

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE ENGENHARIA DA PRODUÇÃO**

ARILDO DIRCEU CORDEIRO

**GERADOR INTELIGENTE DE SISTEMAS COM
AUTO-APRENDIZAGEM PARA GESTÃO DE
INFORMAÇÕES E CONHECIMENTO**

**FLORIANÓPOLIS
2005**

ARILDO DIRCEU CORDEIRO

**GERADOR INTELIGENTE DE SISTEMAS COM AUTO-
APRENDIZAGEM PARA GESTÃO DE INFORMAÇÕES E
CONHECIMENTO**

Tese apresentada ao Programa de Pós-Graduação em Engenharia da Produção da Universidade Federal de Santa Catarina como requisito parcial para obtenção do título de Doutor em Engenharia de Produção.

Orientador: Prof. Rogério Cid Bastos, Dr.

**FLORIANÓPOLIS
2005**

ARILDO DIRCEU CORDEIRO

**GERADOR INTELIGENTE DE SISTEMAS COM AUTO-
APRENDIZAGEM PARA GESTÃO DE INFORMAÇÕES E
CONHECIMENTO**

Esta tese foi julgada e aprovada para a obtenção do Título de Doutor em **Engenharia da
Produção** no **Programa de Pós-graduação em Engenharia da Produção** da
Universidade Federal de Santa Catarina.

Florianópolis, 28 de março de 2005.

Prof. Dr. Edson Pacheco Paladini
Coordenador do Programa

BANCA EXAMINADORA

Prof. Rogério Cid Bastos, Dr.
Universidade Federal de Santa Catarina
Orientador

Prof. Fernando Álvaro Ostuni Gauthier, Dr.
Universidade Federal de Santa Catarina

Profa. Sílvia Modesto Nassar, Dra.
Universidade Federal de Santa Catarina
Mediadora

Prof. Luciano Scandelari, Dr.
Centro Federal de Educação Tecnológica
do Paraná

Profª. Isaura Alberton de Lima, Dra.
Centro Federal de Educação Tecnológica
Do Paraná

À memória de minha mãe – Araci – pela vida e pelos ensinamentos nos primeiros passos desta caminhada juntamente com meu pai, sempre orientando no caminho do bem e do justo.

Aos meus filhos – Arildo Júnior e Licyane, a toda a minha família e amigos com quem tenho a felicidade de compartilhar minha vida.

AGRADECIMENTOS

Uma tese de Doutorado requer esforço substancial de pesquisa e, acima de tudo, disciplina, força de vontade e ajuda de muitas pessoas, de forma que agradeço:

- ao Prof. Dr. Rogério Cid Bastos, meu orientador, por acreditar e pela contribuição e apoios precisos;
- ao Prof. Dr. Edson Pacheco Paladini, Coordenador do Programa de Engenharia da Produção, da Universidade Federal de Santa Catarina;
- aos Profs. Dra. Sílvia Modesto Nassar, Dr. Fernando Ostuni Gauthier, Dra. Isaura Alberton de Lima e Dr. Luciano Scandelari, por contribuírem no processo de socialização dos seus conhecimentos, conduzindo e lapidando os conceitos aqui apresentados;
- a todos que contribuíram com dados e informações, que se consubstanciaram nesta pesquisa em novos conhecimentos.
- a todos os amigos, pelo incentivo incessante e pela ajuda, principalmente nos momentos difíceis.

"A imaginação é mais importante que o conhecimento."

Albert Einstein

"A mente que se abre a uma nova idéia jamais voltará a seu tamanho original."

Albert Einstein

"Se A é o sucesso, então é igual a X mais Y mais Z. O trabalho é X; Y é o lazer; e Z é manter a boca fechada."

Albert Einstein

"Enquanto eu continuar a permanecer no presente, sou feliz por toda a vida: porque toda a vida é sempre o presente.

O presente é simplesmente quem eu sou, apenas como eu sou...agora. E é precioso."

Dr. Spencer Johnson

RESUMO

CORDEIRO, Arildo Dirceu. **Gerador Inteligente de Sistemas com Auto-Aprendizagem para Gestão de Informações e Conhecimento**. 2005. – 162 p. Tese (Doutorado em Engenharia da Produção) - Programa de Pós-Graduação em Engenharia da Produção, UFSC, Florianópolis, SC.

Este trabalho apresenta uma ferramenta que permite ao usuário gerar sistemas inteligentes. Inicialmente é mostrado um sistema Multi-Agentes denominado “SMAGS”, com capacidade de integrar sistemas informatizados diferentes e gerenciar, em tempo real, as informações inerentes ao projeto de interesse, podendo atuar também como um “Gerente Virtual”. Na seqüência são incorporadas novas técnicas de Inteligência Artificial (IA), com interpretação semântica e testes métricos de ontologias, com o objetivo de fornecer ao usuário uma ferramenta CASE (Computer Aided Software Engineering) com capacidade de gerar sistemas de informações informatizados inteligentes. Os sistemas gerados podem aprender a partir de informações oriundas de buscas inteligentes em servidores de ontologias, banco de casos, especialistas na área e mesmo dados fragmentados na *internet*, facilitando e potencializando as tomadas de decisões gerenciais.

Palavras-Chave:

Mineração de Dados e de Texto, Redes *Neuro-Fuzzy*, Web Semântica, Agentes Inteligentes, Ontologias e UNL (Universal Networking Language).

ABSTRACT

CORDEIRO, Arildo Dirceu. **Gerador Inteligente de Sistemas com Auto-Aprendizagem para Gestão de Informações e Conhecimento**. 2005. – Programa de Pós-Graduação em Engenharia da Produção, UFSC, Florianópolis, SC.

This thesis presents a tool for intelligent systems generation. Initially, it's shown a multi-agents system called SMAGS, with ability of integrate distinct information systems and, furthermore, manage the inherent information of the project of interest (like a virtual manager). In the next step, new techniques of Artificial Intelligence are incorporated, adding semantic interpretation concepts and metric tests of ontologies, having the main task of supplying the user with a CASE tool for design and generation of intelligent systems. These generated systems can learn, in real time, from external data sources like enhanced searches over the available ontologies servers, cases databanks, content of experts sites and fragmented data into the Web, becoming it easy and powerful in the managerial decisions process.

Key-words:

Data Mining e Text Mining, Neuro-Fuzzy Nets, Semantic Web, Intelligent Agents, Ontologies and UNL (Universal Networking Language).

SUMÁRIO

AGRADECIMENTOS	5
RESUMO	7
ABSTRACT	8
SUMÁRIO	9
LISTA DE FIGURAS	11
LISTA DE QUADROS	13
LISTA DE TABELAS	14
LISTA DE SIGLAS E ABREVIATURAS	15
1. INTRODUÇÃO	16
1.1. APRESENTAÇÃO	16
1.2. OBJETIVOS	19
1.2.1. Objetivo Geral	19
1.2.2. Objetivos Específicos	19
1.3. JUSTIFICATIVA.....	20
1.4. DELIMITAÇÃO DO TRABALHO	21
1.5. ESTRUTURA DO TRABALHO.....	21
2. REVISÃO DA LITERATURA	23
2.1. MINERAÇÃO DE DADOS	23
2.1.1. Abordagens.....	24
2.3. MINERAÇÃO DE TEXTOS	25
2.3.1. Tipos de Abordagem dos Dados	26
2.3.1.1. Análise Semântica	26
2.3.1.2. Análise Estatística	27
2.3.2. Preparação de Dados	28
2.3.3. Recuperação de Informação	28
2.3.4. Análise dos Dados	29
2.3.5. Dicionário ou <i>Thesaurus</i>	29
2.3.6. <i>Stemming</i>	29
2.3.7. Lista de <i>Stopwords</i>	30
2.3.8. Extração de Características	30
2.3.9. Sumarização, Categorização e <i>Clustering</i>	30
2.4. REDES <i>NEURO-FUZZY</i>	31
2.4.1. Modelo de Rede <i>Neuro-fuzzy</i> ANFIS.....	32
2.4.2. Extração de Regras	34
2.5. AGENTES	36
2.5.1. Agentes com Aprendizagem	37
2.5.2. Sistemas Multiagentes (SMA)	38
2.6. WEB SEMÂNTICA	40
2.7. ONTOLOGIAS	42
2.7.1 Métricas para Medir a Idoneidade de Ontologias	44

2.8. UNL.....	45
2.8.1. Sistema UNL.....	46
2.9. CONSIDERAÇÕES FINAIS	48
3. O SISTEMA MULTIAGENTES SMAGS.....	48
3.1. SMAGS.....	49
3.2. CONSIDERAÇÕES FINAIS	58
4. MODELO PROPOSTO	59
4.1. GISAAGIC.....	59
4.1.1. Subprocesso de Geração do Sistema de Informação.....	61
4.1.2. Subprocesso de Busca.....	62
4.1.2.1. Fluxo para Geração do Grupo de Busca.....	63
4.1.2.2. Fluxo de Processamento dos Resultados.....	68
4.1.3. Subprocesso de Manutenção	78
4.1.3.1. Meta-Agente MARERF	79
4.1.4. Conversão de UNL para Ontologia	88
4.2. CONSIDERAÇÕES FINAIS	91
5. EXEMPLOS DE APLICAÇÃO	92
5.1. EXEMPLO DE APLICAÇÃO PARA A ÁREA DE ENSINO.....	92
5.2.1. Conclusão.....	96
5.3. EXEMPLO DE APLICAÇÃO PARA A ÁREA MÉDICA	97
5.3.1. Subprocesso de Geração do Sistema de Informação.....	97
5.3.2. Subprocesso de Busca.....	107
5.3.3. SubProcesso de Manutenção.....	126
5.3.3.1 Conversão de UNL para Ontologia.....	126
5.3.3.2. Teste Ontométrico	133
5.4. CONSIDERAÇÕES FINAIS	135
6. CONSIDERAÇÕES DO SOFTWARE E OUTROS SISTEMAS.....	136
6.1. CONSIDERAÇÕES PRÁTICAS DO <i>SOFTWARE</i>	136
6.2. COMPARAÇÃO COM OUTROS SISTEMAS	138
7. CONCLUSÕES E RECOMENDAÇÕES	140
7.1. CONCLUSÕES	140
7.2. TRABALHOS FUTUROS	142
REFERÊNCIAS	144
ANEXO I - 41 TIPOS DE RELAÇÕES QUE PODEM OCORRER	151
ANEXO II – REGISTROS DE NÃO CONFORMIDADE, PONTOS FORTES E OBSERVAÇÕES – DIAGNÓSTICO 1.....	154
ANEXO III – REGISTROS DE NÃO CONFORMIDADE, PONTOS FORTES E OBSERVAÇÕES – DIAGNÓSTICO 2.....	159
ANEXO IV – REGISTROS DE NÃO CONFORMIDADE, PONTOS FORTES E OBSERVAÇÕES – DIAGNÓSTICO 3.....	162

LISTA DE FIGURAS

FIGURA 1 - TEMPO DE CICLO DE DECISÕES.....	17
FONTE: ADAPTADO DE (LIAUTAUD E HAMMOND, 2002).....	17
FIGURA 2 – DADO, INFORMAÇÃO E CONHECIMENTO. FONTE: ADAPTADO DE (REZENDE, 2003).	17
FIGURA 3 – ETAPAS DA MINERAÇÃO DE DADOS.....	24
FONTE: ADAPTADO DE: FAYYAD <i>ET AL</i> (1996B) E.....	24
FIGURA 4 – ARQUITETURA DA MINERAÇÃO DE TEXTOS – FONTE: (REZENDE ET AL, 2003).....	26
FIGURA 5 – MODELO DE REDE NEURO-FUZZY CONTENDO DUAS ENTRADAS E TRÊS REGRAS DE INFERÊNCIA.....	33
FIGURA 6 – AGENTES E AMBIENTES. ADAPTADO DE: RUSSEL E NORVIG (1995, 2003).....	36
FIGURA 7 - AGENTE COM APRENDIZAGEM – FONTE: RUSSEL E NORVIG (1995, 2003).	38
FIGURA 8 - ARQUITETURA GERAL DE UM AGENTE – FONTE: (REZENDE, 2003).	39
FIGURA 9 – ARQUITETURA DA WEB SEMÂNTICA PROPOSTA PELO W3C.....	41
FIGURA 10 –DEFINIÇÃO DE UMA META-ONTOLOGIA PARA REPRESENTAR UM MODELO ENTIDADE-RELACIONAMENTO DE UM BANCO DE DADOS	43
FIGURA 11 - ESQUEMA BÁSICO DO PROCESSO ONTOMETRIC	44
FIGURA 12 – SISTEMA UNL.	48
FIGURA 13 – ARQUITETURA DO SMAGS.	50
FIGURA 14 – TELA PARA CONFIGURAÇÃO DO SMAGS. FONTE: CORDEIRO (2001).	52
FIGURA 15 – AGENTE DE MENSAGENS ESCRITAS E FALADAS DO SMAGS – FONTE: CORDEIRO (2001).	57
FIGURA 16 – ARQUITETURA DO GISAAGIC.	60
FIGURA 17 – FLUXO PARA GERAÇÃO DE SISTEMA DE INFORMAÇÕES.	61
FIGURA 18 – PARTE DA ARQUITETURA GERAL ENVOLVIDA NA GERAÇÃO DO SISTEMA DE INFORMAÇÃO.	62
FIGURA 19 – FLUXO PARA A GERAÇÃO DO GRUPO DE TERMOS DE BUSCA.	64
FIGURA 20 – FLUXO GERAL DO SUBPROCESSO DE BUSCA.....	69
FIGURA 21 – PARTES DA ARQUITETURA GISAAGIC ENVOLVIDAS NO PROCESSO DE BUSCA... ..	70
FIGURA 22 – FLUXO DO PROCESSO DE MANUTENÇÃO.....	78
FIGURA 23 – PARTES DA ARQUITETURA DO GISAAGIC ENVOLVIDAS NO SUBPROCESSO MANUTENÇÃO.....	79
FIGURA 24 – SUBPARTE DA ONTOLOGIA NA VIZINHANÇA DE BIBLIOGRAFIA.	81
FIGURA 25 – SUBPARTE DA ONTOLOGIA DE DOMÍNIO NA VIZINHANÇA DE DISCIPLINA.	82
FIGURA 26 – INSTÂNCIAS X E Y LIGADAS A UMA SUBPARTE DA ONTOLOGIA.	83
FIGURA 27 – INSTÂNCIA Z LIGADAS A SUBPARTE DOS CONCEITOS LIVRO E DISCIPLINA.....	84
FIGURA 28 – SUB PARTE DA ONTOLOGIA UTILIZADA PARA O EXEMPLO DE VALIDAÇÃO.....	86
FIGURA 29 - REPRESENTAÇÃO DO PROCESSO DE CONVERSÃO UNL PARA ONTOLOGIA	90
FIGURA 30 – REPRESENTAÇÃO DE PARTE DA ONTOLOGIA PARA O EXEMPLO.....	92
FIGURA 31 - MODELO RELACIONAL PARA O SISTEMA.....	99
FIGURA 32 - TELA PARA DEFINIÇÃO DO DD E A ESTRUTURA DO BD	100
FIGURA 33 - TELA PARA DEFINIÇÃO DE RELACIONAMENTOS ENTRE TABELAS DO BD	101
FIGURA 34 - CONFIGURAÇÃO DO BANCO DE DADOS FÍSICO.	102
FIGURA 35 - MAPEAMENTO DOS DADOS PARA A INTERFACE DE USUÁRIO DO SI.	103
FIGURA 36 - TELAS DO SISTEMA DE INFORMAÇÃO GERADO.	104
FIGURA 37 - REPRESENTAÇÃO DA ONTOLOGIA PARA O EXEMPLO.	105

FIGURA 38 - GEDI – TELA PRINCIPAL.	123
FIGURA 39 - GEDI – MÓDULO DE TREINAMENTO DA REDE <i>NEURO-FUZZY</i>	123
FIGURA 40 - GEDI – PRÉ-CLASSIFICAÇÃO.....	124
FIGURA 41 – CAMADAS DA REDE NEURO-FUZZY ANFIS PROCESSANDO UM DETERMINADO VETOR DE DOCUMENTO.	125
FIGURA 42 - GRAFO UNL REFERENTE AO PARÁGRAFO SUPRACITADO.....	127
FIGURA 43 – ONTOLOGIA TEMPORÁRIA PARA O PARÁGRAFO UNL.....	128
FIGURA 44 – CONEXÃO DA ONTOLOGIA TEMPORÁRIA COM A ONTOLOGIA DE DOMÍNIO.....	133
FIGURA 45 – TESTE ONTOMÉTRICO.	134

LISTA DE QUADROS

QUADRO I - ALGORITMO DO BACKTRACKING.....	35
QUADRO II - EXEMPLO DO PROCESSO DE SILOGISMO HIPOTÉTICO.....	80
QUADRO III - VALORES LINGÜÍSTICOS PARA AS VARIÁVEIS CONSIDERADAS.....	87
QUADRO IV - CORRELAÇÃO UNL -ONTOLOGIA.....	89
QUADRO V - INCLUSÃO DE VIZINHANÇA PELO CRITÉRIO DO BANCO DE DADOS.....	114
QUADRO VI - GERAÇÃO DE GRUPOS DE BUSCA.....	116
QUADRO VII - GERAÇÃO DE GRUPOS DE BUSCA.....	117

LISTA DE TABELAS

TABELA I – PESOS ATRIBUÍDOS AOS CONCEITOS EXISTENTES NA ONTOLOGIA.	93
TABELA II – TF x IDF.	94
TABELA III - TFIDF x FATOR DE GANHO SEMÂNTICO.	95
TABELA IV -	96
VALORES OBTIDOS PARA O GANHO DE INFORMAÇÃO NOS DIFERENTES MÉTODOS.....	96
TABELA V - PESOS ATRIBUÍDOS AOS CONCEITOS EXISTENTES NA ONTOLOGIA DO EXEMPLO.	106
TABELA VI - FREQUÊNCIA ABSOLUTA DE TERMOS.	108
TABELA VII - FREQUÊNCIA RELATIVA DE TERMOS (<i>TFR</i>).	110
TABELA VIII - PONDERAÇÃO ONTOLÓGICA- TFIDF + GS.	112
TABELA IX - ORDENAÇÃO INVERSA PELO CRITÉRIO – ONTOLOGIA DE DOMÍNIO.	115
TABELA X - FREQUÊNCIA ABSOLUTA DE TERMOS.	118
TABELA XI - FREQUÊNCIA RELATIVA DE TERMOS (TF).	119
TABELA XII- TF MAIS IDF (VALORES DE TFIDF).....	120
TABELA XIII - TFIDF COM PONDERAÇÃO ONTOLÓGICA.	121
TABELA XIV- VALORES OBTIDOS PARA O GANHO DE INFORMAÇÃO NOS DIFERENTES MÉTODOS	122

LISTA DE SIGLAS E ABREVIATURAS

ANFIS	– <i>Adaptative Network Fuzzy Inference System</i>
ANVISA	– Agência Nacional de Vigilância Sanitária
CAPES	– Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CASE	– Computer Aided Software Engineering
CEFET-PR	– Centro Federal de Educação Tecnológica do Paraná
CITPAR	– Centro de Integração de Tecnologia do Paraná
CNPq	– Conselho Nacional de Desenvolvimento Científico e Tecnológico
CGEE	– Centro de Gestão de Estudos Estratégicos
FAN	– <i>Free Associative Neurons</i>
FINEP	– Financiadora de Estudos e Projetos
FNDCT	– Fundo Nacional para o Desenvolvimento Científico e Tecnológico
GEDI	– Gerenciador Eletrônico de Documentos Inteligente
IA	– Inteligência Artificial
IBICT	– Instituto Brasileiro de Informação em Ciência e Tecnologia
ISO	– International Organization for Standardization
IFES	– Instituições Federais de Ensino Superior
IPASS	– Instituto Paranaense de Acreditação em Serviços de Saúde
INPI	– Instituto Nacional da Propriedade Industrial
MARERF	– Meta Agente de Raciocínio e Extração de Regras <i>Fuzzy</i>
MCT	– Ministério de Ciência e Tecnologia
MEC	– Ministério da Educação
MIT	– <i>Massachusetts Institute of Technology</i>
NF	– <i>Neuro-Fuzzy</i>
NIT	– Núcleo de Inovação Tecnológica
ONA	– Organização Nacional de Acreditação
PBQP	– Programa Brasileiro de Qualidade e Produtividade
PACTI	– Programa de Apoio a Capacitação Tecnológica Industrial
PADCT	– Programa de Apoio ao Desenvolvimento Científico e Tecnológico
SI	– Sistema de Informação
UFSC	– Universidade Federal de Santa Catarina
UNB	– Universidade Federal de Brasília
UNL	– Universal Network Language

1. INTRODUÇÃO

1.1. APRESENTAÇÃO

Decisões gerenciais inteligentes de assuntos complexos necessitam de informações organizadas e contextualizadas em tempo real para subsidiar as tomadas de decisões. Hammond *et al* (1999, p. 29) afirmam que “os problemas nem sempre são ruins. Na verdade, ao enunciá-los criativamente, é possível convertê-los em oportunidades, abrindo novas, úteis e atraentes alternativas. Albert Einstein: “Em meio à dificuldade se encontra a oportunidade”.”. De acordo com estudos de Liataud e Hammond(2) (2002, p. 123-125), o tempo necessário para tomar as decisões pode ser dividido em duas fases: uma fase de coleta de informações e uma fase de tomada de decisões. A figura 1 mostra como ocorre historicamente esse processo, onde é mostrado que cerca de 80% do tempo que as pessoas gastam para a tomada de decisão são na verdade gastos na coleta de informações. O acesso difícil às informações fazem com que as pessoas tendam a não distinguir estas duas fases distintas. Elas podem pensar que estão trabalhando para tomar uma decisão sobre determinado assunto importante, quando na verdade estão gastando seu tempo e sua energia na coleta de informações: telefonando para vários setores, pedindo diferentes relatórios de ocorrências do assunto, produzindo planilhas para analisar previsões e/ou mesmo analisando velhos relatórios em papel. O fato perigoso dessa situação é que as organizações possuem prazos para decisões e que nem sempre é possível alterar sem outras conseqüências. Os mesmos autores afirmam que as pessoas queimam seu tempo disponível na coleta de informações e em seguida tomam decisões precipitadas ao final do ciclo para cumprir prazo. O que se deseja, não é só tomar decisões rapidamente, mas sim tomar melhores decisões com mais rapidez.

O trabalho desenvolvido por De Ré (2000), identificando que a principal característica de empreendedores, que obtiveram sucesso em seus empreendimentos, é a habilidade em conseguir informações, também comprova e potencializa este trabalho.

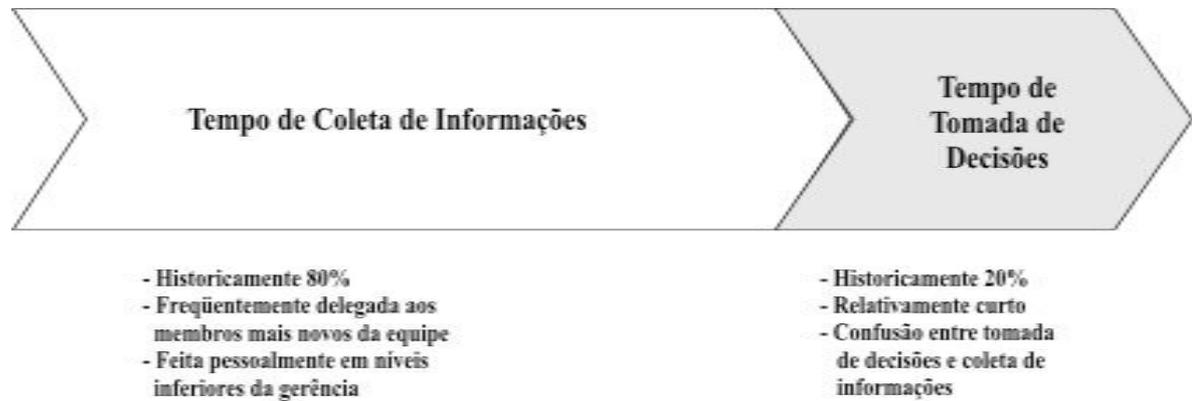


FIGURA 1 - TEMPO DE CICLO DE DECISÕES.
 FONTE: ADAPTADO DE (LIAUTAUD E HAMMOND, 2002).

A informação é o dado analisado e contextualizado, envolvendo a interpretação de um conjunto de dados, cujo fenômeno encontra-se estreitamente vinculado aos processos da cognição humana e é considerada como elemento fundamental para o conhecimento (ALLEN, 1991). O conhecimento é a habilidade de criar um modelo mental que descreva o objeto e identifique as ações a implementar e as decisões a tomar. A compreensão, análise e síntese, necessárias para a tomada de decisões inteligentes, são realizadas a partir do nível do conhecimento (REZENDE, 2003). A figura 2 ilustra essas etapas na forma de uma pirâmide.



FIGURA 2 – DADO, INFORMAÇÃO E CONHECIMENTO. FONTE: ADAPTADO DE (REZENDE, 2003).

Ocorre que os dados, quer sejam nos ambientes internos ou externos das organizações, na maioria das vezes não estão estruturados, o que se constata quando se efetua uma busca na *internet* sobre um determinado assunto e no retorno se verifica que é necessário garimpar em um rol de lixo para encontrar o que realmente interessa.

Do exposto, verifica-se que são necessários sistemas de informações inteligentes com capacidade de fazer buscas inteligentes de dados, nas mais diferentes fontes, quer sejam em ambientes internos ou externos às organizações através da *internet*, estruturando, analisando e contextualizando os dados, os quais são dessa forma transformados em informações úteis para uma determinada área de interesse da organização. Além disso, são necessários mecanismos inteligentes e eficientes para a recuperação de dados, informações e conhecimento em bases de conhecimento que devem estar constantemente atualizados e melhorados com a incorporação de novas tecnologias e novos recursos disponíveis, quer sejam de especialistas da área, grupos, organizações ou outras fontes, e até mesmo com o próprio usuário que estará mais disponível e com conhecimento para escolher entre as melhores opções oferecidas pelo sistema de informação inteligente, em um processo contínuo de auto-aprendizagem em tempo real.

O trabalho desenvolvido por Cordeiro (2001) propõe uma solução para o problema através de um sistema multi-agentes denominado SMAGS que executa automaticamente, de forma inteligente e autônoma, inúmeras tarefas de caráter genérico. O SMAGS é capaz de reconhecer comandos de voz, extrair dados de bancos de dados genéricos, executar programas e extrair dados a distância, em qualquer das estações sob seu monitoramento em uma intranet e *internet*. Entretanto, estes sistemas não são capazes de identificar relações lógicas derivadas do processo de raciocínio humano.

A solução proposta é um gerador inteligente de sistemas que permita o acesso de forma interativa de um especialista de uma área específica, que é orientado para a inserção de todos os dados necessários, dentro uma estrutura e hierarquia de valores estabelecidas, de forma que informações importantes para o projeto em foco não fiquem de fora. Isto poderá ocorrer se o sistema partir de uma base de dados real. Com a incorporação de conceitos derivados de Web Semântica e ontologias, bem como com o uso de ferramentas como redes *neuro-fuzzy*, Mineração de dados e textos e UNL para auxiliar na classificação de padrões e extração de regras, ampliam-se os recursos no modelo inicialmente proposto. Permite-se, então, criar um modelo inicial do caso em foco, do qual se pode extrair características importantes, possibilitando também a criação de ontologias da área de domínio. Estas ontologias podem ser atualizadas e melhoradas através de buscas inteligentes que passam por várias fases até chegar ao conhecimento que é agregado à base de conhecimento do sistema de informação inteligente e distribuído para os usuários através do SMAGS.

Os padrões de entrada são expandidos em uma vizinhança difusa de forma que cada conjunto suporte dessa vizinhança é a combinação dos valores das características próximos dos originais. Assim, dados localizados nessa região de vizinhança se transformam em informações relevantes a partir de ontologias pré-definidas.

As ontologias a serem avaliadas devem ser adicionadas sucessivamente para comparação e um mecanismo de teste métrico é acionado. A valoração obtida é armazenada para uso em futuras avaliações. Como resultado é possível criar uma ontologia de acordo com os conhecimentos adquiridos sobre um determinado assunto de interesse que pode ser por exemplo: uma disciplina ou um conjunto de disciplinas de um determinado curso, cujos conhecimentos são avaliados metricamente em relação a uma ontologia de domínio de uma área específica.

Portanto, este trabalho procura integrar diferentes técnicas e áreas distintas, mas que estão relacionadas pelo objeto de estudo – sistemas de informações inteligentes que possam apoiar e facilitar a tomada de decisões gerenciais que podem resultar em benefício para toda uma organização.

1.2. OBJETIVOS

1.2.1. Objetivo Geral

O presente estudo visa conceber um modelo de ferramenta CASE (*Computer Aided Software Engineering*) para gerar sistemas inteligentes a fim de gerenciar informações e conhecimento, com o objetivo de facilitar tomadas de decisões gerenciais.

1.2.2. Objetivos Específicos

Com base no objetivo geral elaboraram-se os seguintes objetivos específicos:

- Desenvolver uma ferramenta CASE inteligente de forma a interagir com especialistas de assuntos de áreas de interesse específicas;
- Integrar à ferramenta o sistema multi-agentes SMAGS;

- Integrar um módulo GEDI (Gerenciador Eletrônico de Documentos Inteligente), cuja responsabilidade é buscar informações e integrá-las ao sistema gerado;
- Na ferramenta CASE, integrar um módulo de validação pelo usuário das informações processadas pelo GEDI, de forma a agregar os resultados relevantes ao sistema de informação gerado e às ontologias;
- Na ferramenta CASE, agregar um módulo para mecanismos de inferência sobre a ontologia de domínio, mediante extração de regras a partir de definições iniciais do modelo apresentado e também dos dados retornados das buscas subsequentes;
- Validar a ferramenta através da aplicação em exemplos práticos.

1.3. JUSTIFICATIVA

Trabalhos, como De Ré (2000), identificam que a habilidade para conseguir informações necessárias e no tempo que requer determinados problemas, é a chave para o sucesso de muitos empreendedores.

As ferramentas CASE, disponíveis no mercado, procuram traduzir os modelos conceituais para sistemas de informação, restringindo-se a um conjunto de aspectos em detrimento de outros fatos importantes existentes na realidade. É natural que muitos conceitos e relações que não são considerados na fase de modelagem, poderão ser de suma importância na fase de tomada de decisão. A característica estanque de um sistema de informação gerado por uma ferramenta CASE normal, não permite a agregação de tais conceitos ou relações, a não ser que se faça um processo de reengenharia do sistema, o que acaba se tornando dispendioso. As empresas acabam optando por implantar outros sistemas paralelos, incidindo em redundância e perda de produtividade.

A idéia de se utilizar alguns mecanismos e técnicas de Inteligência Artificial, enfocando Web Semântica, Ontologias, Redes *Neuro-Fuzzy*, Mineração de dados e textos e UNL, busca resgatar conceitos e relações reais desconsiderados durante a gênese do sistema, e dessa forma embutir uma função de aprendizado ao sistema de informação.

Outra preocupação é fazer com que o sistema de informação considere o seu ambiente externo (*internet*) como um manancial de dados, onde possa existir informação relevante e que, porventura, possa ser recuperada, tratada e adequada pelo mesmo à base de conhecimento.

O uso de ontologias para expressar um domínio de certa modalidade de negócio também contribui para a fixação de *know-how* na organização, permitindo que as informações transformadas em conhecimento se perpetuem na organização e, desta forma, possam auxiliar *trainees* no aprendizado de suas funções. O aspecto dinâmico do enfoque deste trabalho permite que a ontologia evolua continuamente, adequando-se às mutações do ambiente.

1.4. DELIMITAÇÃO DO TRABALHO

O trabalho foi desenvolvido em várias etapas. Na primeira etapa foram analisadas as vantagens e deficiências do sistema SMAGS e o que mais seria necessário agregar ao sistema para solucionar o problema em tela. Outros sistemas pesquisados não apresentam as mesmas características do sistema SMAGS.

Na seqüência procurou-se vislumbrar uma solução automatizada com melhor adaptação ao problema. Como os dados disponibilizados na internet ainda não estão estruturados, optou-se por uma ferramenta CASE robusta que fosse capaz de gerar sistemas de informações para áreas específicas para as quais os dados podem ser estruturados com a integração do SMAGS, o GEDI e um Agente Buscador.

Na fase de testes foram utilizadas técnicas conjuntas, com Redes *Neuro-Fuzzy* (RNF), em função da origem dos dados, compartilhadas com técnicas de mineração de dados e texto, Ontologias e UNL. As arquiteturas de RNF adotadas foram os modelos ANFIS (*Adaptative Network Fuzzy Inference System*) de (JANG, 1993) e MLP (*Multilayer Perceptron*) na forma proposta por (MITRA e PAL, 1997) e (MITRA, 2003). Estas ferramentas foram integradas ao sistema, e alguns resultados obtidos são mostrados no capítulo relativo à aplicações, através de exemplos.

1.5. ESTRUTURA DO TRABALHO

A tese é composta de sete capítulos, mais as referências.

O capítulo um – Introdução – apresenta aspectos preliminares do trabalho que envolvem o tema e sua importância, os objetivos, a justificativa, a metodologia de desenvolvimento da pesquisa e a estruturação do trabalho.

O capítulo dois – Revisão da Literatura – apresenta uma visão teórica das principais tecnologias envolvidas.

O capítulo três – Sistema Multi-Agentes SMAGS – apresenta a arquitetura do SMAGS com os módulos e suas funções, cuja ferramenta é integrada ao sistema principal.

O capítulo quatro – Modelo Proposto - apresenta o modelo proposto com sua arquitetura e a descrição das funções dos componentes.

O capítulo cinco – Exemplos de Aplicações – são apresentadas os exemplos de aplicações com os respectivos resultados.

O capítulo seis – Considerações do *Software* e Comparação com Outros Sistemas.

O capítulo sete – Conclusões e Recomendações.

Finalmente são apresentadas as referências bibliográficas utilizadas no desenvolvimento do presente trabalho.

2. REVISÃO DA LITERATURA

O presente capítulo apresenta os conceitos das principais técnicas de Inteligência Artificial utilizadas no trabalho: Mineração de dados e de textos, Redes *Neuro-Fuzzy*, Agentes e Sistemas Multiagentes e uma ênfase especial para Web Semântica, ontologias e UNL - *Universal Networking Language*.

As características dos problemas, focados nesta tese, envolvem a busca e estruturação de dados em diferentes fontes, direcionados para uma determinada área de interesse, cujos conhecimentos devem estar formalmente representados. Para que as buscas sejam direcionadas são necessários dispositivos de inferência sobre o conhecimento representado. Os dados, muitas vezes, estão em textos extensos de difícil interpretação. Estes procedimentos envolvem técnicas que sejam capazes de entender a lógica do raciocínio humano e, para isto, é necessário trabalhar com incertezas, grau de pertinência e inferências. As atualizações automáticas direcionadas requerem também a capacidade de auto-aprendizagem.

2.1. MINERAÇÃO DE DADOS

Mineração de dados é o processo de busca de relacionamento e padrões globais existentes nas bases de dados (FAYYAD, PIATETSKI-SHAPIRO, & SMYTH, 1996) e (AMARAL, 2001).

A descoberta do conhecimento em bases de dados é um campo de pesquisa que tem crescido rapidamente, cujo desenvolvimento tem sido dirigido ao benefício de necessidades práticas, sociais e econômicas, entre outras. A motivação para esse crescimento está ligada, principalmente, à existência de uma poderosa tecnologia para coleta, armazenamento e gerenciamento de grande quantidade de dados. Muitos desses dados possuem informações valiosas, como tendências e padrões que poderiam ser usados para melhorar as decisões de negócios, além de outras aplicações (REZENDE *et al*, p.307, 2003).

Para que seja possível a descoberta de novos conhecimentos em padrões escondidos em grandes repositórios de dados, é necessário que sejam utilizadas técnicas de análises por computador de forma ‘inteligente’ e compartilhadas. A partir de dados limpos, agregados e consolidados, são utilizadas ferramentas OLAP(*On-Line Analytical Processing*). Estas ferramentas apresentam facilidades para a realização de consultas complexas em bancos de dados multidimensionais (BRADLEY, FAYYAD & MANGASARIAN, 1998)

2.1.1. Abordagens

Entre as várias abordagens para divisão das etapas do processo de extração ou descoberta do conhecimento em bases de dados (KDD – Knowledge Discovery in Database), destaca-se o proposto em (FAYYAD, PIATETSKI-SHAPIRO, & SMYTH, 1996b), figura 3, com uma divisão em nove etapas.

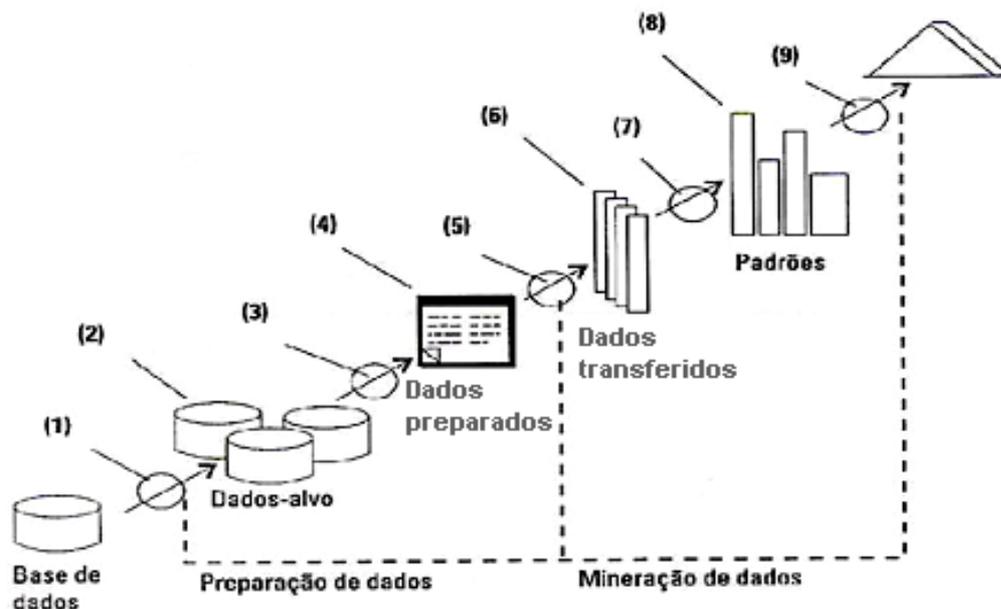


FIGURA 3 – ETAPAS DA MINERAÇÃO DE DADOS
Fonte: Adaptado de: Fayyad *et al* (1996b) e AMARAL (2001).

A fase de preparação dos dados envolve os seguintes passos:

1. Como primeiro passo, é feita a definição do objetivo do problema e do tipo de conhecimento que se deseja extrair do banco de dados, nesta fase é feito o reconhecimento da aplicação e a verificação do conhecimento anterior;

2. No segundo passo, criação de um conjunto de dados-alvo, selecionando-se um conjunto de dados ou focalizando-se em um subconjunto de atributos ou de instâncias de dados, em que a descoberta deverá ser efetuada;
3. No terceiro passo é feita a limpeza e o pré-processamento dos dados.
4. No quarto passo é feita a redução e projeção de dados, consistindo em encontrar as características úteis que representam as dependências dos dados no objetivo do processo.

Já a fase de mineração de dados envolve os seguintes passos:

5. Escolha das tarefas de mineração de dados, onde é decidido qual o objetivo do processo de mineração de dados;
6. Escolha dos algoritmos de mineração de dados, onde são selecionados os métodos para serem usados na busca de padrões dos dados;
7. No sétimo passo é a mineração de dados, caracterizando-se pela busca de padrões de interesse em uma forma particularmente representativa ou em um conjunto dessas representações;
8. Neste passo é feita a interpretação de padrões da exploração, onde os dados de saída, definidos no passo anterior, são analisados e interpretados pelos especialistas do domínio;
9. O último passo consolida o conhecimento descoberto e é incorporado no sistema, em sua base de conhecimento.

Considerando que o processo KDD pode envolver interações significativas, ele pode retornar a qualquer um dos passos, independente da fase a que ele pertença, sendo que a seqüência também pode ser alterada (AMARAL, 2001).

2.3. MINERAÇÃO DE TEXTOS

Os vários tipos de textos que compõem o dia-a-dia das organizações ou pessoas são produzidos e armazenados em meios eletrônicos. Alguns destes textos são gerados e atualizados periodicamente. Até pouco tempo atrás, essas informações em formato de textos não eram usadas para obter algum tipo de vantagem competitiva ou suporte à tomada de decisões. Com o advento da Mineração de Textos, a extração de informações em textos

passou a ser possível e o imenso e crescente mundo dos textos está começando a ser explorado (REZENDE, 2003).

A Mineração de Textos é um conjunto de técnicas e processos que descobrem conhecimento inovador nos textos (REZENDE, 2003).

A figura 4 mostra um esquema com as etapas da Mineração de Textos.

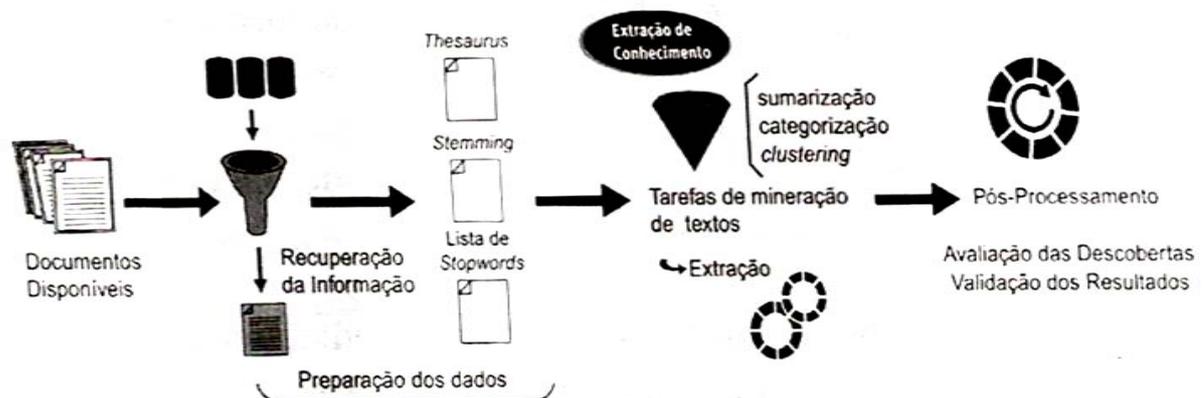


FIGURA 4 – ARQUITETURA DA MINERAÇÃO DE TEXTOS – FONTE: (REZENDE ET AL, 2003)

2.3.1. Tipos de Abordagem dos Dados

São duas as principais formas de abordagem que podem ser utilizadas separadamente ou em conjunto para analisar os dados.

2.3.1.1. Análise Semântica

A análise emprega técnicas que avaliam a seqüência dos termos no contexto da frase, identificando a correta função de cada termo. Esse tipo de análise utiliza fundamentos e técnicas baseadas no processamento de linguagem natural.

