

**UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE INFORMÁTICA E ESTATÍSTICA**

ThiagoAngelo Gelaim

SIMILARIDADE EM LÓGICA DESCRITIVA

Florianópolis,
2013.

Thiago Angelo Gelaim

SIMILARIDADE EM LÓGICA DESCRITIVA

Trabalho de Conclusão de Curso
submetido ao Curso de Ciências da
Computação para a obtenção do Grau
de Bacharel em Ciências da
Computação.

Orientador: Prof. Dr. Ricardo
Azambuja Silveira

Coorientador: Prof. Dr. Elder Rizzon
Santos

Florianópolis,
2013.

Gelaim, Thiago Angelo

Similaridade em Lógica Descritiva / Thiago Angelo Gelaim ; orientador, Ricardo Azambuja Silveira ; co-orientador, Elder Rizzon Santos. - Florianópolis, SC, 2013. 78 p.

Trabalho de Conclusão de Curso (graduação) - Universidade Federal de Santa Catarina, Centro Tecnológico. Graduação em Ciências da Computação.

Inclui referências

1. Ciências da Computação. 2. Lógica Descritiva. 3. Similaridade. 4. Agentes. I. Azambuja Silveira, Ricardo . II. Rizzon Santos, Elder. III. Universidade Federal de Santa Catarina. Graduação em Ciências da Computação. IV. Título.

Thiago Angelo Gelaim

SIMILARIDADE EM LÓGICA DESCRITIVA

Este Trabalho de Conclusão de Curso foi julgado aprovado para a obtenção do Título de “Bacharel em Ciências da Computação”, e aprovado em sua forma final pelo Curso de Ciências da Computação.

Florianópolis, 01 de novembro, 2013.

Prof. Dr. Vitorio Bruno Mazzola
Coordenador do Curso

Banca Examinadora:

Prof. Dr. Ricardo Azambuja Silveira
Orientador

Prof. Dr. Elder Rizzon Santos
Coorientador

Prof.^a Dr.^a Jerusa Marchi

Prof. Dr. Renato Fileto

AGRADECIMENTOS

Aos meus pais, Claodete e Ângelo Gelaim, por razões que não precisam ser descritas.

À minha namorada, Naiane Cristina Salvi, pelo amor incondicional.

Ao meu orientador Ricardo Azambuja Silveira, pela orientação e apoio desde o início da graduação.

Ao meu coorientador Elder Rizzon Santos, pelo apoio e dedicação na minha pesquisa.

Aos membros da banca Jerusa Marchi e Renato Fileto pelas sugestões.

Não menos importantes, aos meus amigos.

RESUMO

No contexto da representação de conhecimento, motores de inferência são utilizados como forma de encontrar relações implícitas a partir de conhecimento explícito. Neste trabalho será desenvolvido um mecanismo para inferir similaridade entre conceitos em lógica descritiva, sendo válido também entre classes OWL. O cenário de testes é a recomendação de serviços similares por um agente facilitador de diretórios, segundo o modelo de referência FIPA.

Palavras-chave: Lógica Descritiva. Ontologia. Similaridade. Motor de Inferência. Raciocínio. Agentes. Facilitador de Diretórios

ABSTRACT

In the context of knowledge representation, inference engines are used as a way of finding implicit relations from explicit knowledge. In this work a mechanism for inferring similarity between concepts in description logics will be developed, which is also valid in the scope of OWL classes. The test scenario is the recommendation of similar services for a directory facilitator agent, according to the FIPA reference model.

Keywords: Description Logics. Ontology. Similarity. Inference Engine. Reasoning. Agents. Directory Facilitator.

LISTA DE FIGURAS

FIGURA 01: distância entre os conceitos Homem (ponto azul) e Mulher (ponto vermelho).....	37
FIGURA 02: Estrutura taxonômica considerando o relacionamento “é um” entre Humano, Mulher e Homem.....	38
FIGURA 03: Representação de Conceito1 e Conceito2 no espaço das características a,b e c.....	41
FIGURA 04: Representação em verde da distância Manhattan e em azul da distância euclidiana.	42
FIGURA 05: Exemplo de estrutura conceitual.	43
FIGURA 06: Modelo de Referência do Gerenciamento de Agentes	51
FIGURA 07: Containers e plataformas (BELLIFEMINE, 2007).	53
FIGURA 08: Grafo em OWL para o exemplo do modelo.	59
FIGURA 09: Instanciação do mecanismo de similaridade e invocação do método que calcula similaridade.	61
FIGURA 10: Método invocado pelo usuário para obter a similaridade.....	62
FIGURA 11: Método que calcula similaridade.	62
FIGURA 12: Segmento do grafo conceitual de uma ontologia de serviço.	63
FIGURA 13: Implementação de um agente que fornece o serviço de BookSelling no DF.	64
FIGURA 14: Implementação de um agente que busca serviços no DF	65

FIGURA 15: Grafo parcial da ontologia *wine*..... 72

FIGURA 16: Grafo parcial da ontologia *food*..... 73

LISTA DE TABELAS

TABELA 01 – SINTAXE E SEMÂNTICA DOS CONSTRUTORES DA LD <i>SROIQ</i>	32
TABELA 02 – SINTAXE E SEMÂNTICA DOS AXIOMAS DA LD <i>SROIQ</i>	32
TABELA 03 - COMPARAÇÃO ENTRE AVALIAÇÕES DE SIMILARIDADE DO MODELO PROPOSTO COM O MODELO DE D'AMATO.....	70
TABELA 04 - SIMILARIDADE ENTRE CONCEITOS OWL NA ONTOLOGIA <i>WINE</i> COM MODELO PROPOSTO.....	71
TABELA 05 - SIMILARIDADE ENTRE CONCEITOS OWL NA ONTOLOGIA <i>FOOD</i> COM MODELO PROPOSTO.....	72
TABELA 06 - SIMILARIDADE ENTRE SERVIÇOS (CONCEITOS).....	73

LISTA DE ABREVIATURAS E SIGLAS

LD – Lógica Descritiva.....	21
OWL – <i>Web Ontology Language</i>	22
FIPA - <i>Foundation For Intelligent Physical Agents</i>	22
DF – <i>Directory Facilitator</i>	22
AP – <i>Agent Plataform</i>	22
AMS – <i>Agent Management System</i>	22
IA – Inteligência Artificial.....	24
AL - <i>Attributive Language</i>	28
DL – <i>Description Logic</i>	31
W3C – <i>World Wide Web Consortium</i>	33
RDF - <i>Resource Description Framework</i>	33
IEEE - <i>Institute of Electrical and Electronics Engineers</i>	50
MTS – <i>Message Transport Service</i>	51
LGPL - <i>GNU Lesser General Public License</i>	52

Sumário

1 INTRODUÇÃO	21
1.1 OBJETIVOS	22
1.1.1 Objetivo Geral	22
1.1.2 Objetivos Específicos	22
1.2 Organização do Trabalho	23
2 LÓGICA	24
2.1 Representação de conhecimento	24
2.2 Lógica Descritiva	27
2.2.1 TBox	27
2.2.2 ABox	29
2.2.3 Raciocínio	30
2.3 A Lógica Descritiva <i>SROIQ</i>	31
2.4 OWL	33
3 SIMILARIDADE	35
3.1 Avaliação de similaridade	35
3.1.1 Modelo Relacional de Tversky (1977)	38
3.1.2 Modelo Geométrico	40
3.1.3 Modelo de Redes	42
3.1.4 Análise dos três modelos apresentados	43
3.1.5 Abordagem quantitativa	44
3.1.6 Abordagem qualitativa	44
3.2 Trabalhos Relacionados	45
3.2.1 Explorando métricas, para descoberta de serviços web semânticos, baseadas em similaridade (DIETZE, et al. 2009).	45
3.2.2 Um modelo híbrido de medida de similaridade conceitual para o ambiente ontológico (DONG, et al. 2009)	46
3.2.3 SIM-DLa : Uma nova medida de similaridade semântica para lógicas descritivas reduzindo similaridade entre conceitos a similaridade entre instâncias (JANOWICZ, et al. 2009).	47
3.2.4 Uma Medida de Similaridade Semântica para Expressivas Lógicas Descritivas (D'AMATO, et al. 2009).	47

4 ESTUDO DE CASO	49
4.1 Agentes	49
4.2 Sistemas Multiagentes	49
4.3 FIPA	50
4.4 JADE	52
4.4.1 O serviço de páginas amarelas	53
4.5 OWL-S	53
5 DESENVOLVIMENTO	56
5.1 Modelo de similaridade proposto	56
5.1.1 Definição da função base	57
5.1.2 Definição do contexto	57
5.1.3 Exemplo	60
5.2 Implementação	61
5.2.1 Implementação do modelo de similaridade	61
5.2.2 Implementação do estudo de caso	62
6 TESTES	66
6.1 Testes do modelo de similaridade	66
6.2 Testes com OWL	71
6.2.1 Testes com ontologias	71
6.2.2 Testes com agentes	73
6 CONSIDERAÇÕES FINAIS	75
6.1 Trabalhos futuros	76
REFERÊNCIAS	77
<i>BERNERS LEE, Tim; HENDLER, James; LASSILA, Ora. The Semantic Web. Scientific American, [S.1.], v.284, n.5, p. 35-43, 2001.</i>	77

1 INTRODUÇÃO

Humanos podem armazenar, pelo menos parcialmente, seu conhecimento em máquinas. Com isso, elas podem utiliza-los e manipula-los em diversos fins, como por exemplo, para inferir novos conhecimentos. Representações de conhecimento são utilizadas para permitir esse mapeamento do conhecimento.

Redes Semânticas e Frames são paradigmas para representação de conhecimento baseados em noções cognitivas, onde os elementos representados e seus relacionamentos formam uma rede. A lógica, como um sistema formal dedutivo, também pode ser utilizada para representar conhecimento. Como exemplo de linguagens lógicas que podem ser utilizadas, temos a lógica proposicional, a lógica de primeira ordem e a lógica descritiva. Essa última estrutura o conhecimento de forma similar as Redes Semânticas e Frames.

Lógica descritiva surgiu a partir da carência de semântica bem definida das Redes Semânticas e dos Frames. O paradigma de representação é similar a esses modelos, porém acresce a semântica definida por segmentos da lógica de primeira ordem. Com isso, embora existam declarações, providas da lógica, o conhecimento também pode ser visualizado como uma estrutura em forma de rede.

O conhecimento em lógica descritiva é armazenado em uma base de conhecimento. Esta, por sua vez, é dividida em duas entidades: O TBox contendo o conhecimento taxonômico, ou seja, definindo os conceitos e regras da representação; e o ABox constituído de asserções sobre indivíduos feitas utilizando as definições do TBox (BADDER, 2003). Sobre esse conhecimento é possível explicitar conhecimento implícito, ou seja, raciocinar. Por exemplo, podemos avaliar a similaridade entre dois conceitos definidos no TBox.

