_identification)
temTarefa(VBE_Characterization , Characterization_of_the_actors)
temTarefa(VBE_Characterization ,
    Definition_of_the_organizational_framework)
temTarefa(VBE_Characterization , Member_profiling)
temTarefa(VBE_Characterization , Definition_of_the_business_strategy)
temTarefa(VBE_Characterization , Proactiveness_measurement_of_members)
temTarefa(VBE_Characterization , Information_forwarding)
temTarefa(VBE_Characterization , Own_initiative_for_Collaboration)
temTarefa(VO_Creation_in_Helice , Identification_of_business_opportunities
)
temTarefa(VO_Creation_in_Helice , VO_partners_selection)
temTarefa(VO_Creation_in_Helice , VO_rough_planning)
temTarefa(VO_Launch , Identification_of_contract_objects)
temTarefa(VO_Launch , Collaborative_agreement)
temTarefa(VO_Launch , Contract_negotiation)
temTarefa(VO_Launch , VO_launch)
temTarefa(VO_Launch , Performance_tests)
temTarefa(VO_Management , VO_Modelling)
temTarefa(VO_Management , VO_Monitoring)
temTarefa(VO_Management , VO_Analysis)

executaTarefa(MembroDoIECOS , New_Applicant)
executaTarefa(AdministradorDoIECOS , Registration_Process)
executaTarefa(AdministradorDoIECOS , Profiling_and_Compentency_Management)
executaTarefa(MembroDoIECOS , Update_Information)
executaTarefa(AdministradorDoIECOS ,
    Assigment_of_Roles_Rights_and_Responsabilities)
executaTarefa(AdministradorDoIECOS , New_Compencies)
executaTarefa(AdministradorDoIECOS , Indicators_Definition)
executaTarefa(AdministradorDoIECOS , Collect_and_Store_Information)
executaTarefa(AdministradorDoIECOS , Performance_Assessment)
executaTarefa(MembroDoIECOS , Collect_and_Store_Trust_Information)
executaTarefa(MembroDoIECOS , Trust_Assessment)
executaTarefa(AdministradorDoIECOS , Performance_Trust_Monitoring)
executaTarefa(AdministradorDoIECOS , Activate_Alarming_Mechanism)
executaTarefa(AdministradorDoIECOS , Analysis_and_Decision_Making)
executaTarefa(MembroDoIECOS , Collaboration_Oportunity_Detection)
executaTarefa(AdministradorDoIECOS , CO_Characterization_Process)
executaTarefa(AdministradorDoIECOS , Material_Selection_Process)
executaTarefa(AdministradorDoIECOS , VO_Rough_Plan_Process)
executaTarefa(AdministradorDoIECOS , Partners_Search_and_Selection)
executaTarefa(AdministradorDoIECOS , Negotiation_Process)
executaTarefa(MembroDoIECOS , Negotiation_Process)
executaTarefa(AdministradorDoIECOS , Contracting:_Send_Purchase_Order)
executaTarefa(MembroDoIECOS , Receive_Purchase_Order)
executaTarefa(AdministradorDoIECOS , VO_Registration)

usaRecurso(Registration_Process , MSMS)
usaRecurso(Profiling_and_Compentency_Management , PCMS)
usaRecurso(Assigment_of_Roles_Rights_and_Responsabilities , MSMS)
usaRecurso(New_Compencies , ODMS)
usaRecurso(Indicators_Definition , DSS)
usaRecurso(Collect_and_Store_Information , DSS)
usaRecurso(Performance_Assessment , DSS)
usaRecurso(Collect_and_Store_Information , Trustman)
usaRecurso(Trust_Assessment , Trustman)
usaRecurso(Performance_Trust_Monitoring , DSS)

```

```

usaRecurso(Activate_Alarming_Mechanism , DSS)
usaRecurso(Analysis_and_Decision_Making , DSS)
usaRecurso(CO_Characterization_Process , COC-Plan)
usaRecurso(VO_Rough_Plan_Process , COC-Plan)
usaRecurso(Partners_Search_and_Selection , PSS)
usaRecurso(VO_Registration , VIMS)
usaRecurso(VO_Modelling , VO-Mod)
usaRecurso(VO_Modelling , SID)
usaRecurso(VO_Monitoring , VO-Mod)
usaRecurso(VO_Monitoring , DI3)
usaRecurso(VO_Monitoring , MAF)
usaRecurso(VO_Analysis , VO-Mod)
usaRecurso(VO_Analysis , DSS)

temOC(IECOS , Centrifuge_Machine)
temOC(IECOS , Punch_for_biopsies)

temCompetencia(IECOS , Manufacturing)
temCompetencia(Helice , Aerospace_Engineering)
temCompetencia(CeBeNetwork , Aerospace_Engineering)

mede(Percentage_of_Qualified_Suppliers , Acquisition)
mede(Cost_per_Invoice , Packaging)
mede(Cost_per_Invoice , Invoice)
mede(Demand/Supply_Planning_Costs , Planning)
mede(Incoming_Production_Material_Quality , Aquisition)
mede(Incoming_Production_Material_Quality , Material)
mede(Order_Management_Costs , Management)
mede(Order_Management_Costs , Orders)
mede(Order_Management_Cycle_Time , Management)
mede(Order_Management_Cycle_Time , Orders)
mede(Package_Cycle_Time , Packaging)
mede(Package_Cycle_Time , Product)
mede(Packaging_Cost , Packaging)
mede(Packaging_Cost , Product)
mede(Cost_of_storage_space , Management)
mede(Cost_of_storage_space , Inventory)
mede(Production_Plan_Adherence , Production)
mede(Production_Process_Validation_Frequency , Production)
mede(Percentage_of_Faultless_production , Production)
mede(Percentage_of_Faultless_production , Material)

ehMedidoCom(Percentage_of_Qualified_Suppliers , Planning)
ehMedidoCom(Cost_per_Invoice , Delivery)
ehMedidoCom(Demand/Supply_Planning_Costs , Management)
ehMedidoCom(Incoming_Production_Material_Quality , Production)
ehMedidoCom(Order_Management_Costs , Management)
ehMedidoCom(Order_Management_Cycle_Time , Management)
ehMedidoCom(Package_Cycle_Time , Delivery)
ehMedidoCom(Packaging_Cost , Delivery)
ehMedidoCom(Cost_of_storage_space , Inventory)
ehMedidoCom(Production_Plan_Adherence , Management)
ehMedidoCom(Production_Process_Validation_Frequency , Management)
ehMedidoCom(Percentage_of_Faultless_production , Management)
ehMedidoCom(Production_Material_Cycle_Time , Production)
ehMedidoCom(Deviation_from_planned_start_date , Management)
ehMedidoCom(Number_of_re-planning , Management)
ehMedidoCom(Problem_Compensation_performance_of_a_VO_member , Management)
ehMedidoCom(Reliability_of_VO_member's_statements/_predictions , Planning
)
ehMedidoCom(Percentage_of_available_of_resources , Planning)
ehMedidoCom(Satisfaction_with_collaboration_performance , Planning)
ehMedidoCom(Number_of_call_backs_as_Percentage_of_total_inquiries ,

```

Management)

```

considera(Percentage_of_Qualified_Suppliers , Reliability)
considera(Cost_per_Invoice , Cost)
considera(Demand/Supply_Planning_Costs , Cost)
considera(Incoming_Production_Material_Quality , Quality)
considera(Order_Management_Costs , Cost)
considera(Order_Management_Cycle_Time , Time)
considera(Package_Cycle_Time , Time)
considera(Packaging_Cost , Cost)
considera(Cost_of_storage_space , Cost)
considera(Production_Plan_Adherence , Flexibility)
considera(Production_Process_Validation_Frequency , Reliability)
considera(Percentage_of_Faultless_production , Quality)
considera(Production_Material_Cycle_Time , Time)
considera(Deviation_from_planned_start_date , Reliability)
considera(Number_of_re-planning , Reliability)
considera(Problem_Compensation_performance_of_a_VO_member , Flexibility)
considera(Reliability_of_VO_member's_statements/_/_predictions ,
Reliability)
considera(Percentage_of_available_of_resources , Productivity)
considera(Satisfaction_with_collaboration_performance , Quality)
considera(Number_of_call_backs_as_Percentage_of_total_inquiries ,
Collaboration)

estaRelacionadoCom(Percentage_of_Qualified_Suppliers , Intra)
estaRelacionadoCom(Cost_per_Invoice , Intra)
estaRelacionadoCom(Demand/Supply_Planning_Costs , Intra)
estaRelacionadoCom(Incoming_Production_Material_Quality , Intra)
estaRelacionadoCom(Order_Management_Costs , Intra)
estaRelacionadoCom(Order_Management_Cycle_Time , Intra)
estaRelacionadoCom(Package_Cycle_Time , Intra)
estaRelacionadoCom(Packaging_Cost , Intra)
estaRelacionadoCom(Cost_of_storage_space , Intra)
estaRelacionadoCom(Production_Plan_Adherence , Intra)
estaRelacionadoCom(Production_Process_Validation_Frequency , Intra)
estaRelacionadoCom(Percentage_of_Faultless_production , Intra)
estaRelacionadoCom(Production_Material_Cycle_Time , Intra)
estaRelacionadoCom(Deviation_from_planned_start_date , IntraInter)
estaRelacionadoCom(Number_of_re-planning , Inter)
estaRelacionadoCom(Problem_Compensation_performance_of_a_VO_member , Inter
)
estaRelacionadoCom(Reliability_of_VO_member's_statements/_/_predictions ,
Inter)
estaRelacionadoCom(Percentage_of_available_of_resources , Inter)
estaRelacionadoCom(Satisfaction_with_collaboration_performance , Inter)
estaRelacionadoCom(Number_of_call_backs_as_Percentage_of_total_inquiries ,
Inter)

```


Apêndice D

Modelo e Regras na sintaxe do Jess

A implementação do modelo de contexto e regras para o protótipo computacional consistiu num arquivo que contém os seguintes elementos:

- Modelo de contexto;
- Ontologia e base de conhecimento;
- Regras e;
- Instâncias de contextos.

Tais elementos são apresentados nas figuras a seguir. Os *templates* para a definição do modelo de contexto e das regras são apresentados a seguir.

```
::: Declaring the templates for the Context Model
-----
(deftemplate FormulatesQuery      (declare (from-class FormulatesQuery)))
(deftemplate IsAbout              (declare (from-class IsAbout)))
(deftemplate HasConstraint        (declare (from-class
      HasConstraint)))
(deftemplate IsIn                 (declare (from-class IsIn
      )))
(deftemplate HasContextElement    (declare (from-class HasContextElement)))
(deftemplate IsAssociatedTo       (declare (from-class
      IsAssociatedTo)))

::: Declaring the templates for the rules
-----
(deftemplate IsFormulatedIn              (
      declare (from-class IsFormulatedIn)))
(deftemplate IsAboutPointedConcept      (declare
      (from-class IsAboutPointedConcept)))
(deftemplate IsAboutPointedConceptFromContext (declare (from-class
      IsAboutPointedConceptFromContext)))
(deftemplate IsInSubClassHierarchy      (declare
      (from-class IsInSubClassHierarchy)))
(deftemplate IsInSuperClassHierarchy    (declare (from-
      class IsInSuperClassHierarchy)))

::: Declaring the triple template -----
(deftemplate triple "Template representing a triple"
```

```
(slot predicate (default ""))
(slot subject (default ""))
(slot object (default ""))
)
```

As regras são mostradas abaixo.

```
::: Declaring the rules for query expansion

(defrule R1
  "associates queries to contexts"

  (FormulatesQuery (user ?u) (query ?q))
  (IsIn (user ?u) (context ?cx))

  =>

  (assert (IsFormulatedIn (query ?q) (context ?cx)))
  (printout t ?q " is formulated in Context " ?cx "(R1)" crlf)
)

(defrule R2
  "deduces new concepts pointed by the ones in the query"

  ; ?q is about ?c1
  (IsAbout (query ?q) (_class ?c1))

  ; ?c1 has a property ?s...
  (triple (object ?c1) (predicate "http://www.w3.org/2000/01/rdf-schema
#domain") (subject ?s))
  ; ?s "" points" to ?c2...
  (triple (subject ?s) (predicate "http://www.w3.org/2000/01/rdf-schema
#range") (object ?c2))
  ; ?c2 is a _class
  (triple (subject ?c2) (predicate "http://www.w3.org/1999/02/22-rdf-
syntax-ns#type") (object "http://www.w3.org/2002/07/owl#Class"))

  ; ?q is not about ?c2
  (not (IsAbout (query ?q) (_class ?c2)))

  ; ?q does not have a constraint containing ?c2
  (and (not (HasConstraint (query ?q) (constraint ?c1 ?s ?c2))) (not(
  IsAboutPointedConcept (query ?q) (constraint ?c1 ?s ?c2))))

  =>

  (assert (IsAboutPointedConcept (query ?q) (constraint ?c1 ?s ?c2)))

  ;(printout t "New constraint to "?q " (R2): " ?c1 " - " ?s " - " ?c2
  crlf)
)
/*
(defrule R3
  "deduces new concepts pointed by the ones in the query, also
  considering the context"

  (IsFormulatedIn (query ?q) (context ?cx))

  (HasContextElement (context ?cx) (contextElement ?e))

  (IsAssociatedTo (contextElement ?e) (_class ?c))
)
```

```

(or (IsInSuperClassHierarchy (class1 ?c) (class2 ?c2)) (triple (
  subject ?c) (predicate "http://www.w3.org/1999/02/22-rdf-syntax-
  ns#type") (object ?c2)))

(IsAboutPointedConcept (query ?q) (constraint ?c1 ?s ?c2))

(not (IsAboutPointedConceptFromContext (query ?q) (constraint ?c1 ?s
  ?c)))

=>

(assert (IsAboutPointedConceptFromContext (query ?q) (constraint ?c1
  ?s ?c)))

(printout t "New constraint to "?q " (context): " ?c1 " - " ?s " - "
  ?c crlf)
)*/

(defrule R3
  "deduces new concepts pointed by the ones in the query, also
  considering the context"

  (IsFormulatedIn (query ?q) (context ?cx))

  (HasContextElement (context ?cx) (contextElement ?e))

  (IsAssociatedTo (contextElement ?e) (_class ?c))

  (or (IsInSuperClassHierarchy (class1 ?c) (class2 ?c2)) (triple (
    subject ?c) (predicate "http://www.w3.org/1999/02/22-rdf-syntax-
    ns#type") (object ?c2)))

  (IsAboutPointedConcept (query ?q) (constraint ?c1 ?s ?c2))

  (not (IsAboutPointedConceptFromContext (query ?q) (subject ?c1) (
    predicate ?s) (object ?c)))

  =>

  (assert (IsAboutPointedConceptFromContext (query ?q) (subject ?c1) (
    predicate ?s) (object ?c)))

  (printout t "New constraint to "?q " (context): " ?c1 " - " ?s " - "
    ?c crlf)
  (add (new IsAboutPointedConceptFromContext ?q ?c1 ?s ?c))
)

(defrule R6
  "says that a class mentioned in the query is subclass and
  superclass of itself"

  (IsAbout (query ?q) (_class ?c))

  =>

  (assert (IsInSubClassHierarchy (class1 ?c) (class2 ?c)))
  (assert (IsInSuperClassHierarchy (class1 ?c) (class2 ?c)))
  ;(printout t "SubClass Hierarchy of " ?c ": " ?c crlf)
  ;(printout t "SuperClass Hierarchy of " ?c ": " ?c crlf)
)

(defrule R7

```

```

    "adds a class to a class hierarchy"

    (IsInSuperClassHierarchy (class1 ?c1) (class2 ?c1))

    (triple (subject ?c1) (predicate "http://www.w3.org/2000/01/rdf-
      schema#subClassOf") (object ?c2))

    (not(IsInSuperClassHierarchy (class1 ?c1) (class2 ?c2)))

    =>

    (assert (IsInSuperClassHierarchy (class1 ?c1) (class2 ?c2)))
    ;(printout t "SuperClass Hierarchy of " ?c1 ": " ?c2 crlf)
  )

(defrule R8
  "adds a class to a class hierarchy"

  (IsInSuperClassHierarchy (class1 ?c1) (class2 ?c2))

  (triple (subject ?c2) (predicate "http://www.w3.org/2000/01/rdf-
    schema#subClassOf") (object ?c3))

  (not(IsInSuperClassHierarchy (class1 ?c1) (class2 ?c3)))

  =>

  (assert (IsInSuperClassHierarchy (class1 ?c1) (class2 ?c3)))
  ;(printout t "SuperClass Hierarchy of " ?c1 ": " ?c3 crlf)
)

(defrule R9
  "adds a class to a class hierarchy"

  (IsInSubClassHierarchy (class1 ?c1) (class2 ?c1))

  (triple (subject ?c2) (predicate "http://www.w3.org/2000/01/rdf-
    schema#subClassOf") (object ?c1))

  (not (IsInSubClassHierarchy (class1 ?c1) (class2 ?c2)))

  =>

  (assert (IsInSubClassHierarchy (class1 ?c1) (class2 ?c2)))
  ;(printout t "SubClass Hierarchy of " ?c1 ": " ?c2 crlf)
)

(defrule R10
  "adds a class to a class hierarchy"

  (IsInSubClassHierarchy (class1 ?c1) (class2 ?c2))

  (triple (subject ?c3) (predicate "http://www.w3.org/2000/01/rdf-
    schema#subClassOf") (object ?c2))

  (not (IsInSubClassHierarchy (class1 ?c1) (class2 ?c3)))

  =>

  (assert (IsInSubClassHierarchy (class1 ?c1) (class2 ?c3)))
  ;(printout t "SubClass Hierarchy of " ?c1 ": " ?c3 crlf)
)

```


As instâncias de contextos utilizadas são apresentadas na listagem a seguir.