Textos em linguagem natural implicam os seguintes tipos de conhecimento relevantes:

- Conhecimento Morfológico:** da estrutura, da forma e das inflexões das palavras;
- Conhecimento Sintático:** das listas de palavras e como as palavras podem ser combinadas para produzir sentenças;
- Conhecimento Semântico:** o que as palavras significam independente do contexto e como significados mais complexos são formados pela combinação de palavras;

- d) **Conhecimento Pragmático:** do uso da língua em diferentes contextos, visualizando como o significado e a interpretação é afetada pelo contexto;
- e) **Conhecimento do Discurso:** como as sentenças imediatamente precedentes afetam a interpretação da próxima sentença;
- f) **Conhecimento do Mundo:** conhecimento geral do domínio ou o mundo que a comunicação da linguagem natural se relaciona.

A análise semântica dos textos procura identificar a importância das palavras dentro da estrutura da oração. Porém, quando se utiliza um único texto algumas funções podem ser identificadas com um grau de importância. Entretanto, para algumas tarefas isso não é suficiente. Como exemplo podem ser citadas as categorizações, onde é interessante analisar um documento comparando-o com bases de conhecimento de diferentes assuntos para descobrir a que categoria ele pertence.

2.3.1.2. Análise Estatística

Nesta análise, a importância dos termos é dada basicamente pelo número de vezes que eles aparecem nos textos. O aprendizado estatístico a partir dos dados inclui os seguintes passos:

- a) **Codificação dos Dados:** é feita uma codificação inicial dos dados baseada na indicação de especialista ou pela maximização de critério que reflita propriedades interessantes dos dados em relação à finalidade da modelagem;
- b) **Estimativa de Dados:** é aplicado um algoritmo de aprendizado ou um método de estimativa, normalmente pela maximização de algum critério;
- c) **Modelos de Representação de Documentos:** realizam-se várias pesquisas de mineração de coleções de documentos, cujos documentos são vistos como *containers* de palavras, sendo também conhecida por codificação *bag of words*, que ignora a ordem das palavras e qualquer informação de pontuação ou estrutural, porém retém o número de vezes que uma palavra aparece.

2.3.2. Preparação de Dados

Esta é a primeira etapa do processo de descoberta de conhecimento em textos e envolve a seleção dos dados que constituirão a base de textos de interesse, procurando-se selecionar o núcleo que melhor expressa o conteúdo dos textos e toda a informação que não refletir nenhuma idéia considerada importante, poderá ser desprezada (REZENDE, 2003).

2.3.3. Recuperação de Informação

Esta área de Recuperação de Informação (RI) desenvolveu modelos para a representação de grandes coleções de textos que identificam documentos sobre tópicos específicos. Um sistema RI atua como se fosse um filtro sobre um conjunto de documentos, retornando ao usuário o resultado de um problema particular (EBECKEN *et al*, 2003).

São dois os modelos relacionados a RI:

- a) **Modelo Booleano:** neste modelo, um documento é representado por um conjunto de termos-índice, sendo que uma consulta ou *query* consiste de um conjunto de termos-índice combinados com operadores de Boole. O modelo é binário e a frequência de um termo não tem efeito. Nesse modelo, a semântica de *query* é bem definida.
- b) **Modelo de Espaço Vetorial:** este Modelo de Espaço Vetorial (*Vectorial Space Model – VSM*) resolve problemas de representação de documentos utilizando representação geométrica. É um modelo usado também quando é preciso encontrar documentos que atendam a um critério, onde a solução desse problema decorre do esquema de representação dos documentos. Documentos são representados como ponto ou vetores em um espaço euclidiano *t*-dimensional em que cada dimensão corresponde a uma palavra (termo) do vocabulário (SALTON, 1989). O *i*-ésimo componente do vetor documento expressa o número de vezes em que a palavra com o índice *i* ocorre em um documento, ou uma função dele. Cada palavra pode ter um peso associado para descrever sua significância. A similaridade entre dois documentos é definida como a distância entre os pontos ou como o ângulo entre os vetores, sendo desconsiderado o comprimento do documento. O modelo de Espaço Vetorial e suas variações têm persistido em avaliações de qualidade, no campo da Recuperação de Informação (BAEZA-YATES & B., 1999).

As ferramentas de RI, normalmente trabalham associadas a técnicas de indexação capazes de acessar e mapear rapidamente documentos em uma base de textos. Podem indexar o texto completo ou por *tags*.

2.3.4. Análise dos Dados

A análise tem como principal objetivo facilitar a identificação de similaridade de significado entre as palavras, apesar das suas variações morfológicas, onde um mesmo termo pode assumir diferentes sufixos como o do *stemming* que é realizado pela extração de cada palavra do texto, considerando aquela palavra isoladamente e tentando reduzi-la à sua provável palavra raiz. A análise também contempla a situação de palavras sinônimas que, apesar de serem morfológicamente diferentes, expressam a mesma idéia.

2.3.5. Dicionário ou *Thesaurus*

O *thesaurus* pode ser definido como um vocabulário controlado que representa sinônimos, hierarquias e relacionamentos associativos entre termos para ajudar os usuários a encontrar a informação que eles precisam (EBECKEN *et al*, 2003). O seu valor vem justamente dos problemas inerentes à procura de indexação da linguagem natural. Usuários diferentes definem a mesma *query* usando termos diferentes. Para resolver tal problema, o *thesaurus* mapeia termos variantes: sinônimos, abreviações, acrônimos e ortografias alternativas, para um termo preferido único para cada conceito. Um *thesaurus* pode também representar a riqueza dos relacionamentos de forma associativa e hierárquica.

2.3.6. *Stemming*

No *stemming* é realizada a extração de palavras do texto, considerando as palavras isoladamente, procurando reduzi-las a sua provável palavra raiz. Os algoritmos correntes não usam informações do contexto para determinar o sentido correto de cada palavra.

Entre os diferentes métodos de *stemming*, pode-se citar o “*Stemmer S*”, que remove apenas poucos finais de palavras como: *ies*, *es* e *s*. Já o método de Porter consiste na identificação das diferentes inflexões referentes à mesma palavra e sua substituição por um radical comum e o método de Lovins que é um método de um único passo, sendo sensível ao

contexto e que usa um algoritmo da combinação mais longa para remover cerca de 250 sufixo diferentes (EBECKEN *et al*, 2003).

2.3.7. Lista de *Stopwords*

A remoção de *stopwords* (textos que aparecem com tal frequência que perdem sua utilidade em termos de precisão) é um dos primeiros passos no processo de preparação dos dados e identifica o que pode ser desconsiderado nos passos posteriores do processamento de dados. É uma tentativa de retirar tudo o que não constitui conhecimento nos textos. Estas palavras são consideradas não relevantes na análise dos textos, isto porque são palavras auxiliares como *e, para, a, eles*, que não fornecem nenhuma informação discriminativa na expressão do conteúdo dos textos. Na construção da lista de *stopwords* são incluídas palavras como: preposições, pronomes, artigos e outras classes de palavras auxiliares.

2.3.8 Extração de Características

Os métodos de extração de características correspondem à extração de termos que podem ocorrer nas bases da informação lingüística estruturada e a seleção de termos que podem ocorrer nas bases de métricas estatísticas, como frequência, informação mútua, entre outras (EBECKEN *et a*, 2003).

2.3.9. Sumarização, Categorização e *Clustering*

A sumarização é um processo de redução da quantidade de texto em um documento, sem perder o seu significado chave (HABN & MANI, 2000). Já a categorização visa identificar os tópicos principais em um documento e associar este documento a uma ou mais categorias pré-definidas (YANG & PEDERSEN, 1997).

O *clustering* é uma técnica útil para agrupar documentos similares. Neste processo é criada uma descrição simplificada do documento para cada texto que vai sendo adicionado aos *clusters*, normalmente um vetor de características (VASILEIOS *et al*, 2000).

2.4. REDES *NEURO-FUZZY*

Uma Rede Neural Artificial (RNA) é uma estrutura que processa informação de forma paralela e distribuída, procurando se assemelhar à estrutura dos neurônios humanos.

São modelos matemáticos que se assemelham às estruturas neurais biológicas e que tem capacidade computacional adquirida por meio de aprendizado e generalização (BRAGA, CARVALHO & LUDEMIR, 2000; HAYKIN, 1994).

A comunicação e aprendizagem em um sentido amplo envolve, entre outras coisas, relações humanas e como tal: “*As relações humanas são, em sua gênese, relações difusas*” BASTOS (1994). Partindo deste princípio, para que seja possível melhorar, aperfeiçoar, agilizar, modernizar e automatizar sistemas de comunicações utilizando as novas tecnologias da informação, é necessário trabalhar também com técnicas que sejam capazes de entender a lógica do raciocínio humano e, para isto, é necessário trabalhar com incertezas, grau de pertinência e inferências.

“Os conjuntos *fuzzy* são utilizados para representar conceitos vagos, imprecisos ou incertos” (EVSUKOFF e ALMEIDA, 2003). Os sistemas *fuzzy* podem ser usados como sistemas de apoio a tomada de decisões; representar o conhecimento de especialistas sobre um determinado assunto e interpolar decisões a partir de entradas contendo incertezas. Permitem a representação de algumas características do raciocínio humano e podem ser usados para o desenvolvimento de sistemas de bases de conhecimento.

Considerando as características destes dois importantes paradigmas, Redes Neurais e Fuzzy, tem-se procurado unir os potenciais de sistemas *fuzzy* e redes neurais, resultando os chamados sistemas híbridos *neuro-fuzzy*. Portanto, estes sistemas podem incorporar conhecimentos empíricos e ter seus parâmetros adaptados através de algoritmos eficientes (EVSUKOFF, A. G.; ALMEIDA, P. E. M.; RAITTZ, R. *et al*, 2003).

Para a classificação dos documentos utilizada neste trabalho, a rede *neuro-fuzzy* aprende a partir de um vetor de características de palavras extraídas de uma série de documentos **TF** (Term Frequency), **IFT** (Inverse Frequency Term), **LF** (Link Frequency) e ponderação por ontologia como dados de entrada, e uma saída binária indicando a relevância do documento para o processo de busca. Os padrões de entrada/saída são apresentados à rede, e assim o aprendizado obtido pela rede *neuro-fuzzy* pode ser reutilizado para classificação *a priori* de novos documentos provenientes de buscas subseqüentes.

2.4.1. Modelo de Rede *Neuro-fuzzy* ANFIS

Um dos modelos de Redes *Neuro-Fuzzy* utilizados para este trabalho é o modelo ANFIS (*Adaptative Network Fuzzy Inference System*) proposto por (JANG, 1993), que se baseia no sistema de regras de inferência de Takagi, Sugeno e Kang (SUGENO & KANG, 1988; TAKAGI & SUGENO, 1985). A utilização deste tipo de rede tem o objetivo de extrair regras de inferência *fuzzy* a partir de um conjunto de padrões de entrada/saída. As regras seguem a forma geral:

$$\text{if } x = A_i \text{ then } y = \phi_i(x) \quad (1)$$

As entradas são definidas pelo vetor x , A_i é um conjunto fuzzy definido sobre o universo da variável x , e $\phi_i(x)$ é a saída da regra. Esta saída geralmente é definida como uma saída polinomial em função da variável de entrada x . Costuma-se denominar de *precedente* da regra a igualdade que vem na cabeça da regra, e de *conseqüente* da regra o polinômio no corpo da regra. Assim, um sistema de inferência *fuzzy* possui um conjunto de regras tais como em (1) e suas camadas têm as seguintes funções:

- 1) Uma camada que transforma os dados de entrada abruptos (*crisp*) em valores *fuzzy* a partir de funções de pertinência;
- 2) A camada seguinte funciona como um operador conjuntivo (geralmente o operador produto) sobre os valores *fuzzy*;
- 3) Uma camada que normaliza os valores após o operador conjuntivo;
- 4) A camada de saída que utiliza os valores normalizados da camada anterior, ponderando o valor do conseqüente da regra ϕ_i ; sendo:

$$\phi_i = \theta_{ij}x_j + \theta_{i0} \quad (2)$$

onde: θ_{ij} são chamados os *parâmetros do conseqüente* das regras;

x_j são os valores assumidos para os vetores de entrada;

θ_{i0} são parâmetros de ajuste fino;

ϕ_i é o conseqüente da regra.

As regras de inferência relativas ao modelo ANFIS podem ser expressas como:

$$\begin{aligned}
 \text{if } x_1 = A_1 \text{ and } x_2 = B_1 \text{ then } \phi_1 &= \theta_{10} + \theta_{11}x_1 + \theta_{12}x_2 & (3) \\
 \text{if } x_1 = A_2 \text{ and } x_2 = B_2 \text{ then } \phi_2 &= \theta_{20} + \theta_{21}x_1 + \theta_{22}x_2 \\
 \text{if } x_1 = A_3 \text{ and } x_2 = B_3 \text{ then } \phi_3 &= \theta_{30} + \theta_{31}x_1 + \theta_{32}x_2
 \end{aligned}$$

A rede ANFIS apresenta então sua característica típica, necessitando de um processo de *fuzzificação* dos dados de entrada, porém não precisando fazer o processo inverso no final, sendo a saída calculada diretamente. O treinamento compõe-se de duas etapas (AZEVEDO, 2000) e (EVSUKOFF, 2003). A primeira envolvendo o algoritmo LSE (*Least Square Estimation*) sobre os parâmetros do conseqüente para cada regra (os parâmetros do conseqüente permanecem fixos nesta etapa). Na segunda etapa, utiliza-se o algoritmo de retropropagação (*backpropagation*) sobre os parâmetros do precedente da regra (aqui, os parâmetros do conseqüente permanecem fixos). O uso de funções de pertinência contínuas permite obter expressões contínuas para a descida de gradiente, necessária ao algoritmo de retropropagação. Este ciclo é executado então para um número de iterações pré-fixadas. Um número excessivo de épocas de treinamento podem levar a uma deformação das funções de pertinência. O uso de restrições sobre a faixa de valores possíveis dos parâmetros do precedente pode ser útil de forma a evitar tal efeito e permitir uma interpretação lingüística aceitável das entradas. A figura 5 apresenta um modelo ANFIS com duas entradas e três regras de inferências.

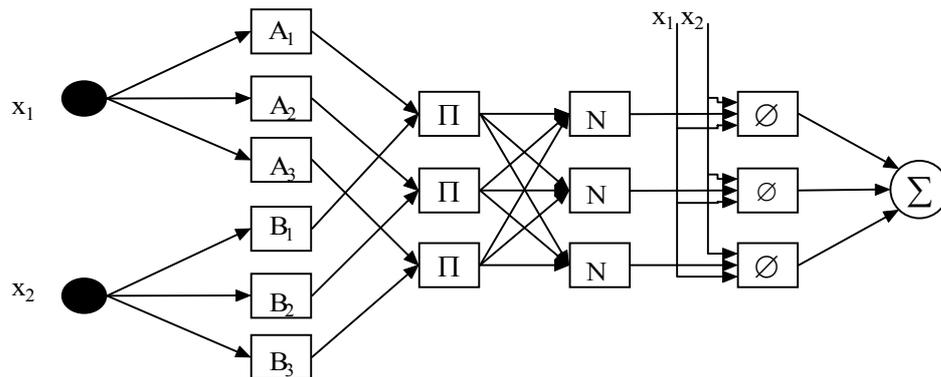


FIGURA 5 – MODELO DE REDE NEURO-FUZZY CONTENDO DUAS ENTRADAS E TRÊS REGRAS DE INFERÊNCIA.

Uma rede deste tipo pode ser utilizada para aproximação de funções e classificação. Para classificação, pode-se determinar limiares (*thresholds*) na função de saída para as classes predefinidas.

2.4.2. Extração de Regras

No escopo desta tese é prevista a utilização de mecanismos para extração de regras através de um agente específico dentro do sistema multiagentes proposto. Através do comportamento de certas variáveis num conjunto de dados previamente escolhidos, testa-se a hipótese de uma certa variável (ou variáveis) que esteja assumindo um determinado comportamento em função de outras variáveis. Uma rede neural comum pode convergir e fazer generalizações após treinamento, mas a identificação de como se chega a um certo resultado de saída a partir de uma dada entrada dificilmente pode ser entendido, visualizando-se diretamente os valores assumidos pelos pesos.

Neste íterim é que entra a abordagem de extração de regras *fuzzy* via uma rede neural MLP (Multilayer Perceptron). De acordo com (MITRA e PAL, 1997) e (MITRA, 2003), é possível extrair regras do tipo **se-então** para uma rede MLP treinada, tomando-se por base que os atributos de entrada possam assumir valores lingüísticos ou *fuzzy*. Através de um algoritmo que efetua um *backtracking* na rede MLP, os pesos e neurônios de entrada com maior impacto na decisão de uma certa saída são identificados e priorizados, possibilitando uma representação em forma de cláusulas e explicando o porquê de certa decisão ter sido tomada.

Uma entrada do tipo lingüística pode ser subdividida em 3 (três), de forma que cada uma represente um valor lingüístico (baixo, médio e alto, por exemplo) ou qualquer tamanho, conforme domínio. Um conjunto n de atributos forma assim um outro conjunto $3n$ que serão efetivamente as entradas da rede MLP. A rede então é treinada e após efetua-se o algoritmo do *backtracking*. Para cada padrão apresentado à rede, a mesma responde com uma saída, gerando uma regra **se-então** para este caso. As regras, que não estão explícitas num primeiro passo, tornam-se transparentes após a aplicação do *backtracking*. Os passos deste algoritmo são mostrados no quadro I a seguir (AZEVEDO, 2000):

Quadro I – Algoritmo do Backtracking

1. Geração das cláusulas por *backtracking*

- a) Escolha do padrão de entrada-saída;
- b) Escolha do neurônio i de uma camada anterior com impacto positivo na conclusão de uma saída qualquer j ;
- c) Definição do conjunto m_i em relação às entradas do neurônio escolhido:

$$m_i = \{a_1, \dots, a_{m_i}\} \quad (4)$$

- d) Determinação do conjunto dos pesos acumulados para o neurônio i , utilizando um limiar para o neurônio de saída:

$$y_i > 0.5 \quad (5)$$

$$wet_i = \max[wet_{a_k} + w_{a_k i}] \quad (6)$$

- e) Determinação dos pesos wet_{ak} :

$$wet_{a_k} = \{w_{ja_1}, w_{ja_2}, \dots, w_{ja_{m_i}}\} \quad (7)$$

- f) Seleção do conjunto de neurônios de entrada m_o :

$$m_o = \{a_1, a_2, \dots, a_{m_o}\} \quad (8)$$

- g) Definição dos pesos do caminho referente ao neurônio j ;

- h) Ordenação em ordem decrescente de impacto na rede (IR) dos elementos do conjunto de pesos obtidos no item anterior tal como:

$$IR_i = y_i * wet_i \quad (9)$$

- i) Seleção dos i_s neurônios de entrada e os i_n neurônios restantes para as cláusulas, tal que:

$$m_o = |i_s| + |i_n| \quad (10)$$

$$\sum_{i_s} wet_{i_s} > 2 \sum_{i_n} wet_{i_n} \quad (11)$$

2. Geração das cláusulas

Para a dedução dos antecedentes da regra:

- a) Definição de u_{s_l} para uma certa entrada lingüística, sendo obtida como (o operador *mod* é entendido como o resto do operando por 3) :

$$u_{s_l} = (l_{s_l} - 1) \bmod 3 + 1 \quad (12)$$

- b) Repete-se os itens anteriores até que todos os $|s|$ (neurônios de entrada selecionados) sejam selecionados pela equação (11)

Para a dedução dos conseqüentes da regra:

- a) Definição da medida de certeza bel_j , dada como:

$$bel_j = y_j - \sum_{i \neq j} y_i \quad (13)$$

- b) Definição dos conseqüentes de uma regra através das seguintes propriedades:

$$prop = \begin{cases} \text{baixo, se } l_{s_l} - 3(u_{s_l} - 1) = 1 \\ \text{médio, se } l_{s_l} - 3(u_{s_l} - 1) = 2 \\ \text{alto, caso contrário} \end{cases} \quad (14)$$

Onde $1 \leq l_{s_l} \leq 3n$ e $1 \leq u_{s_l} \leq n$, com n sendo a dimensão do vetor do padrão de entrada.

3. Retorna-se ao item 1 caso se deseje determinar as regras **se-então** para justificar as saídas de mais padrões, ou mesmo de todos os padrões de treinamento.

2.5. AGENTES

Considerando que os módulos executarão tarefas autônomas e assíncronas, é necessário considerar o conceito de agente inteligente.

Para Jennings (1996), um agente é uma entidade computacional com capacidade para perceber o seu ambiente exterior, possibilitando a intervenção nesse mesmo ambiente de forma flexível e com algum grau de autonomia, guiado por objetivos próprios e com eventual capacidade para aprender com a experiência, de modo a melhorar o seu desempenho.

Etzione, O.; Lesh, N. *et al*, em Souza (2003), definem uma espécie de agente chamado *softbot* – *software robot*(robô): “Um *softbot* é um agente que interage com o ambiente computacional pela emissão de comandos e pela interpretação dos resultados apresentados pelo ambiente. Os atuadores dos *softbots* são comandos, os quais possuem como objetivo promover alterações no estado do ambiente. Os sensores dos *softbots* são comandos que possuem como objetivo produzir informação”.

Uma outra definição mais abrangente, que se aplica tanto para agentes humanos como para robôs e agentes de *software*, afirma que um agente obtém conhecimento do seu ambiente por meio de sensores; age nesse ambiente através de atuadores, conforme ilustrado na figura 6 (RUSSEL & NORVIG, 1995, 2003).

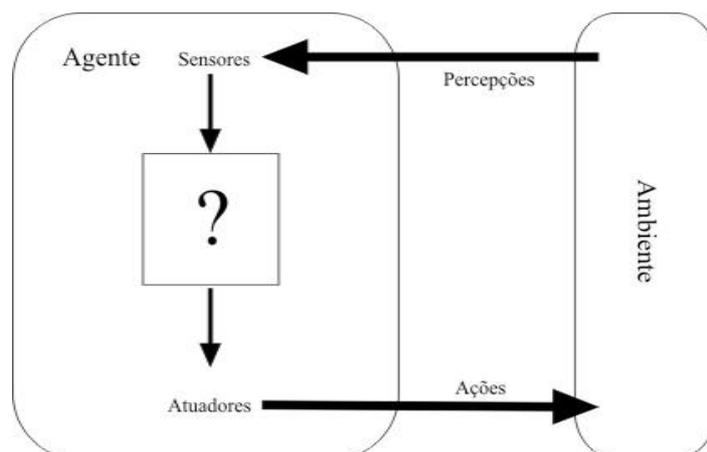


FIGURA 6 – AGENTES E AMBIENTES. ADAPTADO DE: RUSSEL E NORVIG (1995, 2003).

Wooldridge (1995) descreve que: “agente é utilizado para denotar um sistema computacional que goza das seguintes propriedades:

- a) **Autonomia:** agentes operam sem a intervenção direta de seres humanos ou outras entidades e exercem algum tipo de controle sobre as suas ações e estados internos;
- b) **Habilidade social:** agentes interagem com outros agentes;
- c) **Reatividade:** agentes percebem seus ambientes e respondem rapidamente às trocas que neles ocorrem;
- d) **Pró-atividade:** agentes não agem apenas em respostas às alterações dos seus ambientes. Eles são também capazes de exibir um comportamento orientado por objetivos através de iniciativas.

Huhns & Singh (1997) procuram sintetizar o que possa haver de comum nas demais definições, afirmando que agentes de *software* são entidades computacionais persistentes e ativas que agem e se comunicam em um ambiente”.

Cockayne e Zyda (1998) afirmam que: “o conceito de agentes móveis emergiu na década de setenta, de um exame profundo da comunicação entre computadores. Devido ao crescimento exponencial da *internet*, uma nova aproximação foi necessária para incrementar os tipos sofisticados de comunicação (protocolos), sem congestionar os componentes da mesma”.

Pickett (2000), no Dicionário Herança Americano, define um agente como “aquele que age ou tem o poder ou autoridade para agir ou representar outrem”.

2.5.1. Agentes com Aprendizagem

Para Russel & Norvig (1995, 2003), um agente de aprendizado apresenta quatro componentes conceituais, conforme a figura 7, com as seguintes funções:

- a) **Elemento de aprendizado:** responsável pela execução de aperfeiçoamentos;
- b) **Elemento de desempenho:** responsável pela seleção de ações externas;
- c) **Crítico:** informa ao elemento de aprendizado como o agente está se comportando em relação a um padrão fixo de desempenho;

- d) **Gerador de problemas:** é responsável por sugerir ações que levarão a experiências novas e informativas.

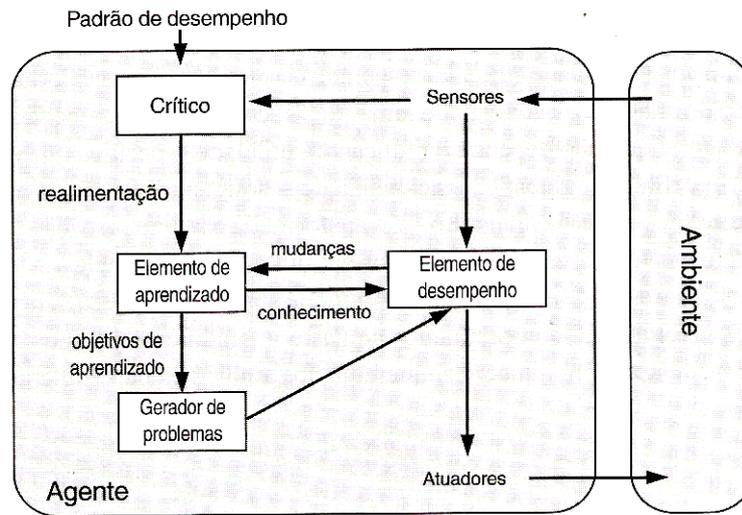


FIGURA 7 - AGENTE COM APRENDIZAGEM – FONTE: RUSSEL E NORVIG (1995, 2003).

2.5.2. Sistemas Multiagentes (SMA)

Para Ferber (1999), o SMA é aplicado a sistema que compreende os seguintes elementos:

- Um ambiente E , representado por um espaço que possui um volume;
- Um conjunto de objetos O . É possível, em um dado momento, associar qualquer objeto com uma posição E . Esses objetos são passivos, ou seja, eles podem ser percebidos, criados, destruídos e modificados pelos agentes;
- Uma comunidade de agentes A , de objetos específicos representando as entidades ativas do sistema;
- Uma comunidade de relações, R , que liga objetos (e portanto agentes) entre si;
- Uma comunidade de operações, Op , que permite aos agentes de A perceberem, produzirem, consumirem, transformarem e manipularem objetos do O ;
- Operações com a tarefa de representar a aplicação dessas operações e a reação do mundo para suas tentativas de modificação.

Souza (2003) define um SMA como “uma rede, fracamente ligada, de agentes que trabalham em conjunto para resolver problemas que ultrapassam a capacidade e o conhecimento de cada um em particular. Eles são especialmente adaptados para representar problemas com múltiplas possibilidades de solução e diversas perspectivas e entidades. Esses sistemas possuem capacidade para resolver problemas tradicionalmente distribuídos e possibilitam sofisticados padrões de interação, tais como:

- a) cooperação (trabalhar em conjunto em prol de um objetivo comum);
- b) coordenação (organizar a resolução de problemas mediante a exploração de interações benéficas e evitando interações prejudiciais);
- c) negociação (tentar atingir um acordo que seja aceitável para todas as partes envolvidas)”.

A figura 8 ilustra a arquitetura geral de um Agente.

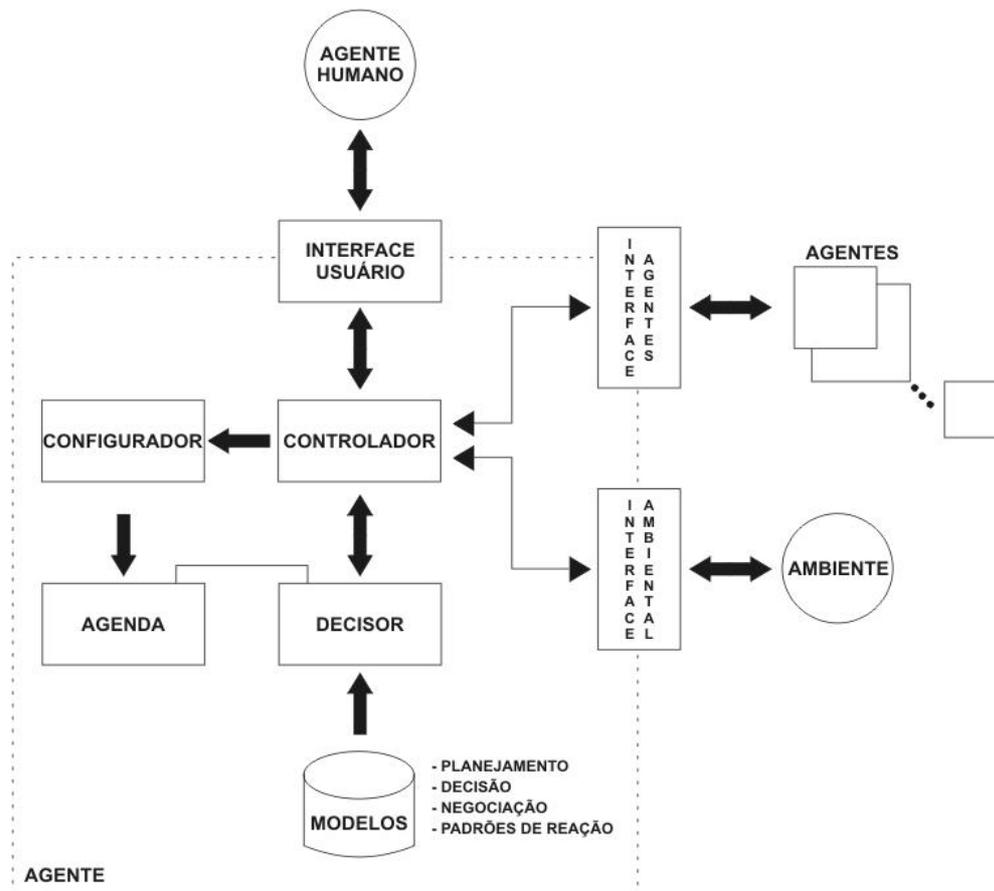


FIGURA 8 - ARQUITETURA GERAL DE UM AGENTE – FONTE: (REZENDE, 2003).

2.6. WEB SEMÂNTICA

A Web Semântica está sendo considerada uma extensão da Web atual, possibilitando a estruturação do conteúdo que está “solto” na *internet*, criando um ambiente no qual agentes poderão perambular de página em página para executar tarefas de interesse para seus usuários.

Esses agentes serão capazes de identificar não apenas o significado de uma palavra, como também as relações lógicas entre várias palavras. A questão que se coloca aqui é a do entendimento de conteúdo. Para isso é necessário ler dados estruturados que são informações colocadas em campos fixos de um arquivo, além de ter acesso a conjunto de regras que ajudem a conduzir seus raciocínios. Neste sentido as páginas Web terão que ser escritas em uma nova linguagem de forma que possam ser compreendidas por diferentes sistemas (BERNERS-LEE *et al*, 2001).

Algumas tecnologias empregadas na Web Semântica ganham espaço. Entre elas o XML (*eXtensible Markup Language*), que permite que os usuários criem *tags* que são marcações de dados inseridos em arquivos e que podem com esta tecnologia serem personalizados. Outra tecnologia é o RDF (*Resource Description Framework*), que forma trios de informações que expressam o significado das *tags* do XML. Cada unidade do trio tem sua própria função, como sujeito, verbo e objeto de uma frase, recebendo uma identificação URI (*Universal Resource Identifier*), que foi a tecnologia usada para criar as conhecidas URL's (*Uniform Resource Locator*). Com tudo isso (parece ser muita complicação) acontece que na linguagem humana uma mesma palavra pode assumir vários significados, causando grandes confusões nos sistemas. A solução é usar uma URI diferente para cada conceito. Nesta fase chega-se ao maior desafio da Web Semântica porque pode acontecer de dois bancos de dados usarem URI's diferentes para um mesmo conceito e neste caso é necessário que o *software* que vai pesquisar saiba quando está tratando da mesma coisa. Uma proposta para este problema é o uso de ontologias que permitem definir relações entre conceitos de diferentes sistemas e vão fornecer o vocabulário necessário para a comunicação entre os agentes e as páginas, bem como mostrar as relações entre os conceitos.

A arquitetura proposta para a Web Semântica está ilustrada na figura 9.

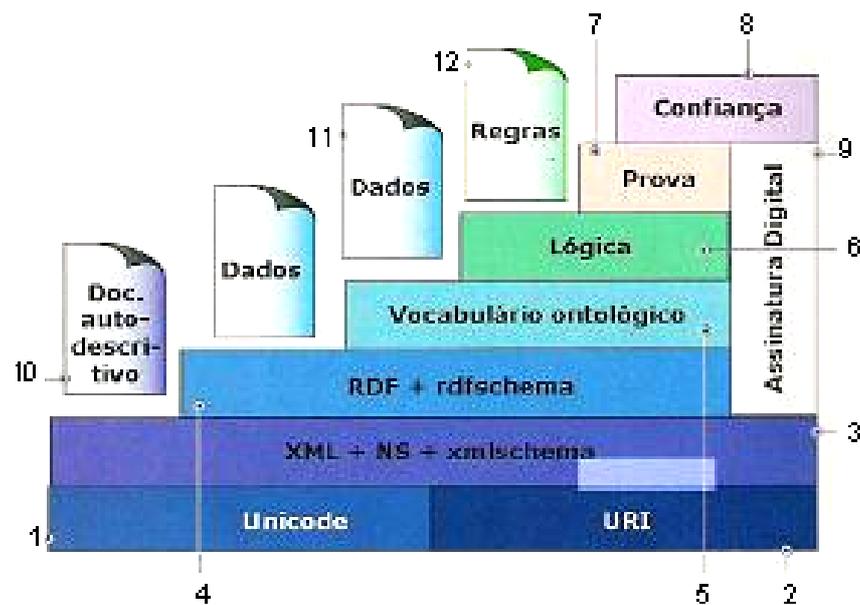


FIGURA 9 – ARQUITETURA DA WEB SEMÂNTICA PROPOSTA PELO W3C
Adaptado de www.com.br: disponível em: <http://www.w3.org/>

Esta estrutura organizada em camadas, tem os seguintes elementos:

- 1) O **Unicode** que é uma codificação usada no lugar do ASCII, cobrindo mais linguagens;
- 2) As **URIs** que possibilitam a marcação de informações que apontam para dados na Web;
- 3) A camada do **XML** que permite a interoperabilidade e é o padrão que serve de sustentação para o formato de descrição **RDF**.
- 4) O **RDF** permite a interoperabilidade entre aplicativos que trocam informações em linguagem de máquina pela Web;
- 5) O **Vocabulário ontológico** permite a descrição detalhada dos termos e suas inter-relações, ultrapassando as descrições básicas habilitadas pelo esquema **RDF**. As ontologias permitem diferenciar as palavras;
- 6) A camada **Lógica** irá fornecer uma linguagem para descrever conjuntos de deduções que podem ser feitas a partir de uma coleção de dados;
- 7) A camada de **Provas** fornecerá um caminho para descrever os passos tomados para alcançar a conclusão a partir de fatos;

- 8) A camada **Confiança** deverá assegurar que as informações são confiáveis e contextualizadas;
- 9) A **Assinatura Digital** garante a procedência de uma informação eletronicamente;
- 10) O **Documento auto-descritivo** possui através do **RDF**, conteúdo e a descrição desse documento;
- 11) Os **Dados** sobre os documentos são passados para as camadas superiores;
- 12) As **Regras** para controlar o relacionamento entre os termos.

2.7. ONTOLOGIAS

Em filosofia, o termo ontologia se define como “a parte da metafísica que trata do ser em geral e suas propriedades transcendentais” (TELLO, 2002). Em Inteligência Artificial tem diferentes conotações (BERNARA, 1996; SWARTOUT, 1997 *et al*) *apud* (TELLO, 2002), definem ontologia como “uma especificação explícita e formal sobre uma conceitualização consensuada”. Uma interpretação desta definição, sugere que as ontologias definem seus conceitos, propriedades, relações, funções, restrições e axiomas de forma “explícita” em alguma linguagem de implementação capaz de conter este conhecimento. O termo “conceitualização” se refere a um modelo abstrato de algum fenômeno no mundo. As ontologias são estabelecidas para serem usadas de forma consensuada e compartilhada por diferentes sistemas. O termo “formal” se refere ao fato de que a ontologia deve ser implementada em alguma linguagem processável pela máquina. Uma ontologia pode ser definida como “um conjunto de termos estruturados hierarquicamente que descrevem um domínio”(SWARTOUT, 1997). As ontologias proporcionam a estrutura taxonômica de um domínio, a qual será especializada com os conhecimentos específicos necessários para a aplicação.

Atualmente, muito tempo é perdido na busca de dados, informações e/ou ontologias disponíveis para utilizar em sua aplicação e ainda não se tem a certeza de que a informação serve para o seu projeto e se não está ultrapassada, de forma que o usuário acaba tomando uma decisão com base na experiência (NASCIMENTO JR e YONEYAMA, 2000), não sendo possível justificar a sua escolha. O trabalho de (TELLO, 2002) propõe o método *OntoMetric* que oferece uma ferramenta de *software* denominada *OntoMetricTool* para facilitar a sua

2.7.1 Métricas para Medir a Idoneidade de Ontologias

O método *OntoMétric* proposto por Tello (2002), cujo esquema básico do processo está ilustrado na figura 11, permite a medida de idoneidade para cada uma das ontologias candidatas que possam ser utilizadas no sistema, onde na seleção de ontologias, oriundas de servidores de ontologias, páginas *web* e bases de conhecimento, as informações armazenadas nestas ontologias devem seguir um marco conceitual denominado ‘marco multinível de características’.

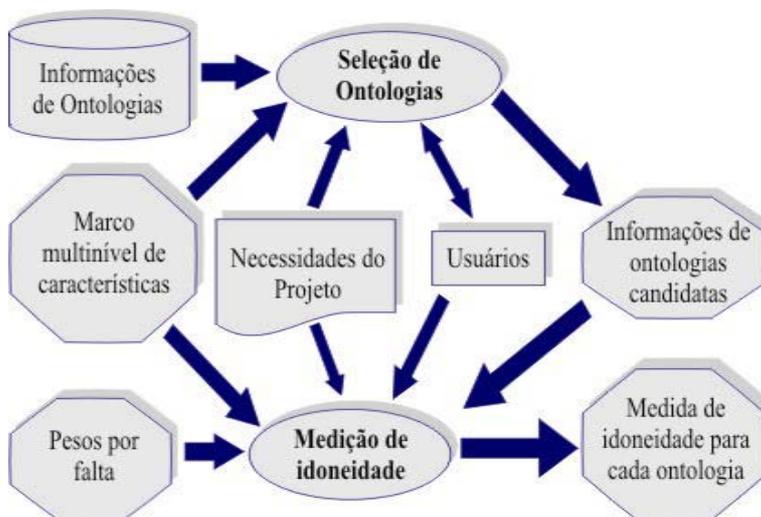


FIGURA 11 - ESQUEMA BÁSICO DO PROCESSO ONTOMETRIC

Fonte: Adaptado de Tello (2002).

Os sub-processos indicados no esquema executam os seguintes procedimentos:

- **Seleção de ontologias:** a seleção é efetuada a partir de informações de ontologias existentes em servidores de ontologias e de páginas *web*. Uma vez identificado o conjunto de dimensões, fatores e características que descrevem o domínio das ontologias com as instâncias de cada uma de suas dimensões, o passo seguinte é escolher as ontologias que vão ser consideradas candidatas para serem usadas no projeto. Este conjunto de ontologias candidatas é formado por todas aquelas que satisfaçam os requisitos estabelecidos pelo usuário, de acordo com as necessidades do sistema. A seleção é feita pelo método de decisão multicritério, utilizando o método das hierarquias analíticas – AHP (Analytic Hierarchy Process), por ser uma maneira poderosa e flexível para a tomada de decisões em problemas multidimensionais ou multicritérios complexos. Este método ajuda as pessoas na aquisição de conhecimento sobre o problema em questão. Quantifica os

julgamentos subjetivos e força a avaliação e comparação de alternativas com respeito aos critérios estabelecidos (TELLO, 2002);

- **Marco multinível de características:** para que o processo se viabilize são necessárias as seguintes entradas:
 - a) O marco multinível de características que forma uma taxonomia de características, agrupadas para usar em um determinado sistema ou projeto;
 - b) Valores obrigatórios exigidos pelas ontologias a serem fornecidos pelo usuário;
 - c) Informação sobre as ontologias existentes que são descritas conforme o marco Multinível de características.
- **Pesos por falta ou defeito:** o *método OntoMetric* oferece ao usuário sem muita experiência uma valoração inicial de importância para os critérios, uma árvore multinível de características com pesos (AMCP) por defeito. Os pesos atribuídos aos nós da AMCP são obtidos a partir de apreciações de importância de dez especialistas, familiarizados no desenvolvimento de aplicações baseadas em ontologias.

2.8. UNL

A UNL (Universal Networking Language) é uma linguagem artificial universal na forma de redes semânticas para computadores para expressar e trocar toda espécie de informação (UCHIDA & ZHU, 2001). É composta por um conjunto de ferramentas desenvolvidas para que os computadores possam entender toda forma de conhecimento: escrito, verbal, gráfico e imagens (BÉRTOLLI JR *et al*, 2003). A UNL tem origem na ONU – Organização das Nações Unidas, mais especificamente na Universidade das Nações Unidas, no ano de 1996. Com essas ferramentas é possível a conversão de informações escritas em linguagem natural para UNL e desta para as demais línguas integrantes do sistema, permitindo a troca de informações entre pessoas na sua língua nativa. A UNL consiste de palavras universais, do inglês Universal Words (UW), Relações, Atributos e Base de Conhecimento (UCHIDA & ZHU, 2001). O computador do futuro necessitará da capacidade de processar conhecimento, sendo que o processamento do conhecimento por meio de um computador leva ao pensamento e julgamento de humanos usando conhecimento de humanos.

Isto é necessário para fazer o processamento baseado em conteúdo. Computadores precisam ter conhecimento para processar conhecimento. Isto é também necessário para ter uma linguagem para processar conteúdos como humanos (UCHIDA & ZHU, 2001).

2.8.1. Sistema UNL

A UNL tem como maior desafio permitir o acesso à informação e conhecimento de pessoas dos mais diferentes países do mundo, representando a informação sentença por sentença, representadas por um grafo que contém conceitos como nós e relações como arcos (UCHIDA & ZHU, 1999).

As UW's constituem o vocabulário da UNL, as quais são unidas semanticamente, formando uma expressão UNL. Cada UW se refere a um conceito e a combinação entre conceitos, ligados por relações e modificados por atributos expressa o significado de uma sentença. As UW's são compostas de uma palavra principal, sem acentuação, seguida de uma lista de restrições e são classificadas como **básicas**, **restritas** e **extras** (BÉRTOLLI JR *et al*, 2003:

- a) **Básicas:** são as palavras principais sem a lista de limites e são expressas por uma palavra em inglês, representando uma gama de conceitos que não precisam necessariamente estar conectados;
- b) **Restritas:** são as palavras principais com a imposição de algum limite que permitam denotar um único conceito para este rótulo;
- c) **Extras:** são palavras universais com conceitos que não são encontrados em inglês e que são introduzidas como categoria extra.