A definição da forma de avaliar quão similares dois conceitos são pode considerar tanto a estrutura da representação quanto as características dos conceitos comparados. (Por característica de um conceito entende-se conceitos e papéis que o definem.)

No contexto dessa pesquisa, os agentes são as entidades que utilizarão o conhecimento junto da função de similaridade desenvolvida. Agente pode ser definido como um sistema computacional situado em um ambiente e capaz de agir autonomamente neste ambiente para atingir seus objetivos de projeto (NORVING; RUSSELL, 2004). Algumas propriedades que são características de agentes são: autonomia, habilidade social, reatividade e pró-atividade (JENNINGS; WOOLDRIDGE, 1995).

Considerando que em um ambiente computacional podem existir diversos agentes, implementados de diversas formas, e que eles provavelmente irão interagir, é necessária a padronização de algumas características, a fim de prover a interoperabilidade entre esse conjunto de agentes. Para isso, a organização FIPA (*Foundation for Intelligent Physical Agents*) da *IEEE Computer Society*, define um conjunto de padrões para a interoperabilidade entre agentes e entre os agentes com outras tecnologias (FIPA, 2013).

No contexto do gerenciamento de agentes a FIPA definiu um modelo de referência, onde os agentes são processos computacionais que habitam uma plataforma de agentes (AP), que é a infraestrutura física para a execução. Na AP existe um sistema de gerenciamento de agentes (AMS) para gerenciar a AP, podendo criar, remover, e monitorar a migração de agentes. Outro componente que pode habitar uma AP é um agente Facilitador de Diretórios (DF), esse agente é responsável pelo fornecimento do serviço de páginas amarelas dentro da AP. Os agentes cadastram nas páginas amarelas os serviços que eles fornecem, para que quando um agente precise de um serviço ele possa perguntar para o DF quem é que fornece o serviço especificado (FIPA, 2013).

Neste trabalho, será desenvolvido um mecanismo para inferir similaridade entre conceitos de ontologias OWL. Esse mecanismo será utilizado pelo agente DF para sugerir serviços similares quando for solicitado.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

Definir um mecanismo de inferência para medição de similaridade de conceitos representados em OWL, analisando sua utilização em agentes, considerando o cenário do modelo de referência FIPA.

1.1.2 Objetivos Específicos

Pesquisar e analisar os mecanismos de inferência relacionados à lógica descritiva;

Pesquisar e analisar os modelos de similaridades que trabalham com a semântica de objetos;

Modelar um motor de inferência de similaridade;

Implementar o motor para a inferência de similaridades de acordo com a especificação estabelecida;

Implementar um agente facilitador de diretórios compatível com o modelo FIPA que utilize o motor de inferência implementado para responder as consultas;

Comparar os resultados do agente facilitador de diretórios utilizando o mecanismo implementado com um agente facilitador de diretórios tradicional.

1.2 Organização do Trabalho

Este trabalho está organizado do seguinte modo:

- O capítulo 2 apresenta uma revisão bibliográfica sobre representação e conhecimento, lógica descritiva e OWL;
- O capítulo 3 apresenta três formas para medição de similaridade;
- O capítulo 4 apresenta o estudo de caso: agentes e sistemas multiagentes;
- O capítulo 5 apresenta o modelo de similaridade proposto, sua implementação e a implementação do estudo de caso;
- O capítulo 6 apresenta os testes feitos;
- O capítulo 7 contém as considerações finais e trabalhos futuros;

2 LÓGICA

Em Inteligência Artificial (IA), o estudo da lógica está vinculado às formas de representar conhecimento por meio de elementos do domínio. Estes elementos podem estar relacionados a entidades inteligentes, que podem utilizá-los como base para tomada de decisões. A Representação de Conhecimento é uma área da Inteligência Artificial dedicada ao estudo das formas para representar esses elementos do domínio.

2.1 Representação de conhecimento

Para Sowa (2000), representação de conhecimento é a aplicação da lógica e da ontologia na tarefa de construir modelos computáveis de algum domínio. Para Brachman (2003) representação de conhecimento é como conhecimento pode ser representado simbolicamente e a partir dessa representação, como ele pode ser manipulado de forma automatizada por programas.

Já Davis et al. (1993), explica a representação de conhecimento a partir de cinco papéis que ela pode representar, sendo que cada um desses papéis produz um conjunto distinto de propriedades que devem estar presentes em uma representação de conhecimento. Os autores, ainda, consideram a lógica apenas como uma de muitas maneiras para representação de conhecimento. Para fins de contextualização, os cinco papéis serão detalhados seguindo a forma em que eles se aplicam na lógica.

- **Papel 1: Uma representação de conhecimento é uma substituta para as coisas do mundo**

Embora o raciocínio seja uma atividade interna dos seres humanos, muitas vezes ele é feito sobre entidades que existem apenas no meio externo. Para raciocinar, é necessário que essas entidades estejam representadas de forma que os seres humanos, ou as máquinas, consigam deduzir conclusões.

Em várias situações, ações no mundo real podem resultar em algo indesejado. Para estimar o resultado de uma ação, podem ser utilizadas representações como substitutas dos objetos do mundo real, e o raciocínio como a entidade responsável pela execução da ação. A partir disso, é desejável que a entidade do mundo e a entidade representada possuam o mesmo valor semântico, e, o que é muito difícil, que a entidade representada seja uma cópia fiel (DAVIS, et al. 1993).

A lógica é utilizada para modelar as coisas, bem como suas relações, do mundo e armazená-las em bases de conhecimento de acordo com a linguagem utilizada. O raciocínio dedutivo é feito com a maquinaria da lógica utilizada e com as informações contidas na base de conhecimento.

- **Papel 2: Uma representação de conhecimento é um conjunto de compromissos ontológicos**

Sabendo que representações são, em sua maioria, imperfeitas, é necessário que elas contenham os aspectos relevantes de cada entidade. O papel dos compromissos ontológicos serve como um guia na modelagem, ele define que as entidades devem conter apenas os aspectos importantes para o domínio. Isso é útil para o raciocínio, uma vez que reduz a quantidade de premissas, e descarta inferências desnecessárias (DAVIS, et al. 1993).

A lógica pode atuar como linguagem para representar os compromissos ontológicos. A forma e a capacidade de representar os elementos do domínio dependem da expressividade da linguagem utilizada. Por exemplo, a lógica proposicional modela conhecimento através de proposições e conectivos lógicos. Já a lógica de primeira ordem, mais expressiva, utiliza predicados (declarações que podem ser verdadeiras ou falsas), funções, constantes, conectivos lógicos e quantificadores.

- **Papel 3: Uma representação de conhecimento é uma teoria incompleta de raciocínio inteligente.**

A teoria do raciocínio inteligente é fragmentada em dois sentidos distintos (DAVIS, et al. 1993): no primeiro a representação incorpora parte do conhecimento ou da crença que a motivou, no segundo, esse conhecimento ou crença é apenas uma parte do fenômeno complexo do raciocínio inteligente.

A teoria do raciocínio inteligente pode ser melhor entendida se analisada a partir de seus três componentes:

O primeiro define uma representação como uma concepção fundamental de raciocínio inteligente. Aqui entram diversas possíveis concepções, dentre elas, existe a derivada da lógica matemática, que supõe raciocínio inteligente como um processo que pode ser capturado de uma descrição formal, precisa e concisa. A concepção proveniente da psicologia, onde o raciocínio é visto como característica do comportamento humano, em estruturas de estados mentais, tais como

objetivos e planos. Do ponto de vista dos seres humanos, o raciocínio é feito de diferentes formas: dedução, indução, associação, etc(DAVIS, et al. 1993).

O segundo componente define o conjunto de inferências que a representação sanciona. Esse conjunto é considerado como conclusão apropriada obtida a partir das informações disponíveis. A definição clássica é fornecida pela lógica formal tradicional, onde só são validadas as inferências que são provenientes de implicação lógica de um conjunto de axiomas (DAVIS, et al. 1993).

O terceiro componente define o conjunto de inferências que a representação recomenda. A partir do conjunto de inferências que a representação sanciona, são escolhidas as inferências que são apropriadas, para isso ser feito é necessário que ao definir a representação, algumas informações sobre o raciocínio sejam fornecidas. Nos sistemas baseados em lógica, esse componente não é muito desenvolvido, uma vez que o segundo componente é poderoso (DAVIS, et al. 1993).

- **Papel 4: Uma representação de conhecimento é um meio para computação eficiente.**

Assumindo que máquinas são utilizadas para representar conhecimento, é necessário que elas o façam de forma eficiente. Representar conhecimento de forma declarativa, tal como é feito pela lógica de primeira ordem, pode ser uma tarefa custosa para um computador, uma vez que uma linguagem designada para expressar fatos de forma declarativa pode não ser boa para expressar a informação imperativa, característica de uma estratégia de raciocínio (DAVIS, et al. 1993).

A utilização de representações pode organizar as informações em conjuntos de forma que a tentativa de inferência seja feita apenas em conjuntos que podem resultar em alguma inferência útil.

- **Papel 5: Uma representação de conhecimento é um meio para expressão humana.**

O último papel caracteriza as representações como uma forma de expressar as coisas do mundo, e assim, como um meio para permitir que as máquinas possam utilizar as representações (DAVIS, et al. 1993). Este é o contexto para a web semântica, onde as máquinas são capazes de compreender a semântica de documentos e dados, e assim, que

computadores e pessoas trabalhem em cooperação (BERNERS-LEE,2001).

2.2 Lógica Descritiva

O item anterior apresentou conceitos que norteiam a representação de conhecimento, neste ponto é apresentada a lógica descritiva como uma linguagem formal para representar conhecimento.

As redes semânticas são estruturas para representação de conhecimento baseadas em redes, onde as entidades são representadas por vértices, e as relações entre elas são representadas por arestas (QUILLIAN,1967). Com estrutura similar, surgiram os *frames* (MINSKY, 1981), que substituem os vértices por *frames* com propriedades das entidades. Por serem sistemas baseados em noções cognitivas possuem uma dificuldade natural para caracterização semântica. Com o objetivo de prover significado declarativo e formal nessas estruturas de representação baseadas em redes, surgiu a lógica descritiva.

Lógica descritiva (LD) é um formalismo para representação de conhecimento baseado em estruturas de redes com ferramentas de raciocínio eficiente. Além disso, é uma família de linguagens, construída a partir de fragmentos, possivelmente decidíveis, da lógica de primeira ordem. E que, assim como a lógica de primeira ordem, busca encontrar consequências implícitas a partir de conhecimento representado explicitamente.