```

::: Instances of the Context Model.
(assert
  ; Context instances

  (HasContextElement (context "VoMember1") (contextElement "VoOIN"))
  (HasContextElement (context "VoMember1") (contextElement "member"))
  (HasContextElement (context "VoMember1") (contextElement "management"))
  (HasContextElement (context "VoMember1") (contextElement "costanalysis
  ")

  (HasContextElement (context "VoMember2") (contextElement "VoOIN"))
  (HasContextElement (context "VoMember2") (contextElement "member"))
  (HasContextElement (context "VoMember2") (contextElement "management"))
  (HasContextElement (context "VoMember2") (contextElement "
  qualityevaluation "))

  (HasContextElement (context "VoManager") (contextElement "VoOIN"))
  (HasContextElement (context "VoManager") (contextElement "manager"))
  (HasContextElement (context "VoManager") (contextElement "VOManagement
  "))
  (HasContextElement (context "VoManager") (contextElement "rescheduling
  "))

  (HasContextElement (context "VbeBroker") (contextElement "VbeHelice"))
  (HasContextElement (context "VbeBroker") (contextElement "Broker"))
  (HasContextElement (context "VbeBroker") (contextElement "VOcreation"))
  (HasContextElement (context "VbeBroker") (contextElement "planning"))

  (HasContextElement (context "VbeAdmin1") (contextElement "VbeHelice"))
  (HasContextElement (context "VbeAdmin1") (contextElement "Admin"))
  (HasContextElement (context "VbeAdmin1") (contextElement "VOcreation"))
  (HasContextElement (context "VbeAdmin1") (contextElement "
  partnerssearch "))

  (HasContextElement (context "VbeAdmin2") (contextElement "VbeIECOS"))
  (HasContextElement (context "VbeAdmin2") (contextElement "Admin"))
  (HasContextElement (context "VbeAdmin2") (contextElement "ManagTemem"))

  (HasContextElement (context "VbeAdmin3") (contextElement "
  VbeCeBeNetwork "))
  (HasContextElement (context "VbeAdmin3") (contextElement "Admin"))
  (HasContextElement (context "VbeAdmin3") (contextElement "ManagTemem"))

  ; Context instance linking to ontology
  (IsAssociatedTo (contextElement "Member") (_class "http://www.gsigma.
  ufsc.br/pi-co#IntraOrganizationalContext"))
  (IsAssociatedTo (contextElement "Delivery") (_class "http://www.gsigma.
  ufsc.br/pi-co#Something_T.2_Delivery"))
  (IsAssociatedTo (contextElement "Packaging") (_class "http://www.gsigma
  .ufsc.br/pi-co#MeasurementObjective_T.2_Delivery"))

  (IsAssociatedTo (contextElement Manager) (_class "http://www.gsigma.
  ufsc.br/pi-co#InterOrganizationalContext"))
  (IsAssociatedTo (contextElement Broker) (_class "http://www.gsigma.ufsc
  .br/pi-co#InterOrganizationalContext"))
  (IsAssociatedTo (contextElement VoCreation) (_class "http://www.gsigma.
  ufsc.br/pi-co#Something_T.7_AnyOrganization"))

```

```
(IsAssociatedTo (contextElement QualityEvaluation) (_class "http://www.
  gsigma.ufsc.br/pi-co#MeasurementObjective_T.3_Management"))
(IsAssociatedTo (contextElement QualityEvaluation) (_class "http://www.
  gsigma.ufsc.br/pi-co#PerformancePerspective_T.3_Quality"))
(IsAssociatedTo (contextElement Rescheduling) (_class "http://www.
  gsigma.ufsc.br/pi-co#MeasurementObjective_T.3_Management"))
(IsAssociatedTo (contextElement Rescheduling) (_class "http://www.
  gsigma.ufsc.br/pi-co#PerformancePerspective_T.1_Reliability"))
(IsAssociatedTo (contextElement Planning) (_class "http://www.gsigma.
  ufsc.br/pi-co#MeasurementObjective_T.1_Planning"))
)
;;; End of Context Model Instances.
```

Apêndice E

Coleções de Teste

E.1 Indicadores de Desempenho

colecoes/IDs/id_01.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1" />
    <feature name="Title" value="Percentage Of Qualified Suppliers which
      Meet Defined Requirements" />
    <feature name="Perspective" value="Reliability" />
  </features>
  <content>The number of qualified suppliers who meet defined
    requirements divided by the total number of qualified suppliers
    used as sources in the measurement period.</content>
</ksDocument>
```

colecoes/IDs/id_02.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="2" />
    <feature name="Title" value="Cost per Invoice" />
    <feature name="Perspective" value="Cost" />
  </features>
  <content>All costs associated with the receipt, review, processing, and
    payment of a supplier's invoice for product received.</content>
</ksDocument>
```

colecoes/IDs/id_03.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="3" />
    <feature name="Title" value="Demand/Supply Planning Costs" />
    <feature name="Perspective" value="Cost" />
  </features>
</ksDocument>
```

```

</features >
<content>Costs associated with forecasting , developing finished goods
or end item inventory plans , and coordinating Demand/Supply
process across entire supply chain , including all channels . (Not
including MIS associated costs).</content>
</ksDocument>

```

colecoes/IDs/id_04.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
<context />
<features >
  <feature name="DocumentId" value="4" />
  <feature name="Title" value="Incoming Production Material Quality" />
  <feature name="Perspective" value="Quality" />
</features >
<content>This indicator measures the quality of raw material to ensure
the correctness of final product. It is calculated by number of
received parts which fail inspection divided by the total Number
of parts received for production sector.</content>
</ksDocument>

```

colecoes/IDs/id_05.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
<context />
<features >
  <feature name="DocumentId" value="5" />
  <feature name="Title" value="Order Management Costs" />
  <feature name="Perspective" value="Cost" />
</features >
<content>The aggregation of the following cost elements (contained in
this glossary): Create Customer Order Costs , Order Entry and
Maintenance Costs , Contract/Program and Channel Management Costs ,
Installation Planning Costs , Order Fulfillment Costs , Distribution
Costs , Transportation Costs , Installation Costs , Customer
Invoicing/Accounting Costs.</content >
</ksDocument>

```

colecoes/IDs/id_06.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
<context />
<features >
  <feature name="DocumentId" value="6" />
  <feature name="Title" value="Order Management Cycle Time" />
  <feature name="Perspective" value="Time" />
</features >
<content>The total amount of time required converting a customer order
into a receipt by the customer.</content>
</ksDocument>

```

colecoes/IDs/id_07.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>

```

```

<context />
<features>
  <feature name="DocumentId" value="7" />
  <feature name="Title" value="Package Cycle Time" />
  <feature name="Perspective" value="Time" />
</features>
<content>The total time required to perform a series of activities that
  containerize completed products for storage or sale to end-users.
  (Within certain industries, packaging may include cleaning or
  sterilization).</content>
</ksDocument>

```

colecoes/IDs/id_08.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="8" />
    <feature name="Title" value="Packaging Cost" />
    <feature name="Perspective" value="Cost" />
  </features>
  <content>The cost to package product as a finished good, not including
    intermediate handling of materials, based on given number of
    Delivered Finished Goods.</content>
</ksDocument>

```

colecoes/IDs/id_09.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="9" />
    <feature name="Title" value="Cost of storage space" />
    <feature name="Perspective" value="Cost" />
  </features>
  <content>Cost of storage space used for the production materials.</
  content>
</ksDocument>

```

colecoes/IDs/id_10.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="10" />
    <feature name="Title" value="Production Plan Adherence" />
    <feature name="Perspective" value="Reliability" />
  </features>
  <content>Production Plan Adherence is calculated at the shippable end-
    product level in units, using the following formula: Production
    Plan - Sum of Variance, Production Plan Where: Production Plan =
    The sum of the units planned to be completed (i.e., placed into
    inventory or shipped) in each month based upon the plan generated
    in the previous month. Sum of Variances = The sum of the absolute
    values, at the end item level, of the differences between each
    month's production plan as defined above and actual production for
    the same month.</content>
</ksDocument>

```

colecoes/IDs/id_11.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features >
    <feature name="DocumentId" value="11" />
    <feature name="Title" value="Production Process Validation Frequency" />
    <feature name="Perspective" value="Reliability" />
  </features >
  <content>The amount of time between reviews of a process. For example,
    Production Process Validation Frequency would refer to the amount
    of time between the reviews of the Production Process. This
    generally would be performed periodically to ensure that the
    process is generating the desired results with the desired inputs
    .</content>
</ksDocument>

```

colecoes/IDs/id_12.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features >
    <feature name="DocumentId" value="12" />
    <feature name="Title" value="Percentage of Faultless production" />
    <feature name="Perspective" value="Quality" />
  </features >
  <content>Number of products produced without error in the manufacturing
    sector. An example of potential defects that can decrease the
    quality is lack of precision in the manufacture process.</content>
</ksDocument>

```

colecoes/IDs/id_13.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features >
    <feature name="DocumentId" value="13" />
    <feature name="Title" value="Production Material Cycle Time" />
    <feature name="Perspective" value="Time" />
  </features >
  <content>Time required moving material to point of use.</content>
</ksDocument>

```

colecoes/IDs/id_14.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features >
    <feature name="DocumentId" value="14" />
    <feature name="Title" value="Deviation from planned start date" />
    <feature name="Perspective" value="Reliability" />
  </features >
  <content>This measures how much the start of a tasks deviates from the
    planned start date.</content>
</ksDocument>

```

colecoes/IDs/id_15.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="15" />
    <feature name="Title" value="Number of re-plannings" />
    <feature name="Perspective" value="Reliability" />
  </features>
  <content>Indicates the reliability of plannings, based upon the
    informations/commitments of the VO Members. If a certain period
    should be analysed, this can be set through the validity dates of
    the indicator.</content>
</ksDocument>

```

colecoes/IDs/id_16.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="16" />
    <feature name="Title" value="Problem Compensation performance of a VO
      member" />
    <feature name="Perspective" value="Flexibility" />
  </features>
  <content>Describes how much (%) of the deviation caused by previous
    partners is compensated. A negative value indicates that the
    deviation has become bigger. Metrics could be indicators for
    deviation already measured.</content>
</ksDocument>

```

colecoes/IDs/id_17.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="17" />
    <feature name="Title" value="Reliability of VO member's statements /
      predictions" />
    <feature name="Perspective" value="Reliability" />
  </features>
  <content>Indicates the reliability of statements/predictions of a VO
    member, e.g. regarding the predicted on-time delivery or
    predictions if budgets are kept.</content>
</ksDocument>

```

colecoes/IDs/id_18.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="18" />
    <feature name="Title" value="Percentage of available resources in
      relation to overall assigned resources" />
    <feature name="Perspective" value="Flexibility" />
  </features>

```

```
<content>Represents the part of the overall resources assigned to a VO
that is actually available for the VO.</content>
</ksDocument>
```

colecões/IDs/id_19.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features >
    <feature name="DocumentId" value="19" />
    <feature name="Title" value="Satisfaction with collaboration
performance of a VO member" />
    <feature name="Perspective" value="Quality" />
  </features >
  <content>Indicates the satisfaction of an asked person/organisation
with the collaboration performance of a certain VO member. Usually
a scale is provided for the answer (e.g. from 1 = very satisfied
to 6 = completely dissatisfied).</content>
</ksDocument>
```

colecões/IDs/id_20.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features >
    <feature name="DocumentId" value="20" />
    <feature name="Title" value="Number of call backs as Percentage of
total inquiries" />
    <feature name="Perspective" value="Collaboration" />
  </features >
  <content>Number of callbacks divided by total inquiries.</content>
</ksDocument>
```


E.2 Documentos

colecoes/docs/doc_1001.xml

```

i»¿ <?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1001" />
    <feature name="Title" value="IECOS (VBE Network)" />
    <feature name="ExtractedFrom" value="D814c - section 2.1" />
  </features>
  <content>IECOS SUMMARY: GLOBAL USE AND RESULTS

```

The taken up of ECOLEAD results in IECOS carried the possibility of achieving important changes not only in its operational process but also in its business model and strategy. Through the utilization and implementation of ECOLEAD framework, IECOS formalized and implemented a new business strategy based on VBE model, where it also takes advantage of new business opportunities and market niches by establishing a new complete business unit reproducing the IECOS-VBE model. Significant improvements in the day-to-day operations at IECOS were possible to be achieved through the implementation of ECOLEAD tools and methodologies.

```

</content>
</ksDocument>

```

colecoes/docs/doc_1002.xml

```

i»¿ <?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1002" />
    <feature name="Title" value="PLANNED PROCESSES AND OBJECTIVES" />
    <feature name="ExtractedFrom" value="D814c - section 2.2.1" />
  </features>
  <content>2.2.1 PLANNED PROCESSES AND OBJECTIVES

```

Three business processes have been identified in IECOS for carrying out ECOLEAD demonstration activities through the implementation of available ECOLEAD tools and methodologies that can support or improve the IECOS business processes selected. The business processes selected for ECOLEAD demonstration activities are:

1. Registration of new members (including profiling and competency management): This demonstration activity is aimed for the creation of a formal registration process for IECOS members and the management of their profiles once they have been registered. The purpose of this business process is to optimize the process for registering new members, considering its evaluation and selection as part of the registration process, and to improve the current management of member's profiles in IECOS.
2. VBE Performance management: This business process is not formally described in IECOS nowadays, but the objective of this demonstration activity is to define and formalize the ideal performance management approach for IECOS using ECOLEAD results (tools and methodologies).
3. VO Creation in IECOS: This demonstration activity focuses on achieving a semi-automatic VO configuration process in IECOS, following

```

    ECOLEAD VO creation framework: collaboration opportunity search ,
    collaboration opportunity characterization , and rough planning and
    partners search.
</content>
</ksDocument>

```

colecoes/docs/doc_1003.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1003" />
    <feature name="Title" value="ECOLEAD MATERIAL PLANNED TO BE USED" />
    <feature name="ExtractedFrom" value="D814c – section 2.2.2" />
  </features>
</content>2.2.2 ECOLEAD MATERIAL PLANNED TO BE USED

```

Following is presented the list of ECOLEAD tools to be used in demonstration activities for the IECOS VBE:

- Membership and Structure Management Systems (MSMS)
- Profiling and Competency Management System (PCMS)
- Ontology and Discovery Management System (ODMS)
- Trust Management System (TrustMan)
- Virtual Organization Information Management System (VIMS)
- Collaboration Opportunity Characterization and Rough Planning (COC–Plan)
- Partners Search and Suggestion (PSS)
- Decision Support Management System (DSS)

```

</content>
</ksDocument>

```

colecoes/docs/doc_1004.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1004" />
    <feature name="Title" value="BUSINESS PROCESS #1" />
    <feature name="ExtractedFrom" value="D814c – section 2.4.3" />
  </features>
</content>BUSINESS PROCESS #1

```

Scenario name: Registration of New Members – Profiles and Competencies Management

Scenario description: This business process was expected to create a registration process supported by standardized mechanisms that would ensure consistency in information about associated network partners . Such mechanisms used the framework developed within ECOLEAD, and adopted of some VBE management functionalities regarding VBE members registration in order to optimize and organize the process for registering a new member (evaluation and selection), as well as profiles and competency management.

Demonstrated Key Results:

KR2.1 – VBE reference framework

- Methodology and mechanism for guiding steps to register a new VBE member:

- Collection and validation of provided information.
- Base trust assessment.

- Methodology and mechanisms for management of membership levels.
- Methodology and mechanism for VBE members updating their profiles.
- Methodology for guiding VBE participants, governance principles, roles, rights and responsibilities.

KR2.2 – VBE management services

- Membership and Structure Management Systems (MSMS)
 - Registration of new VBE members
 - Member's roles, rights and responsibility management
- Profiling and Competency Management System (PCMS)
 - VBE itself profiling and competency management
 - VBE members profiling and competency management
 - VOs profiling and competency management
- Ontology and Discovery Management System (ODMS)

</content>
</ksDocument>

colecões/docs/doc_1005.xml

```

i>¿ <?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1005" />
    <feature name="Title" value="BUSINESS PROCESS #2" />
    <feature name="ExtractedFrom" value="D814c – section 2.4.4" />
  </features>
  <content>BUSINESS PROCESS #2

```

Scenario name: VBE Performance Management

Scenario description: The activities held in this business process were expected to allow IECOS to analyze and evaluate information about each member's performance, considering the assignments that they have in a VO through the utilization of a system(s) that monitor VBE itself, VBE member and VOs performance and notify warning situations.

Besides this, the historical performance record was a base for decision making during the partners selection process, and the main purpose was therefore to control and manage the performance information within the network. By collecting and recording this information, VBE members became solid for IECOS and this represents a competitive advantage that would save time when looking for partners to form a VO.

Demonstrated Key Results:

KR2.2 – VBE Management Services

- Trust Management System (TrustMan)

KR2.3 – Value System and Trust in VBE

```

- Decision Support Management System (DSS)
- VBE Competency Gap Analysis
- Lack of Performance Warning
- Low Trust Level Warning
</content>
</ksDocument>

```

colecoes/docs/doc_1006.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1006" />
    <feature name="Title" value="BUSINESS PROCESS #3" />
    <feature name="ExtractedFrom" value="D814c - section 2.4.5" />
  </features>
</ksDocument>

```

Scenario name: VO Creation in IECOS – supply services

Scenario description: As a result of this business process, it was expected a semi-automated process to model a collaboration opportunity, characterize the required competencies and make the VO rough plan. The collection of VO requirements as a set of necessary competencies to fulfil the product/service specifications and the partners' selection would be obtained through a semiautomated system, saving time and making the operation for IECOS more efficient.

Demonstrated Key Results:

KR2.2 – VBE Management Services

- VO Information Management Systems (VIMS)
- VOs registration within the VBE
- Management of VO inheritance information

KR2.4 – Dynamic VO Creation Assistance Service

- Collaboration Opportunity Characterization and Rough Planning (COC-Plan)

- Partners Search and Suggestion (PSS)

```

</content>
</ksDocument>

```

colecoes/docs/doc_1007.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1007" />
    <feature name="Title" value="OIN (VOM Network)" />
    <feature name="ExtractedFrom" value="D814c - section 6.1" />
  </features>
</ksDocument>

```

6.1 EXECUTIVE SUMMARY: GLOBAL USE AND RESULTS

Dedicated methodologies and tools for the monitoring and control of Virtual Organizations have been developed under the ECOLEAD research project. This part of the document reports on the application of these tools to real life Virtual Organizations in Orona Innovation Network.

The following sections detail the working business processes and scenarios in which tools have been used, the contribution which each tool made to the successful operation of the Virtual Organizations, and the feedback from the participants based on their experiences of working with the software.

Moreover, we must point out that this is the final update of the document and corresponds to the use of all methodologies and tools planned by the final users.

Orona Innovation Network has concentrated its efforts in developing two business scenarios: OIN Roadmapping and VO Management. The first one tried to establish the needed processes based on ECOLEAD concepts. The second one has been dedicated to implement the different tools for VO management; specifically, the tools concerned are VOMod, SID, DI3, MAF and DSS.

In general objectives reached have been very promising for the future management of the different VOs launched from Orona Innovation Network and devoted to Technology Platform and New Product Development.

```
</content>
</ksDocument>
```

colecoes/docs/doc_1008.xml

```
ï»¿<?xml version="1.0" encoding="UTF-8"?>
```

```
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1008" />
    <feature name="Title" value="Business process: OIN Roadmapping" />
    <feature name="ExtractedFrom" value="D814c – section 6.2" />
  </features>
  <content>Business process: OIN Roadmapping
```

The first level of application corresponds to a theoretical foundation whose aim is to deploy the conceptual development of ECOLEAD in the Orona Innovation Network.

In this sense, the scenario known as 'OIN Roadmap' attempts to establish a reference model for the Orona Innovation Network where the role of all the elements identified as pillars in ECOLEAD model was defined. Furthermore, another important objective of the scenario is to set up the mechanisms needed to make the right integration possible between different Virtual Organizations and Professional Virtual Communities.

The development of this scenario and its implementation were identified as a priority by the different managers of Orona Innovation Network. On the one hand, to rely on the deployment of ECOLEAD pillars to the network allows the Technical Committee – a committee created to manage the network with representatives of all partners – to specifically identify the activities to be taken into account in launching Virtual Organizations. On the other hand, establishing a steady model allows Orona Innovation Network not to be restricted to

the implementation and use of the ECOLEAD results while the project is underway and to be able to make use in the future of other methodologies and tools developed in other areas of the project.

To carry out the implementation and take-up of this scenario the need to run a real deployment process covering the definition of a roadmap of the network activities for the period 2006–2010 was established.

</content>
</ksDocument>

colecões/docs/doc_1009.xml

```
<?xml version="1.0" encoding="UTF-8"?>
```

```
<ksDocument>
```

```
<context />
```

```
<features>
```

```
<feature name="DocumentId" value="1009" />
```

```
<feature name="Title" value="Business process: VO Management" />
```

```
<feature name="ExtractedFrom" value="D814c – section 6.2" />
```

```
</features>
```

```
<content>Business process: VO Management
```

The second level of application concerns the use of the results attained for the Virtual Organization Management itself. The aim of the Orona Innovation Network was to facilitate the definition, monitoring and rescheduling of the different Virtual Organizations launched via the network for technology and product development in the strategic scope 2006–2010 defined in the roadmap.

The management of the different Virtual Organizations launched from the network consumes many resources, taking into account that some of the up and running Virtual Organizations rely on the participation of a great number of people and the result (which is not always successful in the area of technological innovation) could affect other Virtual Organizations and consequently the product and marketing strategies of our industrial partners.

Within this application level, three different scenarios have been defined. The "VO-Modelling" scenario aims to help the Virtual Organization manager in defining all the elements required to manage the Virtual Organization. The "VO-Monitoring" scenario seeks to establish the mechanisms needed to pursue activities in the Virtual Organization. And lastly, the "VO-Analysis" scenario tries to offer to the Virtual Organization manager the tools required to respond to any untoward events that may occur during the life of the Virtual Organization.

To proceed with a demonstration of the results of the ECOLEAD project in these scenarios, the Orona Innovation Network has selected the Virtual Organization known as 'ARCA 2010' as a basis for realization of the take-up, created in order to design and develop reliable programmable electronic systems with high availability. The concept of availability is considered to be the relationship between the time of operation without faults and the total time also taking into account the time of replacement.

The Virtual Organization will develop the technology platform that includes the controller, the maintenance terminal, controller testing and control supervision.