Cada sentença é representada por uma expressão em UNL; são compostas de relações binárias entre UW's. As relações são a união entre duas UW's representadas por rótulos compostos por dois ou três caracteres e são chamados de “**relações-rótulo**” ou “**RL**”(relation label).

Para a escolha de um inventário de relações existem vários fatores a considerar:

- a) Uma condição necessária é quando uma palavra universal tem relações entre mais de duas outras UW's, devendo cada rótulo da relação ser ajustado para poder identificar cada uma delas no seu contexto;

b) Uma condição suficiente é quando há relações entre **UW's**, devendo cada rótulo da relação ser ajustado para poder compreender o papel de cada **UW** somente se referindo ao rótulo da relação.

No anexo I são mostrados 41 tipos de **relações** que podem ocorrer entre palavras universais (UCHIDA & ZHU, 2001).

Da mesma forma que as **UW's** e as relações são usadas para representar conceitos, os **atributos** das **UW's** são usadas para dar subjetividade aos conceitos, dando idéia de tempo, posição, ênfase e atitude, qualificando os conceitos.

Os rótulos de atributos ou **AL** (*Attribute Label*) expressam informações adicionais, as quais, em geral, são restritivas, sobre as **UW's**.

Os **AL's** servem para limitar o significado de uma **UW** genérica, particularizando o seu significado. Informações adicionais tais como tempo verbal, aspecto, intenção ou estrutura sentencial são exemplos de atributos específicos de uma **UW**. A representação genérica de um **AL** é dada pela **UW**, seguida por tantos atributos quantos forem os sugeridos na sentença de origem. Cada um deles é identificado pelo símbolo inicial “@”. Desta forma, a representação genérica de uma **UW** com **n** atributos tem a forma:

UW.@atrib1.@atrib2.....@atribn

Os **AL's** limitam o poder de expressão de uma **UW**. Desta forma a sentença, por exemplo “*Pedro fechou a porta*”, é representada em **UNL** por:

agt(closet.@past,Pedro)

obj(shut.@past.door.@def)

Neste caso tem-se que @past indica o tempo passado e o @def indica uma **UW** já referenciada.

A **UNL** possui também uma base de conhecimento para armazenar as possíveis uniões entre conceitos, onde cada entrada recebe um valor chamado grau de certeza que varia de 0 a 255. Quando o grau de certeza for zero, significa que a entrada é falsa, não havendo possibilidade da relação ocorrer e, quando for positivo, a entrada é verdadeira e quanto maior for o seu valor, maior a possibilidade de ocorrência (BÉRTOLLI JR *et al*, 2003).

A **UNL** tem também em sua composição, servidores lingüísticos, editores **UNL** e visualizadores **UNL** (UCHIDA, 2001).

Os servidores são responsáveis pela conversão e desconversão de documentos através do *conversor* “ENCO”, em inglês “ENCOⁿverter” e o *desconversor* “DECO”, ou

“DECOnterter”. Integra também, neste módulo, um dicionário de linguagem e uma gramática. Para escrever um documento em UNL é necessário utilizar um editor UNL que usa o sistema conversor ENCO para transformar o documento em UNL. Para que diferentes pessoas, em diferentes lugares possam visualizar o documento escrito e convertido para UNL, é necessário utilizar o sistema desconversor DECO que transforma o documento em UNL para a linguagem desejada (UCHIDA, 2001) & (UNDL, 2003). A figura 12 mostra o sistema.

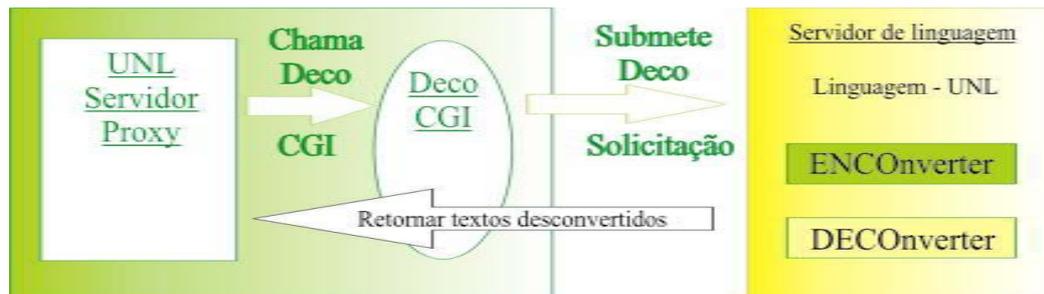


FIGURA 12 – SISTEMA UNL.

2.9. CONSIDERAÇÕES FINAIS

As características dos problemas, focados nesta tese, envolvendo buscas e estruturação de dados fragmentados em diferentes fontes, direcionados para uma determinada área de interesse, cujos conhecimentos devem estar formalmente representados, são melhores trabalhadas com as técnicas de IA discutidas neste capítulo. No próximo capítulo será tratado o sistema Multiagente SMAGS, que é uma ferramenta dotada de recursos para gerenciar, de forma autônoma, serviços especiais que são normalmente desenvolvidos por seres humanos, tais como: análise, pesquisa e classificação de informações, em um ambiente informatizado.

3. O SISTEMA MULTIAGENTES SMAGS

Este capítulo apresenta o Sistema Multiagentes denominado SMAGS, com sua arquitetura e suas funções, integrado ao Gerador Inteligente de Sistemas e executando várias funções de forma compartilhada.

3.1. SMAGS

O SMAGS é uma ferramenta dotada de recursos para gerenciar, de forma autônoma, serviços especiais que são normalmente desenvolvidos por seres humanos, tais como: análise, pesquisa e classificação de informações, em um ambiente informatizado. O SMAGS ou Sistema Multiagentes para Gestão de Sistemas foi desenvolvido e implementado na linguagem Delphi, podendo operar em ambientes Windows, com equipamentos e sistemas operacionais tradicionais, reduzindo custos e outras limitações existentes em outros sistemas. O SMAGS atua sobre o meio através da execução de aplicações especializadas, modificação automática de bancos de dados, geração de consultas, gráficos e relatórios, bem como, a comunicação de mensagens visuais ou sonoras, refletindo as percepções e ações sobre o sistema gerenciado (CORDEIRO, 2001).

O SMAGS, desenvolvido em Cordeiro (2001), apresenta características, contendo vários módulos ou subsistemas, onde o subsistema principal, é um *gerente de aplicações*, denominado Agente Gerente Virtual, cuja célula pode ser aperfeiçoada, (o que está acontecendo no presente trabalho) com a agregação de outros agentes com recursos inteligentes mais avançados. Não será apresentado neste capítulo um fluxograma das rotinas e sub-rotinas do SMAGS, porque as possibilidades seriam infinitas, em função da flexibilidade do sistema para cada aplicação específica.

Os elementos que constituem o SMAGS, são mostrados na sua arquitetura na figura 13, sendo que todos eles cooperam mutuamente, através de eventos locais ou externos por meio da rede (*intranet* ou *internet*).

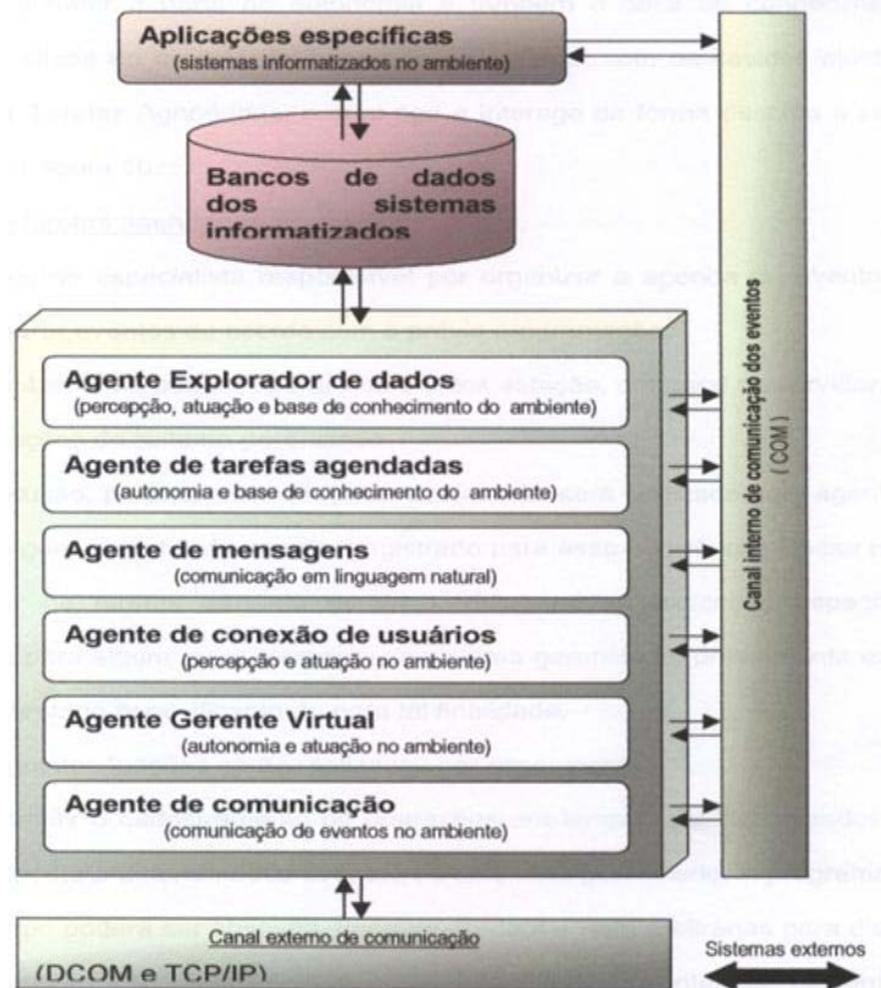


FIGURA 13 – ARQUITETURA DO SMAGS.

A sensibilidade do SMAGS às ocorrências no sistema gerenciado é dada através de eventos especiais, que podem ser disparados por qualquer um de seus elementos, tais como: por aplicações especiais e pelos próprios agentes do SMAGS. Esses eventos são comunicados, em primeira instância, dentro do sistema operacional, sensibilizando os subsistemas que estiverem registrados para determinados eventos. Em segunda instância, os eventos são comunicados ao longo da rede, através do agente de comunicação de eventos que, desse modo, converte eventos locais em eventos externos e vice-versa, a fim de propagar as informações para todos os destinatários registrados.

O subsistema principal é o *gerente de aplicações*, que deve ser programado ou configurado para monitorar e atuar sobre determinadas aplicações no sistema gerenciado. Os outros subsistemas são elementos genéricos que podem existir em qualquer ambiente gerenciado, de modo a colaborar com o gerente de aplicações, fazendo dele um subsistema

mais especializado e mais fácil de construir e adaptar, para cada caso específico de gerenciamento.

O *gerente de aplicações* é notificado sobre ocorrências especiais no seu domínio, ou ambiente, de gerenciamento. Esses eventos são classificados em alguns tipos elementares, que serão ainda subclassificados em operações. Os principais eventos que podem transitar no ambiente gerenciado são:

- Evento de notificação genérica: usado por aplicações específicas do ambiente gerenciado;
- Evento de notificação de tarefa agendada: usado para informar que uma tarefa agendada está na hora de ser executada;
- Evento de transmissão de dados: usado para transferir resultados de consultas do subsistema explorador de dados;
- Evento de transmissão de mensagem: usado para comunicar mensagens para usuários;
- Evento de controle de conexão de usuário: usado para notificar uma conexão ou desconexão de usuário ao sistema;
- Evento de execução de operação: usado para executar tarefas especializadas nos subsistemas constituintes.

Todos os eventos podem alcançar o *gerente de aplicações*; desse modo ele poderá ser treinado para responder automaticamente a eles, ou, então, pode ser programado para reagir de forma especializada a determinadas situações.

As operações de configurações mencionadas são realizadas através de uma tela específica do SMAGS, conforme é mostrado na figura 14.

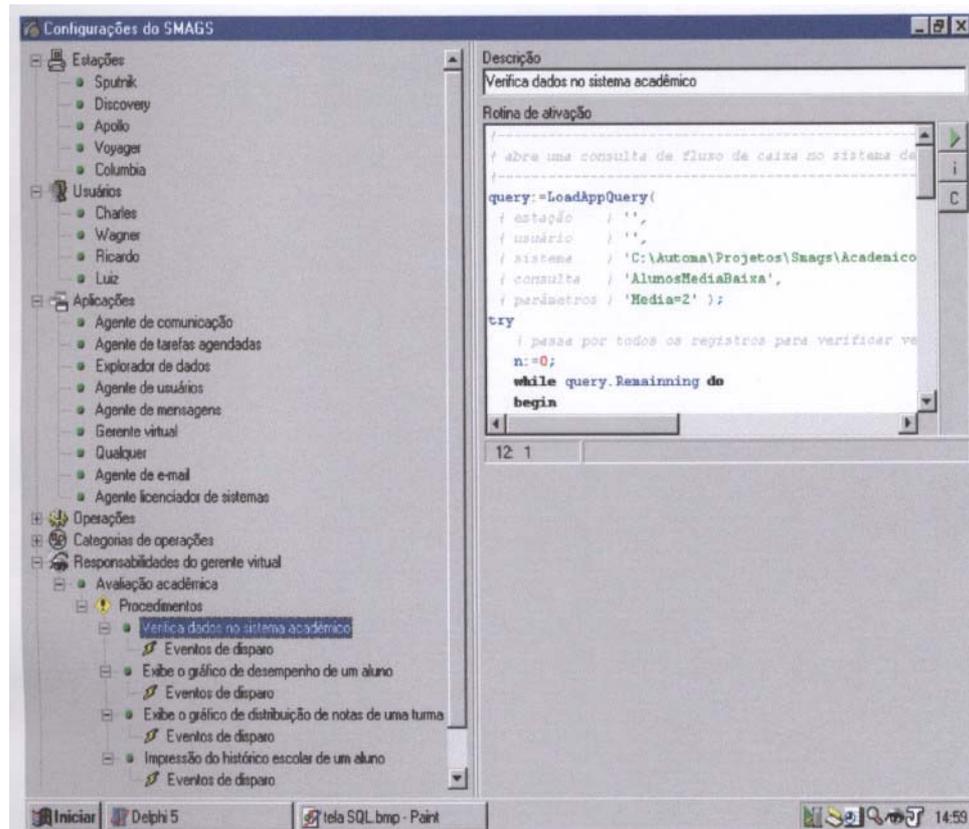


FIGURA 14 – TELA PARA CONFIGURAÇÃO DO SMAGS. FONTE: CORDEIRO (2001).

A seguir são apresentados detalhes de cada um dos elementos que constituem o SMAGS.

- **Aplicações Específicas:** as aplicações específicas funcionam na rede, no ambiente gerenciado, consultando e modificando os dados de bancos de dados, bem como realizando operações não dedicadas a bancos de dados. Essas aplicações não precisam, necessariamente, ser modificadas para funcionarem de modo integrado ao SMAGS, haja vista que sua integração pode ocorrer através da simples execução destes programas ou através do banco de dados que eles operam. O explorador de dados poderia, assim, extrair dados e modificar dados que estão sendo compartilhados por tais aplicações em uso no ambiente gerenciado. Por outro lado, com adaptações nessas aplicações específicas, pode-se tornar a interatividade do gerente de aplicações muito mais intensa, permitindo, por exemplo, que ele realize operações especiais dentro dessas aplicações ou perceba operações especiais que estão sendo feitas nessas aplicações;

- **Agente Explorador de Dados:** é a aplicação de comunicação com bancos de dados, responsável pela extração, modificação e comunicação de informações do banco de dados. O explorador de dados é o módulo ou subsistema que desempenha as seguintes funções no ambiente gerenciado:

a) Permitir que os usuários construam e executem consultas especiais para um banco de dados qualquer, existente no ambiente gerenciado, através de uma interface amigável. Este módulo ou subsistema usa tecnologias diversas de conectividade com bancos de dados através de ODBC (Open DataBase Connectivity) e de bibliotecas específicas para acesso nativo, e ampliável, para diversos tipos de bancos de dados. Alguns tipos de bancos de dados que podem ser nele conectados são: Oracle, SQL Server, SyBase, DB2, InterBase, PostgreSQL, MySql, dBase, FoxPro, Ms-Access, Informix, ZIM e outros;

b) Permitir que os usuários construam formatações de relatórios e gráficos, com auxílio de uma estrutura de menus. As consultas, relatórios e gráficos são construídos por pessoas que têm conhecimento da estrutura do banco de dados e são montados de tal forma que os usuário possam executá-las parametrizadamente, sendo assim mecanismos genéricos de extração de dados. Estes mecanismos também poderão ser disparados, automaticamente, por eventos do SMAGS;

c) Permitir que os usuário construam e executem rotinas de modificação de dados, baseadas em parâmetros que podem ser solicitados para o usuário ou recebidos através de eventos do SMAGS, de modo a realizar alguma função específica no sistema gerenciado, atualizando seus dados. A operação do extrator de dados pode ser feita independentemente dos outros subsistemas do SMAGS. Mas quando feita em conjunto, permite que o SMAGS perceba as operações que o usuário está fazendo, correlacionando-as com outros dados, permitindo assim algum tipo de aprendizado ou treinamento. O extrator também poderá receber comandos do SMAGS e transmitir informações do banco de dados, através dos eventos locais e externos. O extrator de dados também poderá integrar aplicações específicas do ambiente gerenciado, de modo a desempenhar funções que poderiam ser programadas dentro de tais sistemas, mas que, de modo integrado ao SMAGS, torna a interface com o banco de dados mais homogênea e permite que essas aplicações específicas tenham interatividade com o SMAGS. O explorador de dados opera em uma estação da rede, em geral no servidor, podendo, entretanto, operar em outras estações;

- **Agente de Tarefas Agendadas:** é o agente especialista responsável por organizar a agenda de eventos, bem como disparar eventos de acordo com a prévia programação. Este

subsistema pode operar em uma única estação ou em várias estações do sistema gerenciado. A execução, propriamente dita, das tarefas não é realizada pelo agente, mas sim, por algum receptor de eventos registrado para essa ocorrência. Desse modo, o agendador de tarefas torna-se genérico, delegando a execução específica de atividades para algum outro elemento do sistema gerenciado, previamente existente ou implementado especificamente para tal finalidade. As seguintes funções serão realizadas por esse agente:

a) Permitir o cadastramento de operações em tempos pré-determinados ou em resposta a determinados eventos do ambiente gerenciado. A programação de tempo poderá ser absoluta, informando data e hora arbitrárias para disparar a operação; ou, então, poderá atuar, informando um intervalo de tempo que sucederá algum evento específico do sistema gerenciado, como, por exemplo, uma determinada execução de atualização de dados, uma execução de um determinado programa, a ocorrência de algum feriado, final de semana ou outro evento de calendário, etc.;

b) Registrar e perceber eventos associados ao disparo de tarefas agendadas;

c) Cronometrar o tempo para o disparo de tarefas agendadas;

d) Notificar o sistema, através de eventos específicos, informando que uma tarefa está na hora de ser executada;

e) Operar de forma autônoma, através de eventos do SMAGS;

- **Agente de Mensagens:** é o agente especialista responsável por elaborar, armazenar, classificar e entregar mensagens de usuários na rede. Este subsistema é análogo a um cliente, integrado a um servidor de *e-mail* da *internet*. Porém, este subsistema realiza funções adicionais especializadas para operar em cooperação com o SMAGS, atuando em todas as estações do ambiente gerenciado. As seguintes funções são desempenhadas por esse agente:

a) Permitir a pré-definição de mensagens especiais (visuais ou sonoras), que serão usadas para notificar os usuários do sistema gerenciado, acerca das ocorrências percebidas tanto por pessoas, quanto pelo SMAGS;

b) Permitir a elaboração de quaisquer mensagens, cujo remetente é o usuário ativo no agente de conexão de usuários, e o destinatário é o usuário registrado como usuário do sistema gerenciado;

c) Enviar e receber mensagens entre usuário do sistema gerenciado, utilizando eventos do agente de comunicação;

d) Notificar, classificar e arquivar mensagens recebidas localmente e recebidas no destinatário;

e) Arquivar mensagens enviadas para cada destinatário;

f) Operar autonomamente através de eventos disparados por outros subsistemas do SMAGS;

- **Agente de Conexão de Usuários:** é o agente especialista responsável por ativar e monitorar as conexões de usuários na rede. Esse agente mantém registros dos usuários autorizados para operar o SMAGS. Ele desempenha as seguintes funções no ambiente gerenciado:

a) Solicitar senha para ativar a conexão de um usuário em uma estação qualquer;

b) Permitir desativação do usuário corrente e reconexão de outro usuário na mesma estação;

c) Monitorar os usuários conectados no sistema;

d) Registrar, opcionalmente, um *log* das operações que estão sendo realizadas pelos usuários no sistema gerenciado;

e) Realizar todas as suas operações, de modo autônomo, através de eventos recebidos e enviados no SMAGS.

- **Agente Gerente Virtual ou Gerente de Aplicações:** este é o agente principal do sistema, responsável por integrar as aplicações e outros agentes que fazem parte do sistema. Este agente pode ser aperfeiçoado em nível de “inteligência”. Ele é constituído de uma base de conhecimento, que por sua vez foi montada através de regras lógicas e um dispositivo de inferência, que analisa as regras, de modo a reagir automaticamente aos eventos que estiverem trafegando no sistema. Esse agente é executado em qualquer ponto do sistema gerenciado. Sua concepção é essencialmente cliente servidor, porque deve ser capaz de centralizar sua base de conhecimento e distribuir suas ações na rede. Suas atribuições são:

a) Registrar, receber e codificar os eventos de entrada que são relevantes para suas ações. Geralmente, todos os eventos devem chegar até o gerente;

b) Aplicar as entradas codificadas, eventos de operações e eventos de dados à etapa de resolução, aprendizado e reação autônoma, avaliando numericamente as informações presentes, mudando seu estado quando necessário e decodificando seus resultados para agir sobre o meio;

c) Agir sobre o sistema gerenciado, executando outras aplicações, produzindo eventos para executar operações nos subsistemas do SMAGS, realizando tarefas específicas do ambiente gerenciado;

d) Permitir, opcionalmente, que algum operador humano ou outra aplicação possa interferir em sua autonomia ou monitorar seu estado, quando possível e desejado. Isso pode ser feito através de uma interface de usuário ou mesmo através de eventos especiais do SMAGS.

- **Agente de Comunicação:** é o agente especialista responsável por converter e entregar eventos entre aplicações e outros agentes na rede. Este subsistema funciona em segundo plano, de modo autônomo e transparente, em todos os pontos do ambiente gerenciado; embora apenas um destes pontos seja configurado como servidor de comunicação, todos os pontos podem receber eventos locais e propagá-los para que o servidor os entregue ao destinatário, quando existirem receptores de eventos interessados externamente. Cada ponto da rede mantém uma lista de receptores externos e o endereço do servidor. Assim, apenas eventos que devem ser entregues em pontos externos trafegam na rede, do remetente para o destinatário diretamente, caso essa topologia seja mais interessante. Os eventos locais são pacotes de dados que trafegam através de algum sistema de comunicação do sistema operacional. Os eventos externos são pacotes de dados que trafegam através da rede local ou *internet*. O agente de comunicação realiza as seguintes funções no ambiente gerenciado:

- a) Registrar receptores, ou ouvintes, para determinadas categorias de eventos;
- b) Receber eventos locais, identificar o destinatário e avaliar se é local ou externo;
- c) Propagar eventos locais no sistema operacional, quando o remetente e o destinatário forem locais;
- d) Converter eventos locais em eventos externos, quando o remetente for local e o destinatário for externo;
- e) Converter eventos externos em locais, quando o remetente for externo e o destinatário for local;
- f) Monitorar a presença do servidor e dos outros clientes.

Este agente trabalha em conjunto com agente de mensagens.

Do exposto, constata-se uma outra característica importante do SMAGS que é a sua capacidade de aproveitar os *softwares* existentes nas diferentes organizações, bem como os dados armazenados nos respectivos locais, o que amplia a capacidade do sistema e economiza recursos.

O SMAGS é dotado de um sistema de reconhecimento de voz e as mensagens resultantes de suas ações são comunicadas de forma escrita ou falada por um agente especial que pode, também, assumir fisionomia característica, de acordo com a natureza da notícia, conforme está ilustrado na figura 15.

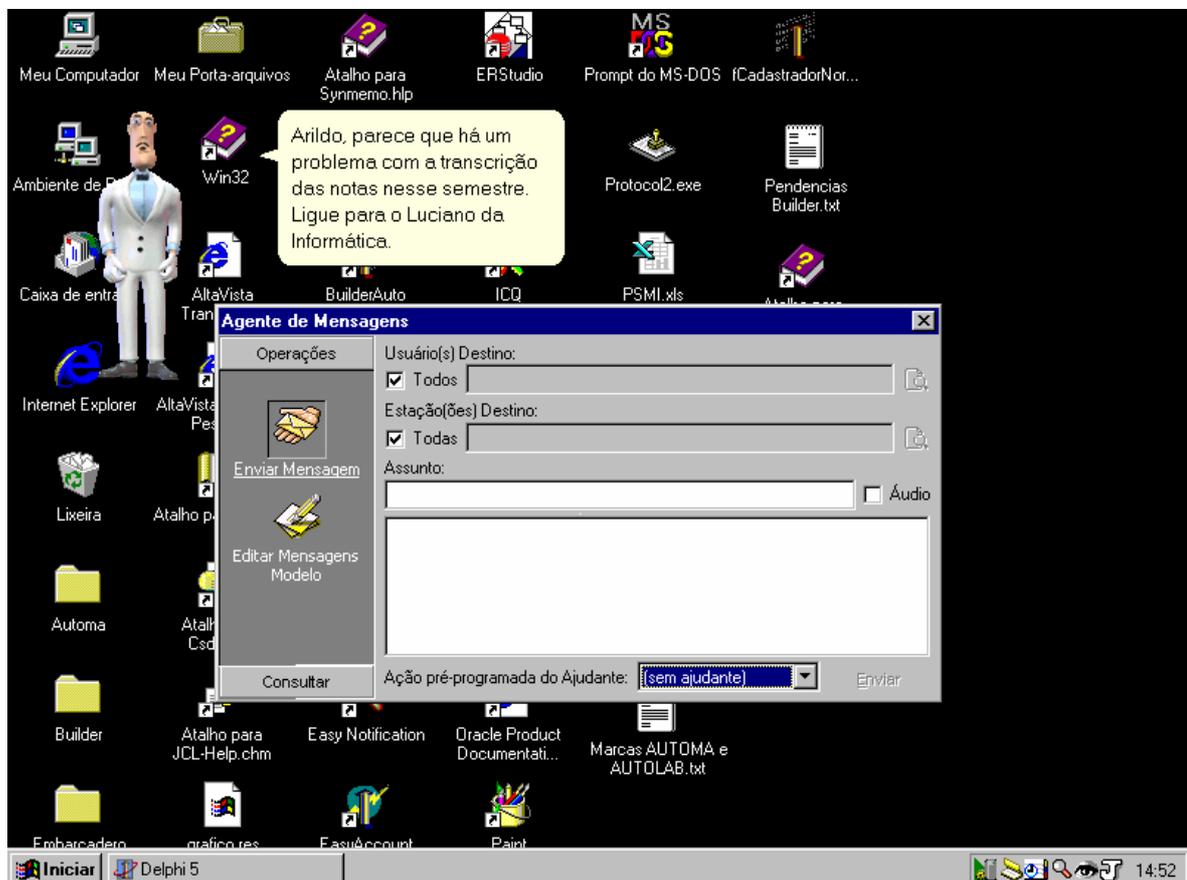


FIGURA 15 – AGENTE DE MENSAGENS ESCRITAS E FALADAS DO SMAGS – FONTE: CORDEIRO (2001).

3.2. CONSIDERAÇÕES FINAIS

Esta capítulo apresentou o sistema Multiagentes SMAGS. O próximo capítulo apresentará o modelo proposto, em que ao SMAGS foram incorporados os agentes GEDI (Gerenciador Eletrônico de Documentos Inteligente), MARERF (Meta Agente de Raciocínio e Extração de Regras *Fuzzy*) e um Agente Buscador Inteligente.

4. MODELO PROPOSTO

O objetivo do presente capítulo é apresentar o Gerador Inteligente de Sistemas com Auto-Aprendizagem para Gestão de Informações e Conhecimento, para apoio a tomadas de decisões, denominado “GISAAGIC”.

4.1. GISAAGIC

O GISAAGIC proposto é uma ferramenta CASE que integra outras ferramentas e incorpora várias técnicas de IA (Inteligência Artificial) e WS (*Web Semântica*), tornando-a capaz de gerar sistemas de informações inteligentes. Esta ferramenta possui uma modelagem orientada a objetos com conceitos de Data e Text Mining (Mineração de Dados e de Textos), Redes *Neuro-Fuzzy*, *Web Semântica* e Ontologias. A tecnologia empregada no desenvolvimento do *software* possui elementos de expansão que permitem acomodar novos sistemas e novos elementos na modelagem, viabilizando, no caso em foco, a construção de uma ferramenta inteligente com vários novos recursos de Tecnologia da Informação (TI). O objetivo principal é o de fornecer informações organizadas e atualizadas em tempo real que possam apoiar as tomadas de decisões gerenciais com informações estruturadas e contextualizadas. Estas informações permitem a formulação de hipóteses de possíveis resultados, tendo como base resultados anteriores e as novas situações de ocorrências no contexto atual, com dados e conhecimentos que justifiquem as hipóteses formuladas a partir de novos conhecimentos adquiridos através de buscas inteligentes. A ferramenta é capaz de aprender com o usuário, agregando novos conhecimentos, facilitando a recuperação de casos de sucesso e as suas comparações com novas possibilidades que os avanços tecnológicos viabilizam. A sua arquitetura está representada na figura 16.

O GISAAGIC gera aplicações de forma interativa, orientada e estruturada de acordo com as necessidades da área de interesse do usuário ou de sua organização. O sistema incorpora as tecnologias do SMAGS, as metodologias de um Gerenciador Eletrônico de Documentos Inteligente (GEDI) que trabalha de forma integrada com um Meta-Agente de Raciocínio e Extração de Regras *Fuzzy* (MARERF) que incorpora também conceitos de UNL, cujas ações compartilhadas viabilizam as tarefas do agente buscador.

eventos no ambiente) ou através de um estímulo verbal que é dado pelo usuário através de uma linha de busca.

As ferramentas foram implementadas na filosofia multiplataforma, na linguagem Delphi, podendo operar com diversos bancos de dados e sistemas operacionais, Oracle, Sql Server, Interbase, Ms-Access e outros, nos sistemas Windows e Linux.

Para facilitar o entendimento de todas as fases do processamento, o sistema foi subdividido em três subprocessos, os quais serão descritos a seguir.

4.1.1. Subprocesso de Geração do Sistema de Informação

Após a modelagem do sistema de informação a partir do modelo relacional proposto para o objetivo considerado e do dicionário de dados, o sistema é gerado e está apto à entrada de dados e ao processamento das informações.

Neste sub-processo acontece a geração do sistema de informação, que é peculiar ao GISAAGIC e que prepara as bases para o subprocesso de busca. A figura 17 mostra o fluxo para geração do sistema de informações.

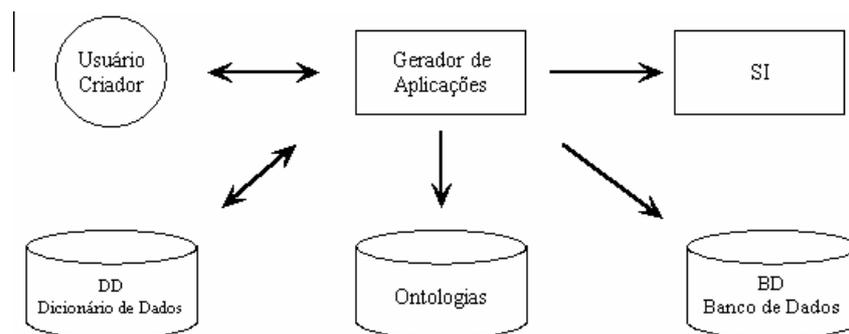


FIGURA 17 – FLUXO PARA GERAÇÃO DE SISTEMA DE INFORMAÇÕES.

O **Usuário Criador** utiliza o Gerador de Aplicações, convertendo o modelo relacional proposto e armazenando as informações no Dicionário de Dados; Informações complementares, tais como o *design* da interface do usuário, são adicionadas ao **Dicionário de Dados (DD)**; A partir disto, o Gerador de Aplicações gera o **SI**, bem como o **BD**

necessário às operações do mesmo. O Gerador também fornece o protótipo da ontologia de domínio a partir das definições existentes no **DD**.

A figura 18 mostra a parte da Arquitetura Geral envolvida na geração do Sistema de Informação.

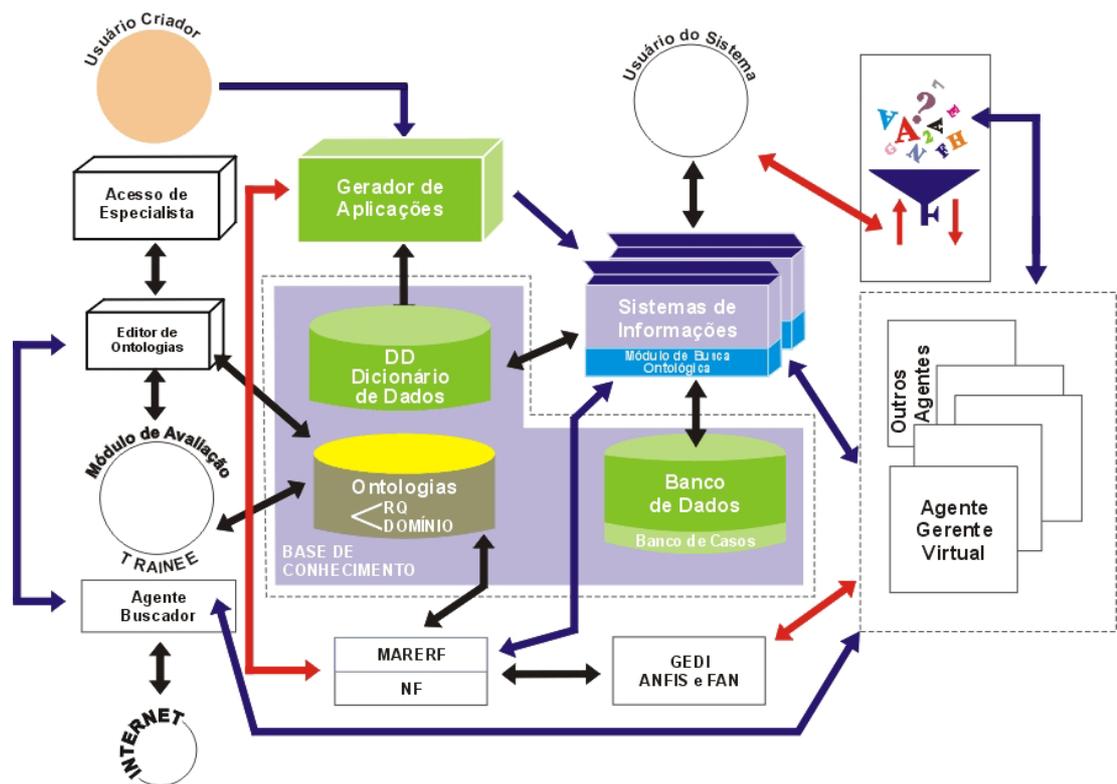


FIGURA 18 – PARTE DA ARQUITETURA GERAL ENVOLVIDA NA GERAÇÃO DO SISTEMA DE INFORMAÇÃO.

4.1.2. Subprocesso de Busca

Neste subprocesso estão considerados os elementos e interações necessárias a buscas inteligentes. Os termos que foram considerados para a criação do sistema constituem um conjunto que é considerado o ponto de partida para um *glossário* do sistema de informação (não confundir com o vocabulário do processo de busca relatado mais adiante), mesmo que não se disponha de uma ontologia do domínio neste instante.

Após a geração do sistema, o **SI** (Sistema de Informação) está à disposição do usuário e os dados começam a ser cadastrados no banco de dados do sistema, ou seja, o sistema de informação começa a ser utilizado para o objetivo para o qual foi criado. Desta forma, os dados que são cadastrados, juntamente com os *metadados* (utilizados para a criação do sistema), constituem o conjunto de termos utilizados para direcionar as buscas.

A seguir, o conjunto de termos ou glossário é resumido e ordenado pela frequência de ocorrência no escopo do **SI** (Sistema de Informação).

O agente MAREF obtém as informações do banco de dados (**BD**), acontecendo a extração de características (filtragem e preparação) para a criação de “*grupos*” de busca.

4.1.2.1. Fluxo para Geração do Grupo de Busca

Este fluxo explica o tratamento interno sobre os dados existentes no **BD** e no **DD** de forma a obter-se palavras ou frases válidas para alimentar o processo de busca, acontecendo no sentido do ambiente interno do GISAAGIC para o ambiente externo. A geração dos termos para alimentar a busca mediada pelo GEDI é feita de acordo com o fluxo mostrado na figura 19. Tanto os dados existentes nos registros do **BD** quanto as definições do dicionário de dados são aproveitados para gerar o que é denominado de **Grupo de Busca**. As fases ou algoritmos definidos para este fluxo são:

- 1) Processamento de Termos do Dicionário de Dados;
- 2) Preparação dos Campos do Banco de Dados;
- 3) Extração de Características;
- 4) Montagem do Conjunto-Base de Termos;
- 5) Ponderação Ontológica;
- 6) Inclusão de Vizinhança;
- 7) Ordenação Inversa pelo Critério de Importância;
- 8) Geração do Grupo de Busca;
- 9) Inclusão na Agenda do GEDI.

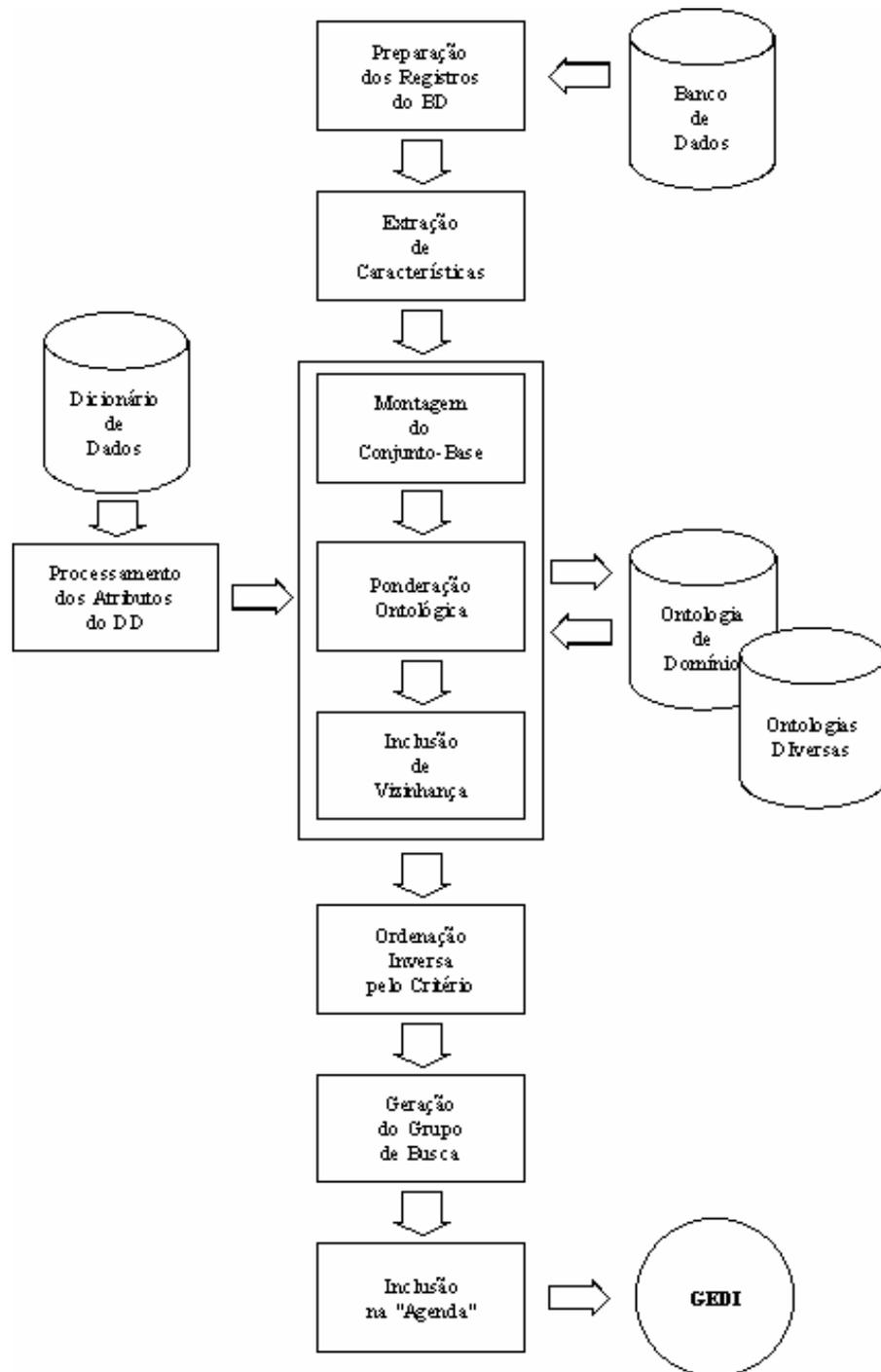


FIGURA 19 – FLUXO PARA A GERAÇÃO DO GRUPO DE TERMOS DE BUSCA.

A seguir são mostrados os detalhes de cada um dos processos:

1) Processamento de Termos do Dicionário de Dados:

Para a geração do Sistema de Informação, é necessário alimentar o Gerador de Aplicações com o Dicionário de Dados. O Gerador a partir daí pode construir a interface GUI (Graphical User Interface) para interação com o usuário do **SI** e o Banco de Dados para a aplicação. O **DD** constitui-se também em um conjunto de termos considerados para a fase da *inclusão de vizinhança*. No instante da Geração do **SI**, aproveita-se também o **DD** para gerar um protótipo da ontologia de domínio. Desta forma, se possui um grau de importância dos termos presentes no **DD**, facilitando a montagem do **conjunto-base** de termos. Caso sejam utilizadas outras ontologias, é preciso passar pela fase de montagem do conjunto-base para haver a ponderação ontológica.

2) Preparação dos Campos do Banco de Dados

Outra fonte de dados para a geração do Grupo de Busca são os registros existentes no **BD**. Os campos disponíveis são configurados em uma interface de acordo com certos critérios que serão definidos manualmente, tais como:

- a) Atribuição de prioridade de tabelas;
- b) Atribuição de prioridade de campos;
- c) Utilização de um registro por vez, um grupo de acordo com um atributo-chave ou todos os registros na tabela do banco de dados;
- d) Quais atributos serão participantes do processo;
- e) Número mínimo e máximo de termos por atributo;
- f) Utilização de conteúdo de registros atendendo determinada condição.

Esta configuração antecede o próximo passo que vem a ser a Extração de Características.

3) Extração de Características

A extração de características visa processar o conteúdo já preparado proveniente do banco de dados. Esta extração encerra a metodologia de *Text Mining* e é semelhante àquela feita pelo Buscador e o GEDI no retorno das informações obtidas externamente. Supressão de

stopwords e símbolos, substituição de siglas, limiar de poda de termos, conteúdo em branco, são executadas e no final (da mesma forma na qual é gerado o Vocabulário) o **conjunto-base** de termos.