Uma base de conhecimento LD é composta por conhecimento intensional e extensional, sendo que o conhecimento é separado em duas estruturas distintas, TBox e ABox. O conhecimento intensional é definido pelo TBox, que contém o vocabulário de uma aplicação de domínio. O conhecimento extensional é definido pelo ABox e contém asserções sobre indivíduos em termos desse vocabulário (BAADER,2003).

2.2.1 TBox

O TBox é composto por conceitos e papéis. Conceitos denotam conjuntos de indivíduos e papéis denotam o relacionamento binário entre indivíduos. As formas mais comuns de raciocínio sobre o TBox são: determinar se uma descrição é satisfazível (não contraditória); determinar se uma descrição é mais genérica que a outra, isto é, se uma subsume a outra.

Os elementos base da construção e definição do TBox são os conceitos atômicos, os papéis atômicos e a linguagem descritiva utilizada. Os conceitos atômicos são utilizados como abreviações, ou nomes, para conceitos complexos. Papéis atômicos podem ser utilizados com outros papéis ou conceitos para formar novos papéis.

A linguagem \mathcal{AL} (Linguagem Atributiva) utilizada para contextualização desta pesquisa foi introduzida pelos autores Schmidt-Schauß e Smolka (1991) como uma linguagem mínima para interesses práticos. Assumindo C e D como descrição de conceitos arbitrários, A e B como conceitos atômicos, R e S como papéis atômicos:

$C, D \Rightarrow A$ | (conceito atômico)
 \top | (conceito universal)
 \perp | (conceito vazio)
 $\neg A$ | (negação atômica)
 $C \sqcap D$ | (intersecção)
 $\forall R.C$ | (restrição de valor)
 $\exists R.T$ | (quantificador existencial limitado)

Onde \top representa um conceito equivalente a todos conceitos do domínio, e $\forall R.C$ representa que o papel R é válido para todo o conceito C . Assumindo agora que *Feminino* e *Pessoa* são conceitos atômicos e *possuiFilho* é um papel atômico, é possível construir intuitivamente conceitos complexos da forma:

$\text{Feminino} \sqcap \text{Pessoa}$, para designar aqueles que são pessoas do sexo feminino. $\text{Pessoa} \sqcap \exists \text{possuiFilho}$. \top , para representar as pessoas que possuem filhos. $\neg \text{Feminino} \sqcap \text{Pessoa}$, para representar as pessoas que não são do sexo feminino.

Essas descrições complexas podem receber nomes simbólicos, e assim, formarem novos conceitos atômicos. Cabe aqui, então, a introdução de axiomas terminológicos como sendo, em sua maioria, axiomas de equivalência e axiomas de inclusão, e sendo representado respectivamente por $C \equiv D (R \equiv S)$ e $C \sqsubseteq D (R \sqsubseteq S)$. Essas relações são válidas para conceitos e papéis, porém a linguagem \mathcal{AL} não suporta construtores de papéis. Seja o axioma terminológico de equivalência:

$\text{Mulher} \equiv \text{Feminino} \sqcap \text{Pessoa}$

A partir desse exemplo caracteriza-se *definição*. Uma definição é o lado esquerdo de um axioma de equivalência, deve ser um conceito atômico, como no caso acima - Mulher. Definições são utilizadas para introduzir nomes simbólicos para descrições complexas.

Uma terminologia, ou TBox, é um conjunto de definições não ambíguas, ou seja, se para todo conceito atômico A existe no máximo um axioma em que o lado esquerdo da definição é A.

Supondo que não seja possível definir o conceito Criança, porém sabe-se que toda criança é uma Pessoa, utilizando o axioma de inclusão Criança pode ser descrita como uma especialização de Pessoa:

$$\text{Criança} \sqsubseteq \text{Pessoa}$$

Se uma terminologia aceita construções do tipo acima, então ela recebe o nome de *terminologia generalizada*, porém, o lado esquerdo de qualquer axioma de qualquer construção terminológica não deve ser repetido. Em teoria, a adição de axiomas de inclusão não adiciona expressividade em terminologias, entretanto, na prática, eles são um meio conveniente para introdução de termos que não podem ser completamente definidos (BAADER,2003).

2.2.2 ABox

A Descrição do Mundo, ou ABox, é formada por um conjunto finito de asserções sobre indivíduos utilizando os conceitos e papéis definidos no TBox. Por exemplo, Mulher(MARIA), representa um indivíduo com nome MARIA e pertence ao conceito Mulher. A asserção possuiFilho(MARIA,JOÃO), representa que o indivíduo MARIA possui um filho JOÃO.

De forma simplista, um ABox pode ser visto como uma instância de banco de dados relacional constituída apenas de relações unárias e binárias. O identificador de um indivíduo é o seu nome, ou seja, asserções com um mesmo nome representam o mesmo indivíduo. No exemplo acima, a MARIA que é Mulher é a mesma MARIA que possui o filho JOÃO.

Enquanto que em uma instância de banco de dados relacional, apenas uma interpretação de cada objeto é possível - mundo fechado -, um ABox pode representar diversas interpretações para um mesmo objeto - mundo aberto. Por exemplo, supondo que a única asserção seja

`possuiFilho(MARIA,JOÃO)`, para um banco de dados relacional, MARIA possui apenas um filho, JOÃO. Entretanto, para um ABox, JOÃO pode ser o único filho de MARIA, mas pode também não ser. Ou seja, sabe-se apenas que JOÃO é filho de MARIA, mais nada a respeito disso pode ser afirmado como verdadeiro ou falso.

2.2.3 Raciocínio

Uma base de conhecimento de LD - composta do TBox e do ABox - é equivalente semanticamente a um conjunto de axiomas da lógica de primeira ordem, assim, caso exista conhecimento implícito, é possível explicitá-los a partir de inferências lógicas.

O raciocínio em sistemas de LD é definido como um conjunto de inferências lógicas. As inferências podem ser sobre conceitos, TBox, ABox e sobre TBox e ABox.

Seja \mathcal{T} uma terminologia, I uma interpretação dessa terminologia. As principais formas de raciocínio sobre conceitos são:

Satisfazibilidade: Um conceito C é *satisfazível* em relação à \mathcal{T} , se existe um modelo I de \mathcal{T} tal que C^I é um conjunto não vazio.

Subsunção: Um conceito C^I é *subsumido* por um conceito D em relação à \mathcal{T} se $C^I \subset D^I$ para todo modelo I de \mathcal{T} .

Equivalência: Dois conceitos C e D são *equivalentes* em relação à \mathcal{T} se $C^I = D^I$ para todo modelo I de \mathcal{T} .

Disjunção: Dois conceitos C e D são *disjuntos* em relação à \mathcal{T} se $C^I \cap D^I = \emptyset$ para todo modelo I de \mathcal{T} .

O mecanismo básico para raciocínio em sistemas de LD é a subsunção, uma vez que ela é suficiente para implementar as outras inferências.

As inferências sobre o TBox consistem em considerar ele como vazio, expandir as definições para expressões com apenas símbolos base (o lado direito do axioma de equivalência), e então deduzir fatos sobre as expansões, a principal delas é a verificação de consistência. Por exemplo:

Mulher \sqcap Homem

Pode ser expandido para:

Feminino \sqcap Pessoa $\sqcap \neg(\text{Feminino} \sqcap \text{Pessoa})$

e concluído que Mulher e Homem são disjuntos. A expansão de conceitos pode ser computacionalmente custosa, no pior caso com complexidade exponencial. Por isso deve ser evitada.

A inferência mais básica que pode ser feita sobre um ABox é a verificação de instância, ou seja, checar se uma asserção é implicada por um ABox.

Considerando agora uma base de conhecimento como uma forma para salvar informações sobre indivíduos. Dada uma descrição de um conceito, é desejado buscar na base todos os indivíduos que são caracterizados por aquele conceito. A forma mais simples, porém mais custosa, é verificar entre todos os indivíduos, os que são instâncias do conceito descrito (BAADER,2003).

2.3 A Lógica Descritiva *SROIQ*

Até aqui foram caracterizados alguns elementos básicos das lógicas descritivas. Nesse ponto inicia-se o estudo da LD *SROIQ*, uma vez que seu modelo teórico semântico é compatível com a linguagem padrão utilizada para representar conhecimento na web, OWL DL, que será descrita em 2.4.

A linguagem *SROIQ* suporta construtores de papéis, cabe aqui, então, a definição dos construtores de papéis: Todo nome de papel é uma descrição de papel (papel atômico), e se R e S são descrições de papéis, então $R \sqcap S$ (interseção), $R \sqcup S$ (união), $\neg R$ (complemento), $R \circ S$ (composição), R^+ (fecho transitivo) e R^- (inverso) também são descrições de papéis.

Papéis são relações binárias que podem ser consideradas axiomas de relações (RBox). Supondo que JOÃO é irmão de JOSÉ, e JOSÉ é pai de PEDRO, intuitivamente pode-se concluir que JOÃO é tio de PEDRO. A partir dos papéis *irmãoDe*, *paiDe* e *tioDe*, é possível construir o axioma:

$$\text{irmãoDe} \circ \text{paiDe} \sqsubseteq \text{tioDe}$$

A Sintaxe e a semântica dos construtores da LD *SROIQ* são exibidos na tabela 1. Onde I representa o domínio da interpretação. Por exemplo, a representação sintática $\neg C$ significa que dado todos os conceitos da interpretação Δ^I e o conceito C, $\neg C$ representa todos aqueles conceitos que estão na interpretação do domínio, menos C.