```
</content>  
</ksDocument>
```

colecoes/docs/doc_1010.xml

```

i»¿ <?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
<context />
<features>
  <feature name="DocumentId" value="1010" />
  <feature name="Title" value="Results used in OIN / VO Management" />
  <feature name="ExtractedFrom" value="D814c - section 6.2.2" />
</features>
<content>Key Results used in OIN for VO Management:
- VO operational governance models (KR3.2): A framework defining the
  basis for the management of distributed collaborative organizations.
- Dynamic VO management models (KR3.3): Definition of models supporting
  the operative management of dynamic VO.
- VO Process Supervision & Performance Measurement (KR3.1):
  Functionalities to support performance based VO supervision and
  management.
- VO Management e-Services (KR3.4): e-Services developments to be
  implemented as software tools:
  - VOM VO-Model
  - DSS Decision Support System
  - SID Supporting Indicator Definitions
  - DI3 Distributed Indicator Information Integrator
  - MAF Monitor and Finance
</content>
</ksDocument>

```

colecoes/docs/doc_1011.xml

```

i»¿ <?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
<context />
<features>
  <feature name="DocumentId" value="1011" />
  <feature name="Title" value="VOM TOOLKIT COMPONENT: VO-MOD" />
  <feature name="ExtractedFrom" value="D34.6 - section 3.1" />
</features>
<content>3.1 VOM TOOLKIT COMPONENT: VO-MOD

```

This section describes the upgrades to the VO-Mod component in the last period. The VO-Mod is a set of tools and Web services devoted to the control of the VO information. In particular it provides instruments for developing and managing VO models during the entire VO lifecycle. The VO-Mod plays a fundamental role as central point for the integration of the VOM platform in two different ways: first with the VO-Mod Wizard which orchestrates the usage of the other VOM tools (following specific VO processes), then with a set of Web services that give access to the VO data from external tools.

3.1.1 The VO-Mod Wizard 2.0

In the last period the VO-Mod Light Wizard has increased its functionalities and now covers entirely the VO lifecycle. The interface follows the four macro processes that collect all the specific VO activities, providing a usable interface able to support the VO management in following his workflow. In this perspective the VO-Mod Wizard is the real "core" of the platform, providing a single gateway for accessing the other tools of the platform on the

base of the business processes specifically designed for the VO management.

In particular the VO lifecycle is mapped in four main macro processes:

1. Initiation: In this phase the VO manager define the initial VO model.
2. Operation: The VO manager sets up the monitoring environment and tracks the VO actual status.
3. Evolution: This stage allows modifying elements of the VO models in accordance with the changes occurred during the Operation phase.
4. Dissolution: The VO manager closes the VO and stores the VO history to the VBE environment in order to support further analysis

</content>
</ksDocument>

colecões/docs/doc_1012.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1012" />
    <feature name="Title" value="VOM TOOLKIT COMPONENT: DSS" />
    <feature name="ExtractedFrom" value="D34.6 - section 3.5" />
  </features>
  <content>3.5 VOM TOOLKIT COMPONENT: DSS

```

This section presents concept of the Decision Support System (DSS) for operational management of the Virtual Organizations (VO).

The DSS concentrates on operation phase of the VO lifecycle. It offers user (the VO manager) with various possibilities how to configure and reconfigure VO and its schedule. Manual as well as semi-automated and automated (re)scheduling is supported. The core feature of the system is the what-if analysis based on a future performance simulation. As an input for the simulation the actual state of the VO, which is represented by a VO configuration and schedule, and to be simulated occurrences influencing them are used. The other simulation parameters configure using of the rescheduling, ability to move tasks to the earlier day if necessary and number of configuration rounds. The simulation provides adapted VO configuration how it would have been in case of occurrences injected to the simulation.

The DSS consists of various modules that together form a system supporting the VO manager in VO simulations and what-is analyses. The analysis is provided by the simulation environment, which applies the "ifs" to the simulated VO and present the "whats". Next to the simulation there is a VO configuration management that allows using up-to-date information about the VO status (only if connected to the rest of the VOM) and manual creation and adaptation of the VO configuration and its schedule. The VO related data are stored in a local repository from that they are provided to the simulation environment. The last main component of the DSS is a management of the DSS users' accounts. This component relates more to the deployment of the system then its value-adding features.

</content>
</ksDocument>

colecões/docs/doc_1013.xml


```

i»¿ <?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1013" />
    <feature name="Title" value="VOM TOOLKIT COMPONENT: MAF" />
    <feature name="ExtractedFrom" value="D34.6 – section 3.4" />
  </features>
  <content>3.4 VOM TOOLKIT COMPONENT: MAF EXTENDED WITH QUALITATIVE
MONITORING

```

This section describes the improvements added in the last period to the MAF platform. The MAF platform is a set of tools devoted to the monitoring of the actual status of a running VO. These tools are used by the VO manager within the Operation phase and allow to control in details the VO activities, maintain updated the VO management situation, monitor the KPIs status, provide an proactive alerting system and manage possible exceptions.

In the last period the MAF Dashboard has been continuously improved and new visual functionalities have been added (details about all the visual features in the annex), in order to increase the support to the visual monitoring of the VO and KPI status. In the figure below is possible to see, for instance, a chart showing an overview of all the KPI values reported. The VO management support is fully integrated with the VO-Mod modelling platform and it allows the constantly update of the VO models information.

```

</content>
</ksDocument>

```

colecoes/docs/doc_1014.xml

```

i»¿ <?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1014" />
    <feature name="Title" value="VOPM TOOLKIT COMPONENT: SID" />
    <feature name="ExtractedFrom" value="D34.1a – section 3" />
  </features>
  <content>3 VOPM TOOLKIT COMPONENT: SID

```

This chapter will describe the tools of the Supported Indicator Definition (SID) in the second prototype of the VOPM toolkit. The description comprises the available functionalities, the technical realisation/architecture and the handling of the tools (user guide).

3.1 Functionalities

The SID is responsible for the selection and configuration of Performance Indicators (PI). It is keeping PI information and can activate the measurement of PI data and the PI calculation by generating a Measurements Schedule to the DI3. The SID interacts with the DI3 and other VOM tools via web service interfaces.

Four sub-components represent the main blocks of functionality available in the SID at present:

- Catalogue of pre-defined Indicators.
- Indicator Management.
- Impact Analyser.
- Measurement Schedule Generator.

```

</content>

```

```
</ksDocument>
```

colecoes/docs/doc_1015.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1015" />
    <feature name="Title" value="VOPM TOOLKIT COMPONENT: DI3" />
    <feature name="ExtractedFrom" value="D34.1a – section 4" />
  </features>
  <content>4 VOPM TOOLKIT COMPONENT: DI3
```

DI3 (pronounced 'ee-triple-eye') stands for 'Distributed Indicator Information Integrator'. It is a measurement system based on distributed, configurable, interacting components which are situated at multiple VO partner-locations. DI3 provides management-indicators values. Information that is retrieved from the partner locations, is gathered, integrated and indicator values are calculated. The process of fetching the requested data and of indicator values calculation is guided by rules and autonomous mechanisms.

The purpose of DI3 is to provide the requested information by means of measuring, fetching or querying data at the locations of the VO-member (also called VO-partners).

The DI3 components are situated in at one or more VO Partner locations. Depending on the local ICT infrastructure, such a location can be physically situated internally on the floor of the partner or externally, at the location of e.g. a telecom provider of the particular partner.

There are two kinds of components that, together, take care of the flexible information gathering:

- An Information Service (IS), responsible for controlling the measurements, and provisioning the results to the other prototype components.
- A Data Adapter (DA), taking care for technical and semantical translations in the communication with existing systems in particular local domains.

```
</content>
</ksDocument>
```

colecoes/docs/doc_1016.xml

```
<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1016" />
    <feature name="Title" value="PCMS" />
    <feature name="ExtractedFrom" value="D22.4a – section 1.1.1" />
  </features>
  <content>1.1.1 PCMS
```

Profile and Competency Management System (PCMS) is designed in order to support the VBE with management of its constituent profiles and

competencies. The VBE profiles provide structured descriptions (of mostly textual content) about the organization entities and the network entities that together represent the VBE (e.g. VBE member organizations, VO-self networks and the VBE-self network), addressing their qualifications, and the records of their related past activities and achievements. The VBE competencies represent a specific part of the profile descriptions of the VBE entities, totally aimed at providing the main characteristics of the VBE entities that are directly / indirectly needed to evaluate organizations for VO creation. The PCMS and its related services are addressed in chapter 2.

```
</content>
</ksDocument>
```

colecoes/docs/doc_1017.xml

```
ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1017" />
    <feature name="Title" value="ODMS" />
    <feature name="ExtractedFrom" value="D22.4a - section 1.1.2" />
  </features>
</content>1.1.2 ODMS
```

Ontology and Discovery Management System (ODMS) is designed in order to support the VBE actors with the manipulation of the VBE ontologies, which are needed for the successfully VBE management. The definition of an ontology for the VBE aims at achieving the following: (1) providing the common understanding of the VBE-related concepts, (2) facilitating the reusability of knowledge accumulated in one VBE with other VBEs, (3) providing the formal classification of the PCMS's knowledge (e.g. competency) in order to facilitate the knowledge processing at VBE by software, and (4) supporting knowledge interoperability both intra-VBE (to support varied forms of collaboration), and inter-VBEs (through sharing the unified models of PCMS information/knowledge). The ODMS and its related services are addressed in chapter 3.

```
</content>
</ksDocument>
```

colecoes/docs/doc_1018.xml

```
ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1018" />
    <feature name="Title" value="TrustMan" />
    <feature name="ExtractedFrom" value="D22.4a - section 1.1.3" />
  </features>
</content>1.1.3 TrustMan
```

The main purpose of TrustMan system is to provide semi-automated services for supporting the management of trust and the related processes in VBEs. The system is implemented based on SOA and specifically web services. Therefore, the functionalities implemented to automate these processes are here referred as services. The complete set of services is implemented to support various actors in the VBE to achieve the following main trust objectives:

- 1) Creating trust and establishing trust relationships among organizations .
- 2) Creating trust of a VBE member to the VBE administration .
- 3) Creating trust of an external organization to the VBE.

To address these main trust objectives the system is constituted by two main manipulation services and two main information provision services. The manipulation services are for the assessment of base trust level of organizations and evaluation of specific trustworthiness of organizations. The information provision services are for the provision of information to convince a member organization to trust the VBE administration and an external organization to the VBE. TrustMan system and its related services are addressed in chapter 4.

```
</content>
</ksDocument>
```

colecões/docs/doc_1019.xml

```
ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features >
    <feature name="DocumentId" value="1019" />
    <feature name="Title" value="Membership and Structure Management
      System – MSMS" />
    <feature name="ExtractedFrom" value="D22.2b – section 1.3" />
  </features>
  <content>1.3 Membership and Structure Management System – MSMS
```

Functionalities for MSMS support the VBE administrator, its members, and those organizations which want to become members by providing functionalities to submit, store, access, delete, and modify the member's company and contact information. The management of the structure of the VBE and the company-related information of its members is a crucial cornerstone for supporting management of all the VBE related activities. Primary functions of VBE membership management are to register and delete members to and from the VMS database; to reward members in case of proactive and positive behavior, and to manage the rights, roles, and responsibility of the VBE members.

```
</content>
</ksDocument>
```

colecões/docs/doc_1020.xml

```
ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features >
    <feature name="DocumentId" value="1020" />
    <feature name="Title" value="VO Information Management System – VIMS"
      />
    <feature name="ExtractedFrom" value="D22.2b – section 1.4" />
  </features>
  <content>1.4 VO Information Management System – VIMS
```

In this section we will address functionalities for management of VO related information.

Purpose of the Functionality As easily to be seen from the underlying concepts of Virtual Breeding Environments and Virtual Organizations, it is beneficial for the VBE to incorporate experiences from previous VO's into the creation of new ones [D21.1]. Developing thorough processes and guidelines on how to use this information in the process of VO Creation is subject to WP2.3 (VO Creation Framework), but the management and provision of VO related data is subject to the VIMS. Consequently, the functionalities for management of VO information provide mechanisms for storing information on created VO's in the VMS data-structure such as structuring, storing, and providing inheritance information to the VO Creation process.

VO related information will be needed by the VO Initiator as repository of experiences with certain partners and combinations of partners in the past [D23.1]. S/he will need this information as input to the decision on which of two or more competing partners to engage in the VO.

```
</content>
</ksDocument>
```

colecoes/docs/doc_1021.xml

```
ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features >
    <feature name="DocumentId" value="1021" />
    <feature name="Title" value="Decision Support System – DSS" />
    <feature name="ExtractedFrom" value="D22.2b – section 1.9" />
  </features>
  <content>1.9 Decision Support System – DSS
```

The Management Decision Support functionality shall help VBE Administrator and VBE Members to keep high level of competitiveness and performance of VBE.

The Management Decision Support functionality basically provides two means for assisting users in their decision and management process:

- Data analysis – processing the data stored in the VBE Management system and providing user with easy to understand results
- Pro-active notification – the VBE Management System can automatically perform selected data analysis regularly and notify user by email when the results of analysis require user attention

The decision support functionality covers the following areas: (a) VBE competency gap analysis, (b) Lack of performance warning, (c) Low trust level warning.

```
</content>
</ksDocument>
```

colecoes/docs/doc_1022.xml

```
ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features >
    <feature name="DocumentId" value="1022" />
    <feature name="Title" value="PARTNERS SEARCH AND SUGGESTION" />
  </features>
  <content>
```

```

    <feature name="ExtractedFrom" value="D23.4b – section 3" />
  </features>
  <content>3 PARTNERS SEARCH AND SUGGESTION

```

The purpose of the partners search and suggestion (PSS) tool is to assist the VO Planner in the selection of the most suitable members for a VO regarding the requirements of a given collaboration opportunity (CO). These requirements are received from the previous VO creation phase "CO characterization and rough planning" (COC-Plan), which will provide a VO macro structure concerning the CO specifications and modality.

The output of the PSS tool is a list of potential "instantiations" (configurations) of the VO structure including an associated risk analysis and the configurations' expected performance with respect to selection criteria. These possible VO configurations are presented to the VO Planner for a further decision making and final VO composition.

Partners are selected based on a sort of criteria. Besides traditional elements like price, delivery date and quality level, performance indicators will be applied both in the searching (filtering inadequate organizations) and in the suggestion (electing the ones that better fit the desired indicators) steps so that the results can be achieved in a faster and potentially better ways. Along the searching step a negotiation process is triggered in order to drive the agreement on the several contract's negotiation objects.

The mentioned indicators will be further used in the VO management phase for monitoring partners' performance along the VO operation. Yet, log-type information about the PSS process will be generated in order to provide historical data for VO inheritance management.

```

</content>
</ksDocument>

```

colecões/docs/doc_1023.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1023" />
    <feature name="Title" value="AGREEMENT NEGOTIATION WIZARD" />
    <feature name="ExtractedFrom" value="D23.4b – section 3" />
  </features>
  <content>4 AGREEMENT NEGOTIATION WIZARD

```

The process of creating a VO is not entirely linear. There are different conditions that might influence its creation as definition of tasks, seeking for appropriate partners, verifying partners' availability, reaching agreements among partners, etc. Therefore, it urges the necessity to obtain a tool that can assist the actors in the process of a VO creation, especially in what concerns the negotiation among the potential partners of the new VO (Camarinha-Matos et. all, 2005).

There are two different situations where negotiation might be required: to select the appropriate partners to compose the VO, and to reach agreements on the details of the VO. The WizAN tool is then intended to provide functionalities for both, being the main result of it the "agreement" reached among the partners, which covers the related issues of the VO.

Although WizAN is intended to provide some degree of automation, it is not intended to cover a fully automated process but to leave the decisions and important inputs to the human user (VO planner). In this case, what is addressed is not a complex e-contracting process where the system is capable of automatically interpret, execute, and manage a contract, but rather, a system that can receive and store inputs into an electronic source for later interpretation, guiding the user through the process (Camarinha-Matos & Oliveira, 2006).

```
</content>
</ksDocument>
```

colecoes/docs/doc_1024.xml

```
ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1024" />
    <feature name="Title" value="CO Characterization and Rough Planning (
      COC-Plan)" />
    <feature name="ExtractedFrom" value="D23.4b - section 1" />
  </features>
  <content>CO Characterization and Rough Planning (COC-Plan)
```

The CO Characterization and Rough Planning (COC-Plan) allows a user to make CO decomposition, identify needed competencies and make a rough planning of the VO. The CO decomposition is done in terms of products and services on a first level, and then the user can make a deeper decomposition of a product into sub-products, assemblies, subassemblies and components; and of a service into sub-services and activities. When the decomposition is done, the tool aids the user to indicate the needed competencies and finally to make a rough plan using a methodology in accordance to the collaboration modality.

```
</content>
</ksDocument>
```

colecoes/docs/doc_1025.xml

```
ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1025" />
    <feature name="Title" value="VIRTUELLE FABRIK (VOM Network)" />
    <feature name="ExtractedFrom" value="D81.4c - section 5" />
  </features>
  <content>5 VIRTUELLE FABRIK (VOM Network)
```

As industrial ECOLEAD partner Virtuelle Fabrik has been very interested in the implementation and demonstration of ECOLEAD methodologies and tools. Especially in the range of VO Management the tool suite has proved to be very useful. In the development phase we had the possibility to influence the design of the different software tools. Our business case was part of the definition for a generic VO model within the development. In the implementation phase we had the possibility to map our structures and processes in a test bed which was provided by the different developers. In the final take-up phase we did the last customizing effort to adapt the tools to our specific end-user's needs. The final demonstration phase shows a WP2 and WP3 integrated scenario, because both work packages has reached

```

    a satisfactory maturity level. Furthermore, much effort is
    undertaken to design post-ECOLEAD activities throughout work
    packages.
  </content>
</ksDocument>

```

colecoes/docs/doc_1026.xml

```

i>¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1026" />
    <feature name="Title" value="PLANNED PROCESSES AND OBJECTIVES" />
    <feature name="ExtractedFrom" value="D81.4c - section 5.2.1" />
  </features>
  <content>5.2.1 PLANNED PROCESSES AND OBJECTIVES

Our objectives in the WP3 are aligned to our global interests in the
Ecolead project. Virtuelle Fabrik is targeting the following VO
Management objectives:

1. VO modeling

This activity of demonstration should show a more efficient way of
modeling a VO and creating a well structured process of order
fulfillment within a VO. The process will be tried out with project
managers of Virtuelle Fabrik.

2. VO performance measurement

VO performance measurement is one of the base functions of VO management.
The efficient
gathering of data from all participants of a VO is one of the key
functions in this process. This will be tried out in a real VO.

3. VO monitoring activities and finance

Full monitoring of the activities and financial evolution of a VO will be
tried out in a real case.

4. Integrated VO Management with VO Management eServices

Integrated trial of the full VO Management eServices in a real VO. This
trial should show the impact for VO Management in terms of
efficiency, speed and overview of the progress of all relevant data
of a VO.
  </content>
</ksDocument>

```

colecoes/docs/doc_1027.xml

```

i>¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1027" />
    <feature name="Title" value="Scenario name: Performance Management"
      />
    <feature name="ExtractedFrom" value="D81.4c - section 5.4.2" />
  </features>

```



```
<content>VO: Virtuelle Fabrik
```

Scenario description: The VO Process Supervision & Performance Measurement is a definition of VO performance, its measurement processes and their relation to the supervision of the VO processes.

Demonstrated Key Results: KR3.1 – VO Process Supervision & Performance Measurement

```
</content>
</ksDocument>
```

colecoes/docs/doc_1028.xml

```
ï»¿<?xml version="1.0" encoding="UTF-8"?>
```

```
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1028" />
    <feature name="Title" value="Scenario name: Integrated VO Management" />
    <feature name="ExtractedFrom" value="D81.4c – section 5.4.2" />
  </features>
  <content>Scenario name: Integrated VO Management
```

VO: Virtuelle Fabrik

Scenario description: The VO Management e-Services will provide an integrated software prototype for easy and efficient modelling of a VO as well as Comprehensive VO Management through a set of tools enabling a permanent overview of ongoing work and performance data. This will be supported by distributed data gathering.