Este conjunto-base é mais complexo que o gerado via **DD**, e pode ser entendido como uma matriz de *termos* por *subconjuntos*. A composição do conjunto-base é feita desta forma por *subconjuntos representados pelos registros* (como se cada subconjunto representasse um “documento“, mantendo-se uma abordagem paralela com o processo da busca inteligente).

Como os subconjuntos são compostos por registros, na definição dos mesmos devem ser definidos critérios para que o número de subconjuntos não aumente demasiadamente, diminuindo assim a performance de cálculo. Tais critérios podem envolver:

- Proporção do número de registros por subconjuntos;
- Quantidade fixa de subconjuntos definidos aleatoriamente;
- Limiar de poda de subconjuntos.

4) Montagem do Conjunto-Base de Termos

Na montagem do conjunto-base calculam-se a frequência de termos (**TF**) e a frequência inversa de termos (**IDF**), bem como a composição **TFIDF** nos moldes já definidos anteriormente. Ao conjunto-base agregam-se desta forma os dados quantitativos.

5) Ponderação Ontológica

A ponderação ontológica segue também os passos do processo da busca inteligente. A idéia agora é valorizar os termos do conjunto-base existentes na ontologia de domínio (ou ontologias diversas) para reforço (ou penalização) de acordo com o contexto estabelecido. Esta fase também é parametrizada para escolha da ontologia a ser aplicada.

6) Inclusão de Vizinhança

Nesta fase é feita a comparação dos termos com a ontologia e verifica-se então a vizinhança dos termos coincidentes. Aos termos no conjunto-base adicionam-se os termos que são vizinhos na ontologia. Alguns critérios devem ser levados em conta para a definição da inclusão:

- Quantidade limite de termos vizinhos;
- Nível de profundidade da vizinhança;
- Tipo de vizinhança (*is-a*, *part-of*).

A ontologia pode também conter termos que não fazem parte do contexto do domínio, (denominados termos fora de contexto). São também expressos como conceitos, porém mantendo relações de negação ou complementar (*not is-a*) com o termo em questão. Tais termos podem ser bem utilizados para refinar o processo da busca mais tarde.

7) Ordenação Inversa pelo Critério de Importância

Neste passo o conjunto-base é ordenado pelos termos de maior ao menor grau de ponderação obtido na montagem do conjunto-base, agora em conjunto com os termos vizinhos inclusos, de forma a dar prioridade ao processo de busca.

8) Geração do Grupo de Busca

O conjunto-base se transforma agora no **grupo de busca**. O grupo de busca conterá os termos do conjunto-base, porém agora são explicitadas as combinações possíveis com os termos vizinhos inclusos. Esta combinação pode ser parametrizada de acordo com os critérios:

- Combinação n-a-n termos;
- Quantidade máxima de termos para combinação;
- Uso de operador para agrupar os termos (espaço em branco, '+' (mais) ou '-' (menos), de acordo com a sintaxe dos *search sites*).

O termo fora de contexto pode ser relacionado a um termo principal no grupo de busca mediante o uso de ‘-’ (menos), de forma a direcionar a futura busca em documentos excluindo o termo fora de contexto.

9) Inclusão na Agenda do GEDI

O grupo de busca agora está pronto para ser adicionado à agenda do GEDI. De acordo com a configuração deste agente, cada tarefa de busca agendada será feita no período determinado, dando a partida a todo o processo de busca inteligente.

4.1.2.2. Fluxo de Processamento dos Resultados

Este fluxo trata das operações envolvidas com o processamento de resultados obtidos, acontecendo no sentido do ambiente externo para o interno ao sistema. O agente MARERF interage com a ontologia de domínio para obter a ponderação ontológica e fazer a inclusão de vizinhança nos grupos de busca. A mensagem é enviada através do Agente Gerente Virtual, agendando tal mensagem e direcionando-a para o GEDI para execução do processo de busca. O conteúdo da mensagem é o grupo de busca. A mensagem é enviada através do Agente Gerente Virtual. Cada uma das mensagens contém uma entrada do grupo de busca, as quais são agendadas e direcionadas ao buscador inteligente.

1) Agente Buscador Inteligente

As buscas inteligentes envolvem as seguintes operações:

- a) O agente buscador inteligente executa a busca lançando mão dos recursos de busca programados no mesmo e retorna o resultado em forma de mensagem ao Agente Gerente Virtual, direcionada ao GEDI;

- b) O agente GEDI processa o resultado das buscas, colocando a relevância dos documentos, considerando aspectos de mineração de textos e ponderação ontológica, sendo o aprendizado armazenado em sua rede neural (ANFIS ou FAN) para futuras classificações de relevância;
- c) Os resultados relevantes são enviados como mensagens para o Agente Gerente Virtual direcionar para o MARERF;
- d) O agente MARERF recebe as mensagens das buscas efetuadas e armazena os links relevantes no **BD**;
- e) O agente MARERF compara os termos ou conjunto de termos mais solicitados para busca, e envia tais termos para a ontologia a fim de serem considerados como novos conceitos.

O fluxo geral do subprocesso de busca, mostrando o sentido da geração e o sentido do processamento dos resultados, está ilustrado na figura 20.

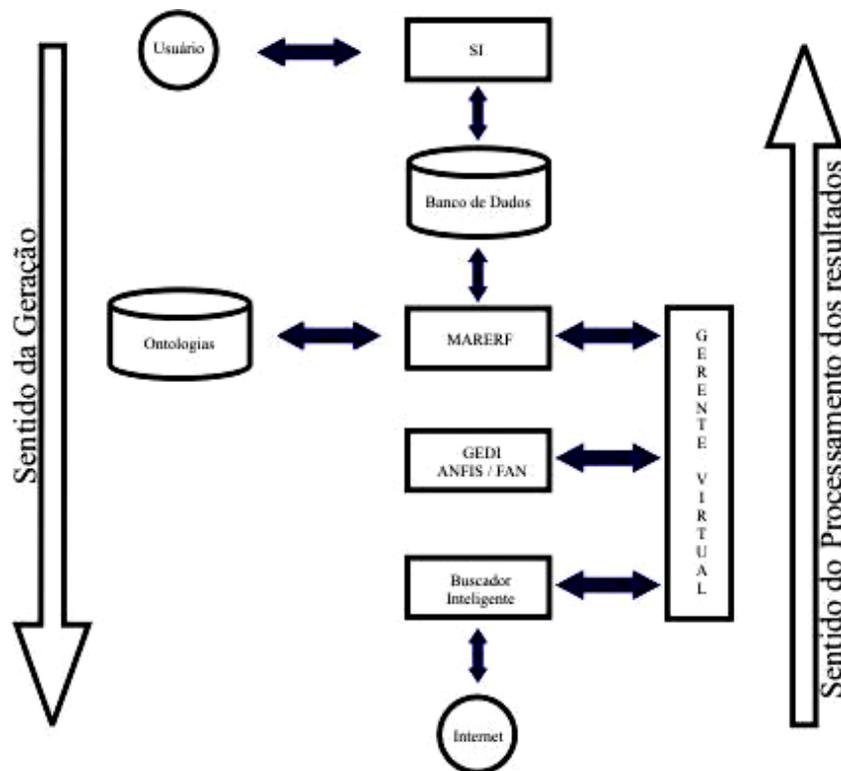


FIGURA 20 – FLUXO GERAL DO SUBPROCESSO DE BUSCA.

Os dados brutos são refinados, utilizando conceitos de *Data Mining e Text Mining*, para que fiquem apenas os dados pertinentes ao assunto de interesse.

a) Limpeza de Documentos e Extração de Características

Como acontece nos processos de *Data Mining*, os dados brutos precisam ser refinados para que forneçam os dados pertinentes ao problema, conforme (PYLE, p.92, 1999) com algumas adaptações. A limpeza de documentos deve extrair certas características, como por exemplo: *tags* e símbolos de acentuação presentes em um documento **HTML** e deixar apenas o conteúdo de texto significativo. A limpeza envolve, ainda, a extração de símbolos que não são importantes e a conversão de certos padrões erráticos (caracteres de acentuação trocados) que podem estar presentes.

Outro aspecto explorado é a interpretação do documento **HTML** e as referências existentes de *links* de outras páginas que são armazenadas pelo GEDI, de modo a proporcionar outros *sites* para efetuar a busca. Desde que se considere a hipótese de que um *site* terá mais valor quanto mais novas **URL**'s ele proporcionar ao agente de busca, esta forma de valoração pode ser considerada como parâmetro para avaliação da relevância de um documento. Denomina-se de frequência de *links* (*Link Frequency*) **LF** este parâmetro que quantificará a ramificação de um documento em termos de produzir novos documentos. Se for considerado dentro de um documento **HTML** com código em JavaScript, o parâmetro **URL** da função **window.open** poderá fornecer também *links* para serem utilizados pelo agente buscador.

b) Extração de *Stopwords*

Stopwords são palavras não relevantes no processamento do texto para o significado da busca, de forma que o GEDI utiliza uma lista previamente codificada pelo especialista (EBECKEN. N. F. F.; LOPES, M. C. S. *et al*, p. 347). Os termos presentes nesta lista são extraídos do texto normal. Termos como pronomes, artigos e preposições são retirados do texto. Outras variantes podem ser incluídas como o resgate de termos ou siglas que poderão ser deixadas no documento.

c) Cálculo da Frequência de Termos

Após a limpeza e extração de *stopwords*, calcula-se a Frequência de Termos (term frequency ou TF), (SALTON, 1989; EBECKEN *et al*, p.355, 2003; CHOUDARY &

BHATTACHARYYA, p.1, 2002; SHAH *et al*, p.2, 2002). A TF se refere aos dados quantitativos absolutos e relativos dos termos apurados no documento. A idéia é considerar termos que têm relevância em termos da sua frequência de ocorrência no documento. Esta técnica, apesar de ser simples, apresenta uma complexidade computacional aproximadamente linear e facilmente escalável. A TF faz parte do que se denomina de técnicas tradicionais de mineração de textos.

d) Poda de Termos não Representativos

A *poda* de termos que estão abaixo de certo limiar absoluto ou relativo em termos de TF é feita de forma a simplificar o conjunto de palavras para o espaço de representação posterior do documento. A suposição desta técnica é que termos que aparecem pouco não são representativos para predizer a categoria do documento em uma posterior classificação.

e) Elaboração do Vocabulário

Na elaboração do vocabulário, dois ou mais termos que aparecem em mais de um documento servem para uma lista destes termos presentes no conjunto de documentos. Palavras que são coincidentes em mais de um documento promovem uma espécie de *poda*, de forma a reduzir o espaço de representação dos documentos para a entrada da rede *neuro-fuzzy*.

f) Comparação do Vocabulário com Ontologia de Domínio

O vocabulário é interessante também para que os termos retornados de um certo conjunto de documentos sejam comparados com termos da ontologia de domínio do problema considerada para o sistema de informação. Enquanto vocabulário, os conceitos estão soltos, não caracterizando uma ontologia. Porém, conceitos novos que não estejam presentes na ontologia poderão integrá-la, de forma a caracterizar uma evolução ou aprendizado da ontologia. Conceitos novos ficarão à espera de uma posterior classificação na ontologia, de forma que o especialista irá ligar ou não tais conceitos mediante relações apropriadas (existentes ou não). A comparação pode ser efetuada em um passo anterior à *poda*, para evitar perda de termos que seriam relevantes na consideração do documento.

g) Cálculo da Frequência Inversa de Documento

O cálculo da Frequência Inversa de Documentos (*Inverse Document Frequency*) ou **IDF** fornece um parâmetro a mais para classificar documentos, agora no sentido de que os termos que aparecem poucas vezes em um documento são significativos para a classificação em um vetor de documentos que irá servir de conjunto de treinamento para uma rede *Neuro-Fuzzy*. Porém, o índice **IDF** não serve como um valor representativo isolado, mas sim quando combinado com **TF**. Assim como a **TF**, a **IDF** também faz parte das técnicas tradicionais.

É utilizada uma combinação dos métodos denominados tradicionais **TF** e **IDF** em conjunto com uma representação de ontologia de domínio. A representação de uma ontologia traz aspectos significativos em termos de busca. Um *link* entre dois conceitos, conforme o seu tipo, pode refletir em maior ou menor importância para os conceitos relacionados. Um conceito com vários *links* que entram ou saem do mesmo é entendido como tendo bastante importância dentro de uma ontologia. Neste caso, uma ontologia em RDF ou DAML-OIL (ou até mesmo em UNL), que represente um domínio no qual o termo de busca está inserido, irá agregar mais informação ao processo, possibilitando uma busca de documentos mais refinada.

Estes conceitos embasam a proposição para este trabalho, da combinação também das técnicas tradicionais e da ontologia com a técnica de frequência de *link* (**LF**), rede *neuro-fuzzy* com inferências e graus de pertinências.

A contagem e valoração relativa dos conceitos de acordo com seus *links* será então multiplicada pelo parâmetro **TFIDF** (multiplicação de **TF** por **IDF**) e também pelo parâmetro **LF**, de forma a obter-se um indicador para os termos considerados no vocabulário que servirão para a entrada da rede no processo de treinamento.

O escore para o vetor de representação é obtido a partir da multiplicação direta da frequência de termos **TF** pela frequência inversa **IDF**, denominado na literatura de **TFIDF**. Alguma ponderação pode ser considerada em termos de favorecer um ou outro parâmetro na participação deste parâmetro.

h) Classificação *a Priori* pelo Usuário

A classificação *a priori* do usuário significa que, num primeiro momento de uma pesquisa a respeito de um conjunto de termos-chave, o usuário classifica os documentos retornados pelo agente de busca de acordo com sua importância. A classificação é feita de forma binária (ou seja, atribui-se '1' ao documento relevante e '0' ao documento não relevante). Outras formas de valoração da saída são exploradas (uma faixa de valores *fuzzy*

pode ser atribuída na prioridade de classificação). Esta pré-classificação serve para formar o conjunto de treinamento da rede *neuro-fuzzy*. Os parâmetros obtidos do documento, ou seja, **TF**, **IDF**, **TFIDF** e **LF**, serão as entradas da rede, e a classificação binária será a saída.

i) Ganho Semântico

Após o uso dos métodos tradicionais, faz-se a verificação de cada documento com relação à ontologia do domínio do assunto de interesse. Assim, a partir de uma ontologia definida, faz-se a valoração dos termos presentes na mesma. Esta valoração irá ponderar os termos comuns no documento, aproximando um determinado documento do seu domínio conforme mais próximo ele esteja da ontologia.

No processo, para cada termo presente em um documento é verificada a sua existência na ontologia. Caso esteja presente, o peso respectivo é acumulado no “*fator de ganho semântico*” do documento. Se não estiver presente, este fator de ganho semântico será penalizado por um fator constante.

A valoração das relações entre termos de um documento **UNL** é feita pelo método de (SHAH, CHOUDHARY & BHATTACHARYYA, 2002). Como em UNL tem-se uma variedade de relações que pode existir para as Universal Words **UW's** e, desde que se considere um documento **UNL** como sendo uma ontologia, cada relação pode vir a aumentar ou não ter efeito na valoração de um conceito dentro da ontologia.

Finalmente, o fator de ganho semântico é multiplicado para cada escore obtido pelo termo do documento na composição da **TF** mais a **IDF** (ou mais considerando a **LF**), e assim obtém-se o vetor do documento que o irá representar na coleção.

A medida da qualidade de representação de vetores é feita pelo método da informação mútua indicado pelos autores já citados e também por (JIN e HAUPTMANN, 2001), que se baseia na teoria da informação. O valor da informação mútua entre considerar uma variável aleatória para o conteúdo total de um documento e, num passo seguinte, outra variável aleatória para o vetor do documento que o represente, fornece o ganho de informação, sendo este maior quanto mais representativo for tal vetor.

A metodologia, a seguir exposta, resume um melhoramento da representação de vetores de documentos, partindo das técnicas usuais de *text mining* e adicionando o que se denominou de *ganho semântico* a partir de uma ponderação de termos presentes em uma ontologia correlacionada ao contexto de uma certa coleção de documentos.

Assume-se, então, uma coleção de documentos \mathbf{D} , contendo os documentos d_1, d_2, \dots, d_n . O conteúdo dos documentos representados pela totalidade de termos existentes denomina-se de vocabulário. Um vocabulário V irá conter os termos t_1, t_2, \dots, t_m . Têm-se na totalidade m termos e n documentos.

Ao valor absoluto de ocorrências de um certo termo v_i em um documento d_j será representado por \mathbf{TF}_{ij} (*Term Frequency*). Entende-se que \mathbf{TF}_{ij} seja assim uma matriz com m termos dispostos em linhas e n documentos dispostos em colunas.

Para o processo de obtenção de termos do vocabulário, pode-se adotar a restrição de poda de termos não representativos. Assim, define-se uma nova matriz \mathbf{TFA}_{ij} (*Term Frequency Absolute*) onde:

$$\mathbf{TFA}_{ij} = \{v_i \in V \mid \mathbf{TF}_{ij} \geq \delta\}, \forall d_j \in D \quad (15)$$

Ou seja, para um certo termo v_i , o conjunto \mathbf{TFA}_{ij} terá apenas elementos nos quais a frequência absoluta seja maior que um valor de restrição δ .

A seguir, obtém-se a expressão para o cálculo da matriz de frequência relativa de termos, ou seja, \mathbf{TFR}_{ij} (*Term Frequency Relative*).

$$\mathbf{TFR}_{ij} = \frac{\mathbf{TFA}_{ij}}{\sum_{i=1}^m \mathbf{TFA}_{ij}}, \forall d_j \in D \quad (16)$$

Agora procede-se ao cálculo da frequência inversa de termos \mathbf{IDF}_i . A frequência inversa está relacionada ao conjunto de documentos \mathbf{D} , e quanto menor a ocorrência n_t de um certo termo v_i na coleção de n documentos, maior será o fator de frequência inversa. Assim:

$$\mathbf{IDF}_i = \log\left(\frac{n}{n_t}\right) \quad (17)$$

Com as matrizes \mathbf{TFR}_{ij} e \mathbf{IDF}_i , pode-se agora obter a matriz do índice conjunto \mathbf{TFIDF}_{ij} :

$$TFIDF_{ij} = TFR_{ij} \cdot IDF_i \quad (18)$$

ou expressando de forma matricial,

$$\mathbf{TFIDF} = \mathbf{IDF}^T \mathbf{TFR} \quad (19)$$

Até aqui foi descrito o processo comum para extração de informações de um documento. A partir daqui, começa-se a envolver uma ontologia agregada ao contexto da coleção de documentos. Define-se então uma ontologia \mathbf{O} contendo conceitos o_1, o_2, \dots, o_n entre os conceitos. As relações entre os conceitos representam *links* entre os mesmos. Cada conceito o_k da ontologia contém uma frequência f_k de *links*. Esta frequência inclui os *links* de um conceito com os outros conceitos ligados diretamente e indiretamente. O peso w_k será uma quantidade que é definida como o inverso da frequência f_k de *links*

$$w_k = \frac{1}{f_k}, \quad \forall o_k \in \mathbf{O} \quad (20)$$

A idéia desta técnica é que apesar de existirem conceitos que estão fortemente relacionados na ontologia com outros termos, estes conceitos não irão agregar mais informação ao que existe. Mas os termos que estão fracamente conectados poderão agregar mais alguma informação sobre como um certo documento poderá ser melhor representado. Define-se então o ganho semântico GS_j como sendo o coeficiente dado por:

$$GS_j = \begin{cases} \varphi \sum w_k & , o_k \in D \\ GS_j \cdot \kappa & , o_k \notin D \text{ com } 0 < \kappa < 1 \end{cases} \quad (21)$$

Quando $o_k \in \mathbf{D}$, ou seja, quando o termo da ontologia \mathbf{O} estiver presente no conjunto de documentos \mathbf{D} , obtém-se o somatório dos pesos w_k dos termos existentes para o ganho semântico, multiplicado por um fator de escala φ . Em contrapartida, se $o_k \notin \mathbf{D}$, aplica-se um fator de penalização κ sobre o ganho semântico.

Com o ganho semântico definido, pode-se agora obter a matriz $TFIDFG_{ij}$, onde:

$$TFIDFG_{ij} = TFIDF_{ij} \cdot GS_j \quad (22)$$

ou na forma matricial:

$$\mathbf{TFIDFG} = \mathbf{TFIDF} \cdot \mathbf{GS}^T \quad (23)$$

Cada coluna da matriz $TFIDFG_{ij}$ agora pode ser considerada como um vetor do documento d_j , modificado pela análise quantitativa efetuada sobre uma ontologia que esteja agregada ao contexto da coleção de documentos, de forma a melhorar a qualidade de representação de tais vetores.

j) Atuação de Rede *Neuro-Fuzzy*

A rede *Neuro-Fuzzy* é estruturada de forma a conter o número de entradas correspondente ao número de palavras do vocabulário criado. Em se tratando da rede ANFIS, cada entrada possui um conjunto de funções de pertinência *fuzzy*. Estas funções começam o treinamento com valores padrão, de forma que é possível se determinar, a partir de valores de entrada, se o mesmo está contido na faixa lingüística (baixa, média, alta). Um número maior de funções de pertinência pode ser também explorado. O treinamento da rede ANFIS permite a atualização das funções de pertinência de cada entrada, sendo que, após o treinamento, tais funções assumem comportamentos diferentes para cada uma.

Na saída utiliza-se, como visto no tópico anterior, valores binários. Após o treinamento, a tendência é que as classificações posteriores assumam valores intermediários. A determinação de um limiar (ou mesmo outro conjunto *fuzzy*) para a saída permite definir a classificação.

Devido à peculiaridade da rede ANFIS em ser definida como um aproximador universal de funções (Azevedo, 2000), estipula-se uma saída escalonada, de forma a permitir mais níveis na identificação da relevância de um documento, além da forma binária, explicada anteriormente. Além do mais, a própria estrutura da rede permite a identificação, através de regras **se-então** dos termos e seus valores lingüísticos, aqueles que estão contribuindo para a relevância de um documento.

O algoritmo de treinamento utilizado baseia-se no filtro de Kalman estendido, sendo adequado para tratar com dados temporais. Os processos de consultas subseqüentes irão

contribuir para treinamentos posteriores da rede ANFIS, evidenciando assim o caráter adaptativo da mesma.

Outros tipos de rede *Neuro-Fuzzy* também foram considerados para esta fase, tal como o FAN.

A rede treinada proporciona novas buscas pelo agente e, portanto, uma diretriz de classificação. Assim, a rede aprende com as consultas subseqüentes de acordo com a relevância dada pelo usuário no *feedback* da consulta.

4.1.3. Subprocesso de Manutenção

No subprocesso de manutenção acontece a manutenção das ontologias, ocorrendo com a interferência do usuário especialista. A figura 22 mostra o fluxo do processo.

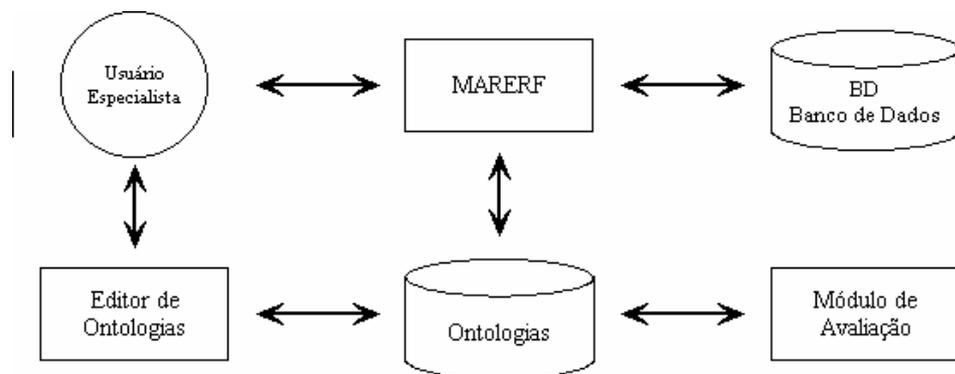


FIGURA 22 – FLUXO DO PROCESSO DE MANUTENÇÃO.

Este subprocesso acontece da seguinte forma:

- O agente **MARERF**, através das buscas, acrescenta novos conceitos à ontologia e executa a função de validação para *feedback* ao Usuário Especialista;
- O agente **MARERF**, solicitado pelo Usuário Especialista, pode executar mecanismos de extração de regras para identificar padrões existentes no contexto (BD mais ontologias);

- c) O **Usuário Especialista** faz as modificações necessárias nas ontologias através do **Editor de Ontologias**; tais modificações podem incluir novas relações aos conceitos provindos do MARERF, bem como conceitos e relações alimentados pelo especialista.
- d) O **Módulo de Avaliação** permite que se compare a evolução da ontologia ao longo do tempo, utilizando processos de métrica de ontologias.

A figura 23 mostra em destaque as partes envolvidas na arquitetura principal do GISAAGIC.

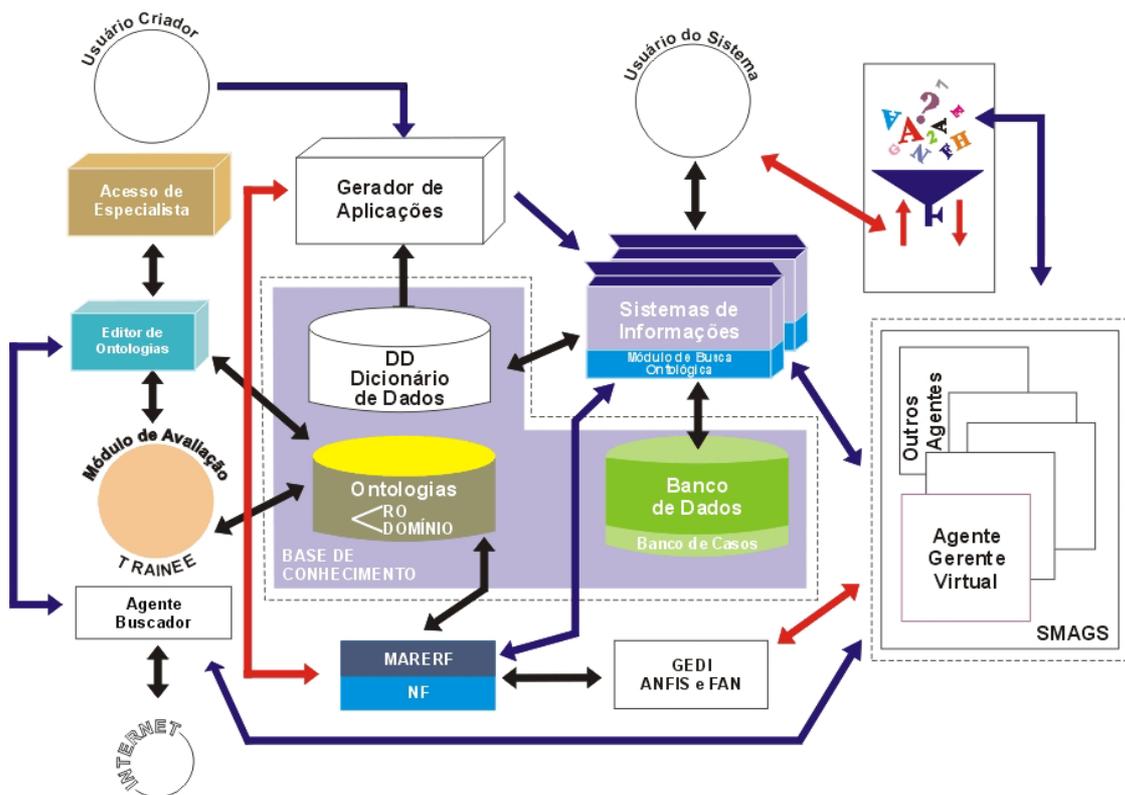


FIGURA 23 – PARTES DA ARQUITETURA DO GISAAGIC ENVOLVIDAS NO SUBPROCESSO MANUTENÇÃO.

4.1.3.1. Meta-Agente MARERF

O Meta-Agente de Raciocínio e Extração de Regras *Fuzzy* MARERF é responsável pelas atividades de raciocínio que vão além da responsabilidade do GEDI. O GEDI pode, através da rede *neuro-fuzzy* e utilizando as características obtidas quantitativamente da ontologia de domínio, sugerir a relevância de um certo documento ou não dentro de consultas subsequentes geradas pelo agente buscador. Porém, para que sejam feitas inferências sobre os conceitos

existentes na ontologia, outro agente deve estar identificado no conjunto. Assim, o MARERF executa, de forma autônoma, as chamadas *atividades de raciocínio*. Tais atividades incluem de forma geral:

- 1) Silogismo Hipotético
- 2) Identificação de Similaridade
- 3) Validação de Representação
- 4) Extração de Regras

1) Silogismo Hipotético

Na medida em que a ontologia se desenvolve, os novos conceitos que vão sendo agregados e também as instâncias que são adicionadas podem ter novas relações com outros conceitos já existentes. O silogismo hipotético permite que sejam inferidas novas associações a partir de uma associação já existente utilizando a lógica de primeira ordem. A regra geral indica que tendo os elementos **a**, **b** e **c**, e se $a \rightarrow b$ e $b \rightarrow c$, necessariamente $a \rightarrow c$. Tomando-se o exemplo de um planejamento acadêmico, na figura 24 tem-se uma subparte da ontologia indicando **X** como sendo uma instância do conceito Livro. O sistema pode sugerir que uma instância de um livro conste como bibliografia básica ou complementar. Este agente infere então para saber se o livro **X** pode vir a ter tal característica. Através da ontologia, o MARERF parte dos seguintes fatos e regras:

Quadro II - Exemplo Do Processo De Silogismo Hipotético.

FATO	REGRA	RESULTADO
X = LIVRO	LIVRO part_of BIBLIOGRAFIA	X = BIBLIOGRAFIA
	BIBLIOGRAFIA is_a BÁSICA	X = BÁSICA
	BIBLIOGRAFIA is_a COMPLEMENTAR	X = COMPLEMENTAR

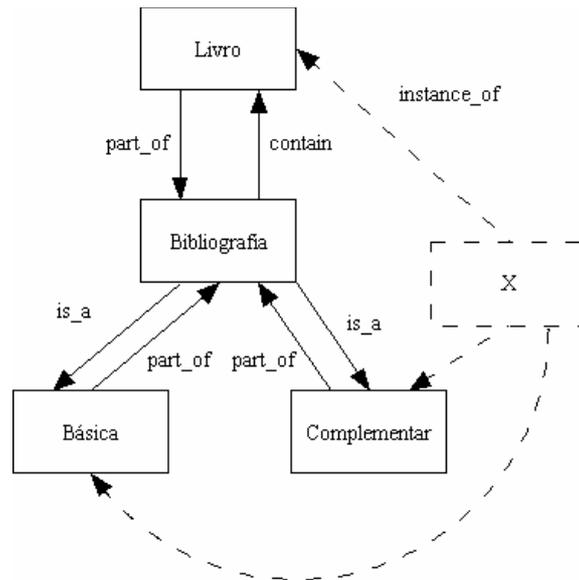


FIGURA 24 – SUBPARTE DA ONTOLOGIA NA VIZINHANÇA DE BIBLIOGRAFIA.

2) Identificação de Similaridade

A identificação de relações indiretas não evidenciadas entre conceitos e instâncias dentro de uma ontologia necessita de uma forma de raciocínio mais apurada. Tomando-se por base dois conceitos distintos quaisquer na ontologia, deve-se saber qual o grau de similaridade que existe entre os mesmos. O agente de raciocínio MARERF faz a pesquisa que pode ser direcionada ou não pelo usuário. O algoritmo básico desta pesquisa deve ter os seguintes passos:

a) **Algoritmo de caminhamento** – deve ser possível o caminhamento em uma ontologia de forma que dois conceitos deverão ser alcançados (ou visitados) nos dois sentidos. O algoritmo de busca pode ser efetuado simultaneamente a partir dos dois conceitos (como os nós de um grafo) até encontrarem um caminho

b) **Função de similaridade** – O conteúdo das instâncias referentes a dois conceitos deve ser comparado. Vários graus de similaridade podem ser obtidos ou calculados. Pode-se estipular um limiar de importância acima do qual o agente sugere a existência de similaridade. São excluídas do algoritmo as *stopwords*.

No escopo do exemplo do planejamento acadêmico, o MARERF pode efetuar aleatoriamente várias pesquisas elaborando pares de conceitos, tais como:

(Ementa, Ementa)
(Livro, Ementa)
(Ementa, Programação)
(Programação, Programação)
(Livro, Livro)

Por exemplo, o agente pode retornar informações na pesquisa para o par *(Ementa, Ementa)* a respeito de duas instâncias de disciplinas **X** e **Y** que são afins entre si, ou seja, possuem semelhanças nas ementas. Na figura 25 tem-se a representação da subparte da ontologia de domínio. As linhas pontilhadas indicam a instanciação dos conceitos. Utilizando a notação de propriedades da orientação a objetos, cada instância possui seu conceito relacionado separado por um ponto (.).

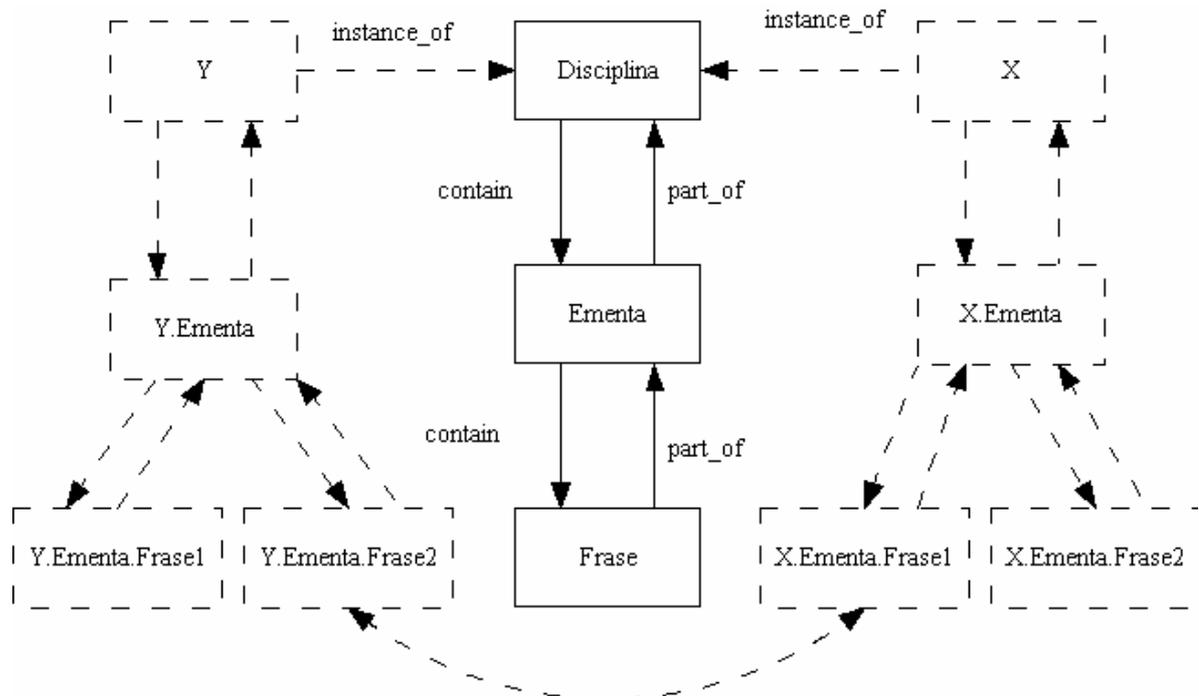


FIGURA 25 – SUBPARTE DA ONTOLOGIA DE DOMÍNIO NA VIZINHANÇA DE DISCIPLINA.

Desde que a representação das instâncias **X** e **Y** esteja correta e completa, cada frase que faz parte de uma ementa de certa instância é visitada, para cada instância de disciplina, e o grau de similaridade é calculado. Um valor próximo de zero indica que não existe nenhuma

similaridade. Um grau próximo de 1 indica igualdade. Como a igualdade entre duas instâncias não é esperada (duas disciplinas exatamente iguais), o valor indica uma inconsistência. Se um valor acima de zero é obtido, o algoritmo mostra também as frases em comum. A similaridade evidenciada pode sugerir indicação de interdisciplinaridade para as disciplinas.

A pesquisa pode ser aprofundada em termos da programação detalhada dos conteúdos das disciplinas. Com o par de conceitos considerado (*Programação, Programação*), procura-se, assim, identificar se existem conteúdos afins. O grau de afinidade indica então a existência de itens de programação assemelhados entre si. Na figura 26 tem-se a representação das instâncias **X** e **Y** novamente, agora com a subparte da ontologia que relaciona disciplina e programação (ou itens de programação). As linhas pontilhadas indicam a instanciação, e *X.Programação1* indica instância de um item de programação (tomando por empréstimo novamente a notação para propriedades de orientação a objeto, separando por um ponto os dois conceitos). O teste de similaridade é feito então entre duas instâncias de itens de programação.

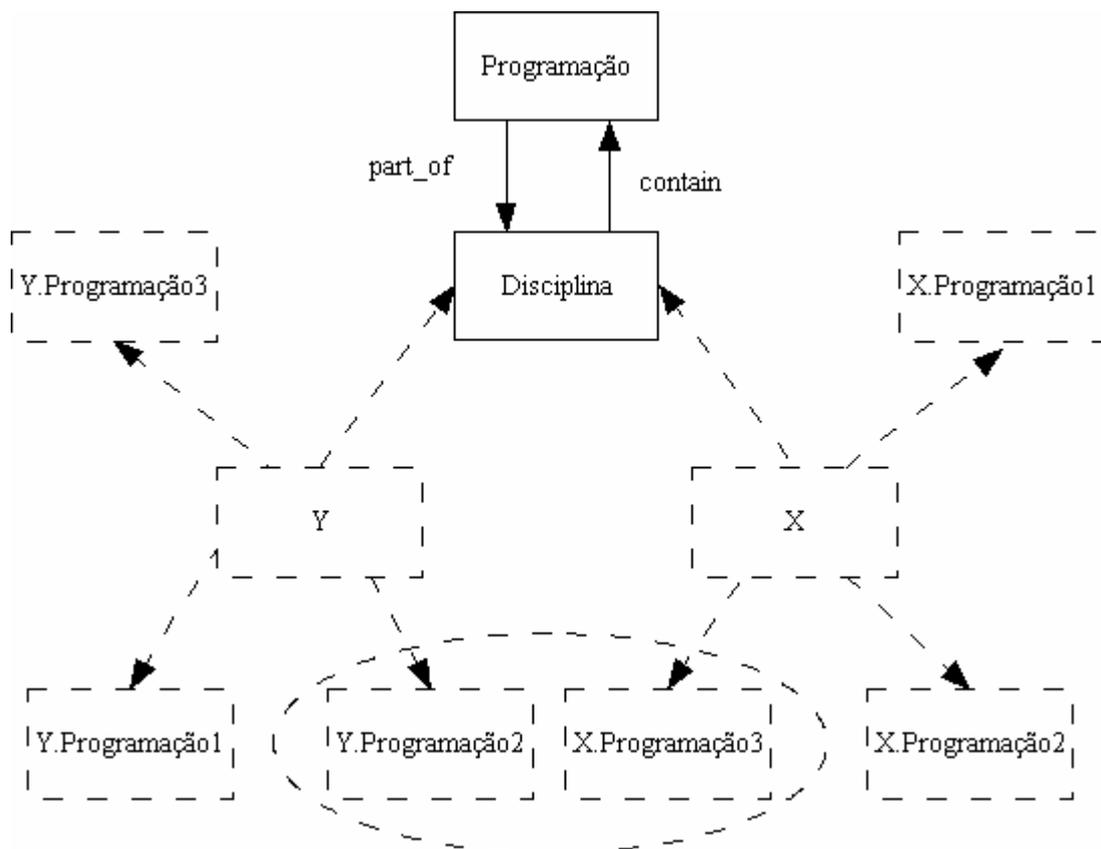


FIGURA 26 – INSTÂNCIAS **X** E **Y** LIGADAS A UMA SUBPARTE DA ONTOLOGIA.

A compatibilidade de um livro com uma determinada ementa de uma disciplina pode ser evidenciada pela identificação de similaridade para o par (*Livro, Ementa*). Desde que se

considere o conceito *Abstract* (ou ainda o conceito *Palavra-chave*), e de acordo com a subparte da ontologia descrita na figura 27, a instância **Z** de um livro contendo o *abstract* pode ser comparada com as frases-chave de uma ementa da disciplina **Y**. Para cada frase-chave pertencente à ementa da disciplina **Y**, o teste de similaridade é feito com o *abstract* do livro **Z**. Novamente tem-se um grau quantitativo que representa quanto similar é um conceito com outro.

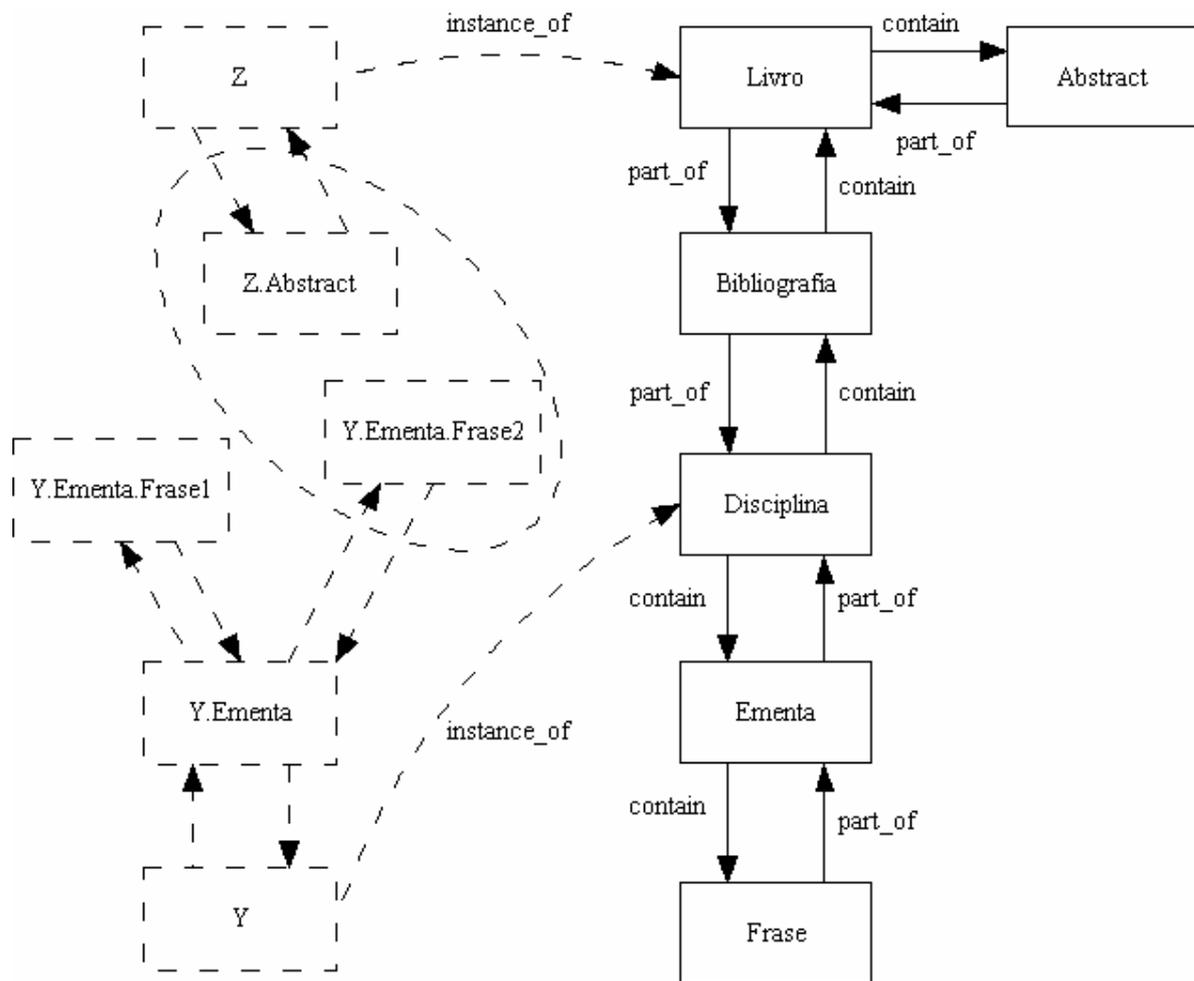


FIGURA 27 – INSTÂNCIA **Z** LIGADAS A SUBPARTE DOS CONCEITOS LIVRO E DISCIPLINA.