	Syntax	Semantics
<i>Individuals:</i>		
individual name	a	a^I
<i>Roles:</i>		
atomic role	R	R^I
inverse role	R^-	$\{\langle x, y \rangle \mid \langle y, x \rangle \in R^I\}$
universal role	U	$\Delta^I \times \Delta^I$
<i>Concepts:</i>		
atomic concept	A	A^I
intersection	$C \sqcap D$	$C^I \cap D^I$
union	$C \sqcup D$	$C^I \cup D^I$
complement	$\neg C$	$\Delta^I \setminus C^I$
top concept	\top	Δ^I
bottom concept	\perp	\emptyset
existential restriction	$\exists R.C$	$\{x \mid \text{some } R^I\text{-successor of } x \text{ is in } C^I\}$
universal restriction	$\forall R.C$	$\{x \mid \text{all } R^I\text{-successors of } x \text{ are in } C^I\}$
at-least restriction	$\geq n R.C$	$\{x \mid \text{at least } n R^I\text{-successors of } x \text{ are in } C^I\}$
at-most restriction	$\leq n R.C$	$\{x \mid \text{at most } n R^I\text{-successors of } x \text{ are in } C^I\}$
local reflexivity	$\exists R.\text{Self}$	$\{x \mid \langle x, x \rangle \in R^I\}$
nominal	$\{a\}$	$\{a^I\}$

where $a, b \in N_I$ are individual names, $A \in N_C$ is a concept name, $C, D \in \mathbf{C}$ are concepts, $R \in \mathbf{R}$ is a role

TABELA 01 – SINTAXE E SEMÂNTICA DOS CONSTRUTORES DA LD *SROIQ*

A sintaxe e a semântica dos axiomas da LD *SROIQ* estão na tabela 2, ela descreve operações que podem ser feitas no TBox, ABox e RBox:

	Syntax	Semantics
<i>ABox:</i>		
concept assertion	$C(a)$	$a^I \in C^I$
role assertion	$R(a, b)$	$\langle a^I, b^I \rangle \in R^I$
individual equality	$a \approx b$	$a^I = b^I$
individual inequality	$a \neq b$	$a^I \neq b^I$
<i>TBox:</i>		
concept inclusion	$C \sqsubseteq D$	$C^I \subseteq D^I$
concept equivalence	$C \equiv D$	$C^I = D^I$
<i>RBox:</i>		
role inclusion	$R \sqsubseteq S$	$R^I \subseteq S^I$
role equivalence	$R \equiv S$	$R^I = S^I$
complex role inclusion	$R_1 \circ R_2 \sqsubseteq S$	$R_1^I \circ R_2^I \subseteq S^I$
role disjointness	$\text{Disjoint}(R, S)$	$R^I \cap S^I = \emptyset$

TABELA 02 – SINTAXE E SEMÂNTICA DOS AXIOMAS DA LD *SROIQ*

2.4 OWL

Neste ponto é introduzida a OWL, uma linguagem com semântica equivalente a lógica descritiva. Com isso, os conceitos relacionados à lógica descritiva, bem como, o mecanismo de similaridade proposto nesta pesquisa são válidos tanto para lógica descritiva quanto para OWL.

A OWL (Web Ontology Language) é uma linguagem recomendada pela W3C para definir e instanciar ontologias na Web. A primeira versão foi recomendada em fevereiro de (W3C, 2004) com o objetivo de prover uma linguagem para descrever classes e seus relacionamentos em documentos Web e aplicações. A versão atual, OWL 2, foi padronizada em (W3C, 2009).

Uma ontologia OWL é equivalente a uma base de conhecimento de LD. Os blocos principais da OWL são similares aos das LDs, a principal diferença é a nomenclatura. Conceitos são chamados de classes e papéis são chamados de propriedades.

A atribuição de significado (W3C, 2012) em ontologias OWL 2 pode ser feita de duas maneiras: Semântica Direta e Semântica Baseada em RDF (*Resource Description Framework*). A atribuição por meio de Semântica Direta é feita a partir da semântica do modelo teórico da LD *SRIOIQ*. Porém, essa semântica exige algumas restrições sintáticas da OWL, essas restrições geram a sublinguagem OWL DL. Axiomas OWL DL podem ser lidos como axiomas *SRIOIQ*. Por exemplo, seja o axioma OWL:

```
EquivalentClasses (Mulher
ObjectIntersectionOf (Feminino Pessoa))
```

Ele pode ser escrito em *SRIOIQ* como:

$$\text{Mulher} \equiv \text{Feminino} \sqcap \text{Pessoa}$$

Existem algumas diferenças entre a OWL DL e a linguagem *SRIOIQ*. Em nível sintático, a OWL fornece mais operadores, embora sejam logicamente redundantes. Por exemplo, construtores especiais para quantidade de elementos que certas propriedades têm.

Em relação à expressividade, a OWL possui suporte a *datatypes* e literais de *datatype*. O comportamento destes é o mesmo de classes e

nomes de indivíduos, porém com a interpretação fixada e predefinida. Por exemplo, o *datatype* para valores booleanos possui exatos dois elementos - *true* e *false* - em qualquer interpretação. Embora linguagem *SRDIQ* descrita acima não implemente *datatypes*, ela possui a extensão *SRDIQ^(D)*, que suporta *datatypes* (o *D* do *SRDIQ^(D)* deve-se aos *datatypes*).

A última característica lógica presente na OWL e ausente na *SRDIQ* é a definição de *Keys*. *Keys* são usados para integração de dados. De forma simplista, uma *key* especifica que dois nomes de indivíduos implicam serem iguais se compartilham determinadas propriedades e são membros de mesmas classes.

Além dos aspectos lógicos, a OWL inclui outras características que não são consideradas na lógica descritiva. Por exemplo, a OWL permite a importação de axiomas de uma ontologia para outra. Para isso ser feito em uma base de conhecimento LD, cuidados como verificação de ambiguidades em axiomas devem ser tomados.

3 SIMILARIDADE

A similaridade atua como um princípio organizacional, o qual orienta a classificação, formação de conceitos e generalizações (TVERSKY, 1977), podendo assim, ser considerada uma forma de raciocínio. Esta seção apresentará de forma superficial como a similaridade pode ser avaliada de forma genérica, e também no contexto de lógica descritiva.

Existem dois aspectos importantes na medição de similaridade. Um qualitativo, ou seja, busca estabelecer em quê dois objetos são similares, e outro quantitativo, o qual define o quanto dois objetos são similares. Mesmo sendo útil em diversos problemas de psicologia e ciências da computação ela é difícil de modelar (GOODMAN, 1972).

Dizer que dois objetos são similares por possuírem pelo menos uma característica em comum, considerar similares dois objetos que compartilham todas características, ou ainda, considerar que dois objetos a e b , são mais similares que outros dois objetos c e d se a e b possuírem mais propriedades em comum que c e d , pode ser errôneo. Para o primeiro caso, similaridade acaba se tornando um conceito universal. No segundo caso, similaridade torna-se muito restritiva, e assim, resultando relações vazias. Para o terceiro caso, dois pares de objetos podem conter a mesma quantidade de características comuns. Uma estratégia útil pode ser a seleção não apenas das características importantes, mas também a ponderação da importância relativa de cada característica, além da variação em relevância.

3.1 Avaliação de similaridade

Definir quão similares dois objetos são é subjetivo. O julgamento de similaridade é feito a partir do contexto e da estrutura da referência (TVERSKY, 1977). Por exemplo, Santa Catarina é mais similar ao Rio Grande do Sul do que a São Paulo se for considerada a distância geográfica entre os estados. Porém, considerando apenas os nomes dos estados, Santa Catarina pode ser mais similar a São Paulo por ambos se referirem a nomes santos.

Em seu trabalho, Tversky (1977) mostrou evidências empíricas de que o julgamento de similaridade pode ser assimétrico, ou seja, é possível que um objeto a seja mais parecido a um objeto b do que o objeto b seja a a . Por exemplo, é possível considerar *papagaio* mais similar a *ave* do que *ave* a *papagaio*.

Segundo (SCHWERING, 2008), as duas principais formas de mensurar a similaridade entre conceitos são através da consideração das características que cada um possui, ou da distância entre eles.

Para o caso das características, quanto mais elementos em comum, mais similares os conceitos são. Por exemplo, supondo os conceitos Homem, Mulher e Cavalo, e as atribuições Humano, Equídeo, Masculino, Feminino, Mamífero, Bípede, Quadrúpede. Em lógica descritiva os conceitos podem ser definidos como:

$$\begin{aligned} \text{Homem} &\equiv \text{Humano} \sqcap \text{Masculino} \sqcap \text{Mamífero} \sqcap \text{Bípede} \\ \text{Mulher} &\equiv \text{Humano} \sqcap \text{Feminino} \sqcap \text{Mamífero} \sqcap \text{Bípede} \\ \text{Cavalo} &\equiv \text{Equídeo} \sqcap \text{Masculino} \sqcap \text{Mamífero} \sqcap \\ &\text{Quadrúpede} \end{aligned}$$

Onde o símbolo \equiv equivale a dizer que o conceito do lado esquerdo é equivalente a descrição do lado direito, e \sqcap representa a interseção entre o que o precede e o sucede. Os conceitos Homem e Mulher possuem três características em comum (Humano, Mamífero, Bípede), Homem e Cavalo duas (Masculino, Mamífero) e Mulher e Cavalo apenas uma (Mamífero). A partir disso, Homem e Mulher são mais similares entre si do que Homem e Cavalo, Mulher e Cavalo. Isso ocorre por Homem e Mulher possuírem mais características em comum (três) e menos características incomuns (duas). Já Mulher e Cavalo possuem a menor similaridade por compartilhar uma característica (Mamífero), e divergir em seis (Humano, Feminino, Bípede, Equídeo, Masculino, Quadrúpede).

Partindo agora da distância semântica como uma forma de medir similaridade, o primeiro item a ser considerado é que normalmente distância semântica mede a dissimilaridade entre dois conceitos. Características incomuns afastam os conceitos e o cálculo da similaridade é feito a partir das divergências entre eles. Por exemplo, reescrevendo os conceitos Homem e Mulher:

$$\begin{aligned} \text{Homem} &\equiv \text{Humano} \sqcap \text{Masculino} \\ \text{Mulher} &\equiv \text{Humano} \sqcap \neg\text{Masculino} \end{aligned}$$

Duas formas simplistas para caracterizar a utilização de distância para medir similaridade com esses dois conceitos são colocadas a seguir. Na primeira, cada característica distinta dos conceitos é considerada uma coordenada (Humano e Masculino). O valor da coordenada pode variar, assumindo 1 para o caso em que a característica está presente no conceito, e 0 caso contrário. É possível construir um gráfico como ilustrado na figura 1, onde o conceito Masculino é representado por um ponto azul, e o conceito Mulher é representado com um ponto vermelho. A distância entre os conceitos é dada pela distância entre os pontos.

Essa estratégia mede a dissimilaridade entre os conceitos, uma vez que são as características incomuns que os distanciam. Ao aumentar o número de características, aumenta o número de coordenadas, com isso, de um lado dificulta a visualização do local dos pontos já que é um espaço n dimensional, e de outro enriquece a avaliação da distância entre conceitos pela mesma razão.

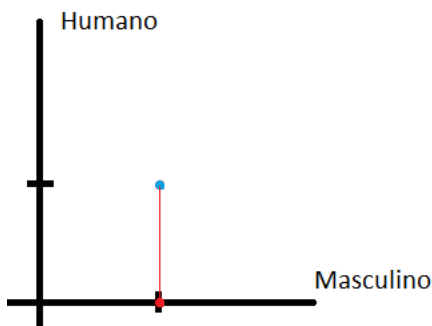


FIGURA 01: distância entre os conceitos Homem (ponto azul) e Mulher (ponto vermelho).