Demonstrated Key Results: KR3.4 – VO Management E-Services

```
</content>
</ksDocument>
```

colecoes/docs/doc_1029.xml

```
ï»¿<?xml version="1.0" encoding="UTF-8"?>
```

```
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1029" />
    <feature name="Title" value="OIN PRESENTATION" />
    <feature name="ExtractedFrom" value="D81.2c – section 8.3" />
  </features>
  <content>8.3 OIN PRESENTATION
```

ORONA is the leading Spanish company in the vertical transport industry. ORONA encouraged firms belonging to the MCC Business group (made of 220 companies and entities is the leading group in the Basque Country and 7th in the ranking in Spain) to join the research consortium: ORONA Innovation Network, whose purpose and objectives includes:

- sharing the cost and risk of research,
- pooling scarce expertise and co-specialization,
- performing pre-competitive research
- gaining new competencies through learning and collaboration in networks
- increasing trust and reducing transaction cost among members
- creating innovative products for the short distance transportation sector.

This association is largely a response to the competitive threat of the biggest manufacturers of elevators in the world.

```
</content>
</ksDocument>
```

colecões/docs/doc_1030.xml

```
ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features >
    <feature name="DocumentId" value="1030" />
    <feature name="Title" value="OIN Partners" />
    <feature name="ExtractedFrom" value="D81.2c - section 8.3" />
  </features >
  <content>OIN Partners
```

The ORONA INNOVATION NETWORK (OIN) can be considered as a Virtual Breeding Environment (VBE) consisting of a number of positions or nodes, occupied by individuals, firms business units, universities, customers or other actors and links and interactions between these nodes. It is composed by the following members:

- Orona
- EIC
- IKERLAN
- MFE
- ELECTRA
- MONDRAGON
- Herreros
- ZABALA

```
</content>
</ksDocument>
```

colecões/docs/doc_1031.xml

```
ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features >
    <feature name="DocumentId" value="1031" />
    <feature name="Title" value="ORONA" />
    <feature name="ExtractedFrom" value="D81.2c - section 8.3" />
  </features >
  <content>ORONA
```

Spanish leading company in the vertical transport industry. Is the leader of the Network. The main interest of ORONA in the VO-OIN is based on the fact that the network leads the development of the New Products and technologies for ORONA as this is one of the pillars of its strategic plan.

```
</content>
</ksDocument>
```

colecões/docs/doc_1032.xml

```
ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features >
```

```

    <feature name="DocumentId" value="1032" />
    <feature name="Title" value="EIC" />
    <feature name="ExtractedFrom" value="D81.2c – section 8.3" />
  </features>
</content>EIC

```

ORONA eic works on product development for the lift and elevator sector. Given that the mission of EIC is to provide solutions and knowledge to ORONA S. Coop., the VO-OIN is the core fabric to EIC's mission, especially in the ambit of technology capture and the identification of new external partners.

```

</content>
</ksDocument>

```

colecoes/docs/doc_1034.xml

```

ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1034" />
    <feature name="Title" value="ELECTRA" />
    <feature name="ExtractedFrom" value="D81.2c – section 8.3" />
  </features>
</content>ELECTRA

```

ELECTRA Vitoria Group has managed to become an international benchmark in the development of vertical transport systems. ELECTRA's interest in the VO-OIN is the same as for ORONA, with a much more incipient character, furthermore, its participation in the VO-OIN could help its mission to narrow the gap and become closer to ORONA thus forming part of the MCC Industrial Group.

```

</content>
</ksDocument>

```

colecoes/docs/doc_1035.xml

```

ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1035" />
    <feature name="Title" value="MONDRAGON CORPORACION COOPERATIVA (MCC)" />
    <feature name="ExtractedFrom" value="D81.2c – section 8.3" />
  </features>
</content>MONDRAGON CORPORACION COOPERATIVA (MCC)

```

MCC is the largest business corporation in the Basque Country and the seventh largest in SPAIN, organized in three sectorial groups: Financial, Industrial and Distribution, together with the Research and Training areas.

```

</content>
</ksDocument>

```

colecoes/docs/doc_1038.xml

```

ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />

```

```

<features >
  <feature name="DocumentId" value="1038" />
  <feature name="Title" value="ISOIN / HELICE.NET" />
  <feature name="ExtractedFrom" value="D81.2c – section 6.1.4" />
</features >
<content>ISOIN / HELICE.NET

```

Organizations in the aeronautic sector currently operate under an Extended Enterprise model, with common supporting ICT infrastructure, derived from an ERP model (SAPECMA and SAPORTAL), together with methodologies, services, and tools to facilitate the delivery of supplied parts. Tools developed in ECOLEAD will be inserted in the existing ICT infrastructure managed by HELICE.NET to test and validate the VBE formation and VO creation and launch.

Three business processes will be implemented, from the formation of a VBE from the existing cluster to the creation and launch of resulting VOs. The pilots will be run in the real life environment of four members of HELICE.NET: MEUPE, INESPASA, AEROSUR and GHESA. The required implementation plan is assisted by ISOIN, the core technological partner of HELICE.

```

</content >
</ksDocument >

```

colecões/docs/doc_1039.xml

```

ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features >
    <feature name="DocumentId" value="1039" />
    <feature name="Title" value="HELICE.NET Members" />
    <feature name="ExtractedFrom" value="D81.2c – section 6.1.3" />
  </features >
  <content>HELICE.NET Members

```

From the diversity of companies that form the cluster, some of them have been selected to create a multidisciplinary working group. All of them have previous experience in the development of projects in cooperation. They can be classified into four main groups:

- Main contractor: EADS–CASA
- Manufacturing and assembly SMEs: MEUPE, INESPASA and AEROSUR. They have previous experience in the preparation of joint contracts to main contractors.
- Engineering company: GHESA which has had international experiences in the naval engineering field. The internationalisation of their aeronautical field through cooperation with CeBeNetwork is a cornerstone.
- Supporting entity: Regional Development Agency. It is responsible for the promotion of the advanced collaborative Paradigm in strategic industrial clusters as a driver to innovation in the Regional RTD Program for Innovation (PIMA 2005–2010).

They all have experience in cooperation activities. Furthermore, MEUPE and INESPASA have a collaboration agreement in place, to rapidly answer business opportunities with joint offers. An Introductory workshop was held on 28th July 2006 in INESPASA headquarter with the premise of establishing the working group and define the key objectives and tasks in the ECOLEAD demonstration activities.

```

</content>
</ksDocument>

```

colecoes/docs/doc_1040.xml

```

i»¿ <?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1040" />
    <feature name="Title" value="CeBeNetwork" />
    <feature name="ExtractedFrom" value="D81.2c - section 6.1.3" />
  </features>
  <content>CeBeNetwork

```

CeBeNetwork VBE was founded in 1996 and has grown rapidly since with a turnover of more than 30M euros and about 200 in 2005. It provides the aeronautical industry with engineering services where AIRBUS is the main customer. Typically for the aeronautical industry, also CeBeNetwork customers reduced the number of suppliers starting in 2003. The remaining suppliers where claimed to deliver the whole spectrum of engineering services necessary for the development of aircrafts. Furthermore the request was raised to be present at various customer sites throughout Europe. CeBeNetwork needed to manage this challenge to stay in this field and become one of the key suppliers. But this task could not be handled successfully by CeBeNetwork alone. It required the accomplishment of motivated partners, therefore CeBeNetwork have been collaborating with others in the past few years.

```

</content>
</ksDocument>

```

colecoes/docs/doc_1041.xml

```

i»¿ <?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1041" />
    <feature name="Title" value="CeBeNetwork Members" />
    <feature name="ExtractedFrom" value="D81.2c - section 6.2" />
  </features>
  <content>CeBeNetwork Members

```

The attempt of CeBeNetwork is to unify the companies for large projects and set up tools to facilitate this process. For the ECOLEAD project four companies were selected (EuroEngineering aerospace, P+Z, Aerocon and AIDA Development) to participate in the development and training activities. The lead in the introduction and training activities is takes by CeBeNetwork.

```

</content>
</ksDocument>

```

colecoes/docs/doc_1042.xml

```

i»¿ <?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1042" />

```

```

    <feature name="Title" value="Key personel – Carmen" />
    <feature name="ExtractedFrom" value="D81.2c – section 6.3.2" />
  </features>
  <content>
Carmen M. Aguilera, MEng in Telecommunication Sciences, heads the
  Innovation and Technology Department in ISOIN since 2002. She owns a
  Master on 3G technologies e-services and mobile applications. She
  is specialized in collaborative engineering, concurrent enterprise,
  communication networks and ambient intelligence. She is currently
  working as head unit and coordinator for both customer oriented and
  RTD projects in national and EU programs. Most relevant FP5 and FP6
  projects: OLIMP 3D and CADPIPE, distributed design and simulation in
  virtual teams; ECOLEAD, dissemination on collaborative networked
  organization in the aeronautic cluster; "Global Performance
  Management in SMEs" (GPM-SME), with pilots in aeronautical and
  automotive sectors. National research projects: "Set up of advanced
  industrial cluster in the aeronautical sector", "Life-cycle
  Management System for SMEs" (SIDAPRO); "Concurrent design in SMEs"(
  SIGESDA), RFID based quality tracking in manufacturing (INSOTRAZA);
  Distributed design (SIDIFA); Collaborative Design in distributed
  teams (DISCO); Aml Platform supporting the "Annette & Salomon"
  ISTAG scenario (ANNETTE), "Set up of advanced industrial cluster in
  the aeronautical sector. She is an accredited strategic and
  technical consultant in the Regional Government of Andalusia and
  actively collaborates in research activities and joint initiatives
  with the University of Seville.
  </content>
</ksDocument>

```

colecões/docs/doc_1043.xml

```

ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features>
    <feature name="DocumentId" value="1043" />
    <feature name="Title" value="Key personel – Jose" />
    <feature name="ExtractedFrom" value="D81.2c – section 6.3.2" />
  </features>
  <content>
Jose A. Parejo, MEng in Computing Sciences, heads the Development
  Department in ISOIN since 2002. He is currently focused on project
  management in the fields of distributed services for virtual
  communities and e-Business: "System for Evaluation and Decision
  Management" (SIMITS), "Distributed simulation in logistic and
  services for supply chain" (GESTRA), "ISOIN eLearning Platform" (
  ISOIN-VIRTUAL). He has managed experience from dissemination and
  commercialization of results from research projects in the company.
  He has a deep background on distributed computing, artificial
  intelligence (Cool language) and advance optimization techniques (
  Metaheuristics) with a large number of publications in international
  conferences and scientific journals, as well as active
  collaborations in research activities and PhD programs in the
  University of Seville.
  </content>
</ksDocument>

```

colecões/docs/doc_1044.xml

```

ï»¿<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>

```

```

<context />
<features>
  <feature name="DocumentId" value="1044" />
  <feature name="Title" value="Key personel – Ricardo" />
  <feature name="ExtractedFrom" value="D81.2c – section 6.3.2" />
</features>
<content>
Ricardo Galan, PhD in Industrial Engineering from University of Seville .
  His major research interests include reconfigurable and automated
  manufacturing systems , production control systems , virtual
  organisations , process engineering and concurrent enterprising. Dr.
  Galan has written many international publications and participated
  in industrial projects at national and international levels .
</content>
</ksDocument>

```

colecoes/docs/doc_1045.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
<context />
<features>
  <feature name="DocumentId" value="1045" />
  <feature name="Title" value="Key personel – Simon" />
  <feature name="ExtractedFrom" value="D81.2c – section 6.3.2" />
</features>
<content>
Simon Vazquez, Industrial Engineer from the University of Sevilla , is
  directly involved in the development of AEROPOLIS–Andalusian
  Aeronautical Cluster since its creation as Andalusia Aeroespacial
  and its current form of AEROPOLIS. He has been acting as project
  manager for AEROPOLIS in the fields of Virtual and Extended
  Enterprises and advanced Information Technology Systems and services
  for the Andalusian Aeronautical Sector .
</content>
</ksDocument>

```

colecoes/docs/doc_1046.xml

```

<?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
<context />
<features>
  <feature name="DocumentId" value="1046" />
  <feature name="Title" value="Key personel – Berthold" />
  <feature name="ExtractedFrom" value="D81.2c – section 6.3.2" />
</features>
<content>
Dipl.–Ing. Berthold Tiefensee, studied production technologies at the
  University of Bremen. After finishing his studies in 1989 he became
  research engineer at BIBA–Institute. During this time he worked in
  quality management and co–operation projects. In 1999 he started to
  work for the company M.H Wilkens&SÄ¶hne GmbH. As the head of
  quality management he implemented the integrated quality management
  system. The main aim of the system was to increase the quality of
  the international manufactured products. The system focuses on the
  production systems of the suppliers manufacturing in Asia and east
  Europe. Additionally he was in response to organise the production
  cooperation between M.H.Wilkens&SÄ¶hne and an Estonian company.
  In 2004 he started his activities at the CeBeNetwork GmbH. He
  supported the foundation of the co–operation network. Now he is

```

responsible to all related business processes and the related customer support. His interest in the last two years was to support the development of CeBeNetwork to an European first level engineering supplier for Airbus and build up the CeBeNetwork Co-Operation.

```
</content>
</ksDocument>
```

colecões/docs/doc_1047.xml

```
ï»¿<?xml version="1.0" encoding="UTF-8"?>
```

```
<ksDocument>
```

```
<context />
```

```
<features >
```

```
<feature name="DocumentId" value="1047" />
```

```
<feature name="Title" value="Key personel – Daniel" />
```

```
<feature name="ExtractedFrom" value="D81.2c – section 6.3.2" />
```

```
</features >
```

```
<content >
```

Dipl.–Ing. Daniel Laessig started studying International Sales and Purchase Engineering at the University of Applied Sciences in Kiel in 2001. During several internships amongst others in the Procurement Departments at Airbus in Germany as well as in the UK he gained experience in dealing with Supplier Networks from the buyers' side. In his Diploma Thesis he investigated the development and different ways of working of the so called "Strategic Suppliers" of the Engineering Services Provider for Airbus Germany. After finishing his studies in summer 2005 he started working in the Co-Operation Business at CeBeNetwork GmbH. Here he works in close contact to the Airbus Procurement, organises and coordinates projects within the internal network and supervises them. Moreover he is involved in the establishment of a CoOperation network in the UK.

```
</content >
</ksDocument>
```

colecões/docs/doc_1048.xml

```
ï»¿<?xml version="1.0" encoding="UTF-8"?>
```

```
<ksDocument>
```

```
<context />
```

```
<features >
```

```
<feature name="DocumentId" value="1048" />
```

```
<feature name="Title" value="Business process 1: Characterization of the HELICE VBE" />
```

```
<feature name="ExtractedFrom" value="D81.4c – section 3.2.1" />
```

```
</features >
```

```
<content>Business process 1: Characterization of the HELICE VBE and the CeBeNetwork VBE
```

The goal is to design the organizational and technical framework to support the operation of the VBE according to the ECOLEAD VBE Ontology. In parallel, a structure for the cooperation of Helice and CeBeNetwork will be analyzed.

VBE Management Software used:

- Membership and Structure Management Systems (MSMS)
- Profiling and Competency Management System (PCMS)
- Trust Management System (TrustMan)

```
</content >
</ksDocument>
```


colecoes/docs/doc_1049.xml

```

i»¿ <?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
  <context />
  <features >
    <feature name="DocumentId" value="1049" />
    <feature name="Title" value="Business process 2: VO Creation in
      HELICE VBE" />
    <feature name="ExtractedFrom" value="D81.4c – section 3.2.1" />
  </features>
  <content>Business process 2: VO Creation in HELICE VBE

```

The goal is to develop and implement the required infrastructures to support a single company to identify collaboration opportunities with other companies to perform larger packages than those that can be performed alone.

Software Used in this Process

- Collaboration Opportunity Characterization and Rough Planning (COC-Plan)
 - Partners Search and Suggestion (PSS)
- ```

</content>
</ksDocument>

```

## colecoes/docs/doc\_1050.xml

```

i»¿ <?xml version="1.0" encoding="UTF-8"?>
<ksDocument>
 <context />
 <features >
 <feature name="DocumentId" value="1050" />
 <feature name="Title" value="Business process 3: VO Launch in HELICE"
 />
 <feature name="ExtractedFrom" value="D81.4c – section 3.2.1" />
 </features>
 <content>Business process 3: VO Launch in HELICE

```

The objective of the VO launch process is to support the firms within the cluster to produce joint commercial offers and at the same time to guarantee the working activities of the firms. The goal is to optimize the effort of creating a project agreement and the related working mechanisms when a business opportunity is identified.

Software tools used

- Partners Search & Suggestion Tool (PSS).
  - Negotiation Wizard Tool (WizAN).
- ```

</content>
</ksDocument>

```


Apêndice F

Ferramentas de Suporte

F.1 Geração de Relatórios de Desempenho de Sistemas de RI

Esta ferramenta foi utilizada na etapa de avaliação do protótipo onde, depois de executados os experimentos, os resultados das consultas, juntamente com os resultados esperados, são processados para a geração dos relatórios de avaliação. A metodologia para avaliação pode ser vista na seção 7.2.1 (página 124), onde esta ferramenta é mencionada.

Ela funciona em dois modos de avaliação:

- Avaliação simples, ou seja, de um único sistema e;
- Comparação entre dois sistemas.

F.1.1 Avaliação Simples

Este modo de avaliação toma como base um arquivo contendo os resultados obtidos, um arquivo com os resultados esperados, e gera um relatório de desempenho que é armazenado num terceiro arquivo. A figura F.1 mostra a tela da ferramenta, onde os campos para esses três arquivos devem ser especificados e a execução é feita pelo botão “*Perform Evaluation*”.

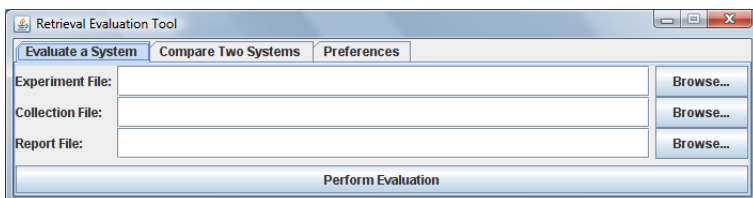


Figura F.1: Tela da ferramenta de avaliação (modo de avaliação simples).

Formatos dos Arquivos

Os arquivos seguem um formato especificado em XML, onde os arquivos de entrada têm uma estrutura bem parecida, composta por uma lista de consultas e os respectivos resultados. A figura F.2 mostra um exemplo de arquivo contendo os resultados esperados para uma coleção (*Exemplo 1*) que contém uma única consulta.