Deve-se notar que, na identificação de similaridade, não é necessária a interferência do usuário. O próprio MARERF Agente de Raciocínio efetua as pesquisas percorrendo a ontologia de forma exaustiva. Os resultados obtidos deverão passar por uma validação *a posteriori* do usuário e indicar as linhas de ação que deverão ser decididas para o processo de interdisciplinaridade.

3) Validação da Representação

Um mecanismo de validação verifica se os conceitos representados na ontologia possuem a semântica correta. Certas linguagens de ontologias prevêm a utilização de axiomas embutidos na própria representação da ontologia, que permitem inferir sobre a validade dos conceitos e instâncias representados. O MARERF Agente de Raciocínio executa buscas exaustivas na ontologia, de forma a identificar inconsistências de representação de instâncias. Tais buscas são parametrizadas pelo usuário. A princípio, uma inconsistência pode ocorrer quando uma certa instância possui conceitos relacionados que não estão previstos na ontologia do domínio.

As inconsistências são afinal reportadas para o usuário, que deve tomar as ações corretivas necessárias. As modificações que se fizerem necessárias devem ser feitas a partir do Editor de Ontologias. Uma inconsistência pode também indicar a falta de uma representação na ontologia do domínio.

Um exemplo, ainda referenciando sobre planejamento acadêmico, é descrito a seguir. Conforme a figura 28, tem-se a subparte da ontologia indicando que uma disciplina possui um professor (ou mesmo professores). A busca efetuada pelo MARERF Agente de Raciocínio de uma instância de disciplina X contendo um professor indica que este mesmo professor é autor de livro, estando relacionado ao conceito *Autor* pela instância de livro Y. Não existindo a relação entre *X.Professor* e *Y.Autor*, o MARERF Agente de Raciocínio indica uma inconsistência de representação. O usuário deve explicitar a relação entre os dois conceitos através de associações *is_a*, ou mesmo introduzir um novo conceito denominado *Pessoa*, relacionando os conceitos de *Professor* e *Autor* a este novo conceito através de associações *is_a*.

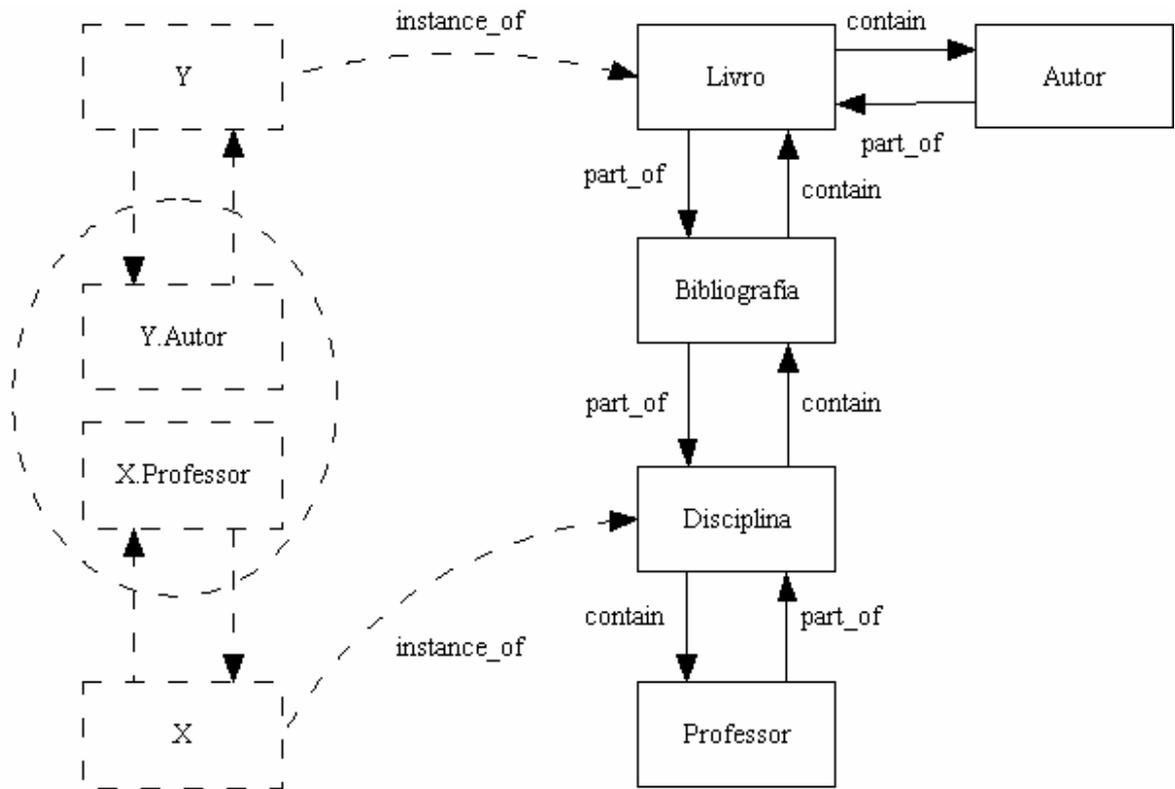


FIGURA 28 – SUB PARTE DA ONTOLOGIA UTILIZADA PARA O EXEMPLO DE VALIDAÇÃO.

4) Extração de Regras

A extração de regras é uma das atividades desenvolvidas pelo MARERF que utiliza os dados providos do banco de dados do SI, da Ontologia e também do GEDI. O objetivo aqui é o de detectar certos padrões subjacentes nos dados na forma de regras, de maneira que possam ser relatadas ao usuário. O módulo extrator de regras contém uma rede *neuro-fuzzy* que, por sua vez, é treinada com as informações de entrada e saída selecionadas pelo usuário. Após o treinamento é feito o processo de extração de regras utilizando a abordagem proposta por (MITRA e PAL, 1997) e MITRA, 2003), e pela abordagem do algoritmo FUZZYRULEXT, (AZEVEDO *et al*, p.241, 2003), que executa um *backtracking* sobre os pesos da rede *neuro-fuzzy*, identificando na forma de regras porque a rede inferiu uma certa saída para um determinado padrão apresentado na entrada. Desta forma, as regras vão sendo extraídas para cada padrão do conjunto de treinamento, e são expressas na forma *se-então*. O processo acontece de forma automatizada, sem interferência do usuário na fase de geração de regras.

Como exemplo para o planejamento acadêmico, o usuário pode definir como atributos de entrada como sendo a titulação e experiência do professor num conjunto de várias disciplinas, grau de dificuldade das disciplinas e recursos didáticos utilizados. Para cada amostra, define-se a saída como sendo o aproveitamento dos alunos em termos de média de notas e faltas. Após o treinamento, a extração de regras pode evidenciar algumas informações, tais como:

SE EXPERIÊNCIA = ALTA E RECURSOS = MUITOS ENTÃO NOTA = ALTA

SE TITULAÇÃO = MÍNIMA E GRAU DE DIFICULDADE = ALTO ENTÃO NOTA = BAIXA E FALTA=MÉDIA

As variáveis utilizadas podem ser convertidas, se for o caso, nos seguintes valores lingüísticos:

Quadro III - Valores Lingüísticos Para As Variáveis Consideradas.

Atributo	Valor
EXPERIÊNCIA	BAIXA, MÉDIA, ALTA
TITULAÇÃO	MINIMA, MÉDIA, MÁXIMA
GRAU DE DIFICULDADE	BAIXO, MÉDIO, ALTO
RECURSOS	POUCOS, MUITOS
NOTA (Valor Médio)	BAIXA, MÉDIA, ALTA
FALTA (Valor Médio)	BAIXA, MÉDIA, ALTA

Como se trata de extração de regras *fuzzy*, no método proposto o conseqüente das regras ainda possui, após o processo, um grau de confiança para as regras produzidas. O grau de confiança representa um limiar abaixo do qual as regras não podem ser conclusivas.

Convém ressaltar que, para diferentes domínios, o processo de extração de regras pode variar de acordo com os objetivos definidos para cada sistema. O processo pode utilizar em maior ou menor grau os dados que provêm das diferentes fontes: banco de dados, ontologias ou GEDI.

4.1.4. Conversão de UNL para Ontologia

Os documentos em formato UNL encontrados pelo buscador inteligente são armazenados pelo GEDI para que a sua estrutura seja adaptada conforme os critérios de construção definidos para a ontologia de domínio. A tarefa de conversão é atribuída ao MARERF, que utiliza as informações referentes às palavras universais UW's, as etiquetas de relação "Relation Label" (RL) e também as etiquetas de atributo "Attribute Label" (AL), conforme o diagrama da figura 29. Nesta adaptação, as etiquetas de relações são convertidas diretamente em relações utilizadas na ontologia de domínio, conforme uma tabela de correlação. Para o problema da tradução das palavras universais, dois caminhos são possíveis:

- 1) Cada palavra universal é traduzida conforme um dicionário acoplado ao sistema, de forma que uma "ontologia temporária" é montada;
- 2) A ontologia temporária é montada com as palavras em inglês, e adota-se a utilização de uma ontologia de domínio que comporte o conceito em português e inglês ao mesmo tempo em sua estrutura.

O MARERF então verifica, para cada conceito existente na ontologia temporária, se o mesmo está presente na ontologia de domínio. Em caso positivo, pode-se adotar duas estratégias:

- 1) Conectar automaticamente os conceitos, convertendo as relações à ontologia de domínio para posterior manipulação pelo usuário especialista. Porém, enquanto não forem validados por este usuário, tais conceitos não poderão ser utilizados para os processos de busca inteligente;
- 2) Agregar o conceito sem relacionar, deixando a critério do usuário especialista a posterior conexão à ontologia de domínio, com a possibilidade de renomeação das relações.

Ainda com relação às palavras universais, certos termos poderão ser traduzidos como verbos expressos no modo infinitivo. Para a colocação na ontologia é preferível substituí-los pelo substantivo que expressa a ação.

As palavras universais podem ser genéricas ou limitadas (OLIVEIRA, 2001). As genéricas são expressas de forma simples. As limitadas possuem o conceito restrito a outro conceito que ajuda a representar melhor o significado. Portanto, as palavras universais limitadas poderão gerar mais de um conceito na ontologia temporária. Um exemplo é o uso da etiqueta de relação **icl** em palavras universais, que indica a inclusão de um conceito em outro.

Quadro IV - Correlação UNL – Ontologia.

Relação UNL	Relação da Ontologia
nam, mod	<i>é_um (relação inversa)</i>
pof, ptn, cag,ins	<i>parte_de</i>
cnt	<i>mesmo_que</i>
gol, pur, to	<i>para</i>
plc, ppl, plt,plf, via,opl,scn	<i>localizada_em, em, onde</i>
obj, pos	<i>possui</i>
qua	<i>cardinalidade</i>
bas	<i>baseado_em</i>

Ainda pode-se correlacionar a etiqueta de relação **or** para introduzir axiomas de disjunção na ontologia de domínio. A relação **con** pode ser representada na ontologia como um axioma de condição, semelhante a uma regra de inferência.

O conjunto de etiquetas de relação, que não constam no quadro, produzem relações não nominadas na ontologia. Estas relações deverão ser atualizadas ou mesmo modificadas pelo usuário especialista na intervenção via editor de ontologias.

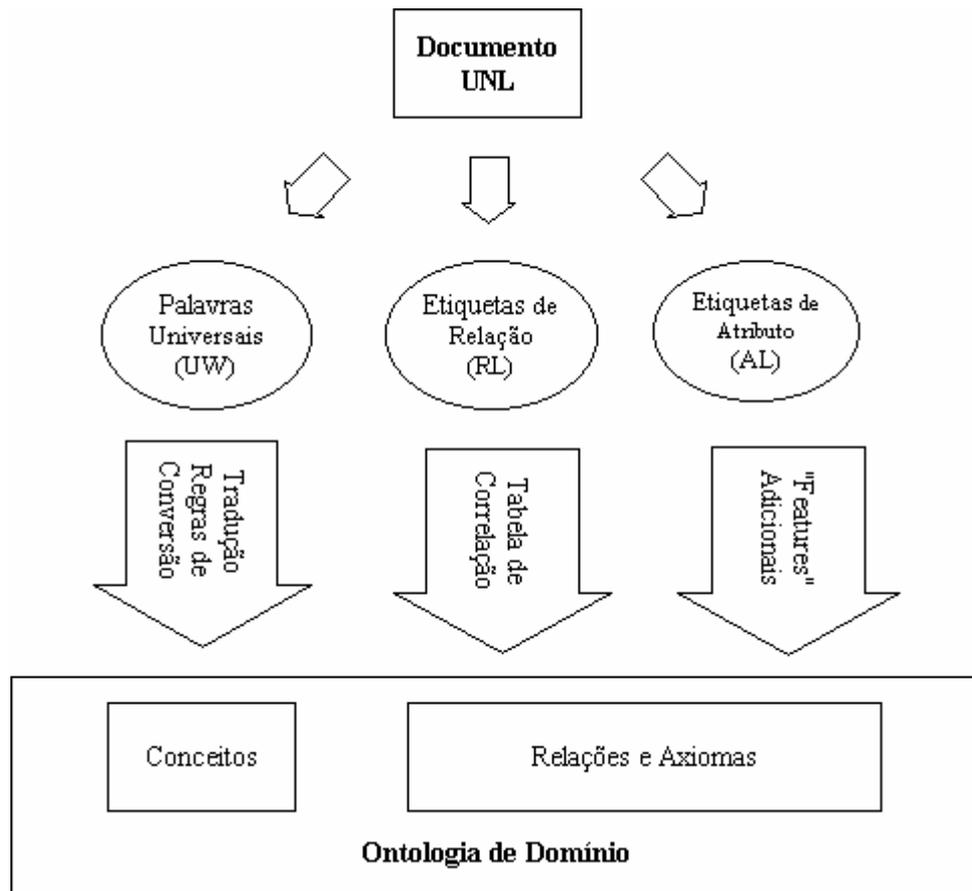


FIGURA 29 - REPRESENTAÇÃO DO PROCESSO DE CONVERSÃO UNL PARA ONTOLOGIA

Quanto ao uso de etiquetas de atributo, pode-se descrever as seguintes opções:

- O MARERF começa a partir do conceito da palavra universal possuidora da etiqueta de atributo **.@entry**, que contém o ponto de entrada para uma representação UNL;
- A etiqueta de atributo **.@pl** pode indicar que o conceito na ontologia terá cardinalidade mínima unitária;
- Conceitos que durante o processamento já tenham sido utilizados, que contenham a etiqueta de atributo **.@def** não precisarão ser agregados novamente;
- A etiqueta de atributo **.@not** fornece a informação sobre complementaridade de um conceito, podendo ser adicionado à ontologia como um axioma de negação.

4.2. CONSIDERAÇÕES FINAIS

Este capítulo apresentou o Gerador de Sistemas Inteligente para Gestão de Informações e Conhecimento com a sua Arquitetura e as funções de seus componentes e ainda como eles serão viabilizados com a aplicação das técnicas específicas para cada caso.

Foi apresentada também uma solução para o caso das buscas inteligentes retornarem um documento na representação UNL, cuja representação, pela sua riqueza de dados na forma de representar as relações semânticas em um documento, facilita o relacionamento com a ontologia de domínio, a qual recebeu melhorias na sua forma de representação, agregando os princípios da representação UNL.

No próximo capítulo serão apresentadas as aplicações práticas.

5. EXEMPLOS DE APLICAÇÃO

Neste capítulo são apresentados dois exemplos de aplicações, sendo um na área de ensino, envolvendo a disciplina empreendedorismo na parte do conteúdo programático referente a plano de negócio. A aplicação foca a métrica para avaliação da qualidade de informações contidas em documentos oriundos de diferentes fontes; considera as relações semânticas entre as palavras e sentenças, sem destacar as demais etapas do processo, para facilitar o entendimento desta etapa do processo e destacar o caráter genérico do sistema e um exemplo de aplicação para a área médica, envolvendo a subárea de infecção hospitalar, focalizando as etapas desde o preparo e geração do sistema de informação até a informação estruturada e contextualizada para apoio às tomadas de decisões gerenciais.

5.1. EXEMPLO DE APLICAÇÃO PARA A ÁREA DE ENSINO

Considere a parte de uma ontologia descrita na figura 30. Ela se refere ao conjunto de termos relacionados com as palavras “Plano de Negócios” e “Análise Ambiental”.

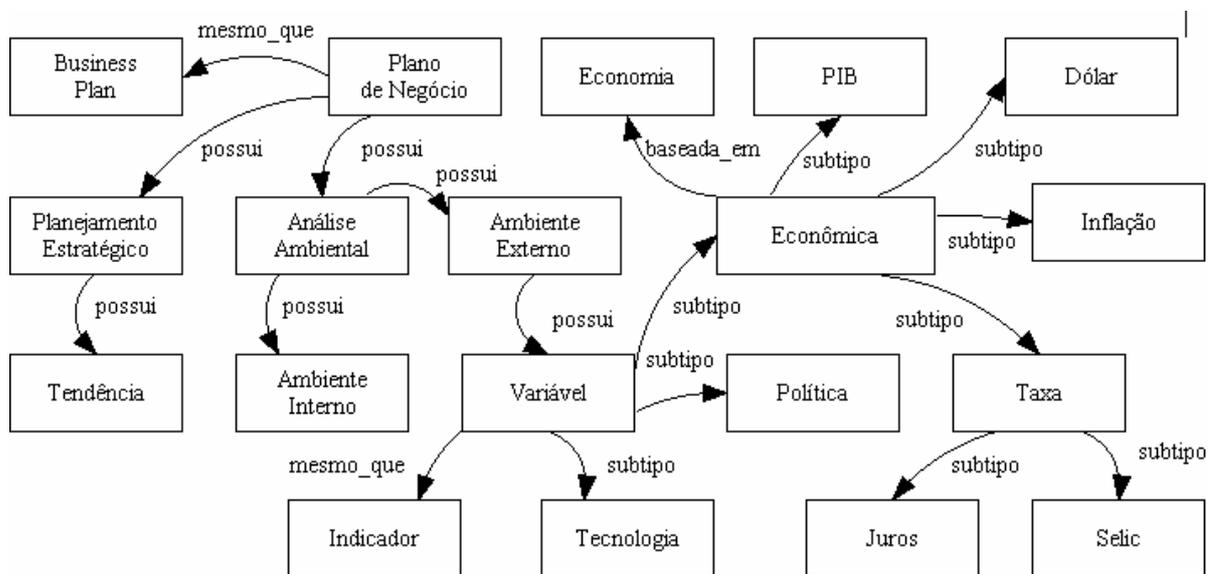


FIGURA 30 – REPRESENTAÇÃO DE PARTE DA ONTOLOGIA PARA O EXEMPLO.

Os conceitos desta ontologia são representados pelos termos, e os *links* são as relações entre tais conceitos.

Para valorar os termos presentes nesta ontologia, tomam-se os *links* que conectam cada conceito e executa-se o somatório, fazendo este total pertencer ao conceito.

Tabela I – Pesos Atribuídos aos Conceitos Existentes na Ontologia.

Conceito	Links	Peso
Plano de Negócio	15	0,06667
Business Plan	15	0,06667
Análise Ambiental	14	0,07142
Ambiente Externo	11	0,09090
Variável	10	0,10000
Indicador	10	0,10000
Econômica	7	0,14285
Economia	7	0,14285
Taxa	3	0,33333
Planejamento Estratégico	2	0,50000
Ambiente Interno	1	1,00000
PIB	1	1,00000
Inflação	1	1,00000
Dólar	1	1,00000
Juros	1	1,00000
Selic	1	1,00000
Tendência	1	1,00000
Tecnologia	1	1,00000
Política	1	1,00000

De forma hierárquica, o processo acontece indo dos nós mais externos à ontologia, até chegar no conceito principal. A tabela I mostra como fica tal somatório. Na última coluna efetua-se o cálculo inverso de cada total, e o resultado significa o peso de cada termo da ontologia.

Destacando-se o termo “economia” da ontologia da figura 30, verifica-se que todos os *links* relacionados a este conceito totalizam 7 (sete). O peso a ser considerado para este termo será então o valor inverso (0,14285). Cada vez que o termo aparece no documento, este valor irá acumulado para o fator de ganho semântico que é normalizado, para que possa ser aplicado posteriormente ao vetor do documento.

Na tabela II, estão ilustrados os valores TF multiplicados pela IDF, referente a 38 termos presentes no vocabulário, dentre os 4 documentos. Cada coluna pode ser entendida como o vetor de cada documento.

Tabela II – TF x IDF.

Termo	1	2	3	4	IDF
área	0,001	0,002	0,001		0,288
produção	0,001	0,002		0,001	0,288
setor	0,005	0,004			0,693
investimento	0,001		0,003	0,000	0,288
site	0,002		0,005		0,693
negócio	0,001		0,002	0,001	0,288
valor	0,001		0,001	0,001	0,288
internet	0,002		0,003		0,693
capital	0,001		0,001	0,000	0,288
taxa	0,003		0,002		0,693
produto	0,002		0,001	0,001	0,288
sindicato	0,002			0,001	0,693
filho	0,002			0,001	0,693
insumos	0,002			0,001	0,693
ambiental		0,004		0,014	0,693
desenvolvimento		0,006		0,004	0,693
estado		0,011		0,001	0,693
projeto			0,001	0,024	0,693
economista			0,005	0,018	0,693
commodities			0,001	0,009	0,693
econômico			0,006	0,006	0,693
mercado			0,014	0,006	0,693
financeiro			0,011	0,004	0,693
análise			0,005	0,003	0,693
consultoria			0,001	0,002	0,693
gestão			0,001	0,002	0,693
são_paulo			0,002	0,002	0,693
economia			0,019	0,002	0,693
econômica			0,005	0,002	0,693
profissional			0,002	0,002	0,693
empresa			0,004	0,002	0,693
dados			0,003	0,001	0,693
internacional			0,001	0,001	0,693
sistema			0,002	0,001	0,693
brasileira			0,002	0,001	0,693
administração			0,001	0,001	0,693
renda			0,003	0,001	0,693
política			0,003	0,001	0,693

Na tabela III, estão os valores da TFIDF mais a ponderação ontológica através de um fator de ganho semântico, que aparece na primeira linha, aplicado para cada vetor de documento.

Tabela III - TFIDF x Fator De Ganho Semântico.

Termo	1	2	3	4	IDF
Ganho Semântico (GS)	5,672	0,005	13,100	11,434	
área	0,841	0,001	0,803		0,288
produção	0,504	0,001		0,939	0,288
setor	3,646	0,002			0,693
investimento	0,504		3,786	0,512	0,288
site	1,215		7,187		0,693
negócio	0,504		2,753	0,598	0,288
valor	0,504		1,950	0,768	0,288
internet	1,013		3,593		0,693
capital	0,588		1,147	0,512	0,288
taxa	2,026		2,488		0,693
produto	1,009		0,688	0,939	0,288
sindicato	1,013			1,440	0,693
filho	1,013			1,028	0,693
insumos	1,215			1,028	0,693
ambiental		0,002		16,455	0,693
desenvolvimento		0,003		4,319	0,693
estado		0,005		1,440	0,693
projeto			1,935	27,356	0,693
economista			6,910	20,363	0,693
commodities			1,382	10,284	0,693
econômico			7,740	7,199	0,693
mercado			17,691	6,788	0,693
financeiro			14,374	4,319	0,693
análise			6,081	3,702	0,693
consultoria			1,382	2,674	0,693
gestão			1,935	2,674	0,693
são paulo			2,488	2,468	0,693
economia			24,878	2,263	0,693
econômica			6,634	2,057	0,693
profissional			3,041	2,057	0,693
empresa			5,805	1,851	0,693
dados			3,593	1,646	0,693
internacional			1,935	1,440	0,693
sistema			2,488	1,234	0,693
brasileira			2,764	1,234	0,693
administração			1,659	1,234	0,693
renda			3,317	1,234	0,693
política			3,317	1,028	0,693

A idéia desta ponderação da ontologia é fazer com que os termos, que aparecem um número menor de vezes, tenham uma valoração maior, de forma semelhante à IDF (frequência inversa de termos), reforçando assim a representatividade do vetor do documento. Na seqüência, cada termo presente em um documento é comparado com a sua existência na ontologia e, desta forma, calcula-se o fator de ganho semântico.

Na tabela IV foi considerado o uso dos métodos da TF, TFIDF, TF com ponderação ontológica e TFIDF com ponderação ontológica. Considerou-se um conjunto de 4 documentos referentes ao domínio dos termos “Análise Ambiental” em uma busca efetuada na WEB. Verifica-se, então, que a ponderação pela ontologia do domínio contribuiu para aumentar a qualidade da representação dos vetores de documentos.

O ganho semântico foi calculado pela equação 21:

$$GS_j = \begin{cases} \varphi \sum w_k & , o_k \in D \\ GS_j \cdot \kappa & , o_k \notin D \text{ com } 0 < \kappa < 1 \end{cases}$$

Onde:

w_k = É o peso dos termos existentes em uma coleção de documento D que é expresso pelo inverso da frequência de *links*:

$$w_k = \frac{1}{f_k}, \forall o_k \in O, \text{ sendo } f_k = \text{frequência de links},$$

O_k = Conceito da ontologia,

φ = Fator de escala,

κ = Fator de penalização.

Tabela IV - Valores Obtidos para o Ganho de Informação nos Diferentes Métodos.

Método	I(C D)
TF	1,4654
TF com ganho semântico	1,6426
TFIDF	1,5152
TFIDF com ganho semântico	1,6640

5.2.1. Conclusão

Este exemplo de aplicação prática demonstrou a validade da ferramenta para agregar informações, ressaltando que o processo evolui na seqüência de forma automática.

Verificou-se também a validade desta nova metodologia para avaliar o quanto um determinado conjunto de vetores é representativo para certo conjunto de documentos, de acordo com o critério fornecido pela teoria da informação. Verificou-se também que uma ontologia agregada ao contexto onde se enquadra o conjunto de documentos pode melhorar tal representatividade. Tal aspecto fornece importante subsídio, visto que os conjuntos de vetores serão utilizados para treinamento da rede *neuro-fuzzy* com objetivo de classificação de novos documentos *a posteriori*, implementados em buscadores inteligentes. Técnicas mais refinadas podem ser obtidas, considerando mais aspectos relativos a conceitos e relações em ontologias.

A prática demonstrou também que o uso de diferentes fatores de escala e penalização presentes na equação 24 geram diferenças na obtenção do ganho de informação, sendo que valores adequados, ou mesmo maximizados, devem ser utilizados conforme o contexto de aplicação. Embora à primeira vista os ganhos referenciados na tabela IV sejam relativamente pequenos, o processo irá acontecer de forma contínua.

A aplicação para a área de ensino, envolvendo a disciplina de empreendedorismo, demonstra a importância do uso do processo com ganho semântico e da utilização da metodologia.

5.3. EXEMPLO DE APLICAÇÃO PARA A ÁREA MÉDICA

Neste exemplo de aplicação serão apresentados todos os passos realizados na prática, desde a geração do SI com a assistência de um especialista da área de interesse, no caso Infecção Hospitalar, elaboração do grupo de busca com dados reais de hospitais do Brasil, fornecidos pelo IPASS – Instituto Paranaense de Acreditação de Serviços de Saúde, que apesar de ser um órgão regional, trabalha com dados de todo o país. Este Instituto é subordinado à ONA – Organização Nacional de Acreditação, que por sua vez é ligada à ANVISA – Agência Nacional de Vigilância Sanitária e ao Ministério da Saúde. Os dados foram incorporados à ontologia de domínio para realização automática das buscas inteligentes e comprovação com métricas específicas para avaliar a qualidade das informações resultantes.

Para facilitar o entendimento, todo processamento foi subdividido em 3 subprocessos a saber:

- Subprocesso de Geração do Sistema de Informação,
- Subprocesso de Busca,
- Subprocesso de Manutenção de Sistema.

5.3.1. Subprocesso de Geração do Sistema de Informação

Inicialmente é feita a modelagem do Sistema de Informação para o objetivo considerado; no caso em tela, “Infecção Hospitalar”, a partir do modelo relacional proposto, conforme é mostrado na figura 31.

O usuário criador configura o Gerador de Aplicações, sendo as informações necessárias à criação do sistema alimentadas ao dicionário de dados (**DD**).

O Gerador de Aplicações gera o **SI** utilizando a ferramenta CASE “Automa Bulder”, cujas telas são mostradas nas figuras 33 a 36. A partir do modelo relacional na figura 32, a estrutura do **BD** e os relacionamentos são montados na ferramenta (figuras 32 e 33), gerando o modelo de **BD** lógico e físico (figura 34) necessário às operações do **SI**. O Gerador, ainda, permite a definição da interface do usuário do **SI** para manipulação dos dados, sendo o mapeamento feito automaticamente a partir do **DD** (figura 35). O Gerador fornece, também, os conceitos para um protótipo da ontologia de domínio a partir das definições existentes no **DD**. Desta forma, já se possui um grau de importância dos termos presentes no **DD**, o que facilitará a montagem do conjunto-base de termos. Uma parte da interface do **SI** gerado está ilustrado na figura 36.

Os registros existentes no **BD**, também, são preparados para a geração do grupo de busca.

Na seqüência é feita a extração de características visando processar o conteúdo já preparado, proveniente do **BD**. Esta fase encerra a metodologia de *Text Mining* e é semelhante àquela feita pelo buscador e o GEDI no retorno das informações obtidas externamente.

O Gerador verifica a existência de novos conceitos adicionados à ontologia e sugere ao Usuário Criador a possibilidade de inclusão dos mesmos na próxima geração do **SI** e do **BD**.

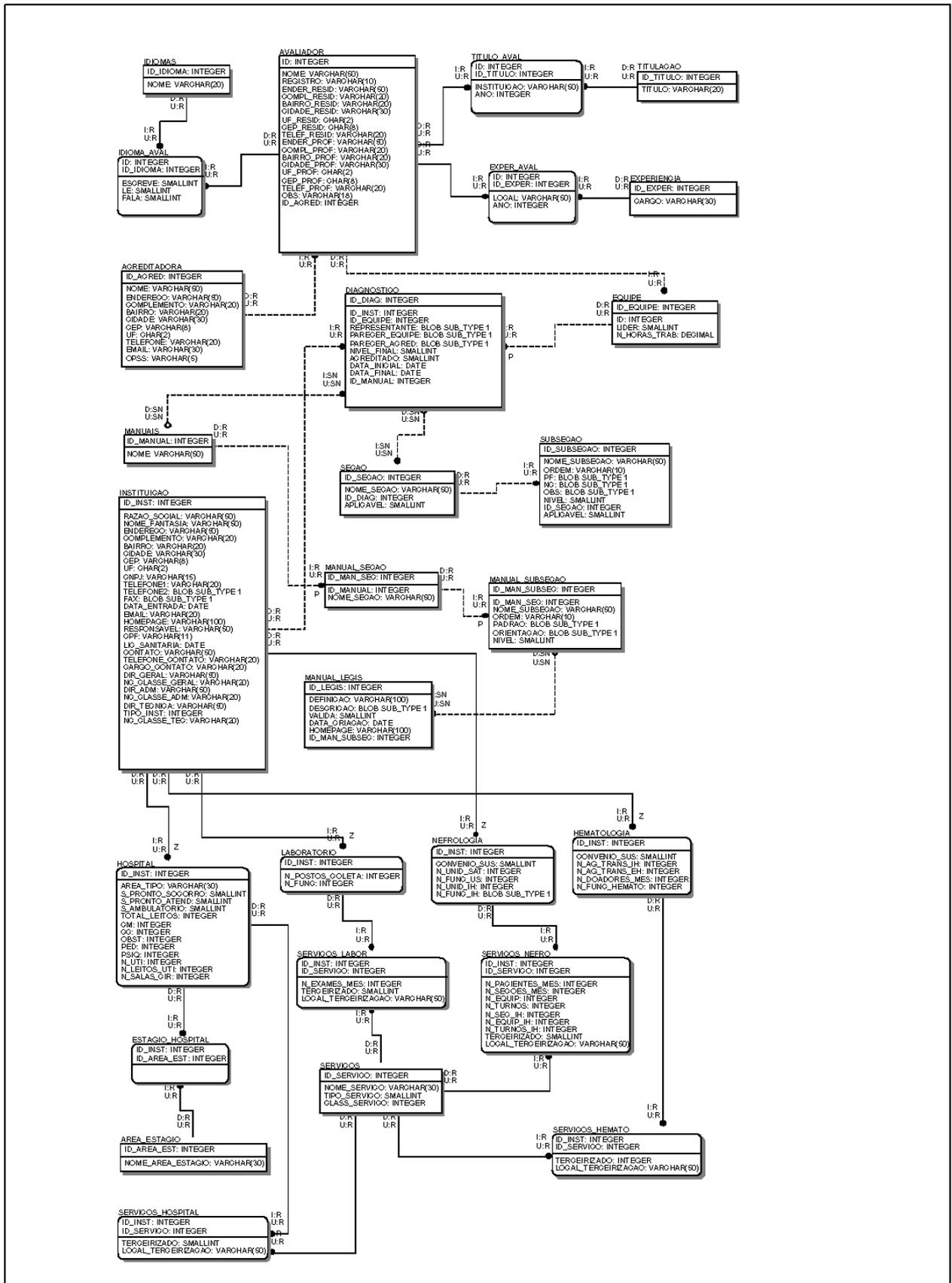


FIGURA 31 - MODELO RELACIONAL PARA O SISTEMA.

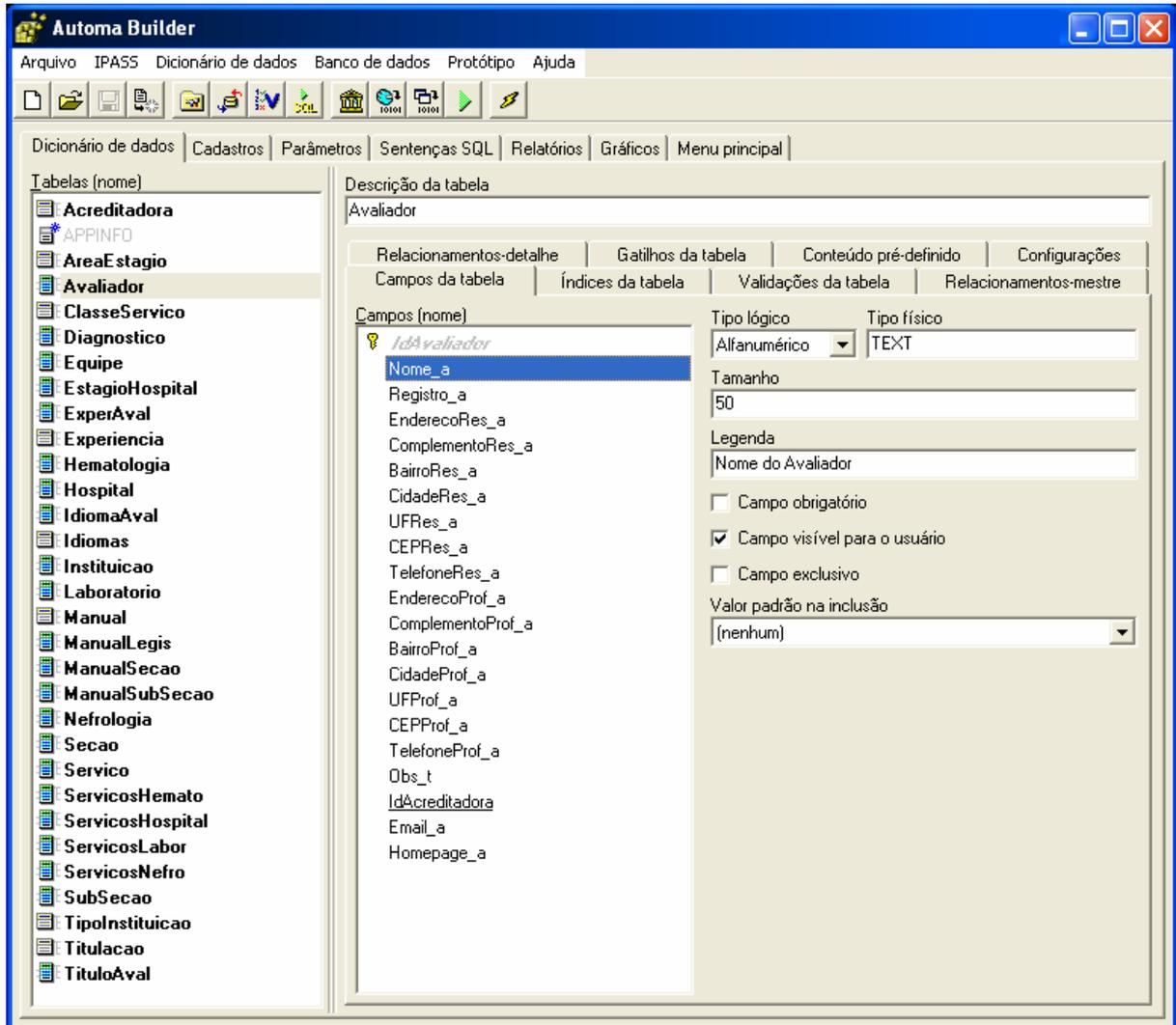


FIGURA 32 - TELA PARA DEFINIÇÃO DO **DD** E A ESTRUTURA DO **BD**.

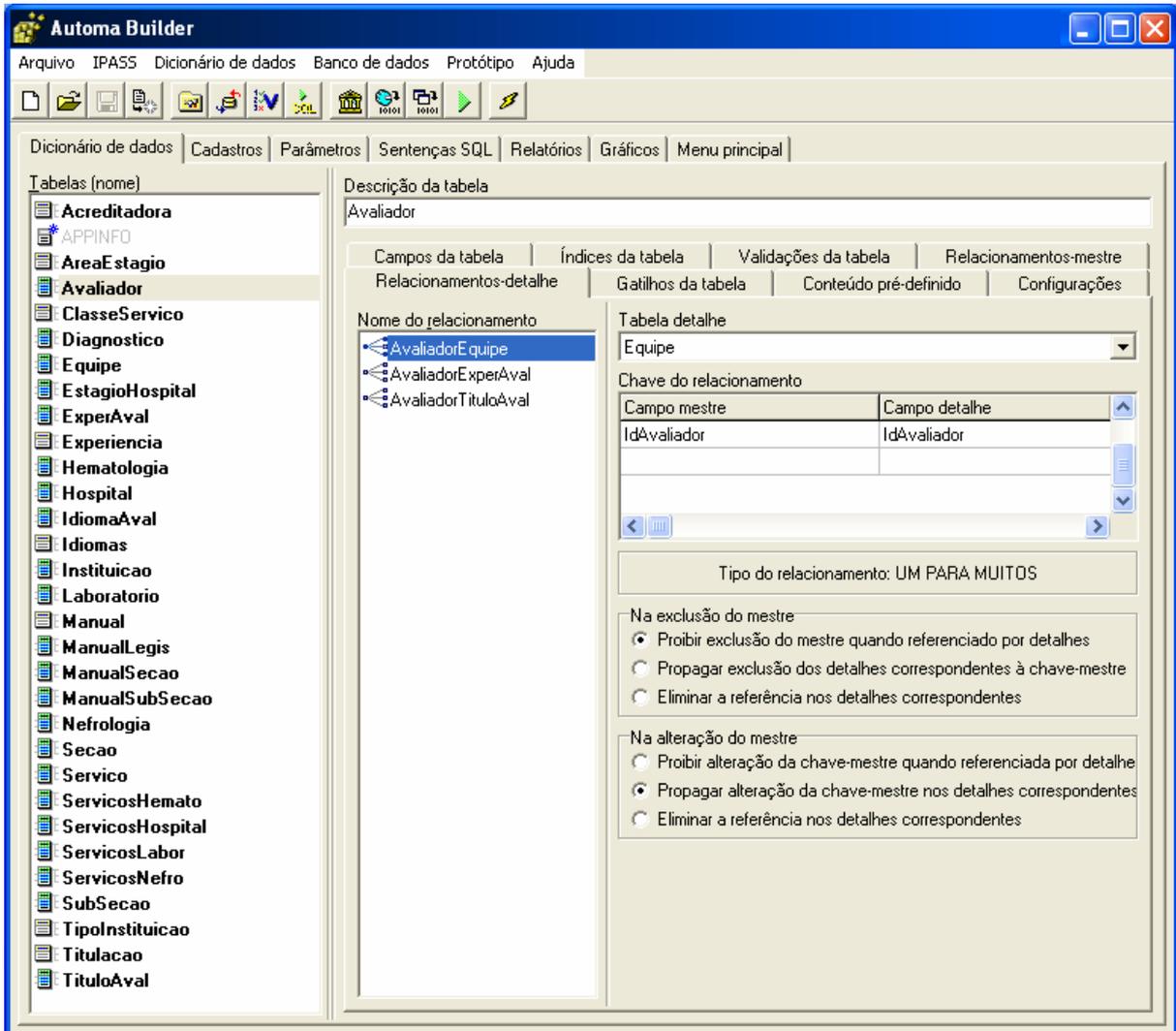


FIGURA 33 - TELA PARA DEFINIÇÃO DE RELACIONAMENTOS ENTRE TABELAS DO **BD**.

Configurações do projeto

Nome do projeto: **IPASS** Título: Sistema de Informações do IPASS

Chave para configurações no registro do Windows: \\Software\\IPASS

Protótipo: Banco de dados

Tipo do DBMS: Ms-Access Localização: c:\\IPASS\\IPASS.mdb

Nome do banco de dados: Apelido da conexão:

Nome do usuário principal: Senha do usuário principal:

Permitir o controle automático de versão do dicionário de dados

Preparar o dicionário de dados e a aplicação para o controle de usuários

Preparar o dicionário de dados para replicação de dados

Utilizar triggers para implementar a integridade referencial entre tabelas

Ok Cancelar

FIGURA 34 - CONFIGURAÇÃO DO BANCO DE DADOS FÍSICO.