Outra forma utilizada para medir distância semântica é encontrada em representações taxonômicas, como árvores. Nestes sistemas, conceitos podem ser representados como nodos e a relação entre dois conceitos pode ser definida por uma aresta, uma forma de medir a distância semântica entre estes é a partir da quantidade de nodos que precisam ser percorridos de um até o outro.

Por exemplo, para os conceitos Homem e Mulher do exemplo anterior, a figura 2 mostra uma possível construção taxonômica. O conceito mais geral na estrutura é Humano. As arestas entre Humano e Mulher, e entre Humano e Homem, referem ao fato de que tanto Mulher quanto Homem “é um” Humano, o que diferencia um Homem

de uma Mulher é a característica, que pode ser uma relação ou atributo, de ser ou não Masculino. A distância semântica entre Homem e Mulher de forma trivial pode ser dada por 1 (considerando apenas Humano), ou 3 (considerando Humano, Mulher e Homem).

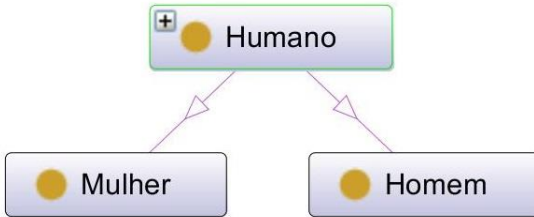


FIGURA 02 Estrutura taxonômica considerando o relacionamento “é um” entre Humano, Mulher e Homem.

A classificação de modelos utilizados para avaliação de similaridade varia de acordo com a forma que é visualizada. (SCHWERING, 2008) Separa os modelos em cinco, três deles consideram distância, e os outros dois características. Já (SÁNCHEZ; BATET; ISERN, 2012), divide em três modelos considerando distância, características e a estrutura da representação. A seguir serão descritos três modelos, um considerando características, um apenas distância e outro distância e a estrutura.

3.1.1 Modelo Relacional de Tversky (1977)

Modelo proveniente da psicologia que, para a avaliação da similaridade, considera as características comuns e incomuns entre os estímulos (conceitos) e o contexto em que eles estão. Antes de fornecer a função do modelo relacional, será apresentado um exemplo de entrada para a função. Este será dividido em duas partes: a primeira construirá os conjuntos de características comuns e diferentes; a segunda definirá o contexto.

Sejam os conceitos:

Pai \equiv Humano \sqcap Masculino \sqcap \exists temFilho.Humano
 Mulher \equiv Humano \sqcap Feminino

Para a parte um, o primeiro passo é extrair as características de cada conceito. Assim, para *Pai* e *Mulher* são, respectivamente, $\{\text{Humano, Masculino, } \exists \text{temFilho.Humano}\}$ e $\{\text{Humano, Feminino}\}$. A seguir, para separar o que é comum e incomum as operações de interseção e subtração da Teoria dos Conjuntos podem ser utilizadas, com isso:

Interseção:

$$\{\text{Humano, Masculino, } \exists \text{temFilho.Humano}\} \cap \{\text{Humano, Feminino}\} = \{\text{Humano}\}$$

Diferença:

$$\{\text{Humano, Masculino, } \exists \text{temFilho.Humano}\} - \{\text{Humano, Feminino}\} = \{\text{Masculino, } \exists \text{temFilho.Humano}\}$$

$$\{\text{Humano, Feminino}\} - \{\text{Humano, Masculino, } \exists \text{temFilho.Humano}\} = \{\text{Feminino}\}$$

Portanto a característica comum é $\{\text{Humano}\}$, as que somente *Pai* tem são $\{\text{Masculino, } \exists \text{temFilho.Humano}\}$ e a que exclusiva de *Mulher* é $\{\text{Feminino}\}$. Estes conjuntos formam as entradas para as funções utilizadas no modelo de Tversky.

A parte dois analisa o contexto. Supondo que as características exclusivas do conceito *Pai* são mais importantes do que as do conceito *Mulher*, com isso o fator α representando o contexto de *Pai* será 2, e o fator β , das características do conceito *Mulher* será 0,5.

Com essas informações é possível avaliar a similaridade entre *Pai* e *Mulher* com a função do Modelo Relacional (*Ratio Model*) de Tversky (1).

$$s(a, b) = \frac{f(A \cap B)}{f(A \cap B) + \alpha * f(A - B) + \beta * f(B - A)} \quad (1)$$

Sendo $\alpha, \beta \leq 0$. a e b os estímulos (conceitos), A e B conjuntos das características dos estímulos.

Mapeando o exemplo para a função, tem-se:

$$a = \text{Pai};$$

$$b = \text{Mulher};$$

$$A = \{\text{Humano, Masculino, } \exists \text{temFilho.Humano}\}$$

$$B = \{\text{Humano, Feminino}\}$$

$$f(A \cap B) = f(\{\text{Humano}\});$$

$$f(A - B) = f(\{\text{Masculino}, \exists \text{temFilho.Humano}\});$$

$$f(B - A) = f(\{\text{Feminino}\});$$

$$\alpha = 2; \beta = 0,5;$$

Para calcular a similaridade é necessário definir a função f , assumindo que f seja a função de cardinalidade, então:

$$f(A \cap B) = 1;$$

$$f(A - B) = 2;$$

$$f(B - A) = 1;$$

Com isso, a similaridade entre Pai e Mulher é dada por:

$$S(\text{Pai}, \text{Mulher}) = 1/(1 + 2*2 + 1*0,5) = 0,1818.$$

Algumas observações sobre esse modelo:

O modelo mede similaridade no intervalo $[0,1]$. Se as características distintas dos conceitos não forem relevantes, então $\alpha = \beta = 0$, com isso, os conceitos terão similaridade 1 o que implica em ser equivalentes. Para os casos em que $\alpha < > \beta$, como no exemplo, a similaridade é assimétrica.

3.1.2 Modelo Geométrico

Modelo com origens na psicologia, anterior ao trabalho de Tversky, com o objetivo de explorar a analogia entre espaço e similaridade. Conceitos são pontos no espaço, e a distância entre os pontos é a dissimilaridade entre eles, já que são as características não comuns que afastam os pontos (conceitos). Cada característica distinta do domínio, é uma dimensão do espaço.

Por exemplo, supondo as características a, b, c e os conceitos Conceito1 e Conceito2, sabe-se que as características de Conceito1 e Conceito2 são respectivamente: a, b e b, c . Com isso, o espaço possuirá três dimensões (uma para representar a , outra para b e outra para c). Os conceitos serão representados como pontos e as características como dimensões, assim os pontos com suas respectivas coordenadas serão: Conceito1 $(a, b, 0)$ e Conceito2 $(0, b, c)$. A representação gráfica é dada na figura 3.

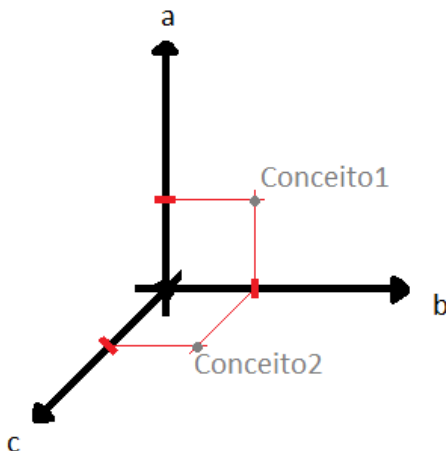


FIGURA 03 Representação de Conceito1 e Conceito2 no espaço das características a, b e c .

Uma vez definidas as coordenadas dos pontos, é possível calcular a distância entre eles. O cálculo pode ser feito a partir da diferença em módulo dos valores dos pontos para cada dimensão. A utilização da medida de distância de Minkowski apresentada na equação (2), permite que a distância entre os pontos seja mensurada utilizando distância euclidiana ou Manhattan.

$$d_{ij} = \left[\sum_{k=1}^n |x_{ik} - x_{jk}|^r \right]^{1/r} \quad (2)$$

Onde i e j são pontos, n é o número de dimensões, X_{ik} é o valor da dimensão k para o ponto i , e X_{jk} é o valor da dimensão k para o ponto j . Com $r = 1$ é calculada a distância Manhattan e $r = 2$, é calculada a distância euclidiana. A figura 4 exibe a distância entre os conceitos considerando a distância euclidiana e a distância Manhattan. A distância Manhattan sempre será maior ou igual à distância euclidiana.

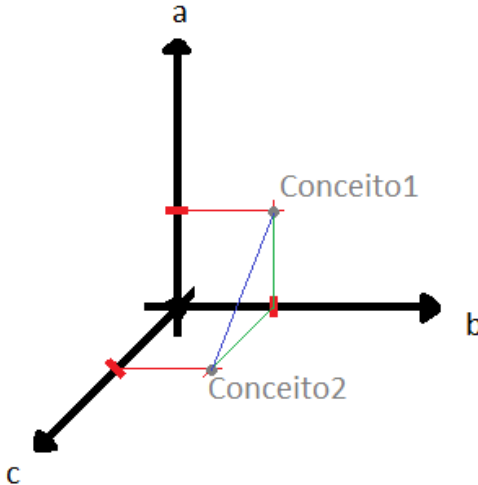


FIGURA 04 Representação em verde da distância Manhattan e em azul da distância euclidiana.

3.1.3 Modelo de Redes

Modelo utilizado para avaliação de similaridade em situações em que o conhecimento é representado de forma estruturada, como por exemplo em Redes Semânticas. Estas foram originadas na psicologia com o propósito de modelar a semântica da memória humana (QUILLIAN,1967). Rede Semântica é uma notação gráfica para representar conhecimento (Sowa, 2013), ela é formada por nodos (representam conceitos, objetos, propriedades) e arestas (relacionamento entre nodos).

A lógica descritiva surgiu para prover significado declarativo e formal para as redes semânticas, preservando o formato estrutural para representação de conhecimento. Desta forma, é válida a utilização do modelo de redes para avaliar similaridade em LD.

A partir do relacionamento "é um", autores seguem diversos caminhos para avaliação da similaridade conceitual. Segundo o modelo proposto por Wu e Palmer (1994), para medição de similaridade entre dois conceitos Conceito1 e Conceito2 são considerados o primeiro ancestral em comum (Conceito3), o número N1 de nodos do caminho entre Conceito1 e Conceito3, o número N2 de nodos do caminho entre Conceito2 e Conceito3 e o número N3 representando a distância entre Conceito3 e a raiz.