```
<?xml version="1.0" encoding="UTF-8"?>
<queryList collection="Exemplo 1">
  <query id="1">
    <docId>3</docId>
    <docId>5</docId>
    <docId>9</docId>
    <docId>25</docId>
    <docId>39</docId>
    <docId>44</docId>
    <docId>56</docId>
    <docId>71</docId>
    <docId>89</docId>
    <docId>123</docId>
  </query>
</queryList>
```

Figura F.2: Formato de arquivo para os resultados esperados.

A razão para se usar XML como formato é que este é um formato relativamente fácil para geração e processamento. Assim, a ferramenta pode avaliar os resultados de qualquer sistema que gere os resultados no formato especificado. A figura F.3 mostra um exemplo de arquivo contendo os resultados obtidos por um certo sistema (*K-search1*) para a coleção *Exemplo 1*.

O Relatório Gerado

O relatório gerado oferece as medidas apresentadas na seção 4.4 (página 56), que inclui:

Medidas gerais / médias:

- Número de consultas;
- Número de documentos recuperados;
- Número de documentos relevantes recuperados;
- Número de documentos relevantes esperados;
- Precisão média para os documentos relevantes observados;
- R-Precisão Média e;

```
<?xml version="1.0" encoding="UTF-8"?>
<experimentResults collection="Exemplo 1" system="K-search1">
  <query id="1">
    <docId>123</docId>
    <docId>84</docId>
    <docId>56</docId>
    <docId>6</docId>
    <docId>8</docId>
    <docId>9</docId>
    <docId>511</docId>
    <docId>129</docId>
    <docId>187</docId>
    <docId>25</docId>
    <docId>38</docId>
    <docId>48</docId>
    <docId>250</docId>
    <docId>113</docId>
    <docId>3</docId>
  </query>
</experimentResults>
```

Figura F.3: Formato de arquivo para os resultados obtidos.

- Precisão nos 11 valores de cobertura padrão, incluindo um gráfico.

Medidas detalhadas (por consulta):

- Número de documentos recuperados;
- Número de documentos relevantes recuperados;
- Número de documentos relevantes esperados;
- Precisão para os documentos relevantes observados;
- R-Precisão e;
- Precisão nos 11 valores de cobertura padrão.

Em termos de formato, o arquivo do relatório também é codificado em XML, conforme mostra o fragmento apresentado na figura F.4.

Graças a uma folha de estilos, definida em XSL, o relatório pode ser visualizado num navegador Web. O gráfico da curva de previsão x cobertura é definido em um arquivo separado, no formato *Scalable Vector Graphics* (SVG)¹. Tanto a folha de estilos quanto o gráfico também são gerados pela ferramenta. A figura F.5 mostra um exemplo de relatório visualizado dessa forma.

¹SVG é um formato para a definição de gráficos vetoriais em XML. Sua especificação é um padrão aberto mantido pelo W3C (Ferraiolo et al., 2009)