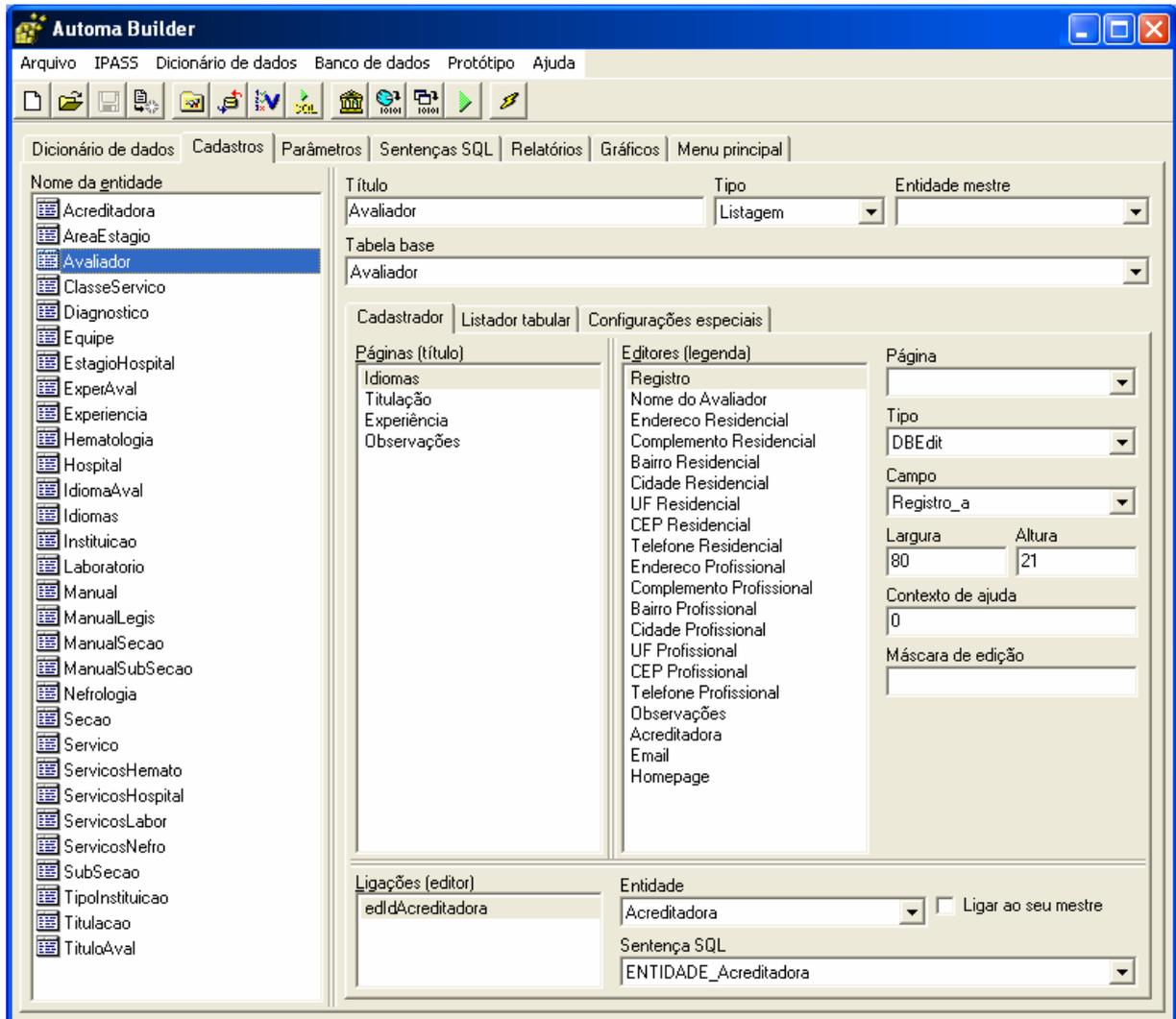


FIGURA 35 - MAPEAMENTO DOS DADOS PARA A INTERFACE DE USUÁRIO DO SI.

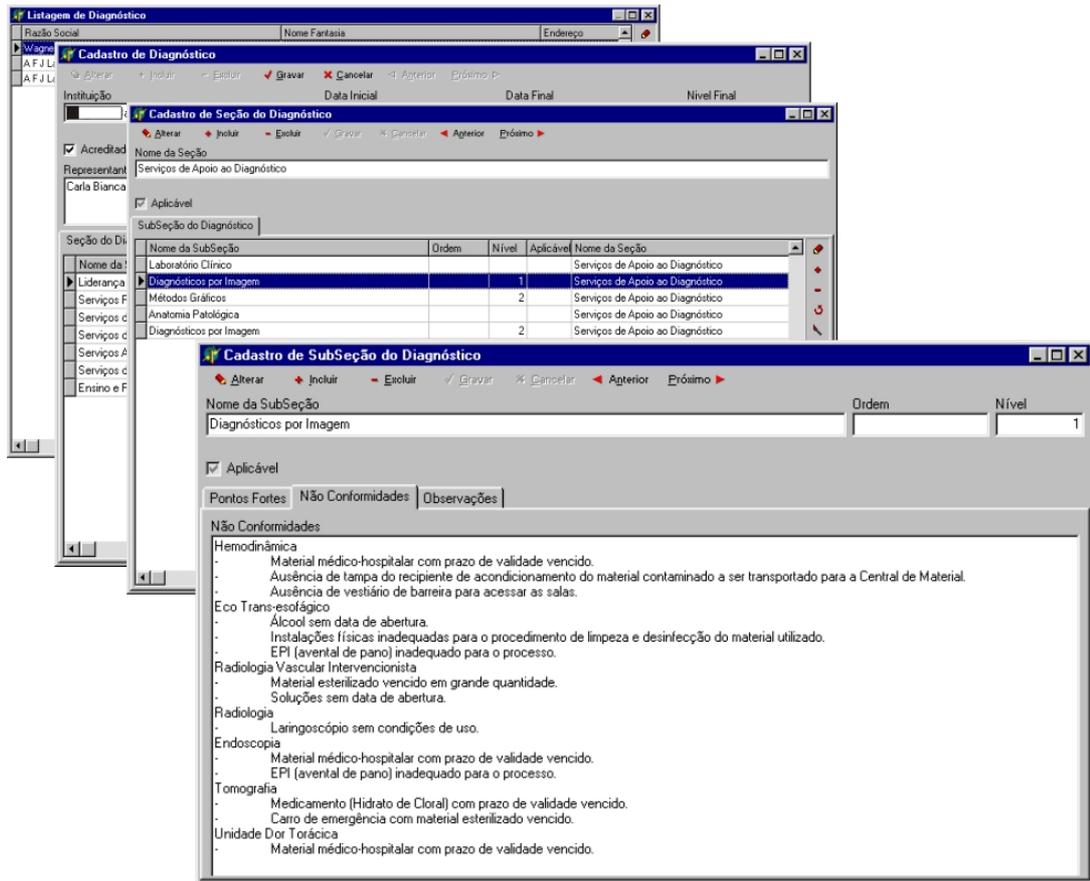


FIGURA 36 - TELAS DO SISTEMA DE INFORMAÇÃO GERADO.

A ontologia gerada está descrita na figura 37. Ela se refere ao conjunto de termos relacionados com as palavras “infecção hospitalar”. Os conceitos desta ontologia são representados pelos termos e os *links* são as relações entre tais conceitos.

Para valorar os termos presentes nesta ontologia, tomam-se os *links* que conectam cada conceito e executa-se o somatório, fazendo este total pertencer ao conceito.

De forma hierárquica, o processo acontece indo dos nós mais externos à ontologia, até chegar no conceito principal.

A tabela V mostra como fica tal somatório. Na última coluna efetua-se o cálculo inverso de cada total, e o resultado significa o peso de cada termo da ontologia.

Na montagem do conjunto-base calculam-se a Frequência de Termos **TF**, a Frequência Inversa de Termos **IDF** e a composição **TFIDF**, agregando-se desta forma os dados quantitativos.

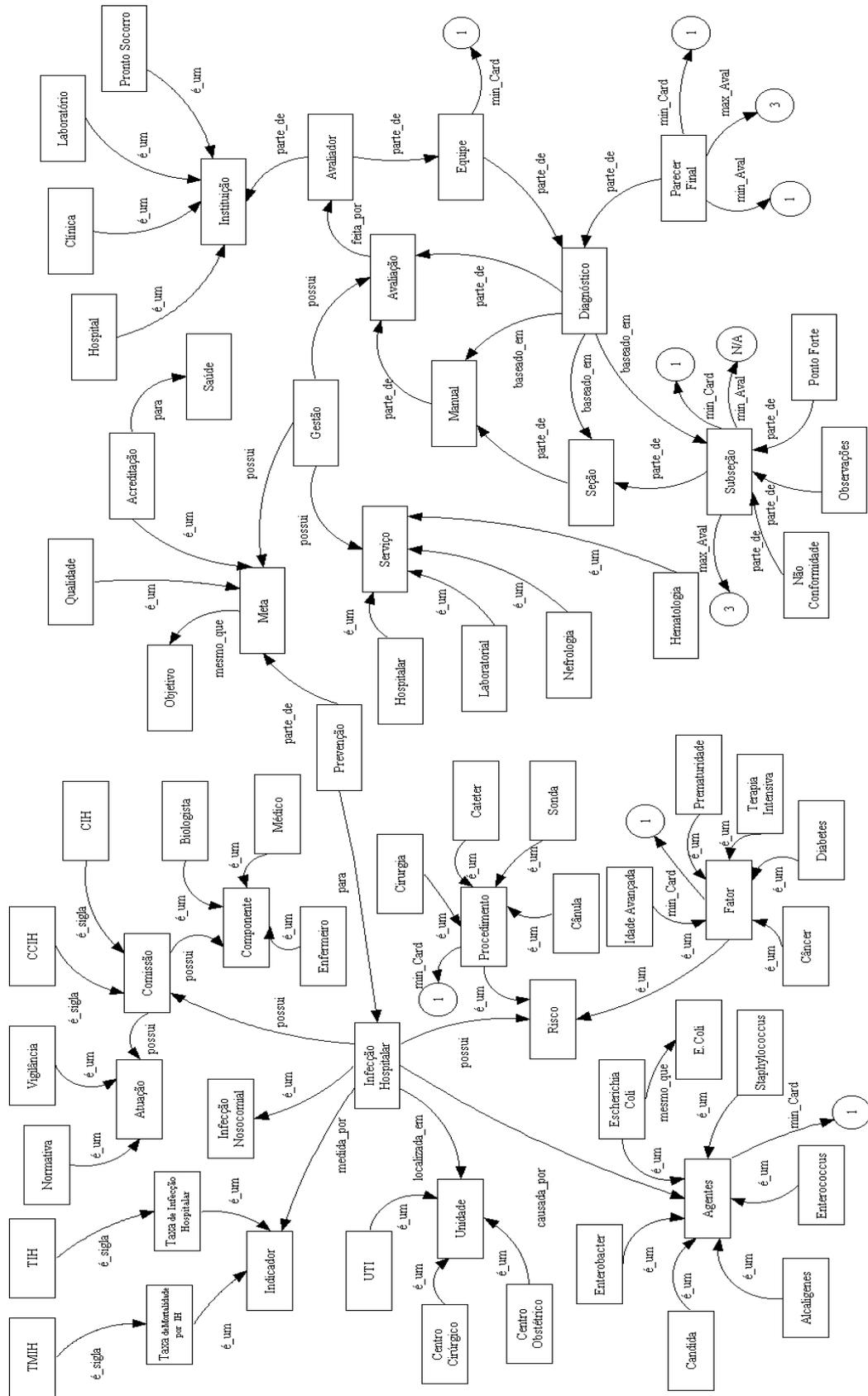


FIGURA 37 - REPRESENTAÇÃO DA ONTOLOGIA PARA O EXEMPLO MÉDICO.

Tabela V - Pesos Atribuídos aos Conceitos Existentes na Ontologia do Exemplo.

Conceito	Total de Links	Peso
gestão	68	0,0147
meta	47	0,0212
prevenção	42	0,0238
infecção hospitalar	40	0,0256
avaliação	16	0,0625
risco	12	0,0833
comissão	10	0,1111
diagnóstico	9	0,1111
agente	8	0,1111
equipe	7	0,1428
avaliador	6	0,1667
fator	6	0,1667
manual	6	0,1667
indicador	5	0,2000
instituição	5	0,2000
procedimento	5	0,2000
seção	5	0,2000
serviço	5	0,2000
componente	4	0,2500
subseção	4	0,2500
unidade	3	0,2500
acreditação	2	0,5000
atuação	2	0,5000
escherichia coli	2	0,5000
alcaligenes	1	1,0000
biólogo	1	1,0000
câncer	1	1,0000
candida	1	1,0000
cânula	1	1,0000
cateter	1	1,0000
CCIH	1	1,0000
centro cirúrgico	1	1,0000
centro obstétrico	1	1,0000
CIH	1	1,0000
cirurgia	1	1,0000
clínica	1	1,0000
diabetes	1	1,0000
E.coli	1	1,0000
enfermeiro	1	1,0000
enterobacter	1	1,0000
enterococcus	1	1,0000
hematologia	1	1,0000
hospital	1	1,0000
hospitalar	1	1,0000
idade avançada	1	1,0000
infecção nosocomial	1	1,0000
laboratorial	1	1,0000
laboratório	1	1,0000
médico	1	1,0000
não conformidade	1	1,0000
nefrologia	1	1,0000
normativa	1	1,0000
objetivo	1	1,0000
observação	1	1,0000
parecer final	1	1,0000
ponto forte	1	1,0000
prematividade	1	1,0000
pronto socorro	1	1,0000
qualidade	1	1,0000
saúde	1	1,0000
sonda	1	1,0000
taxa de infecção hospitalar	1	1,0000
taxa de mortalidade por infecção hospitalar	1	1,0000
terapia avançada	1	1,0000
TIH	1	1,0000
TMIH	1	1,0000
UTI	1	1,0000
vigilância	1	1,0000

5.3.2. Subprocesso de Busca

Nesta fase são considerados os elementos e interações necessárias às buscas inteligentes. Os termos que foram considerados para a criação do sistema constituem um conjunto que é considerado o ponto de partida para um glossário do sistema de informação.

A partir do instante em que o **SI** está à disposição do usuário, os dados começam a ser cadastrados no banco de dados do sistema. Desta forma, os dados que são cadastrados, juntamente com os *metadados* utilizados para a criação do sistema, constituem o conjunto de termos utilizados para direcionar as buscas.

A partir dos registros adicionados ao **BD**, foram utilizados 3 diagnósticos efetuados, dos quais os seguintes campos tiveram os dados utilizados para a geração dos grupos de busca:

- Pontos Fortes,
- Não Conformidades,
- Observações.

Todos os registros referentes às seções e subseções digitadas compuseram cada diagnóstico e cada um foi tratado como um documento exclusivo para efeito da extração de características.

As informações constantes nos anexos II a IV, referem-se aos registros de não conformidades, pontos fortes e observações, respectivamente, de todas as subseções de avaliação dos diagnósticos 1 a 3.

A seguir o conjunto de termos ou glossário é resumido e ordenado pela frequência de ocorrência no escopo do **SI**.

A tabela VI mostra o conjunto base, onde aparece a frequência absoluta de termos ou como já visto do inglês *FTA*, referente aos três diagnósticos com 116 termos. Os quais foram calculados pela equação 15 a seguir:

$$TFA_{ij} = \{v_i \in V \mid TF_{ij} \geq \delta\}, \forall d_j \in D$$

Tabela VI - Frequência Absoluta de Termos.

Palavra	1	2	3	Total
evidência	32	55		87
sistema	9	18	7	34
cliente	3	16	7	26
externo	2	12	5	19
registro	10	12		22
planejamento	2	11		13
análise	27	10		37
interno	1	10	2	13
crítica	4	10		14
avaliação	3	8		11
resultado	1	7		8
área	8	6	3	17
melhoria	26	5		31
educação_continuada	22	4		26
desempenho	1	4		5
treinamento	1	4		5
ausência	11	3		14
atendimento	2	3	2	7
fornecedores	1	3		4
serviço	10	3		13
informação	1	3	3	7
estatística	1	2	1	4
funcionários	1	2	1	4
paciente	3	2		5
terceirizado	2	2		4
técnico	1	2	2	5
atualizados	1	2		3
controle	13	2	2	17
dados	1	2		3
álcool	1	1	1	3
temperatura	3	1		4
mapa	1	1		2
atividades	2	1	2	5
rotina	3	1		4
presença	4	1		5
uso	5	1	1	7
qualidade	1	1	3	5
manutenção_preventiva	1	1		2
saúde	1	1		2
água	2	1		3
fases	1	1		2
processo	3	1	2	6
organizacional	2	1	1	4
financeiro	1	1		2
acompanhamento	1	1		2
habilitação	2	1		3
empresa	1	1		2
administração	2	1	1	4
técnica	3	1		4
parte	1	1		2
circulação	1	1		2
grande	2	1		3
clínico	5	1	1	7
casos	5	1	1	7
esterilizado	3	1		4
material	19	1		20
prazo_de_validade	10	1	1	12
procedimento	1	1		2
data	1	1		2
auditoria	1	1		2
sistemática	2	1	3	6
reunião	1	1	2	4
melhoria_de_processos	5	1		6
trabalho	3		3	6
vencido	14		1	15
validação	1		1	2
produto	2		1	3
participação	1		1	2
coleta	1		1	2
orientação	3		1	4

transporte		2		1		3
manual		5		1		6
clínica		2		1		3
pessoas		1		1		2
metodologia		22		1		23
laboratório		1		1		2
gestão				3		6
via				1		2
exames				2		3
disponibilização				1		2
requisição				1		2
co				1		2
escaneamento				1		2
intranet				1		2
estruturado				1		2
estudante				1		2
ensino				1		2
fomento				1		2
residente				1		2
pesquisa				1		2
bem				1		2
reagentes				2		3
reciclagem				1		2
xilol				1		2
químicos				1		2
consistente				1		2
dinamismo				1		2
empreendedorismo				1		2
premiação				1		2
interna				2		3
amigo				1		2
tomada				1		2
campeão				1		2
semana				1		2
8s				1		2
comparação				1		2
referencial				4		5
liderança				2		3
expressiva				1		2
alta				1		2
programa				4		5
mentoria				1		2
formação				1		2
especialistas				1		2
afim				1		2
discussão				1		2

=====

A tabela VII mostra a Frequência Relativa de Termos ou TFR , calculada pela equação 16 a seguir:

$$TFR_{ij} = \frac{TFA_{ij}}{\sum_{i=1}^m TFA_{ij}}, \forall d_j \subset D$$

Tabela VII - Frequência Relativa de Termos (*TFR*).

Palavra	1	2	3	IDF
evidência	2,728	11,98		10,41
sistema	0,767	3,922	4,118	10,00
cliente	0,256	3,486	4,118	10,00
externo	0,171	2,614	2,941	10,00
registro	0,853	2,614		10,41
planejamento	0,171	2,397		10,41
análise	2,302	2,179		10,41
interno	0,085	2,179	1,176	10,00
crítica	0,341	2,179		10,41
avaliação	0,256	1,743		10,41
resultado	0,085	1,525		10,41
área	0,682	1,307	1,765	10,00
melhoria	2,217	1,089		10,41
educação_continuada	1,876	0,871		10,41
desempenho	0,085	0,871		10,41
treinamento	0,085	0,871		10,41
ausência	0,938	0,654		10,41
atendimento	0,171	0,654	1,176	10,00
fornecedores	0,085	0,654		10,41
serviço	0,853	0,654		10,41
informação	0,085	0,654	1,765	10,00
estatística	0,085	0,436	0,588	10,00
funcionários	0,085	0,436	0,588	10,00
paciente	0,256	0,436		10,41
terceirizado	0,171	0,436		10,41
técnico	0,085	0,436	1,176	10,00
atualizados	0,085	0,436		10,41
controle	1,108	0,436	1,176	10,00
dados	0,085	0,436		10,41
álcool	0,085	0,218	0,588	10,00
temperatura	0,256	0,218		10,41
mapa	0,085	0,218		10,41
atividades	0,171	0,218	1,176	10,00
rotina	0,256	0,218		10,41
presença	0,341	0,218		10,41
uso	0,426	0,218	0,588	10,00
qualidade	0,085	0,218	1,765	10,00
manutenção_preventiva	0,085	0,218		10,41
saúde	0,085	0,218		10,41
água	0,171	0,218		10,41
fases	0,085	0,218		10,41
processo	0,256	0,218	1,176	10,00
organizacional	0,171	0,218	0,588	10,00
financeiro	0,085	0,218		10,41
acompanhamento	0,085	0,218		10,41
habilitação	0,171	0,218		10,41
empresa	0,085	0,218		10,41
administração	0,171	0,218	0,588	10,00
técnica	0,256	0,218		10,41
parte	0,085	0,218		10,41
circulação	0,085	0,218		10,41
grande	0,171	0,218		10,41
clínico	0,426	0,218	0,588	10,00
casos	0,426	0,218	0,588	10,00
esterilizado	0,256	0,218		10,41
material	1,620	0,218		10,41
prazo_de_validade	0,853	0,218	0,588	10,00
procedimento	0,085	0,218		10,41
data	0,085	0,218		10,41
auditoria	0,085	0,218		10,41
sistemática	0,171	0,218	1,765	10,00
reunião	0,085	0,218	1,176	10,00
melhoria_de_processos	0,426	0,218		10,41
trabalho	0,256		1,765	10,41
vencido	1,194		0,588	10,41
validação	0,085		0,588	10,41
produto	0,171		0,588	10,41
participação	0,085		0,588	10,41
coleta	0,085		0,588	10,41

orientação	0,256	0,588	0,41
transporte	0,171	0,588	0,41
manual	0,426	0,588	0,41
clínica	0,171	0,588	0,41
pessoas	0,085	0,588	0,41
metodologia	1,876	0,588	0,41
laboratório	0,085	0,588	0,41
gestão		0,654	1,765 0,41
via		0,218	0,588 0,41
exames		0,436	0,588 0,41
disponibilização		0,218	0,588 0,41
requisição		0,218	0,588 0,41
co		0,218	0,588 0,41
escaneamento		0,218	0,588 0,41
intranet		0,218	0,588 0,41
estruturado		0,218	0,588 0,41
estudante		0,218	0,588 0,41
ensino		0,218	0,588 0,41
fomento		0,218	0,588 0,41
residente		0,218	0,588 0,41
pesquisa		0,218	0,588 0,41
bem		0,218	0,588 0,41
reagentes		0,436	0,588 0,41
reciclagem		0,218	0,588 0,41
xilol		0,218	0,588 0,41
químicos		0,218	0,588 0,41
consistente		0,218	0,588 0,41
dinamismo		0,218	0,588 0,41
empreendedorismo		0,218	0,588 0,41
premiação		0,218	0,588 0,41
interna		0,436	0,588 0,41
amigo		0,218	0,588 0,41
tomada		0,218	0,588 0,41
campeão		0,218	0,588 0,41
semana		0,218	0,588 0,41
8s		0,218	0,588 0,41
comparação		0,218	0,588 0,41
referencial		0,871	0,588 0,41
liderança		0,436	0,588 0,41
expressiva		0,218	0,588 0,41
alta		0,218	0,588 0,41
programa		0,871	0,588 0,41
mentoria		0,218	0,588 0,41
formação		0,218	0,588 0,41
especialistas		0,218	0,588 0,41
afim		0,218	0,588 0,41
discussão		0,218	0,588 0,41

A tabela VIII mostra os valores da *TFIDF* (Frequência de Termos multiplicada pela Frequência Inversa de Termos), com a Ponderação Ontológica ou *TFIDF* mais GS (Ganho Semântico).

Tabela VIII - Ponderação Ontológica- TFIDF + GS.

Palavra	1	2	3	IDF
Ganho Semântico (GS)	9,089	1,239	1,591	
evidência	10,05	6,020		10,41
sistema	6,973	4,859	6,552	10,00
cliente	2,324	4,319	6,552	10,00
externo	1,550	3,239	4,680	10,00
registro	3,142	1,313		10,41
planejamento	0,628	1,204		10,41
análise	8,482	1,095		10,41
interno	0,775	2,700	1,872	10,00
crítica	1,257	1,095		10,41
avaliação	0,942	0,876		10,41
resultado	0,314	0,766		10,41
área	6,198	1,620	2,808	10,00
melhoria	8,168	0,547		10,41
educação_continuada	6,911	0,438		10,41
desempenho	0,314	0,438		10,41
treinamento	0,314	0,438		10,41
ausência	3,456	0,328		10,41
atendimento	1,550	0,810	1,872	10,00
fornecedores	0,314	0,328		10,41
serviço	3,142	0,328		10,41
informação	0,775	0,810	2,808	10,00
estatística	0,775	0,540	0,936	10,00
funcionários	0,775	0,540	0,936	10,00
paciente	0,942	0,219		10,41
terceirizado	0,628	0,219		10,41
técnico	0,775	0,540	1,872	10,00
atualizados	0,314	0,219		10,41
controle	10,07	0,540	1,872	10,00
dados	0,314	0,219		10,41
álcool	0,775	0,270	0,936	10,00
temperatura	0,942	0,109		10,41
mapa	0,314	0,109		10,41
atividades	1,550	0,270	1,872	10,00
rotina	0,942	0,109		10,41
presença	1,257	0,109		10,41
uso	3,874	0,270	0,936	10,00
qualidade	0,775	0,270	2,808	10,00
manutenção_preventiva	0,314	0,109		10,41
saúde	0,314	0,109		10,41
água	0,628	0,109		10,41
fases	0,314	0,109		10,41
processo	2,324	0,270	1,872	10,00
organizacional	1,550	0,270	0,936	10,00
financeiro	0,314	0,109		10,41
acompanhamento	0,314	0,109		10,41
habilitação	0,628	0,109		10,41
empresa	0,314	0,109		10,41
administração	1,550	0,270	0,936	10,00
técnica	0,942	0,109		10,41
parte	0,314	0,109		10,41
circulação	0,314	0,109		10,41
grande	0,628	0,109		10,41
clínico	3,874	0,270	0,936	10,00
casos	3,874	0,270	0,936	10,00
esterilizado	0,942	0,109		10,41
material	5,969	0,109		10,41
prazo_de_validade	7,748	0,270	0,936	10,00
procedimento	0,314	0,109		10,41
data	0,314	0,109		10,41
auditoria	0,314	0,109		10,41
sistemática	1,550	0,270	2,808	10,00
reunião	0,775	0,270	1,872	10,00
melhoria_de_processos	1,571	0,109		10,41
trabalho	0,942		1,139	10,41
vencido	4,398		0,380	10,41
validação	0,314		0,380	10,41
produto	0,628		0,380	10,41
participação	0,314		0,380	10,41
coleta	0,314		0,380	10,41

orientação	0,942	0,380	0,41
transporte	0,628	0,380	0,41
manual	1,571	0,380	0,41
clínica	0,628	0,380	0,41
pessoas	0,314	0,380	0,41
metodologia	6,911	0,380	0,41
laboratório	0,314	0,380	0,41
gestão		0,328	1,139 0,41
via		0,109	0,380 0,41
exames		0,219	0,380 0,41
disponibilização		0,109	0,380 0,41
requisição		0,109	0,380 0,41
co		0,109	0,380 0,41
escaneamento		0,109	0,380 0,41
intranet		0,109	0,380 0,41
estruturado		0,109	0,380 0,41
estudante		0,109	0,380 0,41
ensino		0,109	0,380 0,41
fomento		0,109	0,380 0,41
residente		0,109	0,380 0,41
pesquisa		0,109	0,380 0,41
bem		0,109	0,380 0,41
reagentes		0,219	0,380 0,41
reciclagem		0,109	0,380 0,41
xilol		0,109	0,380 0,41
químicos		0,109	0,380 0,41
consistente		0,109	0,380 0,41
dinamismo		0,109	0,380 0,41
empreendedorismo		0,109	0,380 0,41
premiação		0,109	0,380 0,41
interna		0,219	0,380 0,41
amigo		0,109	0,380 0,41
tomada		0,109	0,380 0,41
campeão		0,109	0,380 0,41
semana		0,109	0,380 0,41
8s		0,109	0,380 0,41
comparação		0,109	0,380 0,41
referencial		0,438	0,380 0,41
liderança		0,219	0,380 0,41
expressiva		0,109	0,380 0,41
alta		0,109	0,380 0,41
programa		0,438	0,380 0,41
mentoria		0,109	0,380 0,41
formação		0,109	0,380 0,41
especialistas		0,109	0,380 0,41
afim		0,109	0,380 0,41
discussão		0,109	0,380 0,41

Na seqüência é feita a inclusão de vizinhança e geração do grupo de busca pelos critérios do **BD** e da ontologia de domínio, comparando os termos com a ontologia e verificando a vizinhança dos termos coincidentes. Aos termos no conjunto-base são adicionados os termos que são vizinhos na ontologia, levando-se em conta a quantidade limite de termos vizinhos, o nível de profundidade da vizinhança e o tipo de vizinhança.

Os valores considerados para esta fase estão ilustrados no quadro V e na tabela IX.

Quadro V - Inclusão de Vizinhança pelo Critério do Banco de Dados.

Geração do Grupos de Busca – Ontologia de Domínio Inclusão de Vizinhança	
Palavras Mestres (indicadas pela ontologia, sempre aparecerão no processo de busca)	Saúde+Gestão Infecção+Hospitalar
Vizinhança de 1º. Nível (palavras mestres mais os nós vizinhos)	Saúde+Gestão+Meta Saúde+Gestão+Serviço Saúde+Gestão+Avaliação Saúde+Gestão+Acreditação Infecção+Hospitalar+Infecção+Nosocomial Infecção+Hospitalar+Prevenção Infecção+Hospitalar+Risco Infecção+Hospitalar+Agente Infecção+Hospitalar+Unidade Infecção+Hospitalar+Indicador Infecção+Hospitalar+Comissão
Vizinhança de 2º. Nível	Saúde+Gestão+Meta+Prevenção Saúde+Gestão+Meta+Objetivo Saúde+Gestão+Meta+Qualidade Saúde+Gestão+Serviço+Hospitalar Saúde+Gestão+Serviço+Laboratório Saúde+Gestão+Serviço+Nefrologia Saúde+Gestão+Serviço+Hematologia Saúde+Gestão+Avaliação+Manual Saúde+Gestão+Avaliação+Diagnóstico Saúde+Gestão+Avaliação+Avaliador Infecção+Hospitalar+Prevenção+Meta Infecção+Hospitalar+Risco+Procedimento Infecção+Hospitalar+Risco+Fator Infecção+Hospitalar+Agente+Enterobacter Infecção+Hospitalar+Agente+Alcaligenes Infecção+Hospitalar+Agente+Staphylococcus Infecção+Hospitalar+Agente+Escherichia+Coli Infecção+Hospitalar+Agente+Candida Infecção+Hospitalar+Unidade+Centro+Cirúrgico Infecção+Hospitalar+Unidade+Centro+Obstétrico Infecção+Hospitalar+Unidade+UTI Infecção+Hospitalar+Indicador+Taxa+de+Mortalidade+por+IH Infecção+Hospitalar+Indicador+Taxa+de+Infecção+Hospitalar Infecção+Hospitalar+Comissão+Atuação Infecção+Hospitalar+Comissão+CCIH Infecção+Hospitalar+Comissão+CIH Infecção+Hospitalar+Comissão+Componente
Vizinhança de 3º. Nível	Saúde+Gestão+Meta+Prevenção+Infecção+Hospitalar Saúde+Gestão+Avaliação+Manual+Seção Saúde+Gestão+Avaliação+Avaliador+Equipe Saúde+Gestão+Avaliação+Avaliador+Instituição Infecção+Hospitalar+Prevenção+Meta+Objetivo Infecção+Hospitalar+Prevenção+Meta+Qualidade Infecção+Hospitalar+Prevenção+Meta+Acreditação Infecção+Hospitalar+Prevenção+Meta+Gestão Infecção+Hospitalar+Risco+Procedimento+Cirurgia Infecção+Hospitalar+Risco+Procedimento+Cateter Infecção+Hospitalar+Risco+Procedimento+Sonda Infecção+Hospitalar+Risco+Procedimento+Cânula Infecção+Hospitalar+Risco+Fator+Idade+Avançada Infecção+Hospitalar+Risco+Fator+Prematuridade Infecção+Hospitalar+Risco+Fator+Terapia+Intensiva Infecção+Hospitalar+Risco+Fator+Diabetes Infecção+Hospitalar+Risco+Fator+Câncer Infecção+Hospitalar+Comissão+Componente+Enfermeiro Infecção+Hospitalar+Comissão+Componente+Médico Infecção+Hospitalar+Comissão+Componente+Biologista

Tabela IX - Ordenação Inversa pelo Critério – Ontologia de Domínio.

Grupo de Busca	Pesos					Soma
Saúde+Gestão+Serviço+Hematologia	1	0,0147	0,2	1		2,2147
Saúde+Gestão+Meta+Objetivo	1	0,0147	0,0212	1		2,0359
Saúde+Gestão+Meta+Qualidade	1	0,0147	0,0212	1		2,0359
Saúde+Gestão+Acreditação	1	0,0147	0,5			1,5147
Saúde+Gestão+Avaliação+Manual+Seção	1	0,0147	0,0625	0,1667	0,2	1,4439
Saúde+Gestão+Avaliação+Avaliador+Instituição	1	0,0147	0,0625	0,1667	0,2	1,4439
Saúde+Gestão+Avaliação+Avaliador+Equipe	1	0,0147	0,0625	0,1667	0,1428	1,3867
Infecção+Hospitalar+Comissão+Componente+Enfermeiro	0,0256	0,1111	0,25	1		1,3867
Infecção+Hospitalar+Comissão+Componente+Médico	0,0256	0,1111	0,25	1		1,3867
Infecção+Hospitalar+Comissão+Componente+Biologista	0,0256	0,1111	0,25	1		1,3867
Infecção+Hospitalar+Risco+Fator+Idade+Avançada	0,0256	0,0833	0,25	1		1,3589
Infecção+Hospitalar+Risco+Fator+Prematuridade	0,0256	0,0833	0,25	1		1,3589
Infecção+Hospitalar+Risco+Fator+Terapia+Intensiva	0,0256	0,0833	0,25	1		1,3589
Infecção+Hospitalar+Risco+Fator+Diabetes	0,0256	0,0833	0,25	1		1,3589
Infecção+Hospitalar+Risco+Fator+Câncer	0,0256	0,0833	0,25	1		1,3589
Infecção+Hospitalar+Unidade+Centro+Cirúrgico	0,0256	0,25	1			1,2756
Infecção+Hospitalar+Unidade+Centro+Obstétrico	0,0256	0,25	1			1,2756
Infecção+Hospitalar+Unidade+UTI	0,0256	0,25	1			1,2756
Infecção+Hospitalar+Risco+Procedimento+Cirurgia	0,0256	0,0833	0,1667	1		1,2756
Infecção+Hospitalar+Risco+Procedimento+Cateter	0,0256	0,0833	0,1667	1		1,2756
Infecção+Hospitalar+Risco+Procedimento+Sonda	0,0256	0,0833	0,1667	1		1,2756
Infecção+Hospitalar+Risco+Procedimento+Cânula	0,0256	0,0833	0,1667	1		1,2756
Saúde+Gestão+Avaliação+Manual	1	0,0147	0,0625	0,1667		1,2439
Saúde+Gestão+Avaliação+Avaliador	1	0,0147	0,0625	0,1667		1,2439
Infecção+Hospitalar+Indicador+Taxa+de+Mortalidade+por+IH	0,0256	0,2	1			1,2256
Infecção+Hospitalar+Indicador+Taxa+de+Infecção+Hospitalar	0,0256	0,2	1			1,2256
Saúde+Gestão+Serviço	1	0,0147	0,2			1,2147
Saúde+Gestão+Avaliação+Diagnóstico	1	0,0147	0,0625	0,1111		1,1883
Infecção+Hospitalar+Agente+Enterobacter	0,0256	0,1111	1			1,1367
Infecção+Hospitalar+Agente+Alcaligenes	0,0256	0,1111	1			1,1367
Infecção+Hospitalar+Agente+Staphylococcus	0,0256	0,1111	1			1,1367
Infecção+Hospitalar+Agente+Candida	0,0256	0,1111	1			1,1367
Infecção+Hospitalar+Comissão+Atuação	0,0256	0,1111	1			1,1367
Infecção+Hospitalar+Comissão+CCIH	0,0256	0,1111	1			1,1367
Infecção+Hospitalar+Comissão+CIH	0,0256	0,1111	1			1,1367
Saúde+Gestão+Meta+Prevenção+Infecção+Hospitalar	1	0,0147	0,0212	0,0238	0,0256	1,0853
Saúde+Gestão+Avaliação	1	0,0147	0,0625			1,0772
Infecção+Hospitalar+Prevenção+Meta+Objetivo	0,0256	0,0238	0,0212	1		1,0706
Infecção+Hospitalar+Prevenção+Meta+Qualidade	0,0256	0,0238	0,0212	1		1,0706
Saúde+Gestão+Meta+Prevenção	1	0,0147	0,0212	0,0238		1,0597
Saúde+Gestão+Meta	1	0,0147	0,0212			1,0359
Infecção+Hospitalar+Infecção+Nosocomial	0,0256	1				1,0256
Infecção+Hospitalar+Agente+Escherichia+Coli	0,0256	0,1111	0,5			0,6367
Infecção+Hospitalar+Prevenção+Meta+Acreditação	0,0256	0,0238	0,0212	0,5		0,5706
Infecção+Hospitalar+Comissão+Componente	0,0256	0,1111	0,25			0,3867
Infecção+Hospitalar+Risco+Fator	0,0256	0,0833	0,25			0,3589
Infecção+Hospitalar+Unidade	0,0256	0,25				0,2756
Infecção+Hospitalar+Risco+Procedimento	0,0256	0,0833	0,1667			0,2756
Infecção+Hospitalar+Indicador	0,0256	0,2				0,2256
Infecção+Hospitalar+Agente	0,0256	0,1111				0,1367
Infecção+Hospitalar+Comissão	0,0256	0,1111				0,1367
Infecção+Hospitalar+Risco	0,0256	0,0833				0,1089
Infecção+Hospitalar+Prevenção+Meta+Gestão	0,0256	0,0238	0,0212	0,0147		0,0853
Infecção+Hospitalar+Prevenção+Meta	0,0256	0,0238	0,0212			0,0706
Infecção+Hospitalar+Prevenção	0,0256	0,0238				0,0494

A seguir é gerado o grupo de busca que conterà os termos do conjunto-base; porém agora são explicitadas as combinações possíveis com os termos vizinhos e são parametrizados de acordo com os critérios estabelecidos para o caso em foco, conforme está sendo mostrado nos quadros VI e VII.

Quadro VI - Geração de Grupos de Busca.

Ontologia de Domínio	Saúde+Gestão+Serviço+Hematologia
	Saúde+Gestão+Meta+Objetivo
	Saúde+Gestão+Meta+Qualidade
	Saúde+Gestão+Acreditação
	Saúde+Gestão+Avaliação+Manual+Seção
	Saúde+Gestão+Avaliação+Avaliador+Instituição
	Saúde+Gestão+Avaliação+Avaliador+Equipe
	Infecção+Hospitalar+Comissão+Componente+Enfermeiro
	Infecção+Hospitalar+Comissão+Componente+Médico
	Infecção+Hospitalar+Comissão+Componente+Biologista
	Infecção+Hospitalar+Risco+Fator+Idade+Avançada
	Infecção+Hospitalar+Risco+Fator+Prematuridade
	Infecção+Hospitalar+Risco+Fator+Terapia+Intensiva
	Infecção+Hospitalar+Risco+Fator+Diabetes
	Infecção+Hospitalar+Risco+Fator+Câncer
	Infecção+Hospitalar+Unidade+Centro+Cirúrgico
	Infecção+Hospitalar+Unidade+Centro+Obstétrico
	Infecção+Hospitalar+Unidade+UTI
	Infecção+Hospitalar+Risco+Procedimento+Cirurgia
	Infecção+Hospitalar+Risco+Procedimento+Cateter
	Infecção+Hospitalar+Risco+Procedimento+Sonda
	Infecção+Hospitalar+Risco+Procedimento+Cânula
	Saúde+Gestão+Avaliação+Manual
	Saúde+Gestão+Avaliação+Avaliador
	Infecção+Hospitalar+Indicador+Taxa+de+Mortalidade+por+IH
	Infecção+Hospitalar+Indicador+Taxa+de+Infecção+Hospitalar
	Saúde+Gestão+Serviço
	Saúde+Gestão+Avaliação+Diagnóstico
	Infecção+Hospitalar+Agente+Enterobacter
	Infecção+Hospitalar+Agente+Alcaligenes
	Infecção+Hospitalar+Agente+Staphylococcus
	Infecção+Hospitalar+Agente+Candida
	Infecção+Hospitalar+Comissão+Atuação
	Infecção+Hospitalar+Comissão+CCIH
	Infecção+Hospitalar+Comissão+CIH
	Saúde+Gestão+Meta+Prevenção+Infecção+Hospitalar
	Saúde+Gestão+Avaliação
	Infecção+Hospitalar+Prevenção+Meta+Objetivo
	Infecção+Hospitalar+Prevenção+Meta+Qualidade
	Saúde+Gestão+Meta+Prevenção
	Saúde+Gestão+Meta
	Saúde+Gestão
	Infecção+Hospitalar+Infecção+Nosocomial
	Infecção+Hospitalar+Agente+Escherichia+Coli
	Infecção+Hospitalar+Prevenção+Meta+Acreditação
	Infecção+Hospitalar+Comissão+Componente
	Infecção+Hospitalar+Risco+Fator
	Infecção+Hospitalar+Unidade
	Infecção+Hospitalar+Risco+Procedimento
	Infecção+Hospitalar+Indicador
	Infecção+Hospitalar+Agente
	Infecção+Hospitalar+Comissão
	Infecção+Hospitalar+Risco
	Infecção+Hospitalar+Prevenção+Meta+Gestão
	Infecção+Hospitalar+Prevenção+Meta
	Infecção+Hospitalar+Prevenção
	Infecção+Hospitalar

Quadro VII - Geração de Grupos de Busca.

<p>Grupo de Busca</p> <p>Ontologia de Domínio</p> <p>+</p> <p>BD</p>	Saúde+Gestão+sistema Saúde+Gestão+evidência Saúde+Gestão+cliente Saúde+Gestão+controle Saúde+Gestão+área Saúde+Gestão+análise Saúde+Gestão+externo Saúde+Gestão+prazo_de_validade Saúde+Gestão+melhoria Saúde+Gestão+educação_continuada Saúde+Gestão+metodologia Saúde+Gestão+material Saúde+Gestão+interno Saúde+Gestão+uso Saúde+Gestão+clínico Saúde+Gestão+casos Saúde+Gestão+vencido Saúde+Gestão+sistemática Saúde+Gestão+processo Saúde+Gestão+registro Saúde+Gestão+informação Saúde+Gestão+atendimento Infecção+Hospitalar+sistema Infecção+Hospitalar+evidência Infecção+Hospitalar+cliente Infecção+Hospitalar+controle Infecção+Hospitalar+área Infecção+Hospitalar+análise Infecção+Hospitalar+externo Infecção+Hospitalar+prazo_de_validade Infecção+Hospitalar+melhoria Infecção+Hospitalar+educação_continuada Infecção+Hospitalar+metodologia Infecção+Hospitalar+material Infecção+Hospitalar+interno Infecção+Hospitalar+uso Infecção+Hospitalar+clínico Infecção+Hospitalar+casos Infecção+Hospitalar+vencido Infecção+Hospitalar+sistemática Infecção+Hospitalar+processo Infecção+Hospitalar+registro Infecção+Hospitalar+informação Infecção+Hospitalar+atendimento
---	--

Com o Sistema de Informações gerado para o caso em foco, em todas as suas etapas, as buscas começam a ocorrer. Embora em um primeiro momento de uma busca, o sistema gerado necessite de uma classificação subjetiva pelo usuário interessado para que o sistema interprete o seu raciocínio, após o treinamento da rede *neuro-fuzzy*, o sistema passa a executar as buscas automaticamente e os conhecimentos do caso em foco ou da organização, representados na ontologia de domínio, vão evoluindo com o tempo.