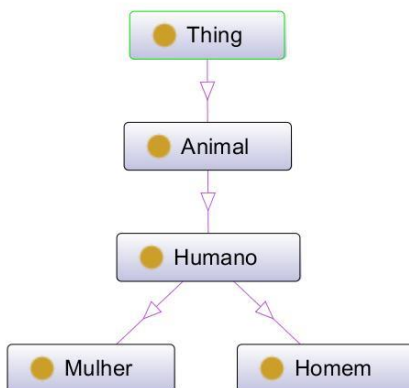


FIGURA 05 Exemplo de estrutura conceitual.

Para o exemplo da figura 05, os valores de N_1 , N_2 e N_3 são, respectivamente, 1, 1 e 2. Para determinar a similaridade entre *Mulher* e *Homem* segundo o modelo de Wu e Palmer, basta substituir os valores na função apresentada em (3), obtendo a avaliação de similaridade igual a 0,66. Observando que, segundo este modelo, conceitos próximos tendem a ter similaridade alta e, quando, além disso, eles ficarem em regiões mais especializadas da árvore, maior será a similaridade.

$$sim(C1, C2) = \frac{2*N3}{N1+N2+2*N3} \quad (3)$$

3.1.4 Análise dos três modelos apresentados

Os três modelos apresentados acima abordam a avaliação de similaridade de forma distinta. Cada modelo tem seu arcabouço matemático. O modelo geométrico é construído a partir de Espaços Multidimensionais, o de características a partir de Teoria dos Conjuntos e o de redes com Teoria dos Grafos. Espaços Multidimensionais e Teoria dos Grafos possuem respectivamente, o ferramental para mensurar distância entre pontos e nodos. Teoria dos Conjuntos para manipular coleções.

O formato dos dados de entrada para o modelo de redes exige que exista uma estrutura envolvendo os dois conceitos, já o modelo geométrico e o de características trabalham apenas com o conceito e suas propriedades. O contexto para o modelo de redes é dado pela

própria estrutura taxonômica, para o modelo de características de Tversky, o contexto é definido por dois parâmetros, já o modelo geométrico considera como contexto apenas as características dos objetos envolvidos, podendo ser mais restritivo.

A lógica descritiva é capaz de estruturar seu conhecimento e, além disso, permite que, dado um conceito, sejam obtidas as suas características. A partir disso, a construção de um modelo que considere os três modelos apresentados de modo a integrar a distância estrutural, a distância de características e as características comuns e distintas dos conceitos para avaliação de similaridade é válida.

Como forma de avaliar os resultados obtidos com o mecanismo de similaridade, pode ser usada abordagem quantitativa ou qualitativa. Em 3.2.1 e 3.2.2 são apresentadas, respectivamente, estas duas abordagens.

3.1.5 Abordagem quantitativa

Esta forma de avaliação de resultados caracteriza-se pelo emprego da quantificação tanto na coleta quanto no tratamento das informações através de técnicas estatísticas. Como por exemplo, percentual, média, desvio padrão, coeficiente de correlação, etc (RICHARDSON, 1985).

Uma estratégia clássica para mensurar similaridade é a transformação dos objetos do domínio em pontos de um espaço de coordenadas, e assim, usar a distância euclidiana como métrica de similaridade. A validação empírica desse modelo foi experimentalmente desafiada, resultando em algumas divergências quanto a esse método. Mesmo assim, em situações como similaridade entre cores, a métrica é utilizada.

São exemplos dessa abordagem: medição de similaridade entre imagens (CHEN; CHU, 2005), onde os autores testam a distância euclidiana, a distância de *Mahalanobi* e a distância de *Chord* para medir similaridade; A abordagem proposta por (JIAN; SANTINI, 1999), é uma abordagem híbrida, construída a partir de lógica Fuzzy para mensurar quantitativamente, e o modelo de Tversky para a parte qualitativa.

3.1.6 Abordagem qualitativa

Ao contrário do método quantitativo, o método qualitativo não utiliza a estatística como base no processo de análise de um problema. É

utilizada em situações complexas ou estritamente particulares. Abordagens qualitativas visam descrever a complexidade de determinado problema, analisar a interação de certas variáveis, compreender e classificar processos dinâmicos (RICHARDSON, 1985).

Objetos são compostos por um conjunto de características, essas podem ser sobre aparências, funções e relações com outros objetos. Sob uma perspectiva qualitativa, tomando por base o trabalho de Tversky (1977) como foi descrito em 3.1.1, objetos são vistos como estímulos, que nada mais são do que a generalização de objetos. A avaliação qualitativa busca, através da avaliação das características, avaliar a qualidade do resultado.

3.2 Trabalhos Relacionados

3.2.1 Explorando métricas, para descoberta de serviços web semânticos, baseadas em similaridade (DIETZE, et al. 2009).

Os serviços web semânticos (SWS) utilizam descrições semânticas que podem ser interpretadas por máquinas para a descoberta automática e para fornecer o serviço apropriado para uma dada requisição.

A similaridade semântica não é uma noção implícita dentro de representações SWS, porém é utilizada na mediação de nível semântico entre diferentes ontologias de serviços web semânticos. O Artigo investiga um mecanismo de mediação baseado em similaridade para superar a necessidade da formalização manual ou semi-automática de um mapeamento um-para-um entre representações distintas de SWS. É proposta uma abordagem de mediação de propósito geral constituída de:

- 1- uma aproximação representacional permitindo implicitamente representar similaridade;
- 2- um mediador de propósito geral para mediação em nível semântico, explorando similaridade através de como foi representado em 1.

É introduzido o conceito de *Mediation Spaces* (MS), que permitem representações implícitas de similaridade semântica entre representações heterogêneas de serviços web semânticos, para isso, as descrições dos SWS são fundamentadas a partir de espaços vetoriais. A mediação visa resolver heterogeneidades existente entre distintos SWS e assim apoiar todos estágios que ocorrem na execução de SWS. O problema da mediação pode ser classificado em dois níveis:

- 1- Nível semântico: lida com a resolução de heterogeneidades entre representações semânticas de serviços concorrentes,
- 2- Nível de dados: faz a mediação entre incompatibilidades relacionais as implementações do serviço web.

O espaço de mediação propõe a representação dos conceitos que são usados como parte das descrições dos SWS como espaços vetoriais geométricos multidimensionais. Instâncias são representadas como vetores, ou seja, pontos em um espaço de mediação, onde a similaridade entre os dois vetores é indicada pela distancia espacial entre eles. A similaridade semântica entre dois membros de um espaço é definida como uma função da distância euclidiana entre pontos representando cada membro.

3.2.2 Um modelo híbrido de medida de similaridade conceitual para o ambiente ontológico (DONG, et al. 2009)

Muitos modelos de similaridade semântica são designados para redes semânticas e para redes de definição (subconjunto das redes semânticas composta apenas pelo relacionamento entre classe e subclasse, ou simplesmente relacionamento "é um". O artigo propõe um modelo híbrido de medida de similaridade considerando o aspecto baseado na medida de similaridade do conteúdo do conceito e o aspecto da medida de similaridade baseada na estrutura da ontologia. Para aplicar o modelo proposto a ontologia precisa ser convertida em um espaço ontológico leve. Isso é feito a partir da transformação dos conceitos da ontologia em pseudo-conceitos, onde os pseudo-conceitos só podem ser ligados pela relação "é um", as demais relações dos conceitos são armazenadas em seus pseudo-conceitos.

O modelo proposto envolve dois submodelos: O primeiro mede a similaridade baseada no conteúdo dos pseudo-conceitos através da abordagem de *Latent Semantic Indexing* (LSI). O segundo submodelo mede a similaridade baseada na estrutura do grafo da ontologia leve por meio de uma abordagem baseada no modelo de medida de similaridade de tópicos. Os produtos desses dois submodelos são duas matrizes conceito: a) conceito que são integrados para obter uma nova matriz conceito, b) conceito que indica a extensão da similaridade entre conceitos.

3.2.3 SIM-DLa : Uma nova medida de similaridade semântica para lógicas descritivas reduzindo similaridade entre conceitos a similaridade entre instâncias (JANOWICZ, et al. 2009).

Para melhorar a interação homem-computador, as medidas de similaridade precisam respeitar dois critérios na recuperação de informação:

- 1- Ser capaz de lidar com a expressividade da lógica descritiva utilizada.
- 2- A classificação da similaridade produzida por essas medidas precisa ser correlata à medidas de similaridades feitas pelos humanos.

O principal desafio para a medição de similaridade semântica é a comparação de significados, sendo oposta a uma comparação puramente estrutural (sintática). Padrões genéricos formam um framework de como similaridade entre conceitos é medida, formado por sete passos:

- 1- Definição de uma área de aplicação e público alvo.
- 2- Seleção da busca (query) e conceitos alvos.
- 3- Transformação dos conceitos para a forma canônica.
- 4- Definição de uma matriz de alinhamento para os descritores dos conceitos.
- 5- Aplicação de funções de construção específicas de similaridade.
- 6- Determinação de similaridade global padronizada.
- 7- Interpretação dos valores de similaridade resultante.

A forma proposta pela SIM-DLa reduz similaridades entre conceitos a similaridades entre instâncias para evitar o problema da canonização (o que é dito, e não como é dito) nas medidas puramente estruturais, através da geração de modelos *proxy* usando o algoritmo de tableau modificado e seguindo os sete passos.

3.2.4 Uma Medida de Similaridade Semântica para Expressivas Lógicas Descritivas (D'AMATO, et al. 2009).

Os autores desenvolveram uma medida de similaridade para lógicas descritivas que pode avaliar similaridade entre conceitos, entre instâncias ou entre instância e conceito. Para o modelo, esses elementos são chamados de objetos. A medida considera as características comuns dos objetos e avalia similaridade no intervalo $[0,1]$, sendo que, quanto

mais próximo de 1 mais similar os objetos são. A função de similaridade é exibida (4).

$$s(C, D) = \frac{|I^I|}{|C^I| + |D^I| - |I^I|} \cdot \max\left(\frac{|I^I|}{|C^I|}, \frac{|I^I|}{|D^I|}\right) \quad (4)$$

onde C e D são conceitos e $I = C \sqcap D$.

4 ESTUDO DE CASO

O conhecimento e o raciocínio descritos nessa pesquisa serão utilizados por agentes. Nesta seção serão apresentados o ambiente, e os elementos que utilizarão, direta ou indiretamente, o mecanismo de similaridade.

4.1 Agentes

A definição do que é um agente ainda é motivo para debates, não existe uma definição universal do termo. Segundo (WOOLDRIDGE,2002) um agente é um sistema computacional que está situado em algum ambiente, e que é capaz de ações autônomas nesse ambiente visando atingir seus objetivos de projeto. Para (RUSSEL; NORVIG, 2003) agente é qualquer entidade capaz de perceber seu ambiente por meio de sensores e agir sobre esse ambiente por intermédio de atuadores.