```

<?xml version="1.0" encoding="UTF-8"?>
<?xml-stylesheet type="text/xsl" href="testReport_novo_ReportFiles\report
.xsl"?>
<performanceReport dateTime="Fri Jan 22 16:51:24 BRST 2010">
  <collectionName>Exemplo</collectionName>
  <systemName>K-search1</systemName>
  <svgFileLocation>testReport_novo_ReportFiles\Exemplo_K-search1.svg</
  svgFileLocation>
  <queryPerformance>
    <summaryTableStatistics>
      <numberOfQueries>1</numberOfQueries>
      <numberOfDocumentsRetrieved>15</numberOfDocumentsRetrieved>
      <numberOfRelevantDocumentsRetrieved>5</
      numberOfRelevantDocumentsRetrieved>
      <numberOfRelevantDocumentsExpected>10</
      numberOfRelevantDocumentsExpected>
    </summaryTableStatistics>
    <averagePrecision-recall>
      <p recallLevel="0.0" precision="1.0" />
      <p recallLevel="0.1" precision="1.0" />
      <p recallLevel="0.2" precision="0.6666667" />
      <p recallLevel="0.3" precision="0.5" />
      <p recallLevel="0.4" precision="0.4" />
      <p recallLevel="0.5" precision="0.33333334" />
      <p recallLevel="0.6" precision="0.0" />
      <p recallLevel="0.7" precision="0.0" />
      <p recallLevel="0.8" precision="0.0" />
      <p recallLevel="0.9" precision="0.0" />
      <p recallLevel="1.0" precision="0.0" />
    </averagePrecision-recall>
    <averagePrecisionAtSeenRelevantDocuments>0.58000004
    </averagePrecisionAtSeenRelevantDocuments>
    <averageR-precision>0.4</averageR-precision>
  </queryPerformance>

```

Figura F.4: Formato de arquivo para o relatório de desempenho (fragmento).

F.1.2 Comparação entre Sistemas

Este modo toma como base dois arquivos contendo os relatórios já obtidos de dois sistemas (individualmente), e armazena o relatório comparativo num terceiro arquivo. A figura F.6 mostra a tela da ferramenta, onde os campos para esses três arquivos devem ser especificados e a execução é feita pelo botão “*Perform Comparison*”.

Este relatório apresenta as seguintes informações (para os dois sistemas):

- Número de consultas;
- Número de documentos recuperados;
- Número de documentos relevantes recuperados;
- Número de documentos relevantes esperados;

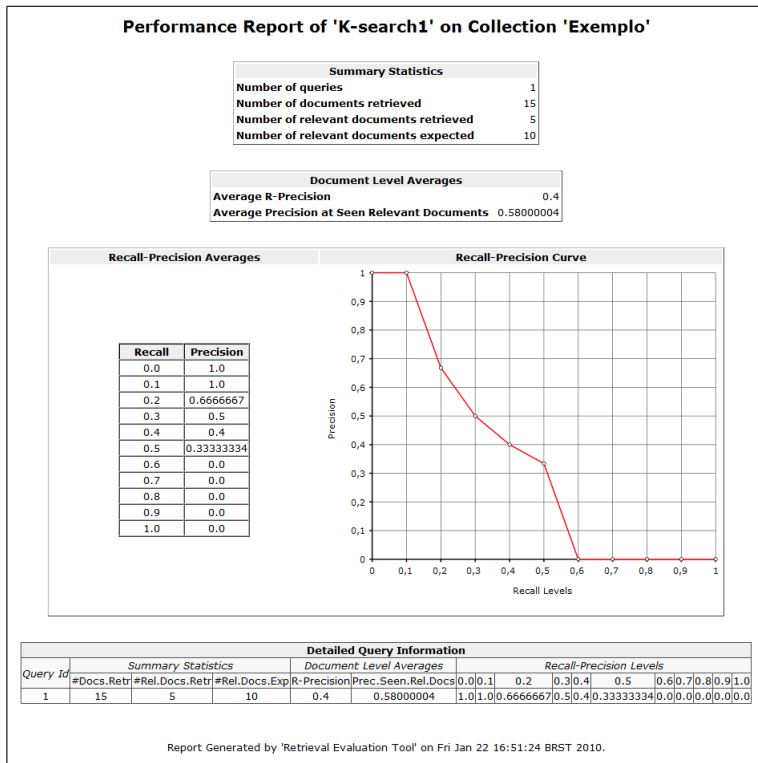


Figura F.5: Relatório de avaliação de um sistema, após a transformação com XSL.

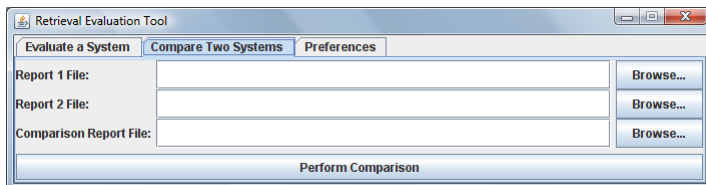


Figura F.6: Tela da ferramenta de avaliação (modo de comparação entre sistemas).

- Precisão média para os documentos relevantes observados;
- R-Precisão Média;
- Precisão nos 11 valores de cobertura padrão, incluindo um gráfico comparativo, e;
- Histograma de R-Precisão.

O formato do relatório comparativo é bastante parecido com o do relatório simples, e também pode ser facilmente visualizado num navegador Web, conforme mostra a figura F.7.

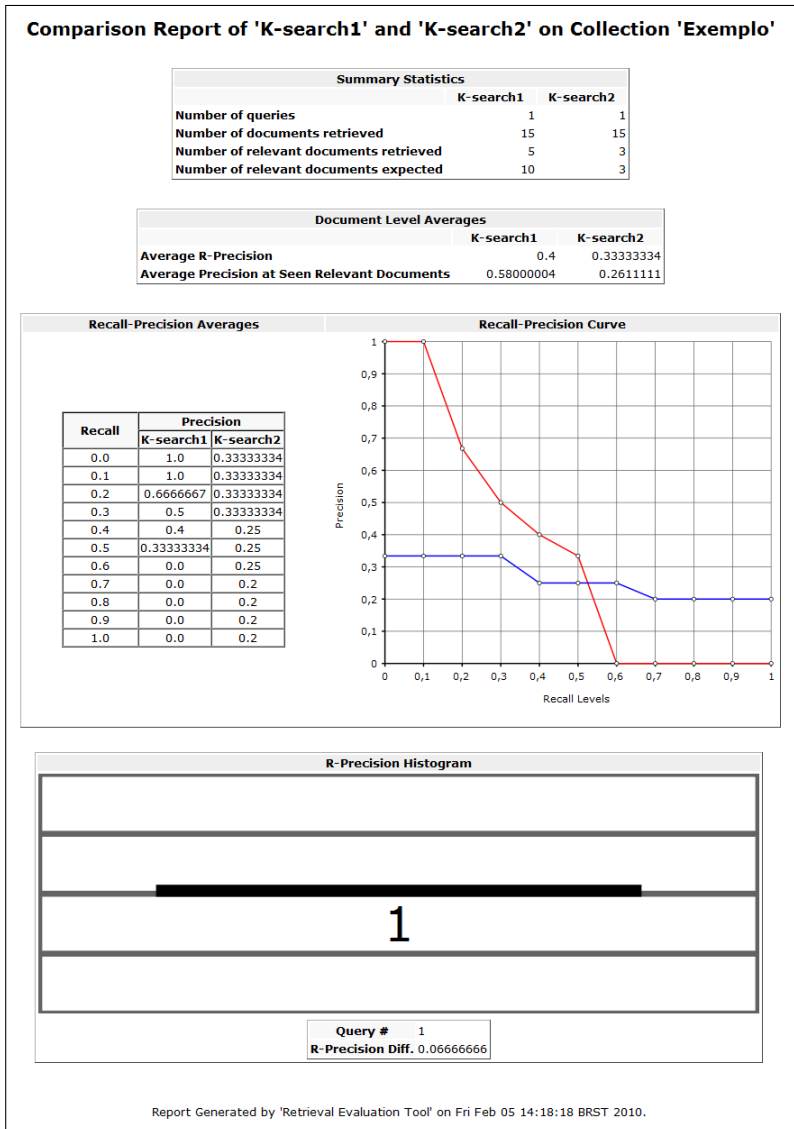


Figura F.7: Relatório de comparação entre sistemas, visualizado num navegador Web.

Apêndice G

Questionário para Definição dos Critérios de Relevância

As páginas seguintes apresentam o questionário utilizado para a definição dos critérios de relevância para as consultas usadas nos experimentos.

Questionário de Avaliação do “Sistema de Busca Baseado em Contexto” – *K-search*

O *k-search* compreende um conjunto de serviços de suporte à busca de conhecimento em redes colaborativas. O objetivo deste questionário é definir o critério de relevância das coleções de teste usadas para avaliar o sistema *k-search*.

A abordagem aqui adotada é diferente da tradicional, onde há uma etapa preliminar de uso do sistema seguida pelo preenchimento do questionário. Aqui o usuário não tem contato com o sistema, participando apenas da segunda etapa. Ou seja, a avaliação acontecerá de forma indireta, sendo os resultados dos questionários preenchidos pelos avaliadores, especialistas do domínio, utilizados como comparativo aos resultados fornecidos pelo sistema.

São apresentados dois cenários de busca, sendo um total de sete consultas. Para cada consulta, uma lista de documentos é apresentada, e cabe ao usuário definir quais desses documentos são relevantes. Tais julgamentos serão comparados às respostas obtidas pelo sistema, sendo então possível fazer uma análise quantitativa de seu desempenho.

Instruções:

Um aspecto essencial a ser considerado pelo usuário é o seu contexto durante a busca. O contexto é um fator que pode auxiliar no refinamento da busca, deixando-a mais específica, ou definindo mais claramente o significado que o termo pesquisado pode ter num contexto específico. Nesta tese, o contexto é definido pelos elementos que compõem o ambiente de trabalho do usuário:

- a rede de organizações da qual faz parte;
- o papel desempenhado pelo usuário (ou a organização para a qual trabalha);
- o processo e a tarefa executados atualmente pelo usuário.

Neste questionário, o contexto atual do usuário é explicitamente determinado para cada consulta. Cada um dos elementos de contexto é explicado brevemente, no intuito de se dar uma idéia dos possíveis interesses do usuário. Adicionalmente, é explicada a situação que levou à necessidade de busca e a consulta é então apresentada.

Com base nessas informações, cabe ao usuário se colocar nesse contexto e fazer os julgamentos de relevância, assinalando um ou mais documentos que julgar relevantes. É importante salientar a necessidade do usuário “entrar” no contexto, ou seja, é preciso considerar tanto a consulta quanto o contexto atual.

Outra consideração é que as consultas aqui não são simples seqüências de palavras-chave. Elas são representadas em um nível mais alto, na forma de frases que especificam a necessidade do usuário. Os termos mais importantes (que são as palavras-chave da busca) são destacados em negrito. É importante também salientar que as consultas são codificadas internamente com o auxílio de uma ontologia. Portanto, o usuário pode levar em consideração os sinônimos dos termos utilizados. Por exemplo, o termo “Accuracy” pode ser interpretado como “Precision”, “Correctness”, “Accurateness”, “Properness”, “Faultless”, “Without Error”, etc.

Os cenários são apresentados a seguir.

Cenário 1: Busca por Indicadores de Desempenho

A busca por indicadores de desempenho é uma atividade essencial em uma rede de organizações, especialmente durante a criação de uma organização virtual, onde parceiros devem ser selecionados e seus desempenhos em negócios passados devem ser analisados como critério para tal seleção. Outra situação que pode utilizar desse tipo de busca é quanto há a necessidade de se tomar alguma decisão (o gestor da organização, por exemplo), e algum indicador pode ser usado nessa avaliação.

De uma maneira simples, um indicador pode ser qualificado por vários atributos. Pode-se dizer, dentre outras coisas, que um indicador (Baldo, 2008):

- mede *algo*: recurso, processo, organização;
- é medido com um certo *objetivo*: planejamento, gestão, produção, distribuição;
- considera uma *perspectiva de desempenho* específica: tempo, custo, confiabilidade, qualidade, flexibilidade.

É importante salientar que os atributos de um indicador podem ter uma relação com o contexto do usuário. Por exemplo, o *objetivo* dá uma idéia do tipo de processo em que o usuário pode estar envolvido.

Consulta 1:

Contexto: O usuário é **membro de uma organização virtual** e está num processo de **gestão (management)** executando a atividade de **análise de custos (cost analysis)**.

Necessidade do usuário: Em um dado momento, ele tem interesse em saber os **custos** relacionados ao processo de **entrega (delivery)**. Ele então define a seguinte consulta:

Consulta: “Indicadores que meçam **entrega (delivery)** e considerem **custo (cost)**”

Resultados:

| | |
|-----|---|
| () | <p>2. Cost per Invoice <i>Perspective:</i> Cost <i>Description:</i> All costs associated with the receipt, review, processing, and payment of a supplier's invoice for product received.</p> |
| () | <p>5. Order Management Costs <i>Perspective:</i> Cost <i>Description:</i> The aggregation of the following cost elements: Create Customer Order Costs, Order Entry and Maintenance Costs, Contract/Program and Channel Management Costs, Installation Planning Costs, Order Fulfillment Costs, Distribution Costs, Transportation Costs, Installation Costs, Customer Invoicing/Accounting Costs.</p> |
| () | <p>6. Order Management Cycle Time <i>Perspective:</i> Time <i>Description:</i> The total amount of time required converting a customer order into a receipt by the customer.</p> |
| () | <p>8. Packaging Cost <i>Perspective:</i> Cost <i>Description:</i> The cost to package product as a finished good, not including intermediate handling of materials, based on given number of Delivered Finished Goods.</p> |
| () | <p>12. Percentage of Faultless production <i>Perspective:</i> Quality <i>Description:</i> Number of products produced without error in the manufacturing sector. An example of potential defects that can decrease the quality is lack of precision in the manufacture process.</p> |
| () | <p>13. Production Material Cycle Time <i>Perspective:</i> Time <i>Description:</i> Time required moving material to point of use.</p> |
| () | <p>15. Number of re-plannings <i>Perspective:</i> Reliability <i>Description:</i> Indicates the reliability of plannings, based upon the informations/commitments of the VO Members. If a certain period should be analysed, this can be set through the validity dates of the indicator.</p> |
| () | <p>18. Percentage of available resources in relation to overall assigned resources <i>Perspective:</i> Flexibility <i>Description:</i> Represents the part of the overall resources assigned to a VO that is actually available for the VO.</p> |

Consulta 2:

Contexto: O usuário é **membro de uma organização virtual** e está num processo de **gestão (management)** executando a atividade de **avaliação de qualidade (quality evaluation)**.

Necessidade do usuário: Este usuário deseja busca indicadores relacionados à **produção**, e faz a seguinte consulta:

Consulta: “indicadores que meçam **produção (production)**”

Resultados:

| | |
|-----|---|
| () | <p>2. Cost per Invoice <i>Perspective:</i> Cost <i>Description:</i> All costs associated with the receipt, review, processing, and payment of a supplier's invoice for product received.</p> |
| () | <p>6. Order Management Cycle Time <i>Perspective:</i> Time <i>Description:</i> The total amount of time required converting a customer order into a receipt by the customer.</p> |
| () | <p>10. Production Plan Adherence <i>Perspective:</i> Reliability <i>Description:</i> It is calculated using the following formula: $\text{Production Plan} - \text{Sum of Variance}$, Production Plan Where: $\text{Production Plan} = \text{The sum of the units planned to be completed in each month based upon the plan generated in the previous month. Sum of Variances} = \text{The sum of the absolute values, at the end item level, of the differences between each month's production plan as defined above and actual production for the same month.}$</p> |
| () | <p>11. Production Process Validation Frequency <i>Perspective:</i> Reliability <i>Description:</i> The amount of time between reviews of a process. For example, $\text{Production Process Validation Frequency}$ would refer to the amount of time between the reviews of the $\text{Production Process}$. This generally would be performed periodically to ensure that the process is generating the desired results with the desired inputs.</p> |
| () | <p>12. Percentage of Faultless production <i>Perspective:</i> Quality <i>Description:</i> Number of products produced without error in the manufacturing sector. An example of potential defects that can decrease the quality is lack of precision in the manufacture process.</p> |
| () | <p>13. Production Material Cycle Time <i>Perspective:</i> Time <i>Description:</i> Time required moving material to point of use.</p> |
| () | <p>14. Deviation from planned start date <i>Perspective:</i> Reliability <i>Description:</i> This measures how much the start of a tasks deviates from the planned start date.</p> |
| () | <p>15. Number of re-plannings <i>Perspective:</i> Reliability <i>Description:</i> Indicates the reliability of plannings, based upon the informations/commitments of the VO Members. If a certain period should be analysed, this can be set through the validity dates of the indicator.</p> |

Consulta 3:

Contexto: O usuário é o **gestor de uma organização virtual** e está num processo de **gestão da OV (VO management)** executando a atividade de **reescalonamento (rescheduling)**.

Necessidade do usuário: O gestor deseja buscar indicadores relacionados à **confiabilidade**.

Consulta: “indicadores que considerem **confiabilidade (reliability)**”

Resultados:

| | |
|-----|--|
| () | 1. Percentage Of Qualified Suppliers which Meet Defined Requirements <i>Perspective:</i> Reliability <i>Description:</i> The number of qualified suppliers who meet defined requirements divided by the total number of qualified suppliers used as sources in the measurement period. |
| () | 5. Order Management Costs / <i>Perspective:</i> Cost <i>Description:</i> The aggregation of the following cost elements: Create Customer Order Costs, Order Entry and Maintenance Costs, Contract/Program and Channel Management Costs, Installation Planning Costs, Order Fulfillment Costs, Distribution Costs, Transportation Costs, Installation Costs, Customer Invoicing/Accounting Costs. |
| () | 10. Production Plan Adherence <i>Perspective:</i> Reliability <i>Description:</i> It is calculated using the following formula: Production Plan - Sum of Variance, Production Plan Where: Production Plan = The sum of the units planned to be completed in each month based upon the plan generated in the previous month. Sum of Variances = The sum of the absolute values, at the end item level, of the differences between each month's production plan as defined above and actual production for the same month. |
| () | 11. Production Process Validation Frequency <i>Perspective:</i> Reliability <i>Description:</i> The amount of time between reviews of a process. For example, Production Process Validation Frequency would refer to the amount of time between the reviews of the Production Process. This generally would be performed periodically to ensure that the process is generating the desired results with the desired inputs. |
| () | 12. Percentage of Faultless production / <i>Perspective:</i> Quality <i>Description:</i> Number of products produced without error in the manufacturing sector. An example of potential defects that can decrease the quality is lack of precision in the manufacture process. |
| () | 14. Deviation from planned start date / <i>Perspective:</i> Reliability <i>Description:</i> This measures how much the start of a tasks deviates from the planned start date. |
| () | 15. Number of re-plannings / <i>Perspective:</i> Reliability <i>Description:</i> Indicates the reliability of plannings , based upon the informations/commitments of the VO Members . If a certain period should be analysed, this can be set through the validity dates of the indicator. |
| () | 17. Reliability of VO member's statements / predictions <i>Perspective:</i> Reliability <i>Description:</i> Indicates the reliability of statements/predictions of a VO member , e.g. regarding the predicted on-time delivery or predictions if budgets are kept. |

Consulta 4:

Contexto: O usuário é um *broker* e participa de um **ambiente de criação de organizações virtuais**. O processo atual é a **criação de uma nova organização virtual (VO creation)**, sendo a tarefa em execução o **planejamento (planning)**.

Necessidade do usuário: o usuário está interessado em indicadores relacionados à **qualidade**, ou seja:

Consulta: “indicadores que considerem **qualidade (quality)**”

Resultados:

| | |
|-----|---|
| () | <p>3. Demand/Supply Planning Costs <i>Perspective:</i> Cost <i>Description:</i> Costs associated with forecasting, developing finished goods or end item inventory plans, and coordinating Demand/Supply process across entire supply chain, including all channels.</p> |
| () | <p>4. Incoming Production Material Quality <i>Perspective:</i> Quality <i>Description:</i> It measures the quality of raw material to ensure the correctness of final product. It is calculated by number of received parts which fail inspection divided by the total Number of parts received for production sector.</p> |
| () | <p>12. Percentage of Faultless production <i>Perspective:</i> Quality <i>Description:</i> Number of products produced without error in the manufacturing sector. An example of potential defects that can decrease the quality is lack of precision in the manufacture process.</p> |
| () | <p>14. Deviation from planned start date <i>Perspective:</i> Reliability <i>Description:</i> This measures how much the start of a tasks deviates from the planned start date.</p> |
| () | <p>15. Number of re-plannings <i>Perspective:</i> Reliability <i>Description:</i> Indicates the reliability of plannings, based upon the informations/commitments of the VO Members. If a certain period should be analysed, this can be set throug the validity dates of the indicator.</p> |
| () | <p>16. Problem Compensation performance of a VO member <i>Perspective:</i> Flexibility <i>Description:</i> Describes how much (%) of the deviation caused by previous partners is compensated. A negative value indicates that the deviation has become bigger. Metrics could be indicators for deviation already measured.</p> |
| () | <p>19. Satisfaction with coollaboration performance of a VO member <i>Perspective:</i> Quality <i>Description:</i> Indicates the satisfaction of an asked person/organisation with the collaboration performance of a certain VO member. Usually a scale is provided for the answer (e.g. from 1 = very satisfied to 6 = completely dissatisfied).</p> |
| () | <p>20. Number of call backs as Percentage of total inquiries <i>Perspective:</i> Collaboration <i>Description:</i> Number of callbacks divided by total inquiries.</p> |

Cenário 2: Busca por Documentos

Ao contrário da busca por indicadores, este cenário apresenta tipo de busca mais geral. Neste cenário, a busca pode ser estruturada com base em diversos tipos de entidades e as relações entre si:

- Organizações;
- Redes de organizações;
- Competências;
- Processos;
- Produtos;
- Recursos;
- Tarefas.

Além disso, a busca envolve uma variedade de tipos de documentos disponíveis no ambiente de uma rede colaborativa, incluindo: descrições de processos de negócio, manuais, tutoriais para *software*, contratos, entre outros.

Consulta 1:

Contexto: O usuário é o administrador de um **ambiente de criação de organizações virtuais** (ACV, ou **VBE**, em inglês) chamado “Helice”. O processo atual é a **criação de uma nova organização virtual** (*VO creation*).

Necessidade do usuário: o usuário deseja analisar os recursos disponíveis nesse ambiente. Um recurso pode ser humano, físico (máquinas, por exemplo) ou tecnológico (*software*).

Consulta: “**recursos pertencentes ao ACV Helice**”

Resultados:

| | |
|-----|---|
| () | <p>1003. ECOLEAD MATERIAL PLANNED TO BE USED</p> <p>Following is presented the list of ECOLEAD tools and to be used in demonstration activities for the IECOS VBE:</p> <ul style="list-style-type: none"> - Membership and Structure Management Systems (MSMS) - Profiling and Competency Management System (PCMS) - Ontology and Discovery Management System (ODMS) - Trust Management System (TrustMan) - Virtual Organization Information Management System (VIMS) - Collaboration Opportunity Characterization and Rough Planning (COC-Plan) - Partners Search and Suggestion (PSS) - Decision Support Management System (DSS) |
| () | <p>1048. Business process 1: Characterization of the HELICE VBE and the CeBeNetwork VBE</p> <p>The goal is to design the organizational and technical framework to support the operation of the VBE according the ECOLEAD VBE Ontology. In parallel, a structure for the cooperation of Helice and CeBeNetwork will be analyzed.</p> <p>VBE Management Software used:</p> <ul style="list-style-type: none"> - Membership and Structure Management Systems (MSMS) - Profiling and Competency Management System (PCMS) - Trust Management System (TrustMan) |
| () | <p>1049. Business process 2: VO Creation in HELICE VBE</p> <p>The goal is to develop and implement the required infrastructures to support a single company to identify collaboration opportunities with other companies to perform larger packages than those that can be performed alone.</p> <p>Software Used in this Process</p> <ul style="list-style-type: none"> - Collaboration Opportunity Characterization and Rough Planning (COC-Plan) - Partners Search and Suggestion (PSS) |
| () | <p>1050. Business process 3: VO Launch in HELICE</p> <p>The objective of the VO launch process is to support the firms within the cluster to produce joint commercial offers and at the same time to guarantee the working activities of the firms. The goal is to optimize the effort of creating a project agreement and the related working mechanisms when a business opportunity is identified.</p> <p>Software tools used:</p> <ul style="list-style-type: none"> - Partners Search & Suggestion Tool (PSS) and Negotiation Wizard Tool (WizAN). |

Consulta 2:

Contexto: O usuário é o administrador de um **ambiente de criação de organizações virtuais** (ACV, ou **VBE**, em inglês) chamado “IECOS”, executando um processo de **gestão** (*VBE management*).

Necessidade do usuário: o usuário busca as **organizações** que participam da **organização virtual** chamada “OIN”.

Consulta: “**organizações** participantes da **OV OIN**”

Resultados:

| | |
|-----|---|
| () | <p>1027. VO: Virtuelle Fabrik</p> <p>Scenario description: The VO Process Supervision & Performance Measurement is a definition of VO performance, its measurement processes and their relation to the supervision of the VO processes.</p> <p>Demonstrated Key Results: KR3.1 – VO Process Supervision & Performance Measurement</p> |
| () | <p>1030. OIN Partners</p> <p>The ORONA INNOVATION NETWORK (OIN) can be considered as a Virtual Breeding Environment (VBE) consisting of a number of positions or nodes, occupied by individuals, firms business units, universities, customers or other actors and links and interactions between these nodes. It is composed by the following members: - Orona, EIC, IKERLAN, MFE, ELECTRA, MONDRAGON, Herreros, ZABALA.</p> |
| () | <p>1031. ORONA</p> <p>Spanish leading company in the vertical transport industry. Is the leader of the Network. The main interest of ORONA in the VO-OIN is based on the fact that the network leads the development of the New Products and technologies for ORONA, as this is one of the pillars of its strategic plan.</p> |
| () | <p>1032. EIC</p> <p>ORONA eic works on product development for the lift and elevator sector. Given that the mission of EIC is to provide solutions and knowledge to ORONA S. Coop., the VO-OIN is the core fabric to EIC's mission, especially in the ambit of technology capture and the identification of new external partners.</p> |