As tabelas X a XIV mostram os resultados de um período de buscas pelo sistema gerado. Foram analisados os resultados de 13 documentos oriundos das buscas, cujas tarefas com os respectivos resultados, foram executadas pelo GISAAGIC.

Tabela XII- TF mais IDF (Valores de TFIDF).

Termo	1	2	3	4	5	6	7	8	9	10	11	12	13	IDF
CCIH	2,024	2,789												1,872
controle	3,761	1,024		0,876	1,801	0,857							2,090	0,773
infecção_hospitalar	0,709	0,304	0,197	0,372	1,018	0,640		0,463	1,060			0,594	0,591	0,262
médico	3,170	1,699		0,831										1,466
comissão	1,585	0,971		1,246										1,466
clínica	3,035	0,930												1,872
hospital	0,669	0,307		1,228	0,769	0,686						0,701	0,558	0,619
hospitalar	1,672			0,438	0,480	0,343		1,819				0,875		0,773
área	4,047			0,581										1,872
enfermagem	2,024											2,119		1,872
enterococcus		0,620	5,629											1,872
vancomicina		1,240	1,876											1,872
outras		0,930	0,938											1,872
uso		1,240	0,938											1,872
resistente		1,859	0,938											1,872
infecções		4,098		1,669	1,098								2,124	1,179
hospitalares		0,976		1,336	0,732								2,124	1,179
profissional		1,171		0,668		0,784						1,779		1,179
normas		0,620		1,061										1,872
infectologista		0,620			0,581									1,872
casos		0,620			0,581									1,872
maior		0,620			0,581									1,872
contato		0,620				1,245								1,872
função		0,930				0,830								1,872
uti		0,620							3,781					1,872
tratamento		0,930										1,413		1,872
cirúrgica		0,620										1,413		1,872
quais		1,240										1,413		1,872
ações		0,930											2,529	1,872
infecção			0,581	2,190	0,480	0,514		1,364				1,167		0,773
hospitais			2,068		3,111	2,352							3,186	1,179
taxa			0,735		2,277				4,443					1,466
estudo			1,876		2,035									1,872
pacientes			0,591		0,732	1,045						0,890		1,179
brasil			2,205		0,683							2,213		1,466
eficácia			0,938		0,581									1,872
doenças			0,735		0,455	0,650								1,466
doença			0,735			2,601						1,107		1,466
parte			1,407			1,245								1,872
bactérias			3,753			1,245								1,872
grande			0,938			0,830								1,872
além			0,735			0,650						1,107		1,466
modo			0,938			0,830								1,872
saúde				0,541	1,039	0,847						1,082	0,861	0,956
cada				0,831	0,455	0,650								1,466
paciente				3,339	0,366	0,784							2,124	1,179
internação				1,083	0,297	0,424						0,721	1,291	0,956
cuidados				1,061		2,075								1,872
pessoas				1,591		1,660								1,872
procedimentos				1,002		1,045						1,334	1,062	1,179
mãos				1,591		1,245								1,872
lavar				1,061		0,830								1,872
prevenção					0,683	0,975						1,107		1,466
programa					1,163	1,245								1,872
problema					0,872	0,830								1,872
centro					0,872			2,202						1,872
média					0,872				20,798					1,872
índices					0,581				3,781					1,872
flora					0,581							1,413		1,872
política					0,581								3,373	1,872

Tabela XIII - TFIDF com Ponderação Ontológica.

Termo	1	2	3	4	5	6	7	8	9	10	11	12	13	IDF
Ganho Semântico (GS)	25,630	76,260	10,130	25,630	21,562	15,249	0,005	5,130	15,130	0,005	0,005	15,249	15,130	
CCIH	5,186	21,270												1,872
controle	9,641	7,810		2,246	3,883	1,307							3,162	0,773
infecção_hospitalar	1,817	2,319	0,200	0,952	2,196	0,976		0,238	1,604			0,906	0,894	0,262
médico	8,126	12,960		2,129										1,466
comissão	4,063	7,405		3,194										1,466
clínica	7,780	7,090												1,872
hospital	1,715	2,345		3,146	1,658	1,047						1,069	0,844	0,619
hospitalar	4,285			1,123	1,035	0,523		0,933				1,335		0,773
área	10,373				1,253									1,872
enfermagem	5,186											3,231		1,872
enterococcus		4,727	5,703											1,872
vancomicina		9,453	1,901											1,872
outras		7,090	0,950											1,872
uso		9,453	0,950											1,872
resistente		14,180	0,950											1,872
infecções		31,251		4,279	2,368								3,213	1,179
hospitalares		7,441		3,423	1,578								3,213	1,179
profissional		8,929		1,712		1,196						2,713		1,179
normas		4,727		2,718										1,872
infetologista		4,727			1,253									1,872
casos		4,727			1,253									1,872
maior		4,727			1,253									1,872
contato		4,727				1,899								1,872
função		7,090				1,266								1,872
uti		4,727						5,721						1,872
tratamento		7,090										2,154		1,872
cirúrgica		4,727										2,154		1,872
quais		9,453										2,154		1,872
ações		7,090											3,827	1,872
infecção			0,589	5,614	1,035	0,784		0,700				1,780		0,773
hospitais			2,095		6,709	3,587							4,820	1,179
taxa			0,745		4,909			6,723						1,466
estudo			1,901		4,387									1,872
pacientes			0,598		1,578	1,594						1,356		1,179
brasil			2,234		1,473							3,375		1,466
eficácia			0,950		1,253									1,872
doenças			0,745		0,982	0,992								1,466
doença			0,745			3,966						1,688		1,466
parte			1,426			1,899								1,872
bactérias			3,802			1,899								1,872
grande			0,950			1,266								1,872
além			0,745			0,992						1,688		1,466
modo			0,950			1,266								1,872
saúde				1,388	2,239	1,292						1,650	1,302	0,956
cada				2,129	0,982	0,992								1,466
paciente				8,558	0,789	1,196							3,213	1,179
internação				2,775	0,640	0,646						1,100	1,954	0,956
cuidados				2,718		3,164								1,872
pessoas				4,077		2,532								1,872
procedimentos				2,567		1,594						2,035	1,607	1,179
mãos				4,077		1,899								1,872
lavar				2,718		1,266								1,872
prevenção					1,473	1,487						1,688		1,466
programa					2,507	1,899								1,872
problema					1,880	1,266								1,872
centro					1,880			1,130						1,872
média					1,880				31,467					1,872
índices					1,253				5,721					1,872
flora					1,253							2,154		1,872
política					1,253								5,103	1,872

A idéia desta ponderação da ontologia é fazer com que os termos que aparecem um número menor de vezes tenham uma valoração maior, de forma semelhante à IDF (frequência inversa de termos), reforçando assim a representatividade do vetor do documento. Na seqüência, cada termo presente em um documento é comparado com a sua existência na ontologia e desta forma calcula-se o fator de ganho semântico.

Na tabela X, está ilustrada a frequência de termos absoluta referente a 59 termos presentes no vocabulário, dentre 13 documentos. Cada coluna pode ser entendida como o vetor de cada documento. Na tabela XI, é apresentada a frequência de termos relativa, com os valores normalizados. Na tabela XII, estão os valores da TFIDF (frequência de termos multiplicada pela frequência inversa de termos), onde na última coluna se encontram os valores da IDF aplicada sobre cada termo. E na tabela XIII, encontram-se os valores da TFIDF mais a ponderação ontológica através de um fator de ganho semântico (conforme a primeira linha) aplicado para cada vetor de documento.

É importante ressaltar que embora o sistema já contenha uma lista de *stopwords* previamente definidas, nas primeiras buscas retornam palavras não previstas inicialmente na lista, como: “outras, maior, quais, parte, grande, além, entre outras”, as quais podem causar ruídos em um primeiro momento, porém, são eliminadas na seqüência das operações.

Na tabela XIV, foi considerado o uso dos métodos da TF, TFIDF, TF com ponderação ontológica e TFIDF com ponderação ontológica. Considerou-se um conjunto de 13 documentos referentes ao domínio dos termos “infecção hospitalar” em uma busca efetuada na WEB. Verifica-se então que a ponderação pela ontologia do domínio contribuiu para aumentar a qualidade da representação dos vetores de documentos.

Tabela XIV- Valores Obtidos para o Ganho De Informação nos Diferentes Métodos

<i>Método</i>	<i>Ganho de Informação</i>
TF	2,4885
TF com ponderação ontológica	2,1620
TFIDF (TF mais IDF)	2,5825
TFIDF com ponderação ontológica	2,6856

A seguir, uma classificação *a priori* do usuário é necessária num primeiro momento de uso do sistema gerado, quando uma pesquisa é efetuada com um conjunto de termos-chave. Os documentos retornados pelo agente de busca são classificados pelo usuário de acordo com

a importância subjetiva para o mesmo. Essa pré-classificação serve para formar o conjunto inicial de treinamento da rede *neuro-fuzzy*.

As figuras 38, 39 e 40 mostram as telas do agente GEDI envolvidas nesta fase. A tela da figura 38 mostra os documentos retornados pelo buscador para a conta de busca 0000002, mostrando a relevância já considerada para cada documento feita pelo usuário.

Este conjunto fará parte do treinamento da rede *neuro-fuzzy*.

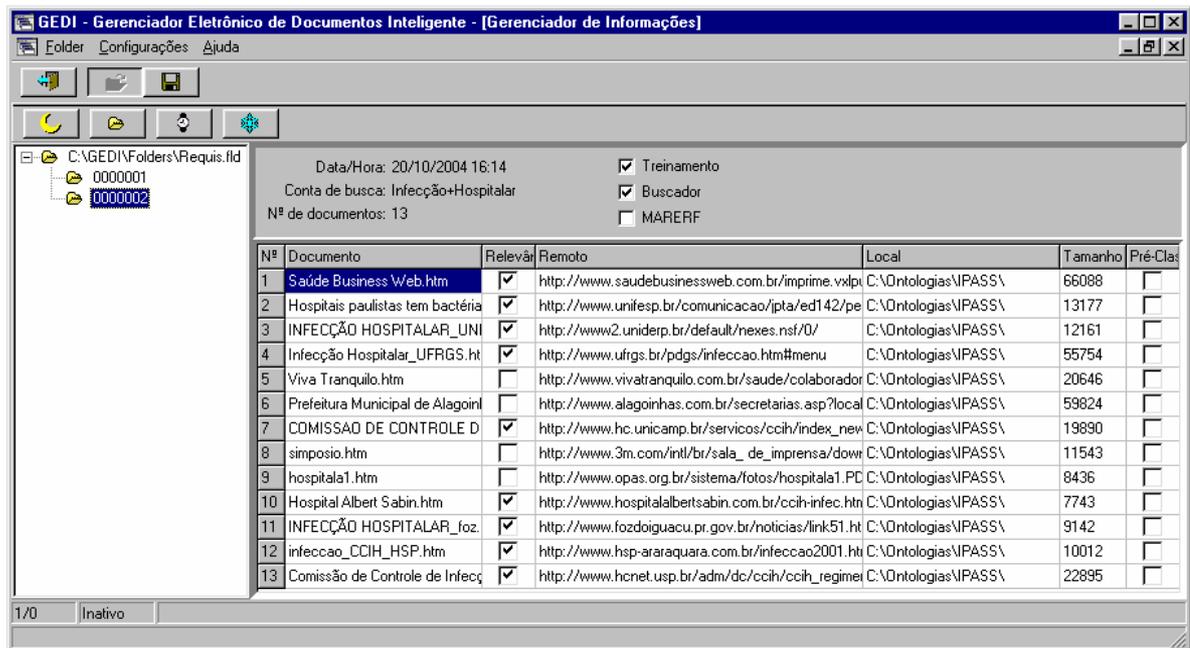


FIGURA 38 - GEDI – TELA PRINCIPAL.

A rede então é treinada com 10 épocas de treinamento (número considerado aceitável em um primeiro momento de buscas), mostrando o erro final de classificação obtido.

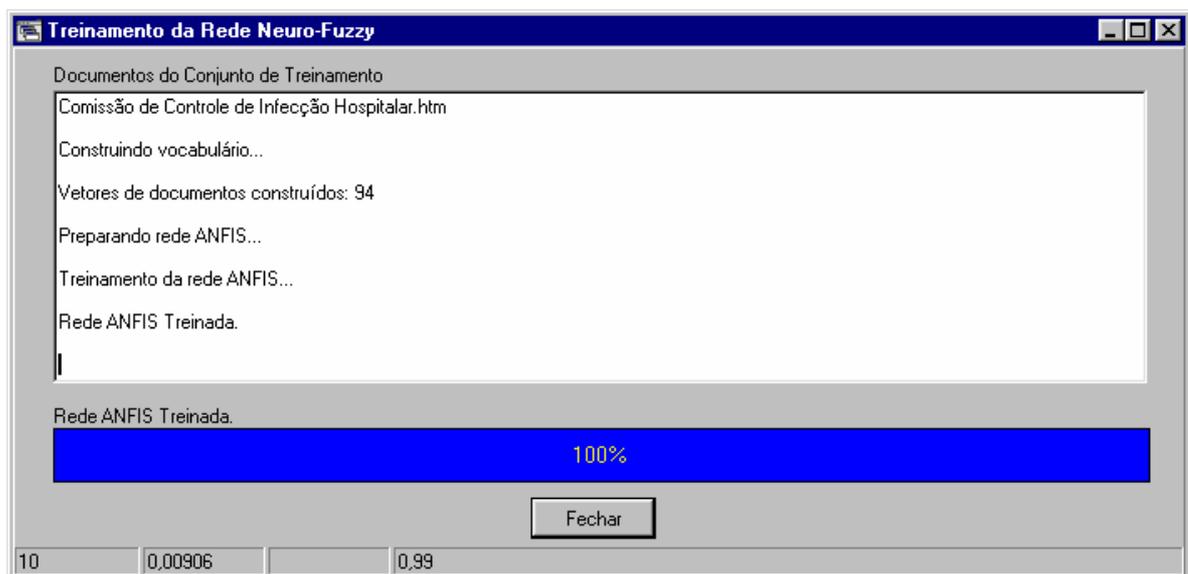


FIGURA 39 - GEDI – MÓDULO DE TREINAMENTO DA REDE *NEURO-FUZZY*

A figura 40 mostra a tela onde aparece o conjunto de documentos pré-classificados de acordo com a rede *neuro-fuzzy*.

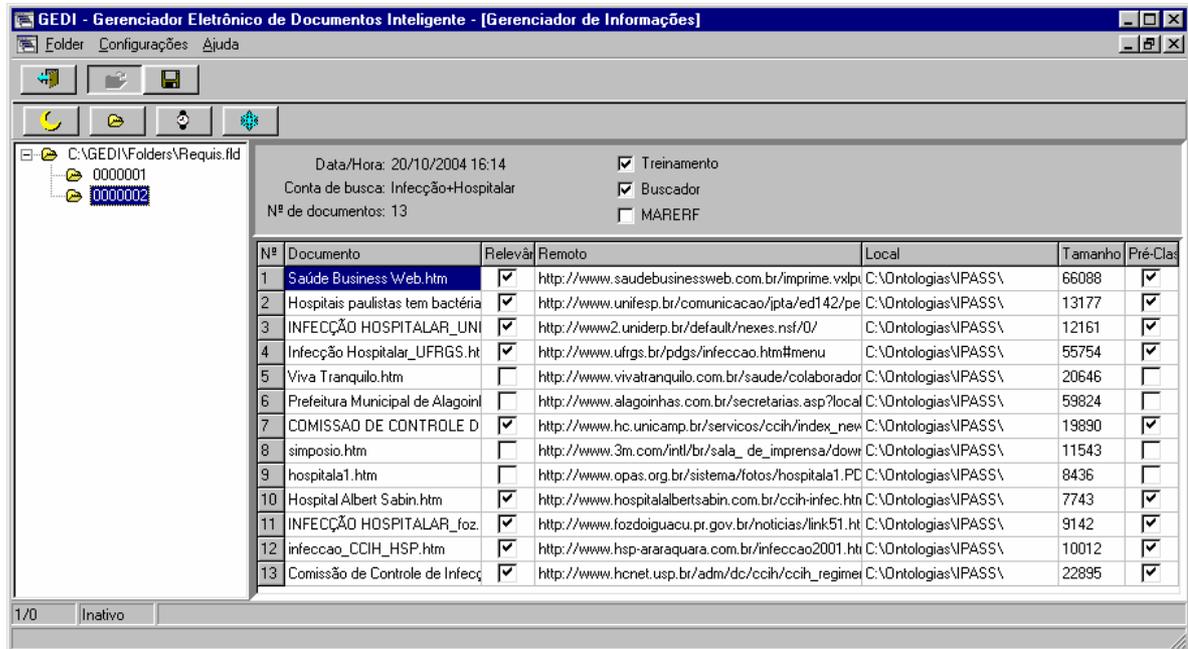


FIGURA 40 - GEDI – PRÉ-CLASSIFICAÇÃO.

A partir deste passo, o sistema tem condições de fazer a tomada de decisão, interpretando o raciocínio do usuário e passando a processar automaticamente as informações. As novas informações relevantes, agora classificadas pelo sistema e validadas pelo usuário, vão sendo agregadas ao sistema, em um processo contínuo de aprendizagem. Na classificação são atribuídos valores, considerando-se um grau de pertinência de valores *fuzzy* de acordo com as prioridades requeridas pelo vocabulário e pela ontologia de domínio.

Na figura 41 tem-se uma rede ANFIS utilizada para a classificação do exemplo, mostrando as camadas da rede processando um vetor de documento. Na camada de entrada são dispostos os valores deste vetor. A rede possui 59 entradas e uma saída. A seguir, na camada de funções de pertinência os valores produzem as saídas *fuzzy* que, por sua vez, são processadas na camada do produto, normalizadas, e alimentam, na seqüência, a camada dos parâmetros do conseqüente. Os dados de entrada são processados também nesta camada e produzem o valor de saída. A rede é treinada (através do método híbrido) com os vetores selecionados pelo usuário, sendo atribuído ao neurônio de saída o valor 1 (um) se o documento é relevante ou 0 (zero) em caso contrário. O teste na figura 41 indica, então, um documento relevante com valor de saída 1,0000.

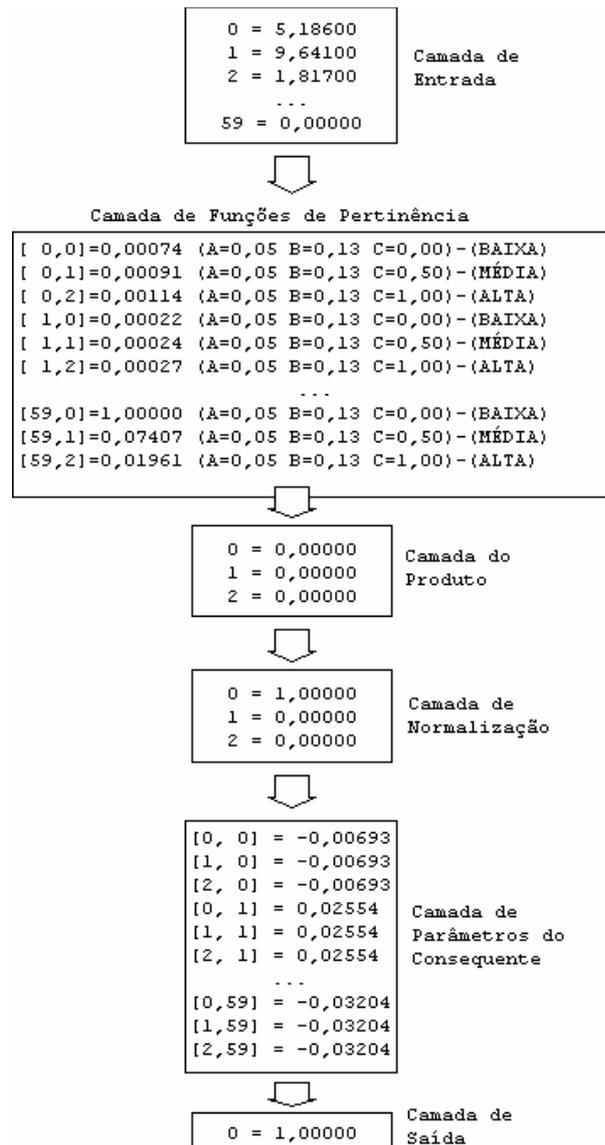


FIGURA 41 – CAMADAS DA REDE NEURO-FUZZY ANFIS PROCESSANDO UM DETERMINADO VETOR DE DOCUMENTO.

Outro aspecto a considerar é o aparecimento de novos termos nas buscas inteligentes que não fazem parte da ontologia de domínio. Após o processamento do GEDI, termos com a relevância aumentada pelo fator de ganho semântico serão disponibilizados como novos conceitos a serem validados pelo usuário para fazer parte da ontologia.

5.3.3. SubProcesso de Manutenção

Neste subprocesso, o agente MARERF executa tarefas de agregação de novos conceitos encontrados nas buscas para a evolução da ontologia de domínio, com algumas fases necessitando de interferência do Usuário Especialista. Após a agregação de novos conceitos, o módulo de avaliação é utilizado para se verificar, através da métrica de ontologias, uma medida quantitativa da evolução da ontologia de domínio.

5.3.3.1 Conversão de UNL para Ontologia

Os documentos retornados pelas buscas inteligentes podem estar representados em UNL. Para este caso, o SI já foi concebido para suportar esta representação, potencializando ainda mais o GISAAGIC em função da riqueza de dados constantes na representação UNL, conforme exposto na apresentação do modelo no capítulo 4.

Na prática, veja-se o caso de um dos documentos retornados nas buscas, o artigo de Ricardo Zorzetto “*Hospitais paulistas têm bactéria que causa infecção hospitalar*”, cujo documento foi focado em função da palavra “*Enterococcus*”, que é um dos conceitos presentes na ontologia de domínio. Para facilitar o entendimento, como a representação UNL utiliza a língua inglesa por se tratar de uma representação universal, na seqüência está sendo mostrado um dos parágrafos do artigo em tela na citada língua e a sua representação em UNL.

*“In a study which mapped the resistance profile of **Enterococcus** bacterium in Latin America, the researcher Rosângela Cereda detected many kinds resistant to the medicines from this microorganism in Brazil and, more precisely, in São Paulo.”*

Uma possível representação UNL deste documento é mostrada no quadro a seguir, com o respectivo grafo UNL representado na figura 42:

```
agt (map (icl>verb) .@entry.@pred.@past,
study (icl>research) .@indef)
opl (study (icl>research) .@indef, Latin America.@def)
obj (map (icl>verb) .@entry.@past, profile.@def)
aoj (profile.@def, resistance)
obj (profile.@def, bacterium (icl>microorganism))
```

```

nam(bacterium(icl>microorganism), Enterococcus)
ptn(study(icl>research).@indef, researcher.@def)
mod(researcher.@def, Rosangela Cereda(icl>woman))
agt(detect.@past, researcher.@def)
plc(detect.@past,
and(Brazil(icl>country), mod(city(icl>place), São
Paulo.@focus)))
obj(detect.@past, kind.@pl)
aoj(kind.@pl, many)
aoj(kind.@pl, resistant.@pl)
gol(resistant.@pl, medicine.@pl)
cob(medicine.@pl, microorganism(icl>living thing))

```

De acordo com o exposto no capítulo 4 sobre a conversão UNL – Ontologia, gera-se uma ontologia temporária para o grafo UNL, a qual está representada na figura 43. Pode-se notar que algumas etiquetas de relação foram convertidas conforme a tabela, existindo, porém, outras que dependem de intervenção do usuário especialista (expressas com um ponto de interrogação). Adotando-se a estratégia de conexão automática, o conceito *Enterococcus* encontra-se presente na ontologia, servindo como ponto de ligação entre a ontologia de domínio e a ontologia temporária (indicado com círculo pontilhado).

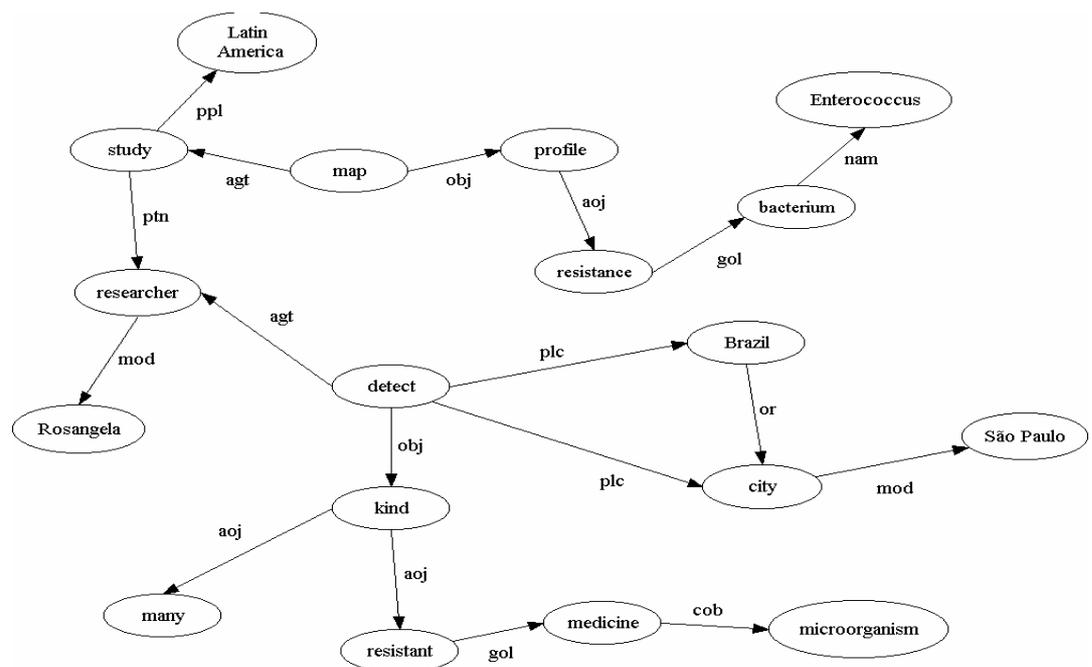


FIGURA 42 - GRAFO UNL REFERENTE AO PARÁGRAFO SUPRACITADO.

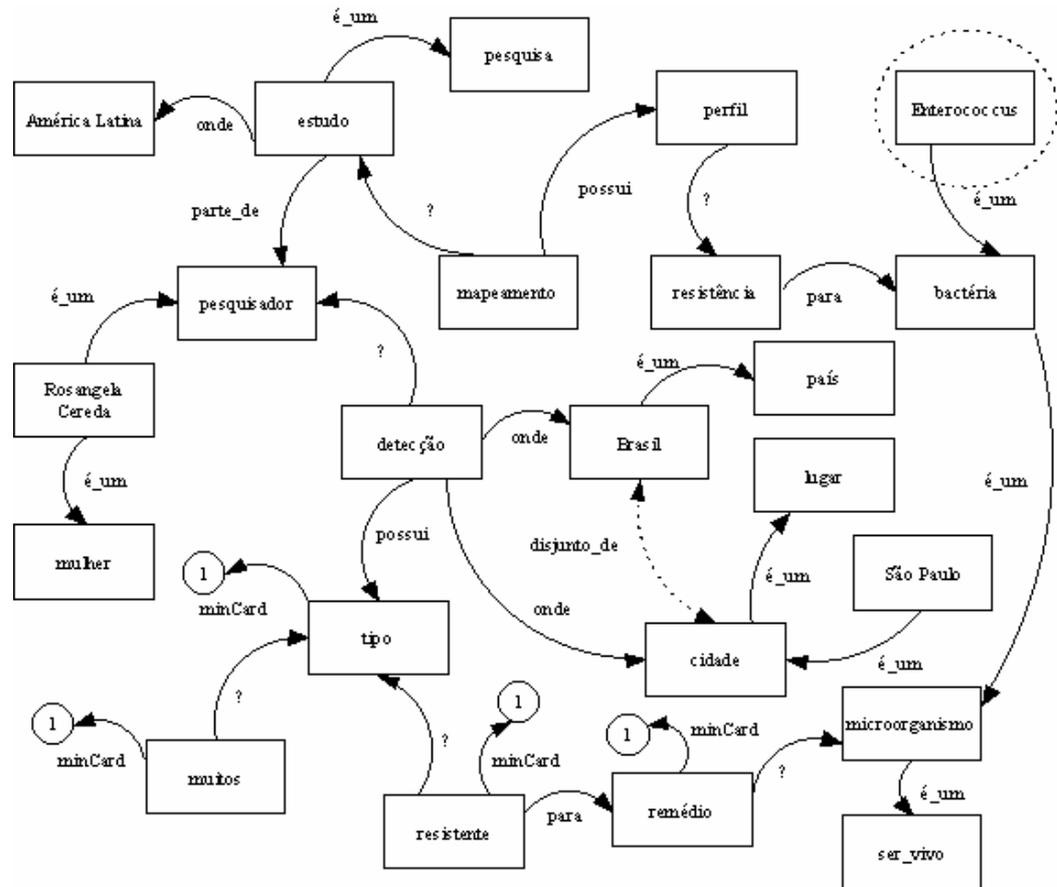


FIGURA 43 – ONTOLOGIA TEMPORÁRIA PARA O PARÁGRAFO UNL.

Para cada palavra universal limitada existente na lista, foram gerados dois conceitos na ontologia temporária. Nota-se também que os verbos “mapear” e “detectar” foram substituídos pelos substantivos “mapeamento” e “detecção” (para eles, a etiqueta de relação **icl** não gerou o conceito “verbo”). Pode-se verificar que para os conceitos com etiquetas de atributo **.@pl** foram geradas relações de mínima cardinalidade unitária. A etiqueta de relação **or** gerou um axioma disjuntivo entre os conceitos “Brasil” e “cidade”.

Em termos de implementação, é gerada uma representação DAML+OIL para esta ontologia temporária como arquivo, ficando disponibilizada ao MARERF para ligação com a ontologia de domínio. Uma versão da ontologia da figura 38 escrita em DAML+OIL é a seguinte:

```
<rdf:RDF
  xmlns:rdf = "http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs = "http://www.w3.org/2000/01/rdf-schema#"
  xmlns:daml = "http://www.daml.org/2001/03/daml+oil#"
  xmlns:xsd = "http://www.w3.org/2000/10/XMLSchema#"
  xmlns:dex = "http://www.daml.org/2001/03/daml+oil-ex#"
```

```

xmlns:exd ="http://www.daml.org/2001/03/daml+oil-ex-dt#"
xmlns     ="http://www.daml.org/2001/03/daml+oil-ex#"
ont       ="Ontology.daml>#"
>

<daml:Ontology rdf:about="">
  <daml:imports rdf:resource="http://www.daml.org/2001/03/daml+oil"/>
</daml:Ontology>

<ont:Enterococcus>
  <rdfs:subClassOf rdf:resource="#bactéria"/>
</ont:Enterococcus >

<daml:Class rdf:ID="bactéria">
  <rdfs:subClassOf rdf:resource="#microorganismo"/>
</daml:Class>

<daml:Class rdf:ID="resistência">
  <daml:Restriction>
    <daml:onProperty rdf:resource="para"/>
    <daml:toClass rdf:resource="#bactéria"/>
  </daml:Restriction>
</daml:Class>

<daml:Class rdf:ID="perfil">
  <daml:Restriction>
    <daml:onProperty rdf:resource="?"/>
    <daml:toClass rdf:resource="#resistência"/>
  </daml:Restriction>
</daml:Class>

<daml:Class rdf:ID="mapeamento">
  <rdfs:subClassOf>
    <daml:Restriction>
      <daml:onProperty rdf:resource="#possui"/>
      <daml:toClass rdf:resource="#perfil"/>
    </daml:Restriction>
  </rdfs:subClassOf>
  <rdfs:subClassOf>
    <daml:Restriction>
      <daml:onProperty rdf:resource="#onde"/>
      <daml:toClass rdf:resource="#América Latina"/>
    </daml:Restriction>
  </rdfs:subClassOf>
</daml:Class>

<daml:Class rdf:ID="perfil">
</daml:Class>

<daml:Class rdf:ID="pesquisa">
</daml:Class>

```

```

<daml:Class rdf:ID="estudo">
  <rdfs:subClassOf rdf:resource="#pesquisa"/>
<rdfs:subClassOf>
  <daml:Restriction>
    <daml:onProperty rdf:resource="#parte de"/>
    <daml:toClass rdf:resource="#pesquisador"/>
  </daml:Restriction>
</rdfs:subClassOf>
</daml:Class>

<daml:Class rdf:ID="América Latina">
</daml:Class>

<daml:Class rdf:ID="pesquisador">
</daml:Class>

<daml:Class rdf:ID="Rosângela Cereda">
  <rdfs:subClassOf rdf:resource="#pesquisador"/>
</daml:Class>

<daml:Class rdf:ID="Rosângela Cereda">
  <rdfs:subClassOf rdf:resource="#mulher"/>
</daml:Class>

<daml:Class rdf:ID="detecção">
  <rdfs:subClassOf>
    <daml:Restriction>
      <daml:onProperty rdf:resource="?"/>
      <daml:toClass rdf:resource="#pesquisador"/>
    </daml:Restriction>
  </rdfs:subClassOf>
</daml:Class>

<daml:Class rdf:ID="tipo">
  <daml:Restriction daml:minCardinality="1">
    <daml:onProperty rdf:resource="#possui"/>
  </daml:Restriction>
</daml:Class>

<daml:Class rdf:ID="muitos">
  <daml:Restriction daml:minCardinality="1">
    <daml:onProperty rdf:resource="?"/>
    <daml:toClass rdf:resource="#tipo"/>
  </daml:Restriction>
</daml:Class>

<daml:Class rdf:ID="resistente">
  <daml:Restriction daml:minCardinality="1">
    <daml:onProperty rdf:resource="?"/>
    <daml:toClass rdf:resource="#tipo"/>
  </daml:Restriction>
</daml:Class>

```

```

<daml:Class rdf:ID="remédio">
  <daml:Restriction daml:minCardinality="1">
    <daml:onProperty rdf:resource="?"/>
    <daml:toClass rdf:resource="#microorganismo"/>
  </daml:Restriction>
</daml:Class>

<daml:Class rdf:ID="microorganismo">
  <rdfs:subClassOf rdf:resource="#ser vivo"/>
</daml:Class>

<daml:Class rdf:about="#detecção">
  <rdfs:subClassOf>
    <daml:Restriction>
      <daml:onProperty rdf:resource="#onde"/>
      <daml:toClass rdf:resource="#Brasil"/>
    </daml:Restriction>
  </rdfs:subClassOf>
</daml:Class>

<daml:Class rdf:ID="Brasil">
  <rdfs:subClassOf rdf:resource="#país"/>
  <daml:disjointWith rdf:resource="#cidade"/>
</daml:Class>

<daml:Class rdf:ID="cidade">
  <rdfs:subClassOf rdf:resource="#lugar"/>
</daml:Class>

<daml:Class rdf:ID="São Paulo">
  <rdfs:subClassOf rdf:resource="#cidade"/>
</daml:Class>

<daml:ObjectProperty rdf:ID="possui">
</daml:ObjectProperty>

<daml:ObjectProperty rdf:ID="para">
</daml:ObjectProperty>

<daml:ObjectProperty rdf:ID="onde">
</daml:ObjectProperty>

<daml:ObjectProperty rdf:ID="parte_de">
</daml:ObjectProperty>

<daml:ObjectProperty rdf:ID="?">
</daml:ObjectProperty>

```

Deve-se notar na representação acima que, na área de espaço de nomes (*xmlns*), já está referida a ontologia de domínio com o prefixo *ont* (tendo o nome de arquivo *Ontology.daml*). Isto é preciso para fazer com que o conceito responsável pela ligação (no caso, o conceito *Enterococcus*) seja referenciado como pertencente à ontologia de domínio e não da ontologia temporária. E o seguinte comando em DAML+OIL:

```
<ont:Enterococcus>
  <rdfs:subClassOf rdf:resource="#bactéria"/>
</ont:Enterococcus >
```

é responsável pela conexão entre as duas ontologias.

Como visto na conversão do parágrafo UNL para a ontologia temporária, a etiqueta de relação **or** gerou um axioma disjuntivo entre os conceitos “Brasil” e “cidade”. Este axioma está representado no código DAML+OIL, na definição do conceito “Brasil”:

```
<daml:Class rdf:ID="Brasil">
  <rdfs:subClassOf rdf:resource="#país"/>
  <daml:disjointWith rdf:resource="#cidade"/>
</daml:Class>
```

A figura 44 mostra a conexão da ontologia temporária com a ontologia de domínio.

Neste trabalho, foram adotados os princípios do método desenvolvido por Tello (2002), cuja seqüência dos passos são mostrados na figura 45.

Seqüência de Passos do Teste Ontométrico



FIGURA 45 – TESTE ONTOMÉTRICO.

Os passos mostrados no fluxograma acima são executados pelos agentes do sistema, os quais executam tarefas de forma compartilhada, cujos resultados são avaliados pelo módulo de avaliação, resultando na indicação da melhor ontologia para o caso em tela.

5.4. CONSIDERAÇÕES FINAIS

Este capítulo apresentou um exemplo de aplicação na área de ensino, envolvendo a disciplina empreendedorismo na parte do conteúdo programático referente a plano de negócio. A aplicação focou a métrica para avaliação da qualidade de informações contidas em documentos oriundos de diferentes fontes, considerando as relações semânticas entre as palavras e sentenças, sem destacar as demais etapas do processo, para facilitar o entendimento desta etapa do processo e destacar o caráter genérico do sistema.

Um exemplo de aplicação para a área médica, envolvendo a subárea de infecção hospitalar, focalizou as etapas desde o preparo e geração do sistema de informação até a informação estruturada e contextualizada para apoio às tomadas de decisões gerenciais.

Foi apresentado, também, um exemplo para o caso de documentos retornados pelo buscador inteligente estarem na forma de representação UNL, o que mostrou a potencialidade da conversão de um documento UNL para geração de ontologias através da diversidade de relações semânticas envolvidas na representação da linguagem UNL.

6. CONSIDERAÇÕES DO SOFTWARE E OUTROS SISTEMAS

6.1. CONSIDERAÇÕES PRÁTICAS DO *SOFTWARE*

Embora na aplicação prática do GISAAGIC na área de saúde, subárea de Infecção Hospitalar, já tenha sido mostrada toda a estrutura de geração, processamento e resultados do sistema, acrescentam-se aqui mais algumas considerações técnicas e comparação com outros sistemas existentes no mercado que executam parte das tarefas habilitadas no sistema proposto.

O GISAAGIC foi desenvolvido na linguagem Delphi, em ambiente Windows. Apesar disto, o conceito pode ser aplicado em ambientes multiplataforma. A implementação é justificada pelo fato dos módulos do SMAGS terem sido desenvolvidos em Delphi no trabalho anterior. O Módulo Buscador, O GEDI e o MARERF foram implementados na filosofia de agentes, ficando em espera de acordo com as mensagens provindas do Gerente Virtual.

As ontologias são criadas e mantidas no formato DAML+OIL no momento atual, e a forma de implementação e carga em memória segue a filosofia de orientação a objetos. As ontologias, os espaços de nomes, conceitos e propriedades são instanciados como classes de coleções e itens de coleções.

O módulo gerador de sistemas utilizado foi o Automa Builder. Esta ferramenta CASE gera sistemas de informações a partir de um dicionário de dados fornecido, e gera código fonte em Delphi, podendo ser configurado para criar banco de dados em formatos como SQL Server, Oracle, Microsoft Access e Interbase. Para a geração do sistema de informações para a aplicação da área da saúde, foi utilizado o banco de dados no formato Microsoft Access. Outras ferramentas CASE poderão ser utilizadas para a geração de sistemas de informação, inclusive em outras linguagens.

O editor de ontologias utilizado para criar a ontologia de domínio foi o OilEd. Na próxima fase deverá ser implementado um editor de ontologias próprio para ser fornecido junto ao GISAAGIC para as tarefas de manutenção das ontologias. Outros editores, como o Protegé, podem ser utilizados para, inclusive, converter a ontologia escrita em DAML+OIL em outros formatos, adicionando portabilidade.

Algumas considerações são oportunas para evidenciar os aspectos práticos encontrados durante o processo de implementação do GISAAGIC.

- 1) No processamento do texto e extração de características, um arquivo grande leva muito tempo para ser processado. Dependendo do computador utilizado, a performance desta fase pode apresentar variações. Um texto padrão ANSI com 154 KB (13.317 palavras) levou em média 78 segundos no processo de *text mining* num computador PIII 800MHz;
- 2) No processamento de páginas HTML acontece menor performance de extração de características em relação a textos puros, devido à presença de *tags*. Assim, foi necessário utilizar um conversor HTML para texto;
- 3) Na conversão de documentos HTML para texto puro, deve ser considerado o padrão do texto sendo utilizado, se Unicode ou ANSI-ISO;
- 4) Foi necessária a utilização de conversor PDF para texto, em função de muitos documentos encontrados já estarem neste formato;
- 5) É passível de serem encontrados erros de lingüística nos textos buscados na *internet*. Este problema levanta a necessidade de se considerar um corretor ortográfico dentro do contexto do sistema ou do GEDI para refinar a confiabilidade dos vetores de documentos;
- 6) Em função do tamanho excessivo de alguns arquivos pode-se estipular um tamanho máximo a ser processado, a ser configurado pelo sistema;
- 7) Os documentos UNL devem apresentar uma espécie de mensagem para o Gerente Virtual de forma a indicar ao usuário que existem ontologias convertidas para UNL em processo de agregação;
- 8) Na conversão para UNL, poderá existir mais de um termo comum. Assim deve ser prevista a definição de prioridades para conexão das ontologias provindas do documento UNL e da ontologia de domínio;
- 9) Na hipótese do conversor ENCO para geração automática de documentos UNL estar disponível para consultas via WEB, um processo em *batch* pode converter os textos de documentos diversos (HTML, PDF, etc.) em sentenças UNL e, assim, possibilitar o uso da estrutura semântica de tais documentos.

6.2. COMPARAÇÃO COM OUTROS SISTEMAS

Os sistemas em uso no mercado, na área de saúde, tais como **MedyDoc**, **SofyMed**, **ArchiMED**, **RIS CEDI 2000** e outros, executam tarefas de gestão de fichas de pacientes, controle de informações de diagnósticos por imagens, controle eletrônico de documentos, controle de medicamentos, agenda eletrônica, diagnósticos, controle do código internacional de doenças CID9 e CID10, controles financeiros, cadastros e outros. A maioria é baseada em sistemas especialistas, sem, no entanto, executarem tarefas inteligentes. Mesmo nos ditos países do primeiro mundo, os sistemas de apoio à decisão médica, a nível profissional, também são baseados em métodos estatísticos e probabilísticos, exceto algumas universidades que oferecem uma disciplina chamada “*Artificial Intelligence in Medicine*” (AIM). Estas desenvolvem os chamados “*Clinical Decision Support Systems*”, baseados em modelos simbólicos que usam técnicas avançadas de Inteligência Artificial, embora com aplicação em partes separadas, sem integração. Existem alguns sistemas que são capazes de fazer diagnóstico e recomendações de terapia, como o da Escola de Medicina de Harvard, no Hospital Geral de Massachusetts, Estados Unidos; contudo, ainda não possuem interatividade com o usuário e nem funções de auto-aprendizagem, além de não serem ainda recomendados para uso profissional, segundo os próprios autores.