Em (WOOLDRIDGE; JENNINGS, 1995), o termo agente é usado de duas maneiras:

- 1 Uma noção fraca: é um hardware ou um sistema de computador baseado em software que: é autônomo (opera sem intervenção direta de humanos ou outros agentes), sociável (interage com outros agentes, ou humanos, através de alguma linguagem de comunicação de agentes), reativo (percebe o ambiente e reagem rapidamente às mudanças que ocorrem nele) e proativo (comportamento direcionado a objetivos).
- 2 Uma noção forte: Além de possuir as propriedades da noção fraca, agentes possuem características comuns a humanos, tais como crenças, desejos, obrigações.

4.2 Sistemas Multiagentes

Segundo (JENNINGS; SYCARA; WOOLDRIDGE, 1998) um sistema multiagente pode ser definido como um conjunto de agentes autônomos e provavelmente heterogêneos fracamente acoplados que interagem entre si para resolver um problema. As características de um sistema multiagente são:

- i) Cada agente possui informação incompleta, ou ainda, capacidade incompleta para resolver o problema;

- ii) Não existe um sistema de controle global;
- iii) Dados são descentralizados;
- iv) Computação é assíncrona.

Por exemplo, (JENNINGS; SYCARA; WOOLDRIDGE) em uma tarefa de monitoramento de veículos distribuídos, os agentes estão distribuídos geograficamente e cada um é capaz de detectar uma parte da área global. Os veículos se movimentam nas áreas monitoradas pelos agentes. Cada agente detecta os sons característicos dos veículos que passam pela sua região de monitoramento, ao analisar a combinação de sons ouvidos no seu local específico, o agente pode desenvolver interpretações de quais veículos poderiam ter criado esses sons. Utilizando o conhecimento sobre as restrições de mobilidade de veículos diferentes, o agente pode gerar mapas preliminares de movimentos de veículos em sua área. Ao comunicar seu mapa com o de outro agente, é possível construir mapas com maior fidelidade, e evitar rastreamento redundante nas regiões sobrepostas.

4.3 FIPA

A FIPA (*Foundation for Intelligent Physical Agents*) é uma organização da IEEE *Computer Society*, que promove a tecnologia baseada em agentes. Ela define um conjunto de padrões para a interoperabilidade tanto entre agentes quanto com outras tecnologias. (FIPA, 2013).

Os principais padrões especificados pela FIPA são o gerenciamento de agentes, uma arquitetura abstrata para os agentes e uma linguagem para comunicação. O modelo para gerenciamento de agentes é composto por agentes, um agente DF, um sistema gerenciador dos agentes, um sistema para transporte de mensagem, plataforma de agentes e software. Na plataforma fica localizado o agente DF que será o objeto de estudo deste trabalho.

A especificação do modelo de referência está na figura 6. As entidades contidas no modelo são serviços e não implicam em nenhuma configuração física. Além disso, os detalhes de implementação de Plataformas de Agentes (AP) e dos agentes, são escolhas de projeto dos desenvolvedores de sistemas de agentes.

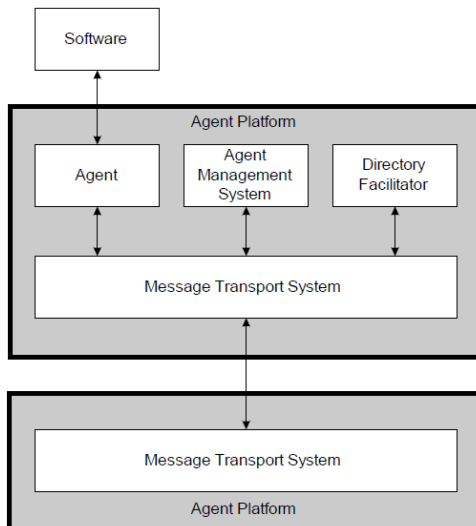


FIGURA 06 Modelo de Referência do Gerenciamento de Agentes

O modelo de referência do gerenciamento de agentes que é apresentado na figura 06 consiste nos seguintes componentes lógicos:

- **Agente (Agent):** processo computacional que implementa a funcionalidade autônoma e comunicativa de uma aplicação.
- **Facilitador de Diretório (Directory Facilitator - DF):** É um componente opcional em uma AP, mas base no estudo de caso. O DF fornece serviços de páginas amarelas para outros agentes. Agentes que desejam publicar seus serviços para que outros agentes possam usar, devem informar o serviço para o DF. Da mesma forma, quando um agente estiver procurando um serviço, deve contatar o DF para descobrir se alguém tem o serviço desejado.
- **Sistema Gerenciador de Agentes (Agent Management System – AMS):** Toda AP possui exatamente um AMS. Entidade responsável pelo gerenciamento das funcionalidades de uma AP, tais como: criação e remoção de agentes; verificação de migração e emigração de agentes na plataforma.
- **Sistema para Transporte de Mensagem (Message Transport System - MTS):** Método de comunicação padrão entre agentes em diferentes PAs.

- **Plataforma de Agentes** (*Agent Platform*): Fornece a infraestrutura física na qual os agentes podem ser implantados. Estar em uma AP não significa estar na mesma máquina local.
- **Software**: Descreve todas as coleções de instruções executáveis acessíveis do agente, mas que não façam parte do mesmo.

4.4 JADE

JADE (*Java Agent DEvelopment Framework*, 2013), é um *software* livre sobre a licença LGPL, distribuído pela Telecom Italia, que visa a simplificação da implementação de sistemas multiagentes através de um mediador (*middleware*) que respeita as especificações da FIPA.

Em uma plataforma de agentes JADE, existem *containers* de agentes que podem ser, ou não, distribuídos pela rede. Na plataforma de agentes existe um *container* principal, que é o primeiro a ser inicializado, e é com ele que todos os outros são registrados. Além disso, o *container* principal contém o agente Facilitador de Diretórios (DF) e o Sistema Gerenciador de Agentes (AMS), tal como especificado na FIPA. A figura 7 mostra containers, plataformas, o DF, o AMS, e agentes A1, A2, A3, A4 e A5 (BELLIFEMINE, 2007)

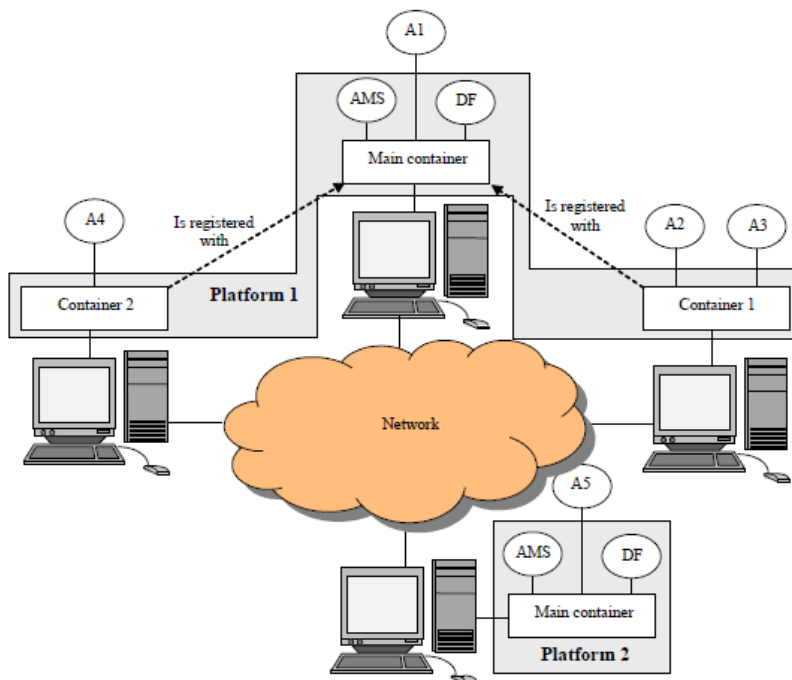


FIGURA 07 Containers e plataformas (BELLIFEMINE, 2007).

4.4.1 O serviço de páginas amarelas

Um agente pode publicar os serviços que ele fornece. Com isso, quando outro agente precisar de um serviço, ele pode consultar um catálogo com serviços publicados, as “páginas amarelas”, para descobrir se alguém fornece a tarefa desejada. Segundo a especificação FIPA, e na implementação JADE, o catálogo das páginas amarelas é fornecido pelo agente DF.

4.5 OWL-S

Nesta seção será apresentada uma ontologia OWL específica para serviços. Os serviços que agentes fornecem implementarão essa ontologia.

Serviços Web são sistemas de *software* projetados para suportar a interação com outras máquinas em uma rede (W3C, 2004). Quem fornece o serviço deve publicar um registro em algum lugar, para que clientes possam localizar o serviço de forma eficiente (MAHMOUD,

2008). Os serviços podem ser publicados através de uma ontologia em um agente DF. Assim, existem agentes que publicam serviços no DF e agentes que buscam serviços no DF.

A *Web Ontology Language for Services* (OWL-S) é uma ontologia de topo (*Upper Ontology*), ou seja, uma ontologia para descrever conceitos gerais. O contexto de uso da OWL-S é a implementação da oferta de serviços na web semântica. Por serviço, entendem-se *sites* web que, não apenas fornecem informações estáticas, mas também permitem fazer alguma ação ou alteração no mundo, como por exemplo, a venda de um produto, ou o controle de um dispositivo físico (W3C, 2004).

Os três componentes principais da OWL-S são: o perfil de serviço (*service profile*) para descrever os serviços que ele fornece, o modelo do serviço (*service model*) para descrever o funcionamento do serviço, e uma base para o serviço (*service grounding*) que detalha como o serviço pode ser acessado. Essa ontologia busca a descoberta, a invocação, a composição e a interoperação automática de serviços na web.

- **Perfil de serviço**

O perfil é adequado para possibilitar a descoberta de serviços (KLUSCHA, 2009), uma vez que ele descreve o que é realizado pelo serviço oferecido, quais as limitações na aplicabilidade e qualidade deste serviço e quais são os requisitos para a execução do serviço, com sucesso.

Um perfil OWL-S descreve um serviço como uma função composta por três tipos básicos de informação: qual organização, ou agente, fornece o serviço; qual função que o serviço computa; e os recursos que especificam as características do serviço.

- **Modelo de serviço**

Um serviço pode ser visto como um processo, não somente sob o ponto de vista computacional, em que um processo é um programa em execução, mas também como uma especificação das maneiras como um cliente deve interagir com um serviço. De acordo com esse ponto de vista, processos podem ter dois propósitos: gerar e retornar alguma nova informação baseada na informação recebida e no estado do mundo; e

podem, eventualmente, produzir também uma mudança no mundo (W3C, 2004).