| () | <p>1034. ELECTRA</p> <p>ELECTRA Vitoria Group has managed to become an international benchmark in the development of vertical transport systems. ELECTRA's interest in the VO-OIN is the same as for ORONA, with a much more incipient character, furthermore, its participation in the VO-OIN could help its mission to narrow the gap and become closer to ORONA thus forming part of the MCC Industrial Group.</p> |
| () | <p>1035. MONDRAGON CORPORACION COOPERATIVA (MCC)</p> <p>MCC is the largest business corporation in the Basque Country and the seventh largest in SPAIN, organized in three sectorial groups: Financial, Industrial and Distribution, together with the Research and Training areas.</p> |

Consulta 3:

Contexto: O usuário é o administrador de um ACV chamado “CeBeNetwork”, executando um processo de **gestão** (*VBE management*).

Necessidade do usuário: o usuário deseja saber mais a respeito de ACVs na área de engenharia **aeroespacial**.

Consulta: “ACVs que tenham competência na área **aeroespacial**”

Resultados:

| | |
|-----|--|
| () | <p>1001. IECOS SUMMARY: GLOBAL USE AND RESULTS</p> <p>The taken up of ECOLEAD results in IECOS carried the possibility of achieving important changes not only in its operational process but also in its business model and strategy. Through the utilization and implementation of ECOLEAD framework, IECOS formalized and implemented a new business strategy based on VBE model, where it also takes advantage of new business opportunities and market niches by establishing a new complete business unit reproducing the IECOS-VBE model. Significant improvements in the day-to-day operations at IECOS were possible to be achieved through the implementation of ECOLEAD tools and methodologies.</p> |
| () | <p>1038. ISOIN / HELICE.NET</p> <p>Organizations in the aeronautic sector currently operate under an Extended Enterprise model, with common supporting ICT infrastructure, derived from an ERP model (SAPECMA and SAPORTAL), together with methodologies, services, and tools to facilitate the delivery of supplied parts. Tools developed in ECOLEAD will be inserted in the existing ICT infrastructure managed by HELICE.NET to test and validate the VBE formation and VO creation and launch.</p> <p>Three business processes will be implemented, from the formation of a VBE from the existing cluster to the creation and launch of resulting VOs. The pilots will be run in the real life environment of four members of HELICE.NET: MEUPE, INESPASA, AEROSUR and GHESA. The required implementation plan is assisted by ISOIN, the core technological partner of HELICE.</p> |
| () | <p>1040. CeBeNetwork</p> <p>CeBeNetwork VBE was founded in 1996 and has grown rapidly since with a turnover of more than 30M€ and about 200 in 2005. It provides the aeronautical industry with engineering services where AIRBUS is the main customer. Typically for the aeronautical industry, also CeBeNetwork customers reduced the number of suppliers starting in 2003. The remaining suppliers were claimed to deliver the whole spectrum of engineering services necessary for the development of aircrafts. Furthermore the request was raised to be present at various customer sites throughout Europe. CeBeNetwork needed to manage this challenge to stay in this field and become one of the key suppliers. But this task could not be handled successfully by CeBeNetwork alone. It required the accomplishment of motivated partners, therefore CeBeNetwork have been collaborating with others in the past few years.</p> |

Apêndice H

Respostas do Questionário

As tabelas a seguir mostram os resultados esperados de cada consulta. A coluna “Doc.” contém os identificadores de documentos e as colunas de “A” a “E” indicam as respostas dos cinco especialistas entrevistados. Finalmente, a coluna “Sel.” contém os documentos que foram selecionados conforme o critério da maioria (3 ou mais).

| (a) Consulta 1. | | | | | | | (b) Consulta 2. | | | | | | |
|-----------------|---|---|---|---|---|------|-----------------|---|---|---|---|---|------|
| Doc. | A | B | C | D | E | Sel. | Doc. | A | B | C | D | E | Sel. |
| 2 | X | X | X | X | X | X | 2 | | | | X | | |
| 5 | X | | | X | | | 6 | | | X | X | | |
| 6 | | | | | | | 10 | X | | X | | | |
| 8 | X | | X | | | | 11 | X | | X | | | |
| 12 | | | | | | | 12 | X | X | X | X | X | X |
| 13 | | | | | | | 13 | | | | X | | |
| 15 | | | | | | | 14 | | | | | | |
| 18 | | | | | | | 15 | X | | X | X | | X |

Tabela H.1: Documentos relevantes para as consultas 1 e 2.

| (a) consulta 3. | | | | | | | (b) Consulta 4. | | | | | | |
|-----------------|---|---|---|---|---|------|-----------------|---|---|---|---|---|------|
| Doc. | A | B | C | D | E | Sel. | Doc. | A | B | C | D | E | Sel. |
| 1 | X | | X | X | X | X | 3 | | | | | | |
| 5 | | | | | | | 4 | X | | | X | | |
| 10 | X | | | X | | | 12 | X | X | | X | X | X |
| 11 | | | | X | X | | 14 | X | | X | X | | X |
| 12 | X | | X | X | | X | 15 | | | X | | | |
| 14 | X | X | | X | | X | 16 | | | X | X | | |
| 15 | X | X | X | X | X | X | 19 | X | X | X | X | | X |
| 17 | X | | X | | | | 20 | X | | | | | |

Tabela H.2: Documentos relevantes para as consultas 3 e 4.

| (a) consulta 5. | | | | | | | (b) Consulta 6. | | | | | | |
|-----------------|---|---|---|---|---|------|-----------------|---|---|---|---|---|------|
| Doc. | A | B | C | D | E | Sel. | Doc. | A | B | C | D | E | Sel. |
| 1003 | | | X | X | | | 1027 | | | | | | |
| 1048 | X | | | X | X | X | 1030 | X | X | X | X | X | X |
| 1049 | X | X | X | X | | X | 1031 | | | X | X | | |
| 1050 | X | | | X | X | X | 1032 | X | | X | X | X | X |
| | | | | | | | 1034 | X | | X | | X | X |
| | | | | | | | 1035 | X | | | | | |

Tabela H.3: Documentos relevantes para as consultas 5 e 6.

| Doc. | A | B | C | D | E | Sel. |
|------|---|---|---|---|---|------|
| 1001 | | | | | | |
| 1038 | X | X | X | | X | X |
| 1040 | X | | X | | X | X |

Tabela H.4: Documentos relevantes para a consulta 7.

Apêndice I

Produção Bibliográfica

Segue abaixo a lista de publicações científicas produzidas como resultado do trabalho apresentado nesta tese:

I.1 Artigos publicados em congressos internacionais

1. Tramontin Jr., R. J. e Rabelo, R. J.; 2007. A knowledge search framework for collaborative networks. Em: *Establishing The Foundation of Collaborative Networks, IFIP TC 5 Working Group 5.5 Eighth IFIP Working Conference on Virtual Enterprises*, September 10-12, 2007, Guimarães, Portugal, IFIP, páginas 573–582. Springer.
2. Tramontin Jr., R. J., Hanachi, C., e Rabelo, R. J.; 2008. A rule-based approach for customizing knowledge search in cnos. Em: *Pervasive Collaborative Networks, IFIP TC 5 WG 5.5 Ninth Working Conference on Virtual Enterprises*, September 8-10, 2008, Poznan, Poland, IFIP, páginas 243–252. Springer.
3. Tramontin Jr., R. J., Rabelo, R. J., e Hanachi, C.; 2009. Using semantics and context for searching in networked organizations. Em: *Proceedings of I2TS'2009 - 8th International Information and Telecommunication Technologies Symposium*, December 9-11, 2009, Florianópolis, Brasil, páginas 17-24. Fundação Barddal de Educação e Cultura.

I.2 Artigos publicados em revistas internacionais

1. Tramontin Jr., R. J., Rabelo, R. J., e Hanachi, C.; 2010. Customizing knowledge search in collaborative networked organisations through context-based query expansion. *Production Planning and Control*, 21:229–246.

Referências Bibliográficas

- Abrol, M., Latache, N., Mahadevan, U., Mao, J., Mukherjee, R., Raghavan, P., Tourn, M., Wang, J., e Zhang, G.; 2001. Navigating large-scale semi-structured data in business portals. Em *Proceedings of the 27th International Conference on Very Large Data Bases, VLDB '01*, páginas 663–666, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Afsarmanesh, H. e Camarinha-Matos, L. M.; 2005. A framework for management of virtual organization breeding environments. Em Camarinha-Matos (2005), páginas 35–48.
- Al-Khalifa, H. S.; 2007. *Automatic Document-level Semantic Metadata Annotation using Folksonomies and Domain Ontologies*. Tese (phd, electronics and computer science), Faculty of Engineering and Applied Science, School of Electronics and Computer Science, University of Southampton, Southampton, United Kingdom.
- Almeida, A., Marreiros, G., e Martins, C.; 2007. Collaboration and adaptation in scheduling. Em Camarinha-Matos et al. (2007a), páginas 611–618.
- Baeza-Yates, R. e Ribeiro-Neto, B., editores; 1999. *Modern Information Retrieval*. Addison Wesley, New York.
- Bahle, D., Williams, H. E., e Zobel, J.; 2002. Efficient phrase querying with an auxiliary index. Em *SIGIR '02: Proceedings of the 25th annual international ACM SIGIR conference on Research and development in information retrieval*, páginas 215–221, New York, NY, USA. ACM Press.
- Bai, J. e Nie, J.-Y.; 2008. Adapting information retrieval to query contexts. *Inf. Process. Manage.*, 44:1901–1922.
- Bai, J., Nie, J.-Y., Cao, G., e Bouchard, H.; 2007. Using query contexts in information retrieval. Em *Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval, SIGIR '07*, páginas 15–22, New York, NY, USA. ACM.

- Bailey, P., Craswell, N., Soboroff, I., e de Vries, A. P.; 2007. The csiro enterprise search test collection. *SIGIR Forum*, 41(2):42–45.
- Balabanović, M. e Shoham, Y.; 1997. Fab: content-based, collaborative recommendation. *Commun. ACM*, 40:66–72.
- Baldauf, M., Dustdar, S., e Rosenberg, F.; 2007. A survey on context-aware systems. *Int. J. Ad Hoc Ubiquitous Comput.*, 2:263–277.
- Baldo, F.; 2008. *Arcabouço para a Seleção de Indicadores de Desempenho para a Busca e Seleção de Parceiros para Organizações Virtuais*. Tese (doutorado em engenharia elétrica), Programa de Pós-Graduação em Engenharia Elétrica, Centro Tecnológico, Universidade Federal de Santa Catarina.
- Baldo, F., Rabelo, R. J., e Vallejos, R. V.; 2009. A framework for selecting performance indicators for virtual organisation partners' search and selection. *International Journal of Production Research*, 47:4737–4755.
- Barrows, R. e Traverso, J.; 2006. Search considered integral. *Queue*, 4(4):30–36.
- Bateman, J.; 2005. John bateman's ontology portal. <http://www.fb10.uni-bremen.de/anglistik/langpro/webpace/jb/info-pages/ontology/ontology-root.htm>, último acesso em: Outubro de 2010.
- Bayerl, P. S., Lüngen, H., Gut, U., e Paul, K. I.; 2003. Methodology for reliable schema development and evaluation of manual annotations. Em *Workshop on Knowledge Markup and Semantic Annotation at the Second International Conference on Knowledge Capture (K-CAP 2003)*, Sanibel, Florida, USA.
- Baziz, M., Boughanem, M., e Aussenac-Gilles, N.; 2005. Conceptual indexing based on document content representation. Em Crestani, F. e Ruthven, I., editores, *CoLLIS*, volume 3507 of *Lecture Notes in Computer Science*, páginas 171–186. Springer.
- Belkin, N. J.; 2008. Some(what) grand challenges for information retrieval. *SIGIR Forum*, 42:47–54.
- Belkin, N. J. e Croft, W. B.; 1992. Information filtering and information retrieval: two sides of the same coin? *Commun. ACM*, 35:29–38.
- Bergman, M. K.; 2011. AI3's Comprehensive Listing of Semantic Web and Related Tools. <http://www.mkbergman.com/sweet-tools/>, último acesso em: Fevereiro de 2011.

- Berners-Lee, T.; 2006. Notation 3 specification. <http://www.w3.org/DesignIssues/Notation3>, acesso: Julho de 2007.
- Berners-Lee, T., Hendler, J., e Lassila, O.; 2001. The Semantic Web. *Scientific American*, 284(5).
- Bhogal, J., Macfarlane, A., e Smith, P.; 2007. A review of ontology based query expansion. *Information Processing and Management*, 43:866–886.
- Bianchini, D., Montanelli, S., Aiello, C., Baldoni, R., Bolchini, C., Bonomi, S., Castano, S., Catarci, T., Antonellis, V. D., Ferrara, A., Melchiori, M., Quintarelli, E., Scannapieco, M., Schreiber, F. A., e Tanca, L.; 2010. Emergent semantics and cooperation in multi-knowledge communities: the esteem approach. *World Wide Web*, 13:3–31.
- Bohanec, M.; 2003. *Data Mining and Decision Support: Integration and Collaboration*, capítulo Decision Support, páginas 23–35. Kluwer Academi Publishers, The Netherlands.
- Bolchini, C., Curino, C. A., Quintarelli, E., Schreiber, F. A., e Tanca, L.; 2007. A data-oriented survey of context models. *SIGMOD Rec.*, 36(4):19–26.
- Bonino, D., Corno, F., e Farinetti, L.; 2003. Dose: a distributed open semantic elaboration platform. Em *ICTAI 2003, The 15th IEEE International Conference on Tools with Artificial Intelligence*.
- Bonino, D., Corno, F., Farinetti, L., e Bosca, A.; 2004. Ontology driven semantic search. *SIGIR Forum*, 1(6):1597–1605.
- Bontcheva, K., Cunningham, H., Kiryakov, A., e Tablan, V.; 2006. Semantic annotation and human language technology. Em Davies et al. (2006), páginas 29–50.
- Booch, G., Rumbaugh, J., e Jacobson, I.; 1999. *The Unified Modeling Language user guide*. Addison Wesley Longman Inc.
- Bouslimi, I., Hanachi, C., Tout, H., e Ghedira, K.; 2008. A coordination framework for cooperative information gathering. *Int. J. Adv. Intell. Paradigms*, 1:60–79.
- Brachman, R. e Schmolze, J.; 1985. An overview of the kl-one knowledge representation system. *Cognitive Science*, 9(2):171–216.
- Braschler, M. e Ripplinger, B.; 2004. How effective is stemming and decompounding for german text retrieval? *Information Retrieval*, 7(3-4):291–316.

- Bratt, S.; 2007. Semantic web, and other technologies to watch. Slides. [http://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/#\(24\)](http://www.w3.org/2007/Talks/0130-sb-W3CTechSemWeb/#(24)), último acesso em: Junho de 2011.
- Bray, T., Hollander, D., Layman, A., e Tobin, R.; 2006a. Namespaces in xml 1.0 (second edition). <http://www.w3.org/TR/REC-xml-names/>, acesso: Julho de 2007.
- Bray, T., Paoli, J., Sperberg-McQueen, C. M., Maler, E., e Yergeau, F.; 2006b. Extensible markup language (xml) 1.0 (fourth edition). <http://www.w3.org/TR/2006/REC-xml-20060816/>, acesso: Julho de 2007.
- Brickley, D. e Guha, R.; 2004. Rdf vocabulary description language 1.0: Rdf schema. <http://www.w3.org/TR/rdf-schema/>, acesso: Julho de 2007.
- Brin, S.; 1998. Extracting patterns and relations from the world wide web. Em *WebDB '98: Selected papers from the International Workshop on The World Wide Web and Databases*, páginas 172–183, London, UK. Springer-Verlag.
- Brin, S. e Page, L.; 1998. The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, 30(1–7):107–117.
- Broekstra, J., Kampman, A., e van Harmelen, F.; 2002. Sesame: A generic architecture for storing and querying rdf and rdf schema. Em Horrocks, I. e Hendler, J. A., editores, *International Semantic Web Conference*, volume 2342 of *Lecture Notes in Computer Science*, páginas 54–68. Springer.
- Camarinha-Matos, L. M., editor; 2005. *Collaborative Networks and Their Breeding Environments: IFIP TC 5 WG 5.5 Sixth IFIP Working Conference on Virtual Enterprises, 26-28 September 2005, Valencia, Spain*. Springer.
- Camarinha-Matos, L. M. e Afsarmanesh, H., editores; 1999a. *Infrastructures for Virtual Enterprises: Networking Industrial Enterprises, IFIP TC5 WG5.3 / PRODNET Working Conference on Infrastructures for Virtual Enterprises (PRO-VE '99), October 27-28, 1999, Porto, Portugal*, volume 153 of *IFIP Conference Proceedings*. Kluwer.
- Camarinha-Matos, L. M. e Afsarmanesh, H.; 1999b. The prodnet goals and approach. Em Camarinha-Matos e Afsarmanesh (1999a), páginas 97–108.
- Camarinha-Matos, L. M. e Afsarmanesh, H.; 1999c. The virtual enterprise concept. Em Camarinha-Matos e Afsarmanesh (1999a), páginas 3–14.
- Camarinha-Matos, L. M. e Afsarmanesh, H., editores; 2004a. *Collaborative Networked Organizations: A Research Agenda for Emerging Business Models*. Kluwer Academic Publishers, Norwell, MA, USA.

- Camarinha-Matos, L. M. e Afsarmanesh, H.; 2004b. The emerging discipline of collaborative networks. Em Camarinha-Matos, L. M., editor, *PRO-VE*, páginas 3–16. Kluwer.
- Camarinha-Matos, L. M. e Afsarmanesh, H.; 2004c. Support infrastructures for new collaborative forms. Em Camarinha-Matos e Afsarmanesh (2004a), páginas 175–192.
- Camarinha-Matos, L. M. e Afsarmanesh, H.; 2006. Collaborative networks: Value creation in a knowledge society. Em Wang, K., Kovács, G. L., Wozny, M. J., e Fang, M., editores, *PROLAMAT*, volume 207 of *IFIP*, páginas 26–40. Springer.
- Camarinha-Matos, L. M., Afsarmanesh, H., Novais, P., e Analide, C., editores; 2007a. *Establishing The Foundation of Collaborative Networks, IFIP TC 5 Working Group 5.5 Eighth IFIP Working Conference on Virtual Enterprises, September 10-12, 2007, Guimarães, Portugal*, volume 243 of *IFIP*. Springer.
- Camarinha-Matos, L. M., Afsarmanesh, H., e Ollus, M.; 2005. Ecolead: A holistic approach to creation and management of dynamic virtual organizations. Em Camarinha-Matos (2005), páginas 3–16.
- Camarinha-Matos, L. M., Afsarmanesh, H., e Ollus, M., editores; 2006. *Network-Centric Collaboration and Supporting Frameworks: IFIP TC 5 WG 5.5 Seventh IFIP Working Conference on Virtual Enterprises, 25-27 September 2006, Helsinki, Finland*, volume 224. Springer.
- Camarinha-Matos, L. M., Afsarmanesh, H., Pereira-Klen, A. A., e Cardoso, T.; 2007b. Deliverable d72.2 the training curriculum on cn (final). Relatório técnico, ECOLEAD Project.
- Camarinha-Matos, L. M., Baldo, F., Oliveira, A. I., Ratti, R., Demsar, D., e Jarimo, T.; 2007c. A Computer-Assisted VO Creation Framework. Em Camarinha-Matos et al. (2007a).
- Camarinha-Matos, L. M., Tschammer, V., e Afsarmanesh, H.; 2004. *Collaborative Networked Organizations: A Research Agenda for Emerging Business Models*, capítulo On Emerging Technologies for VO, páginas 207–224. Em Camarinha-Matos e Afsarmanesh (2004a).
- Castells, P., Fernandez, M., e Vallet, D.; 2007. An adaptation of the vector-space model for ontology-based information retrieval. *IEEE Trans. on Knowl. and Data Eng.*, 19:261–272.
- Cervo, A. L. e Bervian, P. A.; 2002. *Metodologia Científica*. São Paulo: Prentice Hall., 5a. edição.

- Chanana, V., Ginige, A., e Murugesan, S.; 2004. Improving information retrieval effectiveness by assigning context to documents. Em *Proceedings of the 2004 international symposium on Information and communication technologies*, ISICT '04, páginas 86–91. Trinity College Dublin.
- Chituc, C. M. e Azevedo, A. L.; 2005. Multi-perspective challenges on collaborative networks business environments. Em *Camarinha-Matos (2005)*, páginas 25–32.
- Cimiano, P., Handschuh, S., e Staab, S.; 2004. Towards the self-annotating web. Em *WWW '04: Proceedings of the 13th international conference on World Wide Web*, páginas 462–471, New York, NY, USA. ACM Press.
- COIN; 2008. Coin – enterprise collaboration and interoperability. COIN web page. <http://www.coin-ip.eu/>, último acesso em: Fevereiro de 2011.
- Contreras, J., Benjamins, V. R., Blázquez, M., Losada, S., Salla, R., Sevilla, J., Navarro, D., Casillas, J., Mompó, A., Patón, D., Corcho, Ó., Tena, P., e Martos, I.; 2004. A semantic portal for the international affairs sector. Em Motta, E., Shadbolt, N., Stutt, A., e Gibbins, N., editores, *EKAW*, volume 3257 of *Lecture Notes in Computer Science*, páginas 203–215. Springer.
- Crave, S., Bouron, T., e Ladame, S.; 2006. Using social capital as a conceptual framework for professional virtual communities formalization. Em *Camarinha-Matos et al. (2006)*, páginas 371–378.
- Croft, W. B., Turtle, H. R., e Lewis, D. D.; 1991. The use of phrases and structured queries in information retrieval. Em *SIGIR '91: Proceedings of the 14th annual international ACM SIGIR conference on Research and development in information retrieval*, páginas 32–45, New York, NY, USA. ACM Press.
- Cunningham, H.; 2006. Information Extraction, Automatic. Preprint, 18th November 2004, at <http://gate.ac.uk/sale/ell2/ie/main.pdf>. *Encyclopedia of Language and Linguistics, 2nd Edition*, Elsevier, 5:665–677.
- Cunningham, H., Maynard, D., Bontcheva, K., e Tablan, V.; 2002. GATE: A Framework and Graphical Development Environment for Robust NLP Tools and Applications. Em *Proc. of the 40th Anniversary Meeting of the Association for Computational Linguistics (ACL'02)*, páginas 168–175, Philadelphia.
- da Silva, E. L. e Menezes, E. M.; 2005. *Metodologia da Pesquisa e Elaboração de Dissertação*. UFSC, Florianópolis, 4a. edição.

- D'Agostini, C. S., Fileto, R., Dantas, M. A. R., e Gauthier, F. A. O.; 2008. Contextual semantic search - capturing, using the user's context to direct semantic search. Em Cordeiro, J. e Filipe, J., editores, *ICEIS (4)*, páginas 154–159.
- DAMASCOS; 2002. Damascos – dynamic forecast for master production planning with stock and capacity constraints. DAMASCOS web page. <http://www.inescporto.pt/~damascos/>, último acesso em: Maio de 2006.
- Davies, J., Studer, R., e Warren, P., editores; 2006. *Semantic Web Technologies: Trends and Research in Ontology-based Systems*. John Wiley & Sons.
- de Bruijn, J., Ehrig, M., Feier, C., Martín-Recuerda, F., Scharffe, F., e Weiten, M.; 2006. Ontology mediation, merging, and aligning. Em Davies et al. (2006), páginas 95–113.
- de Freitas, F. L. G.; 2007. Ontologias e a web semântica. Tutorial. <http://www.inf.unisinos.br/~renata/cursos/topicosv/ontologias-ws.pdf>, último acesso em: Junho de 2007.
- Decker, S., Melnik, S., van Harmelen, F., Fensel, D., Klein, M. C. A., Broekstra, J., Erdmann, M., e Horrocks, I.; 2000a. The semantic web: The roles of XML and RDF. *IEEE Internet Computing*, 4(5):63–74.
- Decker, S., Mitra, P., e Melnik, S.; 2000b. Framework for the semantic web: An rdf tutorial. *IEEE Internet Computing*, 4(6):68–73.
- Deerwester, S. C., Dumais, S. T., Landauer, T. K., Furnas, G. W., e Harshman, R. A.; 1990. Indexing by latent semantic analysis. *J. Am. Soc. Information Science*, 41(6):391–407.
- Demartini, G.; 2007. Leveraging semantic technologies for enterprise search. Em *PIKM '07: Proceedings of the ACM first Ph.D. workshop in CIKM*, páginas 25–32, New York, NY, USA. ACM.
- Dey, A. K.; 2001. Understanding and using context. *Personal Ubiquitous Computing*, 5(1):4–7.
- Dill, S., Eiron, N., Gibson, D., Gruhl, D., Guha, R., Jhingran, A., Kanungo, T., Mccurley, K. S., Rajagopalan, S., Tomkins, A., Tomlin, J. A., e Zien, J. Y.; 2003. A case for automated large scale semantic annotation. *Journal of Web Semantics*, 1(1):115–132.
- DOSE; 2007. Dose – distributed open semantic platform. <http://dose.sourceforge.net/index.html>, acesso: Julho de 2007.

- Downey, D., Schoenmackers, S., e Etzioni, O.; 2007. Sparse information extraction: Unsupervised language models to the rescue. Em *Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics*, páginas 696–703.
- Du, H. S. e Wagner, C.; 2005. Learning with weblogs: An empirical investigation. Em *HICSS*. IEEE Computer Society.
- Duke, A., Glover, T., e Davies, J.; 2007. Squirrel: An advanced semantic search and browse facility. Em Franconi, E., Kifer, M., e May, W., editores, *ESWC*, volume 4519 of *Lecture Notes in Computer Science*, páginas 341–355. Springer.