O modelo proposto e já parcialmente implementado a nível de protótipo, em aplicações da vida real, é capaz de fazer, além do que os outros já fazem, outras tarefas, incorporando o conhecimento de especialistas da área, mantidos em ontologias de domínio do assunto de interesse. Por sua vez não são estanques e estão em um processo contínuo de evolução e aprendizado, conforme foi mostrado neste trabalho e validado na prática, com resultados que abrem novos caminhos para a área em que o sistema foi testado. As diferentes tecnologias são aplicadas de forma integradas entre elas as ricas formas de representar e interpretar os conteúdos através dos conceitos e relações contidos na representação UNL, as quais são integradas à ontologia de domínio, possibilitando buscas inteligentes, direcionadas para o assunto de interesse, mesmo que sejam oriundas de outros países com outras línguas nativas. O sistema é capaz de estruturar os dados e contextualizar as informações para apoio às decisões gerenciais.

O GISAAGIC mantém, também, um banco de casos de sucessos para subsidiar as decisões e permitir a avaliação comparativa da evolução do sistema, incorporando, em sua

base de conhecimento integrada à ontologia de domínio, os novos conceitos e relações encontrados que possam melhorar o sistema, em processo contínuo de aprendizagem.

O GISAAGIC incorpora tecnologias do SMAGS, metodologias do GEDI com um Meta-Agente de Raciocínio e Extração de Regras *Fuzzy* (MARERF), mais ontologia e conceitos de UNL, viabilizando tarefas do agente buscador inteligente.

O núcleo extrator de informações compartilha ações do Meta-Agente MARERF com uma rede *Neuro-Fuzzy*, combinando aspectos simbólicos e conexionistas provenientes da inteligência artificial.

Todas as ações integradas, em um processo pró-ativo *on line*, viabilizam a tomada de melhores decisões em curtos espaços de tempo. Em certos casos, pode ser em tempo real e executadas pelo próprio sistema, desde que autorizado pelo usuário. Após a modelagem do SI a partir do modelo relacional proposto, o SI é gerado e está apto à entrada de dados e ao processamento das informações. Apesar de seguirem um procedimento genérico, tal interação se modificará de acordo com a implementação na área de interesse.

As informações retornadas, após processadas e validadas irão complementar as informações no banco de dados do SI. Desta forma, informações ou conceitos bem relevantes podem sugerir a geração de uma nova versão do SI, agora com novos conhecimentos agregados. Estas ações justificam o nome de Gerador Inteligente de Sistemas com Auto-Aprendizagem.

Quanto à interação com o processo de busca inteligente, uma classificação *a priori* é necessária no primeiro momento que o sistema é gerado e retorna da primeira pesquisa. Esta pré-classificação serve para formar o conjunto inicial de treinamento da RNF. A partir deste passo, o sistema interpreta o raciocínio e passa a processar automaticamente as informações. As novas informações relevantes, agora classificadas pelo sistema e validadas pelo usuário, vão sendo agregadas ao sistema, em um processo contínuo de aprendizagem.

7. CONCLUSÕES E RECOMENDAÇÕES

7.1. CONCLUSÕES

A ferramenta CASE desenvolvida, denominada GISAAGIC, gerou na prática, com dados reais, um sistema inteligente, híbrido, que utiliza várias técnicas de Inteligência Artificial integradas.

O sistema efetua buscas inteligentes em *intranet* e *internet*, de forma direcionada para a área de interesse que o sistema for gerado, analisando, estruturando e contextualizando os dados para fornecer informações organizadas e atualizadas em tempo real. Estas informações possibilitam formular hipóteses e prever resultados, apoiando a tomada de decisões gerenciais de forma justificada, evitando conflitos e desperdícios de tempo e dinheiro, envolvendo qualidade e gestão moderna com transparência.

O sistema provou ser capaz de estruturar e contextualizar dados oriundos de buscas inteligentes direcionados para áreas de interesse do usuário, transformando-os em informações e conhecimentos que facilitam as tomadas de decisões gerenciais.

O sistema multiagente SMAGS, integrado ao GISAAGIC, facilita a comunicação com sistemas proprietários diferentes e que utilizam diferentes bancos de dados. Facilita, também, a comunicação com o usuário por possuir recursos de reconhecimento de voz, interagindo de forma escrita e falada.

A tecnologia de agentes, combinada com sistemas neurodifusos e apoiados por técnicas de mineração de dados e de textos, potencializa a extração e recuperação de informações através da *internet* ou *intranets*.

A representação do conhecimento de organizações através de ontologias, utilizando conceitos de representação UNL, potencializa e facilita a evolução da ontologia através da agregação de novos conhecimentos oriundos de buscas inteligentes, em função da riqueza de dados presentes, melhorando o entendimento dos conceitos e relações.

A avaliação comparativa entre as novas ontologias, originadas da incorporação de novos conhecimentos fornecidos pelas buscas inteligentes, e uma ontologia de referência

considerada como ontologia de domínio da área de interesse, permitem medir a evolução dos conhecimentos da organização.

Com as informações estruturadas e contextualizadas em tempo real, as decisões gerenciais são antecipadas e justificadas, viabilizando melhores e mais rápidas decisões, dentro de um contexto globalizado em constante mutação.

Uma das dificuldades encontradas para a implantação do sistema foi a necessidade de treinamento de um usuário especialista. O sistema exige que as rotinas de interação (tais como a pré-classificação dos documentos no GEDI e a manipulação da ontologia de domínio) sejam efetuadas periodicamente por um colaborador especialista e não existe disponibilidade de pessoas em tempo integral na instituição para esta tarefa. O sucesso da implantação do sistema exige também um processo de mudança organizacional e comportamental por parte dos seus colaboradores.

Outra dificuldade encontrada foi a inserção dos manuais de avaliação no sistema. No tempo atual de implantação foi possível apenas a colocação de parte de um dos manuais. Aliado a isto, existe ainda a atualização anual dos manuais por parte da ANVISA. Isto evidenciou a necessidade de agregar ao sistema um processo de conversão automática (visto que os manuais podem ser fornecidos em formato PDF) para facilitar o processo de inserção e exigir menos tempo de digitação do usuário.

A aplicação do sistema, na prática, evidenciou os seguintes benefícios:

- O sistema proporcionou a organização das atividades da instituição, a qual geria seus processos de forma isolada e restrita antes da implementação;

- O sistema proporcionou melhor organização do conhecimento, permitindo a consulta rápida aos diagnósticos já efetuados e permitindo a agregação de informações buscadas na *internet*, tais como artigos atuais e informações de cursos oferecidos;

- Permitiu a comparação até mesmo de diagnósticos em momentos diferentes de um mesmo cliente;

- Devido à necessidade da instituição de padronização de acordo com as normas ISO 9002, o sistema melhorou o planejamento e controle dos processos e evidenciou melhorias a serem exploradas;

- A instituição evidenciou a necessidade de expansão do sistema para que seus avaliadores possam fazer a utilização dos processos de busca de forma remota, quando estão

levantando as informações em campo ou relatando tais levantamentos no sistema, o que iria agilizar o processo de montagem do diagnóstico.

7.2. TRABALHOS FUTUROS

Os trabalhos desenvolvidos dentro do contexto desta tese, em particular o modelo proposto, oferecem subsídios ao desenvolvimento de outros projetos de pesquisa que podem melhorar e potencializar ainda mais o sistema de apoio a decisões gerenciais proposto:

- O desenvolvimento de um editor de ontologias agregado ao GISAAGIC que contemple a conversão e tratamento automatizado das relações de ontologias importadas de documentos UNL;

- O desenvolvimento de conversores da ontologia de domínio para documentos UNL, permitindo a portabilidade do conhecimento e publicação em forma de ontologia, bem como a interação com engenhos de busca que utilizem ontologias;

- O desenvolvimento de conversores de documentos em formatos HTML, texto ou PDF (buscados pelo GEDI) para UNL, funcionando de forma agregada aos *softwares* ENCODECO, de forma a proporcionar um meio de geração e disponibilização de documentos UNL para o ambiente externo do sistema de informação;

- O uso de outros modelos de classificadores para o agente GEDI além das redes *neuro-fuzzy*, tais como clusterização, classificadores bayesianos, permitindo inclusive uma avaliação do melhor classificador ao longo do tempo de uso do GEDI;

- Implementação de extratores de regras automáticos agregados ao banco de dados, parametrizados pelo usuário, para mineração e descoberta de novos conhecimentos;

- O uso de módulos de validação automáticos para axiomas de ontologias, extensivo ao processo de evolução das mesmas, de forma a auxiliar no processo de integração das ontologias geradas por UNL à ontologia de domínio;

- Geração de um sistema tutorial inteligente com acesso ao conhecimento da ontologia de domínio para utilização em processos de treinamento de pessoas e usuários ou mesmo ambientação de novos usuários do sistema de informação;

- Criação de um sistema de regras de inferência ou o uso de um modelo conexionista para auxiliar, ou mesmo “aprender”, o processo de renomeação de relações incógnitas provindas de documentos UNL, fornecendo um meio contínuo de automatização e autonomia do processo de evolução da ontologia de domínio;
- A aplicação do GISAAGIC em outras áreas de conhecimento além da área de saúde, no exemplo explorado para aplicação prática, visto que o sistema se propõe a ser genérico;
- O desenvolvimento do modelo do GISAAGIC em outras plataformas além do ambiente Windows, inclusive sendo multiplataforma a nível dos seus próprios elementos;
- O desenvolvimento de novas metodologias além da baseada na teoria da informação, a serem usadas na métrica de avaliação de vetores de documentos;
- O desenvolvimento de conversores automáticos (“wrappers”) para ontologias escritas em linguagens diversas; e
- O desenvolvimento de um módulo de validação de documentos UNL a partir do processo de validação de ontologias do sistema e posterior conversão UNL – ontologia.

REFERÊNCIAS

AGUADO, G. et al. **Ontogeneration: reusing domain in linguistic ontologies for Spanish text generation.** In: EUROPEAN CONFERENCE ON ARTIFICIAL INTELLIGENCE. 13. 1998. Brighton, England, p. 23-28. 1998.

ALBERTAZZI, Liliana. **Formal and material ontology.** In: POLI, Roberto; SIMONS, Peter (Ed) Formal ontology. Dordrecht: Kluwer, p. 199-232. 1996.

ALLEN, B. L. **Cognitive Research in Information Science: Implications for Design.** **ASIS/Learned Information, Inc.:** Annual Review of Information Science and Technology (ARIST), v.26, p.3-24, 1991.

AMARAL, Fernanda Cristina Naliato. **DATA MINING – Técnicas e aplicações para o marketing direto.** São Paulo. Berkeley, 2001.

ANSELMO, Fernando. **Aplicando Lógica Orientada a Objetos em Java.** Florianópolis-SC. Bookstore Livraria Ltda., 2002.

AZEVEDO, F. A.; BRASIL, F. M. e OLIVEIRA, R. C. L. **Redes Neurais com Aplicações em Controle e em Sistemas Especialistas.** Florianópolis. Visual Books, 2000.

BASTOS, Rogério Cid. **“Avaliação de Desempenho em Sistemas Educacionais: uma abordagem utilizando conjuntos difusos”.** Tese de Doutorado, PPGEP, Universidade Federal de Santa Catarina, Florianópolis, 1994.

BAEZA-YATES, R. & R. N. B. **Modern Information Retrieval.** Addison University Press. 1999.

BERNERS-LEE, Tim; HENDELER J. and LASSILA O. “*The Semantic Web*”. Scientific American , May 2001.

BÉRTOLLI JR, Lumar Valmor; Zago, Ricardo Flores e LUZ, Rodolfo Pinto da. **Sistema UNL: Uma Solução para Tradução Automática e Sistemas de Conhecimento**. Instituto UNDL Brasil, 2003.

BAEZA-YATES, R. *et al. Modern Information Retrieval*. Addison Wesley. 1999.

BEUREN, Ilse Maria. **Gerenciamento da informação**. São Paulo. Atlas, 2000.

BRADLEY, P., FAYYAD, U., & MANSASARIAN, O. **Data Mining: Overview and Optimization Opportunities**. *Technical Report MSR-TR-98-04*, Microsoft Research Report, Redmond, WA. 1998.

BRANDÃO, Hugo Pena e GUIMARÃES, Tomás de Aquino. **Gestão de competências e gestão de desempenho**. Artigo apresentado no ENANPAD99. Foz do Iguaçu. 19 a 22 de setembro de 1999.

CARDOSO, H. D. A. **Sistema multiagent para comércio eletrônico**. Dissertação de Mestrado em Inteligência Artificial e Computação, Faculdade de Economia, Universidade do Porto, Lisboa, 1999.

CHOUDHARY, B. e BHATTACHARYYA, P. **Text Clustering Using Semantics**. Em *The Eleventh International World Wide Web Conference*, 2002.

COCKAYNE, W. T.; ZILDA M. **Mobile Agents**. Manning Publications Co, 1998.

CORDEIRO, A. D. “**Concepção e Implementação de um Sistema Multiagentes para Gestão da Comunicação de Dados on line entre Sistemas**,” Dissertação de Mestrado, PPGEP, Universidade Federal de Santa Catarina, Florianópolis, 2001.

CORDEIRO, A. D. & BASTOS, R. C. “**Gestão da Comunicação de Dados On-Line com Sistemas Multiagentes compartilhados com *Neuro-Fuzzy* e Web Semântica – Inteligência Artificial**”, em *Tecnologia e Humanismo – CEFET-PR*, nº 25 p.91-117, 2004.

Darpa Markup Language. www.daml.org, acessado em 04/09/2004.

DE RÉ, Angelita Maria. **Um Método para Identificar Características Predominantes em Empreendedores que Obtiveram Sucesso Utilizando um Sistema Neurodifuso**. Tese de Doutorado, Universidade Federal de Santa Catarina, 2000.

EBECKEN, N. F. F.; LOPES, M. C. *et al.* **Mineração de Textos**, em REZENDE S. O., *Sistemas Inteligentes, Fundamentos e Aplicações*. Barueri-SP. Manole, 2003.

ETZIONE, O.; LESH, N. *et al.* **Building Softbots for UNIX**. In: SOUZA, Luiz Cláudio Guarita. “**Regras de Raciocínio Aplicadas a Ontologias por Meio de Sistemas Multiagentes para Apoio a Decisões Organizacionais**”. Dissertação de Mestrado, Pontifícia Universidade Católica do Paraná, 2003.

EVSUKOFF, A. G. e ALMEIDA, P. E. M. **Sistemas Neuro-Fuzzy**, em REZENDE S. O., *Sistemas Inteligentes, Fundamentos e Aplicações*. Barueri-SP. Manole, 2003

FERBER, Jacques. **Multi-Agent System – an introduction to distributed artificial Intelligence**. Addison Wesley Longma, 1999.

<http://www.daml.org/2000/12/daml+oil-ex.daml>. Acessado em 04/09/2004.

HABN, U. & I. MANI. **The challenges of automatic summarization**. IEEE Computer 33(11), 29-36, 2000.

HUHNS, M. N.; SINGH, M. P. **Readings in Agents**. New York: Morgan Kaufmann, 1997.

HAMMOND, J. S. *et al.* “**Smart Choices**”. Massachusetts: Harvard Business School Press, 1999.

JIN, R.; FALUSOS, C e HAUPTMANN, A. G. **Meta-scoring: automatically evaluating term weighting schemes in IR without precision-recall**. Em *Proceeding of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, p. 83-89. ACM Press, 2001.

LIAUTAUD, Bernard e HAMMOND, Mark. **“Inteligência em e-business”, Transformando Informações em Conhecimento e Conhecimento em Lucro**. Rio de Janeiro. Qualitymark, 2002, pp. 123-125.

LUGER, George F. **Inteligência Artificial: Estruturas e Estratégias para a Solução de Problemas Complexos**. Porto Alegre. Editora Bookmann, 2004.

MITRA, S. & HAYASHI, Y. **Neuro-Fuzzy Rule Generation: Survey in Soft Computing Framework**. IEEE Transactions on Neural Networks, Vol. 11, n°. 3, 2000.

MITRA, S.; PAL, S. K. & MITRA, P. **Rough-Fuzzy MLP: Modular Evolution, Rule Generation, and Evaluation**. IEEE Transactions on Knowledge and Data Engineering. Vol. 15, n°. 1, 2003.

NASCIMENTO JR, C. L. e YONEYAMA, T. **Inteligência Artificial em Controle e Automação**. São Paulo. Editora Edgard Blücher Ltda., 2000.

OLIVEIRA, O. Jr. *et al.* **O uso de interlíngua para comunicação via Internet**. NILC-TR-01-3. Série de Relatórios do NILC-ICMC-SP. Julho, 2001.

Oullet, R. & Oqbuji, U. **Introduction to DAML I**. <http://www.daml.org/2000/12/daml+oil-ex.daml>, acessado em 04/09/2004.

PICKETT, Joseph P. *et al.* ***The American Heritage Dictionary of the English Language***. 4. ed. Boston: Houghton Mifflin Company, 2000.

POLLONI, Enrico Giulio Franco. **Administrando Sistemas de Informação**. São Paulo: Futura, 2000.

PYLE, Dorian. **Data Preparation for Data Mining**. San Diego. Academic Press, 1999.

RAITZ, Roberto. **“FAN 2002: Um Modelo Neuro-Fuzzy para Reconhecimento de Padrões”**. Tese Doutorado, Florianópolis, 2002.

REZENDE, Solange Oliveira. ***Sistemas de Inteligentes, Fundamentos e Aplicações***. São Paulo. Manole, 2003.

RIBEIRO, Lair. **Inteligência Aplicada**. São Paulo. Arx, 2002

RUSSEL, Stuart, NORVIG, Peter. **Artificial Intelligence: A Modern Approach**. New Jersey, USA: Prentice Hall, 1995, 2003.

SALTON, G. **“Editor, *Automatic Text Processing: The Transformation, Analyses and Retrieval of Information by Computer*”**. Addison Wesley, 1989.

SHAH, C.; CHOUDADHARY, B. e BHATTACHARYYA, P., ***Constructing Better Document Vectors Using Universal Networking Language (UNL)***. Proceedings of International Conference on Knowledge-Based Computer Systems (KBCS), 2002.

SOUZA, Luiz Cláudio Guarita. **“Regras de Raciocínio Aplicadas a Ontologias por Meio de Sistemas Multiagentes para Apoio a Decisões Organizacionais”**. Dissertação de

Mestrado, Pontifícia Universidade Católica do Paraná, 2003.

SUGENO, M. & KANG, G. T. **Structure identification of fuzzy model.** *Fuzzy Sets and Systems*, 28:15-33, 1988.

SWARTOUT, B., PATIL R., KNIGHT K. and RUSS T. **“Towards Distributed Use of Large Scale Ontologies”.** AAAI’97 Spring Symposium Series on Ontological Engineering. P. 138-148. 1997.

TAKAGI, T. & SUGENO, M. *Fuzzy identification of systems and its application to modeling and control.* IEEE Trans. On Systems, Man & Cybernetics, 15:116-132, 1985.

TELLO, Adolfo Lozano. **“Métrica de Idoneidad de Ontologias”.** Tese de Doutorado, Universidade Politécnica de Madri, 2002.

TRINDADE, Cristiano. **Desenvolvendo Soluções com a Tecnologia Java.** Florianópolis. Visual Books, 2002.

VASILEIOS, H. *et al.* **An investigation of linguistic features and clusters algorithms for topical document clustering.** Em 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 2000.

UCHIDA, Hiroshi; ZHU, Meiyong. **“The Universal Networking Language Beyond Machine Translation”**, International Symposium on Language in Cyberspace, Seul, Sept 2001.

UNDL, *Foundation.* **“Introduction oh the UNL System”.** Disponível em <http://www.unl.edu/unlsys/introduction.html>.

WOOLDRIDGE, Michael 1 and JENNINGS, N.R.2. *Intelligent agents: Theory and practice–The knowledge Engineering Review*, vol. 10:2, 1995.

1 Departamente of computing, Manchester Metropolitan University, Chester Street, Manchester M1 5GD, UK.

2 Departament Of Eletronic Engineering, Queen Mary & Westfield College, Mile and road, London E1 4NS, UK.

W3C. *Resource Description Framework*. <http://www.w3.org/RDF/>. Acessado em 04/09/2001.

YANG, Y. & J. Pedersen. *A comparative study on feature selection in text categorization*. Em *Fourteenth International Conference on Machine Learning – ICML’97*, 1997.

ZADEH, L. *Commonsense knowledge representation based on fuzzy logic*. *Computer*, 1983.

ZADEH L. *Is Probability Thory Sufficient for Dealing with Uncertainty in AI: A Negative View*. *Uncertainty in Artificial Intelligence*. Editor:L. N. Kanal e J. F. Lemmer. Amsterdam, Elsevier Science Publishers, 1986.

ZADEH, L. and KACPRZYK, J. *Fuzzy Logic for the Management of Uncertainty*. New York, John Wiley e Sons. 1992.

ANEXO I - 41 TIPOS DE RELAÇÕES QUE PODEM OCORRER

ENTRE PALAVRAS UNIVERSAIS

RL	ORIGEM	SIGNIFICADO
agt	agent	= uma coisa que inicia uma ação;
and	conjunction	= relação de conjunção entre conceitos;
aoj	thing with attribute	= uma coisa que é um estado ou tem um atributo;
bas	basis	= uma coisa usada como base (padrão) para expressar um grau;
ben	beneficiary	= um beneficiário não relacionado diretamente ou vítima de um evento ou estado;
cag	co-agent	= a coisa não em foco que inicia um evento implícito que é feito em paralelo;
cao	co-thing with attribute	= uma coisa não em foco é um estado em paralelo;
cnt	content	= um conceito equivalente;
cob	affected co-thing	= uma coisa que é diretamente afetada por um evento fazendo em paralelo ou um estado implícito em paralelo;
con	condition	= um evento não focado ou estado que condiciona um evento focado;
coo	co-occurrence	= um evento de co-ocorrência ou estado para um evento focado ou estado;
dur	duration	= um período de tempo durante a ocorrência de um evento ou estado de existência;
fmt	range	= uma gama entre duas coisas;
frm	origin	= uma origem de uma coisa;
gol	goal/final state	= o estado final do objeto ou a coisa finalmente associada com o objeto de um evento;

ins	instrument	= o instrumento para levar a cabo um evento;
man	manner	= o caminho para levar a cabo um evento ou característica de um estado;
met	method	= um meio para levar a cabo um evento;
mod	modification	= uma coisa restringe uma coisa focada;
nam	name	= um nome de uma coisa;
obj	affected thing	= uma coisa em foco que é diretamente afetada por um evento ou estado;
opl	affected place	= um lugar em foco quando um evento é afetado;
or	disjunction	= relação disjuntiva entre dois conceitos;
per	proportion, rate or distribution	= uma base ou unidade de proporção, taxa ou distribuição;
plc	place	= o lugar onde um evento ocorre ou um estado verdadeiro ou uma coisa existente;
plf	initial place	= o lugar onde um evento começou ou um estado que se tornou verdadeiro;
plt	final place	= o lugar onde um evento terminou ou um estado que se tornou falso;
pof	part-of	= um conceito do qual uma coisa focada é uma parte;
pos	possessor	= o possuidor de uma coisa;
ptn	partner	= um indispensável não focado inicializador de uma ação;
pur	purpose or objective	= a proposta ou um objetivo, agente, evento ou uma proposta de uma coisa que existe;
qua	quantily	= uma quantidade de uma coisa ou unidade;
rsn	reason	= uma razão que um evento ou estado acontece;
sen	scene	= uma virtual palavra que um evento ocorre ou estado é verdadeiro ou uma coisa existe;

seq	sequent	= um evento prioritário ou estado de um evento ou estado focado;
src	source/initial state	= o estado inicial de um objeto ou coisa inicialmente associada como objeto ou um evento;
tim	time	= o tempo um evento ocorre ou um estado é verdadeiro;
tmf	initial time	= o tempo que um evento inicia ou um estado que e tornou verdadeiro;
tmf	final time	= o tempo que um evento termina ou um estado que se tornou falso;
to	destination	= um destino de uma coisa;
via	intermediate place or state	= um lugar intermediário ou estado de um evento.

ANEXO II – REGISTROS DE NÃO CONFORMIDADE, PONTOS FORTES E OBSERVAÇÕES – DIAGNÓSTICO 1

Alguns itens descritos no Manual de normas não estão sendo executados. Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.

N1 - Não evidenciada avaliação do desempenho dos serviços terceirizados.

N2 - Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.

Não estava disponível a habilitação em Administração do Superintendente Planejamento estratégico

N2 - Ausência de manuais de normas, rotinas e procedimentos documentados, atualizados, disponíveis e aplicados ao gerenciamento administrativo e financeiro em alguns setores.

N2 - Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.

n2 - Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.

n2 - Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.

Sistema de análise crítica visando à melhoria da técnica, controle de problemas, melhoria de processos e procedimentos e minimização de riscos em processo de implantação.

n1 - Evolução médica sem assinatura e carimbo.

n2 - não evidenciados protocolos clínicos para as patologias de maior prevalência.

n2 - Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.

n2 - não evidenciados registros de reuniões clínicas periódicas para discutir temas médicos e/ou casos selecionados.

n2 - não evidenciada análise de resultados de auditorias médicas.

n1 - prontuários sem assinatura, carimbo ou número do Conselho de Classe.

n2 - prescrição de enfermagem parcialmente sistematizada, com auditoria relacionada apenas ao faturamento.

n2 - não evidenciados protocolos clínicos para patologias de maior prevalência.

n2 - programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.

- Comissão de ética de enfermagem em fase de implantação.

n1 - materiais esterilizados com prazo de validade expirado em carrinho de emergência.

n1 - medicamentos abertos sem data de abertura.

n1 - banheiros não possuem meios para comunicação com a enfermagem

n1 - unidade One Day sem definição explícita de barreira para pessoas estranhas ao Centro Cirúrgico.

n1 - ausência de padronização da rotina de lavagem e acondicionamento do material contaminado da ser transportado da unidade de internação até a Central de Material.

n1 - não evidenciado registro das orientações dadas a pacientes e familiares

n1 - controle de temperatura de geladeira não disponível no local.

n2 - Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.

- blusa de funcionário em armário de materiais.

n2 - não evidenciada a existência de procedimentos padronizados, como protocolos clínicos para atendimento ao paciente.

n2 - Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.

- desorganização e guarda de materiais em desuso na área designada como Medicina Esportiva.

n1 - triagem de pacientes executada pela recepcionista.

n1 - material esterilizado vencido

n1 - artigo médico hospitalar com prazo de validade vencido.

n1 - medicamentos abertos sem data de abertura.

n1 - medicamento (Hidrato de Cloral) com prazo de validade vencido.

n1 - frasco para umidificação de oxigênio, instalado na parede com prazo de validade vencido.

n1 - artigos médicos hospitalares acondicionados sob sifão da pia.

n1 - não evidenciados sistema de comunicação e articulação com a autoridade judiciária para casos específicos.

n2 - não evidenciados procedimentos adotados nos casos de atendimento a agressão física, moral e psicológica

n2 - Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.

- escalas de algumas especialidades médicas desatualizadas.
- acondicionamento de galões vazios de água na Sala de Gesso.

n1 - mesma área para espera de cirurgias e recuperação pós-anestésica.

n1 - presença de material vencido no setor

n1 - não evidenciado registro de manutenção preventiva dos equipamentos de terceiros

n2 - Não evidência de treinamento para as instrumentadoras.

n2 - Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.

- mapa de cirurgias apenas com o primeiro nome do paciente.
- Sistema de fluxo laminar para as salas cirúrgicas.
- n1 - não evidenciado registro de consulta pré-anestésica em prontuário na unidade de internação.

n1 - não evidência de rotina sistematizada para controle do carrinho de emergência.

n1 - presença de material estéril vencido no setor.

n1 - medicamentos aguardando uso acondicionados em saco plástico no chão.

n1 - estado atual de organização do Serviço quanto à estrutura física e mobiliário dificulta a avaliação das condições estruturais definitivas do Serviço de Hemoterapia.

n1 - A transição do status quo do Serviço de Hemoterapia dificulta a interpretação de sua habilitação para as tarefas as quais está se propondo.

n1 - Não há registro da temperatura do sangue nas fases de saída e chegada.

n1 - material utilizado para o transporte de sangue e seus derivados colocados no chão.

n1 - isopor com bolsa de sangue em local não definido para guarda temporária.

n2 - Não evidência de manuais, normas, rotinas e procedimentos documentados.

n2 - não evidência de documentação dos grupos de trabalho e da análise crítica dos procedimentos realizados

- Avaliação prejudicada por estar o Laboratório em situação de mudança de equipamentos e área física.

- Presença de material no corredor (caixa com aparelho e para arquivo morto), livros e manuais embaixo das bancadas.
 - Teto da sala de administração quebrado.
- Hemodinâmica
- Material médico-hospitalar com prazo de validade vencido.
 - Ausência de tampa do recipiente de acondicionamento do material contaminado a ser transportado para a Central de Material.
 - Ausência de vestiário de barreira para acessar as salas.
- Eco Trans-esofágico
- Álcool sem data de abertura.
 - Instalações físicas inadequadas para o procedimento de limpeza e desinfecção do material utilizado.
 - EPI (avental de pano) inadequado para o processo.
- Radiologia Vasculuar Intervencionista
- Material esterilizado vencido em grande quantidade.
 - Soluções sem data de abertura.
- Radiologia
- Laringoscópio sem condições de uso.
- Endoscopia
- Material médico-hospitalar com prazo de validade vencido.
 - EPI (avental de pano) inadequado para o processo.
- Tomografia
- Medicamento (Hidrato de Cloral) com prazo de validade vencido.
 - Carro de emergência com material esterilizado vencido.
- Unidade Dor Torácica
- Material médico-hospitalar com prazo de validade vencido.
 - Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.
 - Não evidenciado grupos de trabalho para: a melhoria de processos, integração institucional, gerenciamento de riscos e investigação de eventos adversos.
 - Não evidenciado sistema de análise crítica visando à melhoria da técnica, controle de problemas, melhoria de processos e procedimentos e minimização de riscos.
- Hemodinâmica
- Gaveta de psicotrópicos destrancada.
- Mamografia
- Mau funcionamento do sistema de exaustão.
 - Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.
 - Não evidenciado grupos de trabalho para: a melhoria de processos, integração institucional, gerenciamento de riscos e investigação de eventos adversos.
 - Não evidenciado sistema de análise crítica visando à melhoria da técnica, controle de problemas, melhoria de processos e procedimentos e minimização de riscos.
- Unidade Dor Torácica
- Protocolos clínicos não disponíveis no local.
- Central de Entrega de Laudos
- Ausência de Manual de normas, rotinas e procedimentos.
 - Não evidência de rastreabilidade rápida e sistemática do prontuário.
 - Estrutura física pequena.
 - Evidências de melhoria de prontuário e integração com o setor de Faturamento.
 - Situação organizacional favorecendo autonomia do Serviço de Controle de Infecção Hospitalar.
 - Situação organizacional favorecendo autonomia do Serviço de Controle de Infecção Hospitalar

- Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.
 - Não evidenciados procedimentos de orientação ao cliente/paciente.
 - Não evidenciados procedimentos para a continuidade de cuidados ao cliente e seguimento de casos.
 - Sistema de fármaco-vigilância iniciante.
 - Ausência de carimbo no prontuário referente à orientação nutricional e visita clínica.
 - Dieta enteral sem registro no Ministério da Saúde.
 - Manipulação direta de alimentos já preparados (montagem de pratos) não observando o Manual de Boas Práticas.
 - Armário de estoque de produtos de nutrição enteral com umidade acima do preconizado
 - Equipamentos com presença de corrosão e material de acabamento da estrutura física com desgaste.
 - Carro de transporte com sujeira na área de recepção de alimento.
 - Não evidenciada participação das nutricionistas na validação dos fornecedores
 - Ausência de controle de lote dos não perecíveis.
- Cafeteria
- Excesso de gelo no equipamento de refrigeração.
 - Ausência de evidências de controle dos pontos críticos.
- Restaurante Oliveira's
- Alimento industrializado com prazo de validade vencido.
 - Produto exposto em local inadequado, com embalagem suja.
 - Frango acondicionado incorretamente na geladeira.
 - Manual de boas práticas de preparo de alimentos não disponíveis no local
 - Não evidenciados registros de manutenção dos equipamentos.
 - Bancada de buffet sem controle de temperatura dos alimentos quentes e ausência de refrigeração para as saladas e frios.
 - Pia na área da lanchonete sem sabonete líquido para lavagem das mãos.
 - Ausência de local adequado para acondicionamento do lixo.
 - Funcionários de serviços de reforma cruzando fluxo de montagem de pratos.
 - Registro do rol de roupas não evidenciando a garantia do estoque entre as etapas de coleta e recebimento pela empresa terceirizada.
 - Licença sanitária Chanceler vencida em 31/12/2003
- Área Limpa
- Estrutura de madeira na divisão entre áreas (autoclave).
 - Não evidência de indicação de uso e de rotinas de controle das autoclaves de circuito rápido.
 - Materiais esterilizados acondicionados para distribuição com prazo de validade vencido.
 - Materiais esterilizados com ausência de número de lote, etiquetas com prazos de validade rasurados e alterados.
 - Diferentes períodos de validade para o mesmo tipo de processamento e embalagem.
 - Não evidência de rotina sistemática para controle de validade do estoque.
 - Estoque excessivo de material estéril.
 - Materiais de propriedade particular dos médicos acondicionados em prateleira de distribuição com grande parte apresentando data de validade vencida.
 - Armários sem revestimento lavável.
- Área Suja
- Desorganização dos armários.
 - Recipiente plástico no chão com material cirúrgico parcialmente submerso em solução.

- Não evidenciado gerenciamento do setor sobre a qualidade do processamento dos materiais de propriedade particular.
- Material utilizado em vídeo-laparoscopia a ser reprocessado, contendo artigo médico hospitalar de uso único sem a informação de protocolo junto ao Serviço de Controle de Infecção Hospitalar.
- Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.
- Evidenciada funcionária da Higiene abrindo a porta do quarto com a mão enluvada.
- Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.
- Não evidenciados procedimentos voltados para a continuidade de cuidados e seguimento de casos.
- Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.
- Não evidência de acompanhamento técnico da prestação de serviços por terceiros.
- Não evidenciados manuais de normas e rotinas.
- Não evidencia de programa de capacitação.
- Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.
- Manual de procedimentos desatualizado.
- Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.
- O setor não atende os padrões do nível 2.
- Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.
- Apesar do sistema possibilitar a emissão de relatórios dos problemas surgidos e das soluções, esta prática não é realizada, haja vista o alto grau de resolutividade apresentado pelo setor.
- Excelente sistema de documentação, arquivamento, acesso e seguimento longitudinal de projetos, plantas, equipamentos e procedimentos de manutenção.
- Não evidenciado licenciamento sanitário.
- Documentação de planejamento de acessos e circulação não disponível.
- Reservatório de água sem vedação completa
- Porta do abrigo externo de resíduos infectantes mantida aberta.
- Estação de tratamento de efluentes com monitoração
- O setor não atende os padrões do nível 2.
- Não evidenciados procedimentos de orientação ao cliente interno e externo quanto ao uso e conservação da estrutura.
- Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.
- Não evidenciado registro de atividades junto aos serviços terceirizados.
- Programa de capacitação e educação continuada com utilização de metodologia de difícil compilação e análise dificultando evidências de melhoria.
- Não evidenciados registros de atividades junto ao corpo clínico.
- Estatísticas não conferem com dados apresentados.

ANEXO III – REGISTROS DE NÃO CONFORMIDADE, PONTOS FORTES E OBSERVAÇÕES – DIAGNÓSTICO 2

- Não evidência de formalização de co-gestão entre as empresas parceiras (CEDAP, PREVENCEL e CITOCENTRO).
- Não evidência de documentos de habilitação dos profissionais da área técnica e administrativa.
- Liderança expressiva por parte da alta administração.
- Premiação interna para funcionários em dinamismo, empreendedorismo, amigo e campeão da semana "8S".
- Não evidência de planejamento, acompanhamento e avaliação do desempenho organizacional e econômico - financeiro, orientados para o atendimento às necessidades dos clientes.
- Não evidência de avaliação, controle e melhoria de sistemas de aquisição atualizados.
- Não evidência de treinamento e educação continuada, com evidências de resultados.
- Não evidência de plano de metas e indicadores de desempenho para tomada de decisão.
- Não evidência de sistema de avaliação da liderança.
- Não evidência de melhoria nos processos administrativos e gerenciais.
- Não evidência de ciclos de melhoria do sistema de gestão com impacto sistêmico.
- Não evidência de aferição atualizada da satisfação dos clientes externos.
- Não evidência de auditoria interna do sistema de gestão, baseado em indicadores, implantado e consolidado, que gera informações que permitam ações preventivas, melhoria de processos e comparação com referenciais.
- Programa de mentoria para formação de técnicos.
- Reuniões sistemáticas com especialistas de áreas afins para discussão de casos clínicos.
- Não evidência de sistema de aferição atualizada da satisfação dos clientes internos e de avaliação do serviço
 - Não evidência de demonstração de resultados relativos ao programa de treinamento e educação continuada.
- Não evidência de comparações de resultados com referenciais externos.
- Não evidência sistema de aferição atualizada da satisfação dos clientes internos.
- Laminoteca em área de grande circulação comprometendo a segurança das lâminas.
- Procedimento técnico em área administrativa (manuseio de lâminas).
- Ausência de data de esterilização e prazo de validade no material esterilizado.
- Ausência de precauções padrão e rotina de controle de infecção.
- Ausência de mapa de temperatura da estufa.
- Presença de reagentes vencidos.
- Sistema de escaneamento e disponibilização das requisições dos exames via intranet.
- Reciclagem de alguns reagentes químicos como xilol e álcool.
- Não evidência de registros de análise crítica das ações e planejamento e implementação de melhorias.

- Não evidência de programa de treinamento e educação continuada, com evidências de resultado.
- Não evidência de comparações de resultados com referenciais externos.
- Não evidência de sistema de aferição atualizada da satisfação dos clientes internos e externos.
- Não evidência de registros de análise crítica das ações e planejamento e implementação de melhorias.
- Não evidência de programa de treinamento e educação continuada, com evidências de resultado.
- Não evidência de comparações de resultados com referenciais externos.
- Não evidência de sistema de aferição atualizada da satisfação dos clientes internos e externos.
- Não evidência de instrumento formal que assegure as responsabilidades das partes em todas as fases do processo no contrato de serviços terceirizados.
- Não evidência de mecanismos que assegurem o controle de qualidade do serviço terceirizado.
- Não evidência de mecanismos de avaliação do desempenho dos processos de apoio.
- Não evidência de sistema atualizado de aferição de satisfação dos clientes.
- Não evidência de estatísticas voltadas para a melhoria do atendimento ao paciente/cliente.
- Não evidência de registros de análise crítica das ações e planejamento e implementação de melhorias.
- Não evidência de mecanismos de avaliação do desempenho dos processos de apoio.
- Não evidência de sistema atualizado de aferição de satisfação dos clientes.
- Não evidência de estatísticas voltadas para a melhoria do atendimento ao paciente/cliente.
- Não evidência de registros de análise crítica das ações e planejamento e implementação de melhorias.
- Não evidência de relação atualizada de fornecedores qualificados.
- Uso incorreto de EPI's.
- Não evidência de exames atualizados de saúde ocupacional
- Sistema de informação consistente e bem estruturado.
- Não evidência de sistema de informação com dados, índices e indicadores que permitam a avaliação dos fornecedores.
- Não evidência de índices de satisfação dos clientes internos e externos.
- Não evidência de registros de análise crítica das ações e planejamento e implementação de melhorias.
- Não evidência de sistema de informação com dados, índices e indicadores que permitam a avaliação dos fornecedores.
- Não evidência de índices de satisfação dos clientes internos e externos.
- Não evidência de registros de análise crítica das ações e planejamento e implementação de melhorias.
- Não evidência de área apropriada para refeições dos funcionários.
- Não evidência de registros de manutenção preventiva de elevadores.
- Não evidência de registros de análises microbiológica da água reagente.

- Deficiência na iluminação das escadas do edifício.
- Não evidência de registros de análise crítica das ações e planejamento e implementação de melhorias.
- Não evidência de sistema de aferição da satisfação dos clientes internos e externos.
- Não evidência de registros de análise crítica das ações e planejamento e implementação de melhorias.
- Não evidência de sistema de aferição da satisfação dos clientes internos e externos.
- Não evidência de registros de análise crítica das ações e planejamento e implementação de melhorias.
- Não evidência de sistema de aferição da satisfação dos clientes internos e externos.
- Atividades de fomento na área de ensino e pesquisa para residentes e estudantes.
- Não evidência de registros de análise crítica das ações e planejamento e implementação de melhorias.
- Não evidência de sistema de aferição da satisfação dos clientes internos e externos.

ANEXO IV – REGISTROS DE NÃO CONFORMIDADE, PONTOS FORTES E OBSERVAÇÕES – DIAGNÓSTICO 3

- Desenvolvimento de parâmetros de comparação com referenciais externos para gestão de pessoas.
- Liderança expressiva da alta administração.
- Clima Organizacional harmonioso, integrativo.
- Premiação interna para funcionários em dinamismo, empreendedorismo, amigo e campeão da semana "8S".
- Reuniões técnicas e clínicas para tomada de decisões baseada no coletivismo e consenso.
- Aplicação da metodologia dos Seis Sigma.
- Sistema de Informação próprio desenvolvido especialmente para a gestão do Laboratório.
- Divulgação do trabalho e atendimento ao cliente de forma humanizada.
- Responsabilidade social com os Clientes Internos.
- Processo de busca ativa com os Clientes Externos.
- Atividades para a qualidade de vida no trabalho como ginástica laboral e coral.
- Programa de mentoria para formação de técnicos.
- Reuniões sistemáticas com especialistas de áreas afins para discussão de casos clínicos.
- Parcerias para consultorias técnicas.
- Produção e co-autoria de controle externo de qualidade para a Sociedade Científica (CITONET).
- Apresentar a sistemática de validação dos produtos em uso com prazo de validade vencido.
- Sistema de escaneamento e disponibilização das requisições dos exames via intranet.
- Reciclagem de alguns reagentes químicos como xilol e álcool.
- Participação em diversos Sistemas de Controle da Qualidade Externos (nacionais e internacionais)
- Educação para o trabalho de forma planejada e sistemática.
- Produção de Manual Técnico de orientação para a coleta, transporte para os clientes.
- Atendimento ao cliente de forma humanizada.
- Responsabilidade social com os Clientes Internos.
- Processo de busca ativa com os Clientes Externos.
- Sistema de informação consistente e bem estruturado, com série histórica desde 1999.
- Sistema de estatística permitindo cruzamento de variáveis importantes para o sistema de gestão.
- Atividades de fomento na área de ensino e pesquisa para residentes e estudantes.
- Biblioteca com acervo diversificado.