- ECOLEAD; 2004. Ecolead – european collaborative networked organization leadership initiative. ECOLEAD web page. <http://www.ecolead.org>, último acesso em: Maio de 2006.
- Efthimiadis, E. N.; 1996. Query expansion. *Annual Review of Information Systems and Technology (ARIST)*, 31:121–187.
- Ehrig, M., Haase, P., Hefke, M., e Stojanovic, N.; 2005. Similarity for ontologies - a comprehensive framework. Em *ECIS*.
- Ermilova, E. e Afsarmanesh, H.; 2006. Competency and profiling management in virtual organizations breeding environments. Em Camarinha-Matos et al. (2006), páginas 131–142.
- Ermilova, E. e Afsarmanesh, H.; 2008. A unified ontology for vo breeding environments. Em *Proceedings of IEEE International Conference on Distributed Human-Machine Systems (DHMS'08)*, páginas 176–181.
- Fagin, R., Kumar, R., McCurley, K. S., Novak, J., Sivakumar, D., Tomlin, J. A., e Williamson, D. P.; 2003. Searching the workplace web. Em *Proceedings of the 12th international conference on World Wide Web, WWW '03*, páginas 366–375, New York, NY, USA. ACM.
- Fahy, P. e Clarke, S.; 2004. Cass - a middleware for mobile context-aware applications. Em *Workshop on Context Awareness, MobiSys*.
- Fensel, D., Sycara, K. P., e Mylopoulos, J., editores; 2003. *The Semantic Web - ISWC 2003, Second International Semantic Web Conference, Sanibel Island, FL, USA, October 20-23, 2003, Proceedings*, volume 2870 of *Lecture Notes in Computer Science*. Springer.
- Ferraiolo, J., Jun, F., e Jackson, D.; 2009. Scalable vector graphics (svg) 1.1 specification. <http://www.w3.org/TR/2003/REC-SVG11-20030114/>, acesso: Janeiro de 2011.

- Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., e Ruppin, E.; 2001. Placing search in context: the concept revisited. Em *Proceedings of the 10th international conference on World Wide Web*, WWW '01, páginas 406–414, New York, NY, USA. ACM.
- Finkelstein, L., Gabrilovich, E., Matias, Y., Rivlin, E., Solan, Z., Wolfman, G., e Ruppin, E.; 2002. Placing search in context: the concept revisited. *ACM Trans. Inf. Syst.*, 20:116–131.
- Forgy, C.; 1982. Rete: A fast algorithm for the many pattern/many object pattern match problem. *Artificial Intelligence*, 19(1):17–37.
- Fox, E. A.; 1983. *Extending the boolean and vector space models of information retrieval with p-norm queries and multiple concept types*. Tese (phd, computer science), Dept. of Computer Science, Cornell University, Ithaca, N.Y.
- Friedman-Hill, E. J.; 2007. The rete algorithm. Sandia National Laboratories. <http://www.jessrules.com/jess/docs/71/rete.html>, último acesso em: Dezembro de 2008.
- Fu, L., Goh, D. H.-L., e Foo, S. S.-B.; 2005. Cqe: a collaborative querying environment. Em *Proceedings of the 5th ACM/IEEE-CS joint conference on Digital libraries*, JCDL '05, páginas 378–378, New York, NY, USA. ACM.
- Gaizauskas, R. e Wilks, Y.; 1998. Information extraction: Beyond document retrieval. *Journal of Documentation*, 54(1):70–105.
- Gandon, F. L.; 2007. Owl inference engine using xslt and jess. Mobile Commerce Laboratory - School of Computer Science - CMU. <http://www.cs.cmu.edu/~sadeh/MyCampusMirror/OWLEngine.html>, último acesso em: Dezembro de 2008.
- GATE; 2007. Gate home. <http://www.gate.ac.uk/>, acesso: Julho de 2007.
- Gensel, J., Villanova-Oliver, M., e Kirsch-Pinheiro, M.; 2008. Modèles de contexte pour l'adaptation à l'utilisateur dans des systèmes d'information web collaboratifs. Em *Workshop das 8èmes journées francophones*, páginas 5–16.
- Glasgow; 2007. Collections held at glasgow. http://ir.dcs.gla.ac.uk/resources/test_collections/, acesso: Julho de 2007.

- Gonzalez, M., de Lima, V. L. S., e de Lima, J. V.; 2006. Termos, relacionamentos e representatividade na indexação de texto para recuperação de informação. *Letras de Hoje*, 41(2):65–87.
- Gonzalo, J., Verdejo, F., Chugur, I., e Cigarrán, J. M.; 1998. Indexing with wordnet synsets can improve text retrieval. *CoRR*, cmp-lg/9808002.
- Goranson, H. T.; 1999. *The Agile Virtual Enterprise: cases, metrics, tools*. Quorum Books, USA.
- Graser, F., Jansson, K., Eschenbächer, J., Westphal, I., e Negroto, U.; 2005. Towards performance measurement in virtual organizations: Potentials, needs, and research challenges. Em *Camarinha-Matos (2005)*, páginas 301–310.
- Grobelnik, M. e Mladenic, D.; 2006. Knowledge discovery for ontology construction. Em *Davies et al. (2006)*, páginas 9–27.
- Gross, T. e Klemke, R.; 2002. Context modelling for information retrieval - requirements and approaches. Em *Proceedings of the IADIS International Conference WWW/Internet, ICWI 2002*, páginas 247–254.
- Guha, R., McCool, R., e Miller, E.; 2003. Semantic search. Em *Proceedings of the 12th international conference on World Wide Web, WWW '03*, páginas 700–709, New York, NY, USA. ACM.
- Hawking, D.; 2004. Challenges in enterprise search. Em *Proceedings of the 15th Australasian database conference - Volume 27, ADC '04*, páginas 15–24, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.
- Hawking, D., Paris, C., Wilkinson, R., e Wu, M.; 2005. Context in enterprise search and delivery. Em *Proceedings of ACM SIGIR 2005 Workshop on Information Retrieval in Context (IRiX)*, páginas 14–16. Royal School of Library and Information Science, Copenhagen. <http://david-hawking.net/pubs/sigir05-context-wkshp-3.pdf>.
- Heflin, J. e Hendler, J. A.; 2000. Dynamic ontologies on the web. Em *Proceedings of the Seventeenth National Conference on Artificial Intelligence and Twelfth Conference on Innovative Applications of Artificial Intelligence*, páginas 443–449. AAAI Press / The MIT Press.
- Helaakoski, H., Iskanius, P., e Peltomaa, I.; 2007. Agent-based architecture for virtual enterprises to support agility. Em *Camarinha-Matos et al. (2007a)*, páginas 299–306.
- Jess; 2008. Jess, the rule engine for the java platform. Jess web page. <http://www.jessrules.com/jess/>, último acesso em: Fevereiro de 2009.

- Jones, R., Rey, B., Madani, O., e Greiner, W.; 2006. Generating query substitutions. Em *Proceedings of the World Wide Web Conference (WWW 2006)*. ACM.
- Jones, S. e Staveley, M. S.; 1999. Phrasier: a system for interactive document retrieval using keyphrases. Em *SIGIR '99: Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, páginas 160–167, New York, NY, USA. ACM Press.
- Järvelin, K., Kekäläinen, J., e Niemi, T.; 2001. Expansiontool: Concept-based query expansion and construction. *Information Retrieval*, 4:231–255. 10.1023/A:1011998222190.
- Käki, M.; 2005. Findex: search result categories help users when document ranking fails. Em *Proceedings of the SIGCHI conference on Human factors in computing systems, CHI '05*, páginas 131–140, New York, NY, USA. ACM.
- Kalfoglou, Y. e Schorlemmer, M.; 2003. Ontology mapping: the state of the art. *Knowledge Engineering Review*, 18(1):1–31.
- Kalina Bontcheva, John Davies, A. D. T. G. N. K. e Thurlow, I.; 2006. Semantic information access. Em Davies et al. (2006), páginas 139–170.
- Karvounarakis, G., Alexaki, S., Christophides, V., Plexousakis, D., e Scholl, M.; 2002. Rql: a declarative query language for rdf. Em *Proceedings of the 11th international conference on World Wide Web, WWW '02*, páginas 592–603, New York, NY, USA. ACM.
- Köhler, J., Philippi, S., Specht, M., e Egg, A.; 2006. Ontology based text indexing and querying for the semantic web. *Know.-Based Syst.*, 19(8):744–754.
- Kietz, J.-U., Volz, R., e Maedche, A.; 2000. Extracting a domain-specific ontology from a corporate intranet. Em *Proceedings of the 2nd workshop on Learning language in logic and the 4th conference on Computational natural language learning*, páginas 167–175, Morristown, NJ, USA. Association for Computational Linguistics.
- Kifer, M. e Boley, H.; 2010. Rif overview. <http://www.w3.org/TR/2010/NOTE-rif-overview-20100622/>, acesso: Junho de 2011.
- KimDocs; 2010. Extending the kim ontology and knowledge base. <http://ontotext.com/kim/doc/KimDocs-3.0-EN/ExtendInformationExtraction.html>, acesso: Dezembro de 2010.

- Kirsch-Pinheiro, M., Gensel, J., e Martin, H.; 2004. Representing context for an adaptative awareness mechanism. Em *Groupware: Design, Implementation and Use*, volume 3198 of *Lecture Notes in Computer Science*, páginas 339–348. Springer.
- Kirsch-Pinheiro, M., Villanova-Oliver, M., Gensel, J., e Martin, H.; 2005. Context-aware filtering for collaborative web systems: adapting the awareness information to the user's context. Em *Proceedings of the 2005 ACM symposium on Applied computing, SAC '05*, páginas 1668–1673, New York, NY, USA. ACM.
- Kiryakov, A., Popov, B., Ognyanoff, D., Manov, D., Kirilov, A., e Goranov, M.; 2003. Semantic annotation, indexing, and retrieval. Em Fensel et al. (2003), páginas 484–499.
- Kiryakov, A., Popov, B., Terziev, I., Manov, D., e Ognyanoff, D.; 2004. Semantic annotation, indexing, and retrieval. *Journal of Web Semantics*, 2(1):49–79.
- Koivunen, M.-R. e Miller, E.; 2001. W3c semantic web activity. <http://www.w3.org/2001/12/semweb-fin/w3csw>, acesso: Julho de 2007.
- Kraft, R., Chang, C. C., Maghoul, F., e Kumar, R.; 2006. Searching with context. Em *Proceedings of the 15th international conference on World Wide Web, WWW '06*, páginas 477–486, New York, NY, USA. ACM.
- Loss, L.; 2007. *Um Arcabouço para o Aprendizado de Redes Colaborativas de Organizações: Uma Abordagem Baseada em Aprendizagem Organizacional e Gestão de Conhecimento*. Tese (doutorado em engenharia elétrica), Programa de Pós-Graduação em Engenharia Elétrica, Centro Tecnológico, Universidade Federal de Santa Catarina.
- Loss, L., Rabelo, R. J., e Pereira-Klen, A. A.; 2007. Towards learning collaborative networked organizations. Em Camarinha-Matos et al. (2007a).
- Lucene; 2007. Lucene Home. <http://lucene.apache.org/>, acesso: Julho de 2007.
- Maedche, A., Motik, B., Stojanovic, L., Studer, R., e Volz, R.; 2003a. Ontologies for enterprise knowledge management. *IEEE Intelligent Systems*, 18(2):26–33.
- Maedche, A. e Staab, S.; 2001. Ontology learning for the semantic web. *IEEE Intelligent Systems*, 16(2):72–79.

- Maedche, A., Staab, S., Stojanovic, N., Studer, R., e Sure, Y.; 2003b. Semantic portal: The seal approach. Em Fensel, D., Hendler, J. A., Lieberman, H., e Wahlster, W., editores, *Spinning the Semantic Web*, páginas 317–359. MIT Press.
- Mahesh, K., Kud, J., e Dixon, P.; 1999. Oracle at trec8: A lexical approach. Em *TREC*.
- Manola, F. e Miller, E.; 2004. Rdf primer. <http://www.w3.org/TR/rdf-primer/>, acesso: Julho de 2007.
- Mauldin, M. L.; 1991. *Conceptual Information Retrieval: A Case Study in Adaptive Partial Parsing*. Kluwer Academic Publishers, Norwell, MA, USA.
- Mayfield, J. e Finin, T.; 2003. Information retrieval on the Semantic Web: Integrating inference and retrieval. Em *Proceedings of the SIGIR Workshop on the Semantic Web at the 26th Int'l ACM SIGIR Conf. Research and Development in Information Retrieval*.
- Maynard, D.; 2003. Multi-source and multilingual information extraction. *Expert Update*. <http://gate.ac.uk/sale/expertupdate/muse.pdf>.
- McGuinness, D. L. e van Harmelen, F.; 2004. Owl web ontology language overview. <http://www.w3.org/TR/2004/REC-owl-features-20040210/>, acesso: Julho de 2007.
- Mihalcea, R. e Moldovan, D.; 2000. Semantic indexing using wordnet senses. Em *Proceedings of the ACL-2000 workshop on Recent advances in natural language processing and information retrieval*, páginas 35–45, Morristown, NJ, USA. Association for Computational Linguistics.
- Minsky, M.; 1975. A framework for representing knowledge. *The Psychology of Computer Vision*, páginas 211–281.
- Msanjila, S. S. e Afsarmanesh, H.; 2006. Assessment and creation of trust in vbcs. Em Camarinha-Matos et al. (2006), páginas 161–172.
- Mukherjee, R. e Mao, J.; 2004. Enterprise search: Tough stuff. *Queue*, 2:36–46.
- Müller, S., Kritzler, N., Tartakovski, A., Bergmann, R., e Traphöner, R.; 2006. Knowledge search within a company-wiki. Em Althoff, K.-D. e Schaaf, M., editores, *LWA*, volume 1/2006 of *Hildesheimer Informatik-Berichte*, páginas 209–214. University of Hildesheim, Institute of Computer Science.

- MyFashion.eu; 2004. MyFashion.eu – Added Value for the Consumer Through Integrated, Extended Fashion Products. IST - 2001-32560. Projeto apoiado pela Comissão Européia e CNPq. MyFashion.eu web page. <http://www.myfashion.org/>, último acesso em: Outubro de 2004.
- Mylonas, P., Vallet, D., Castells, P., Fernández, M., e Avrithis, Y.; 2008. Personalized information retrieval based on context and ontological knowledge. *Knowl. Eng. Rev.*, 23:73–100.
- Navigli, R. e Velardi, P.; 2003. An analysis of ontology-based query expansion strategies. Em *Proceedings of Workshop on Adaptive Text Extraction and Mining (ATEM) in the 14th European Conference on Machine Learning (ECML)*, páginas 42–49, Cavtat-Dubrovnik, Croatia.
- Navrat, P. e Taraba, T.; 2007. Context search. Em *Proceedings of the 2007 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IATW '07*, páginas 99–102, Washington, DC, USA. IEEE Computer Society.
- Nilsson, K., Hjelm, H., e Oxhammar, H.; 2005. Suis - cross-language ontology-driven information retrieval in a restricted domain. Em *Proceedings of the 15th NODALIDA conference*, páginas 139–145.
- Noy, N. F. e McGuinness, D. L.; 2001. Ontology development 101: A guide to creating your first ontology. Relatório Técnico KSL-01-05, Stanford Knowledge Systems Laboratory.
- ODE; 2011. Apache ODE). <http://ode.apache.org/>, acesso: Janeiro de 2011.
- OntologyMatching; 2006. Ontology matching. <http://ontologymatching.org/>, acesso: Dezembro de 2006.
- Paz-Trillo, C., Wassermanand, R., e Braga, P. P.; 2005. An information retrieval application using ontologies. *Journal of the Brazilian Computer Society*, 11(2):17–31.
- Pires, S. R. I.; 2004. *Gestão da Cadeia de Suprimentos (Supply Chain Management): conceitos, estratégias, práticas e casos*. Atlas S.A., São Paulo.
- Pitkow, J., Schütze, H., Cass, T., Cooley, R., Turnbull, D., Edmonds, A., Adar, E., e Breuel, T.; 2002. Personalized search. *Commun. ACM*, 45:50–55.
- Plisson, J. e Ljubic, P.; 2006. The cno ontology page. <http://ecolead.ijs.si/onto/cn.html>, acesso: Agosto de 2007.

- Popov, B., Kiryakov, A., Kirilov, A., Manov, D., Ognyanoff, D., e Goranov, M.; 2003. Kim - semantic annotation platform. Em Fensel et al. (2003), páginas 834–849.
- Pretschner, A. e Gauch, S.; 1999. Ontology based personalized search. Em *ICTAI*, páginas 391–398.
- PRODNET; 1999. Production planning and management in an extended enterprise. PRODNET II web page. <http://www.uninova.pt/~prodnet/>, último acesso em: Maio de 2006.
- Prud'hommeaux, E. e Seaborne, A.; 2008. Sparql query language for rdf. <http://www.w3.org/TR/rdf-sparql-query/>, acesso: Junho de 2010.
- Rabelo, R. J., Gusmeroli, S., Anara, C., e Nagellen, T.; 2006. The ecolead ict infrastructure for collaborative networked organizations. Em Camarinha-Matos et al. (2006), páginas 451–460.
- Rabelo, R. J., Rodrigo-Castro, M. M., Conconi, A., e Sesana, M.; 2008. The ecolead plug & play collaborative business infrastructure. Em Camarinha-Matos, L. M., Afsarmanesh, H., e Ollus, M., editores, *Methods and Tools for Collaborative Networked Organizations*, páginas 371–394. Springer.
- Raghavan, P.; 2001. Structured and unstructured search in enterprises. *IEEE Data Eng. Bull.*, 24(4):15–18.
- Ratprasartporn, N., Po, J., Cakmak, A., Bani-Ahmad, S., e Ozsoyoglu, G.; 2009. Context-based literature digital collection search. *The VLDB Journal*, 18:277–301.
- Ratti, R. e Rabelo, R. J.; 2007. Deliverable d61.1c ict-i reference framework (version 3). Relatório técnico, ECOLEAD Project.
- Redon, R., Larsson, A., Leblond, R., e Longueville, B.; 2007. Vivace context based search platform. Em *Proceedings of the 6th international and interdisciplinary conference on Modeling and using context*, CONTEXT'07, páginas 397–410, Berlin, Heidelberg. Springer-Verlag.
- Reeve, L. e Han, H.; 2005. Survey of semantic annotation platforms. Em *SAC '05: Proceedings of the 2005 ACM symposium on Applied computing*, páginas 1634–1638, New York, NY, USA. ACM Press.
- Renda, M. E. e Straccia, U.; 2005. A personalized collaborative digital library environment: a model and an application. *Inf. Process. Manage.*, 41:5–21.
- Rick, J. e Guzdial, M.; 2006. Situating coweb: a scholarship of application. *International Journal of Computer-Supported Collaborative Learning (ijCSCL)*, 1(1):89–115.

- Rieh, S. Y. e Xie, H.; 2001. Patterns and sequences of multiple query reformulations in web searching: A preliminary study. Em *Proceedings of the 64th ASIST Annual Meeting*, páginas 246–255.
- Riloff, E. e Lehnert, W.; 1994. Information extraction as a basis for high-precision text classification. *ACM Trans. Inf. Syst.*, 12(3):296–333.
- Rodrigo, M. M. e Ratti, R.; 2007. Deliverable d64.1c ict-i integrated prototype (version 3). Relatório técnico, ECOLEAD Project.
- Salton, G.; 1997. A blueprint for automatic indexing. *SIGIR Forum*, 31(1):23–36.
- Salton, G., Fox, E. A., e Wu, H.; 1983. Extended boolean information retrieval. *Commun. ACM*, 26(11):1022–1036.
- Salton, G., Singhal, A., Mitra, M., e Buckley, C.; 1997. Automatic text structuring and summarization. *Inf. Process. Manage.*, 33(2):193–207.
- Salton, G., Wong, A., e Yang, C. S.; 1975. A vector space model for automatic indexing. *Commun. ACM*, 18(11):613–620.
- Sanderson, M.; 1994. Word sense disambiguation and information retrieval. Em *SIGIR '94: Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*, páginas 142–151, New York, NY, USA. Springer-Verlag New York, Inc.
- Sanderson, M.; 2000. Retrieving with good sense. *Inf. Retr.*, 2(1):49–69.
- Seaborne, A.; 2004. Rdfql - a query language for rdf. <http://www.w3.org/Submission/RDQL/>, acesso: Junho de 2010.
- SEKT; 2006. Semantic knowledge technologies (sekt). SEKT web page. <http://www.sekt-project.com/>, acesso: Julho de 2007.
- Sesame; 2007. Openrdf.org ...home of sesame. <http://openrdf.org/>, acesso: Julho de 2007.
- Seymore, K., McCallum, A., e Rosenfeld, R.; 1999. Learning hidden Markov model structure for information extraction. Em *AAAI 99 Workshop on Machine Learning for Information Extraction*.
- Si, L. e Callan, J.; 2005. Modeling search engine effectiveness for federated search. Em *SIGIR '05: Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, páginas 83–90, New York, NY, USA. ACM Press.

- Silva, M. V. D.; 2010. *Um arcabouço de suporte à tomada de decisão colaborativa para o gerenciamento da evolução de empresas virtuais*. Tese (doutorado em engenharia elétrica), Programa de Pós-Graduação em Engenharia Elétrica, Centro Tecnológico, Universidade Federal de Santa Catarina.
- Singh, M. P. e Huhns, M. N.; 2005. *Service-Oriented Computing: Semantics, Processes, Agents*. John Wiley & Sons, New York, NY, EUA, 1a. edição.
- Sowa, G. e Sniezynski, T.; 2007. Deliverable d61.4 security framework and architecture. Relatório técnico, ECOLEAD Project.
- Sperberg-McQueen, C. M. e Burnard, L.; 1994. A gentle introduction to sgml. <http://www.isgmlug.org/sgmlhelp/g-index.htm>, acesso: Julho de 2007.
- Squiggle; 2007. Semantic search and conceptual indexing. <http://swa.cefriel.it/Squiggle>, acesso: Julho de 2007.
- Stojanovic, L., Maedche, A., Motik, B., e Stojanovic, N.; 2002. User-driven ontology evolution management. Em Gómez-Pérez, A. e Benjamins, V. R., editores, *EKAW*, volume 2473 of *Lecture Notes in Computer Science*, páginas 285–300. Springer.
- Stojanovic, N.; 2005. On the query refinement in the ontology-based searching for information. *Inf. Syst.*, 30(7):543–563.
- Studer, R., Benjamins, V. R., e Fensel, D.; 1998. Knowledge engineering: Principles and methods. *Data Knowl. Eng.*, 25(1-2):161–197.
- Tamine-Lechani, L., Boughanem, M., e Zemirli, N.; 2007. Exploiting multi-evidence from multiple user's interests to personalizing information retrieval. Em *ICDIM*, páginas 7–12. IEEE.
- Teevan, J., Dumais, S. T., e Horvitz, E.; 2005. Personalizing search via automated analysis of interests and activities. Em *Proceedings of the 28th annual international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '05, páginas 449–456, New York, NY, USA. ACM.
- Terziev, I., Kiryakov, A., e Manov, D.; 2005. SEKT project deliverable d1.8.1 - base upper-level ontology (bulo) guidance. Relatório técnico, SEKT Project.
- THINKcreative; 2003. Thinking network of experts on emerging smart organizations (thinkcreative). THINKcreative web page. <http://www.thinkcreative.org/>, último acesso em: Outubro de 2006.

- Thompson, H. S., Sperberg-McQueen, C. M., Gao, S., Mendelsohn, N., Beech, D., e Maloney, M.; 2006. Xml schema 1.1 part 1: Structures. <http://www.w3.org/TR/xmlschema11-1/>, acesso: Julho de 2007.
- TREC; 2004. Text retrieval conference (trec) overview. <http://trec.nist.gov/overview.html>, acesso: Julho de 2007.
- Uren, V., Cimiano, P., Iria, J., Handschuh, S., Vargas-Vera, M., Motta, E., e Ciravegna, F.; 2006. Semantic annotation for knowledge management: Requirements and a survey of the state of the art. *Journal of Web Semantics*, 4(1):14–28.
- Uschold, M.; 2003. Where are the semantics in the semantic web? *AI Magazine*, 24(3):25–36.
- Vaishnavi, V. e Kuechler, W.; 2004. Design research in information systems. January 20, 2004, last updated August 16, 2009. <http://desrist.org/design-research-in-information-systems>, último acesso em: Junho de 2011.
- Van Rijsbergen, C. J.; 1979. *Information Retrieval, 2nd edition*. Dept. of Computer Science, University of Glasgow.
- VOMap; 2004. VOMap Project: Roadmap Design for Collaborative Virtual Organizations in Dynamic Business Ecosystems. VOMap web page. <http://www.vomap.org/>, último acesso em: Outubro de 2006.
- Wang, J., Robertson, S., Vries, A. P., e Reinders, M. J.; 2008. Probabilistic relevance ranking for collaborative filtering. *Inf. Retr.*, 11:477–497.
- Williams, H. E., Zobel, J., e Bahle, D.; 2004. Fast phrase querying with combined indexes. *ACM Trans. Inf. Syst.*, 22(4):573–594.
- Woods, W. A.; 1975. What's in a link: Foundations for semantic networks. *Representation and Understanding Studies in Cognitive Science*.
- Xiang, B., Jiang, D., Pei, J., Sun, X., Chen, E., e Li, H.; 2010. Context-aware ranking in web search. Em *Proceeding of the 33rd international ACM SIGIR conference on Research and development in information retrieval, SIGIR '10*, páginas 451–458, New York, NY, USA. ACM.
- Yu, C., Meng, W., Wu, W., e Liu, K.-L.; 2001. Efficient and effective metasearch for text databases incorporating linkages among documents. *SIGMOD Rec.*, 30:187–198.
- Zara, O.; 2005. *Managing Collective Intelligence: Toward a New Corporate Governance*. Mm2 Editions.

Zhuhadar, L. e Nasraoui, O.; 2008. Semantic information retrieval for personalized e-learning. Em *Proceedings of the 2008 20th IEEE International Conference on Tools with Artificial Intelligence - Volume 01*, páginas 364–368, Washington, DC, USA. IEEE Computer Society.