Processos podem ser atômicos ou compostos. Os atômicos consistem em uma descrição de serviço que recebe uma mensagem e retorna outra como resposta. Os compostos mantêm algum estado e, a cada mensagem que o cliente envia, o processo avança.

- **Base para o serviço** (*grounding*)

Especifica os detalhes de como acessar o serviço: protocolos, formato das mensagens, serialização, transporte e endereço. A base (*grounding*) pode ser vista como uma forma de mapeamento das especificações abstratas (obtidas a partir do *ServiceProfile* e *ServiceModel*) para as especificações concretas. Sua função principal é mostrar de forma concreta como as mensagens são trocadas.

5 DESENVOLVIMENTO

Este capítulo se divide em dois momentos, no primeiro é apresentado a função de similaridade, e no segundo a implementação desta e sua utilização na recomendação de serviços em agentes.

5.1 Modelo de similaridade proposto

O modelo visa utilizar a ideia de distância semântica para definir o contexto e a utilização das características comuns e distintas para avaliar a similaridade. A distância semântica será definida a partir da distância geométrica normalizada (1 quando os dois conceitos forem completamente distintos e 0 caso sejam iguais), somada por um parâmetro da estrutura taxonômica. A similaridade será avaliada através do modelo relacional de Tversky, com o contexto dado pela descrição anterior.

Para exemplificação do modelo, serão considerados o mesmo TBox e os mesmos dois conceitos utilizados no trabalho apresentado em 3.3.4 (D'AMATO, et al. 2009). Os conceitos são Pai e AvôOuAvó, e o TBox é dado abaixo.

Conceitos primitivos: $N_C = \{\text{Feminino, Masculino, Humano}\}$.

Papeis primitivos: $N_R = \{\text{TemFilhoOuFilha, TemPaiOuMãe, TemAvôOuAvó, TemTio}\}$

TBox $\mathcal{T} = \{$
 Mulher \equiv Humano \sqcap Feminino
 Homem \equiv Humano \sqcap Masculino
 PaiOuMãe \equiv Humano \sqcap \exists TemFilhoOuFilha.Humano
 Mãe \equiv Mulher \sqcap PaiOuMãe
 Pai \equiv Homem \sqcap PaiOuMãe
 FilhoOuFilha \equiv Humano \sqcap \exists TemPaiOuMãe.PaiOuMãe
 AvôOuAvó \equiv PaiOuMãe \sqcap \exists TemFilhoOuFilha.
 (\exists TemFilhoOuFilha.Humano)
 IrmãoOuIrmã \equiv FilhoOuFilha \sqcap \exists TemPaiOuMãe.
 (\exists TemFilhoOuFilha \geq 2)
 Sobrinha \equiv Humano \sqcap \exists TemAvôOuAvó.PaiOuMãe \sqcup
 \exists TemTio.Tio

$$\begin{aligned} \text{PrimoOuPrima} &\equiv \text{Sobrinha} \sqcap \exists \text{TemTio.} \\ (\exists \text{TemFilhoOuFilha.Humano}) \\ \} \end{aligned}$$

5.1.1 Definição da função base

Para o modelo proposto, a função de Tversky será reescrita da seguinte forma (5):

$$s(C, D) = \frac{|C \sqcap D|}{|C \sqcap D| + \alpha * |C - D| + \beta * |D - C|} \quad (5)$$

Onde C e D são conceitos, \sqcap é o operador de interseção de lógica descritiva, $|C \sqcap D|$ é a cardinalidade da interseção dos conceitos, $|C - D|$ e $|D - C|$ resultam na cardinalidade da diferença do primeiro com o segundo.

Substituindo C e D por, respectivamente, Pai e AvôOuAvó:

$$\begin{aligned} \text{Pai} &\equiv \text{Homem} \sqcap \text{PaiOuMãe} \\ \text{AvôOuAvó} &\equiv \text{PaiOuMãe} \sqcap \exists \text{TemFilhoOuFilha.} \\ (\exists \text{TemFilhoOuFilha.Humano}) \end{aligned}$$

Ambos os conceitos podem ser reescritos da seguinte forma:

$$\begin{aligned} \text{Pai} &\equiv \text{Humano} \sqcap \text{Masculino} \sqcap \text{Humano} \sqcap \\ \exists \text{TemFilhoOuFilha.Humano} \\ \text{AvôOuAvó} &\equiv \text{Humano} \sqcap \exists \text{TemFilhoOuFilha.Humano} \sqcap \\ \exists \text{TemFilhoOuFilha.} (\exists \text{TemFilhoOuFilha.Humano}) \end{aligned}$$

Assim:

$$\begin{aligned} |\text{Pai} \sqcap \text{AvôOuAvó}| &= |\text{Humano} \sqcap \exists \text{TemFilhoOuFilha.} \\ \text{Humano}| &= 2 \\ |\text{Pai} - \text{AvôOuAvó}| &= |\text{Masculino}| = 1 \\ |\text{AvôOuAvó} - \text{Pai}| &= |\emptyset| = 0 \end{aligned}$$

5.1.2 Definição do contexto

Como dito em 3.1.1, quanto menores forem α e β , menor será o peso para as características distintas. A estratégia para determinar esses fatores é conciliar distância estrutural e não estrutural. O valor estrutural será determinado a partir do modelo de Wu e Palmer, e o valor não estrutural a partir da distância euclidiana.

No modelo de Wu e Palmer (1995), a similaridade é avaliada no intervalo de $[0,1]$. Quanto mais similares dois objetos são, mais próximo de 1 é o valor da similaridade. Esse valor não é apropriado para o α e β . Por exemplo, assumindo os conceitos Mulher e Humano da figura 05, é provável que a similaridade de acordo com esse modelo seja alta. Ao utilizar o resultado da comparação entre Mulher e Humano como α e β , o que estaria sendo maximizado são as características incomuns, ou seja, o comum estaria sendo utilizado em favor do incomum. Porém assumindo que tudo que não é similar, é dissimilar, é possível adaptar a função de modo a utilizar a dissimilaridade (6) como parte de α e β . Assim:

$$dissim(C1, C2) = 1 - \frac{2*N3}{N1+N2+2*N3} \quad (6)$$

Utilizando distância euclidiana, quanto menor a distância entre os conceitos, menos características incomuns eles tem se encaixando assim, no propósito do α e β . Porém, a utilização da distância euclidiana não normalizada resultaria em uma menor importância para a representação estrutural, assim para a determinação do α e β a distância euclidiana será normalizada (7), além disso, essa normalização faz com que todas as características envolvidas sejam consideradas no cálculo, tornando a medida menos restritiva, uma vez que, no formato anterior, a distância entre dois conceitos que possuem 10 características em comum e uma incomum, e dois conceitos que possuem 1 característica em comum e 1 incomum seria a mesma.

$$d_{normalizada} = \frac{d(c1,c2)}{\sqrt{n}} \quad (7)$$

Sendo n o número total de características distintas entre os dois conceitos. Para determinar a distância euclidiana $d(c1,c2)$, como mostrado em 3.1.2, os conceitos $c1$ e $c2$ precisam ser representados em um espaço multidimensional. Cada característica distinta é uma dimensão, e possuir a característica representada por uma certa dimensão é estar localizado na posição 1 e não possuir é estar na posição 0. A função em (8) calcula a distância euclidiana entre dois conceitos.

$$d(c1, c2) = \sqrt{\sum_{i=1}^n |c1_i - c2_i|} \quad (8)$$

onde $c1_i$ e $c2_i$ são os valores de $c1$ e $c2$ para a característica i , 1 caso possua, zero caso contrário.

Assim(9):

$$\alpha = \beta = \left(1 - \frac{2*N3}{N1 + N2 + 2*N3}\right) + \frac{d(c1, c2)}{\sqrt{n}} \quad (9)$$

Retomando o exemplo dos conceitos *Pai* e *AvôOuAvó*, para determinar o valor da primeira fração é necessário construir o grafo dos relacionamentos, este é apresentado na figura 08.

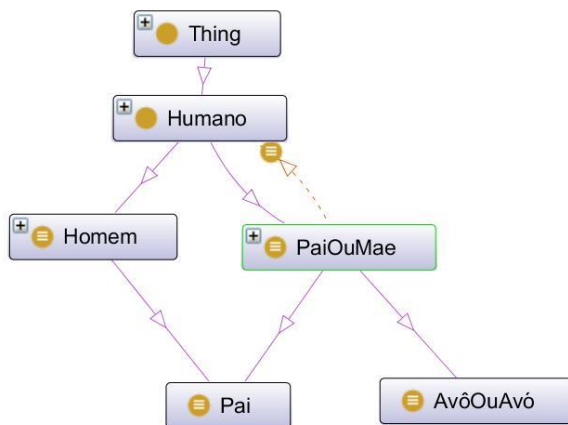


FIGURA 08 Grafo em OWL para o exemplo do modelo.

O primeiro superconceito entre *Pai* e *AvôOuAvó* é *PaiOuMae*, o número de nós $N1$ entre *Pai* e *PaiOuMae* é 1, o número $N2$ entre *AvôOuAvó* e *PaiOuMae* é 1, e o número $N3$ entre *PaiOuMae* e a raiz é 2.

Assim,

$$1 - \frac{2*N3}{N1 + N2 + 2*N3} = \frac{2*2}{1 + 1 + 2*2} = 0,666$$

Determinando $d(c1, c2)$:

Características distintas = $n = \{\text{Masculino, Humano, } \exists \text{TemFilhoOuFilha.Humano}\} = 3$.

Como apenas uma é exclusiva entre os dois conceitos, a distância pode ser reduzida a:

$$d(\text{Pai}, \text{AvôOuAvó}) = \sqrt{1} = 1$$

$$d_{\text{normalizada}} = \frac{1}{\sqrt{3}}$$

Com isso:

$$\alpha = \beta = \frac{1}{\sqrt{3}} + 0,666 \cong 1,2433$$

5.1.3 Exemplo

Com os dados obtidos, basta substituir os valores na função para determinar a similaridade entre Pai e AvôOuAvó :

$$s(C, D) = \frac{|C \cap D|}{|C \cap D| + \alpha * |C - D| + \beta * |D - C|} = \frac{2}{2 + 1 * 1,2433 + 0 * 1,2433} \cong 0,616$$

Comparando com o modelo de (D'AMATO; FANIZZI; ESPOSITO, 2009):

$$s(C, D) = \frac{|C \cap D|}{|C| + |D| - |C \cap D|} \cdot \max\left(\frac{|C \cap D|}{C}, \frac{|C \cap D|}{D}\right) =$$

$$\frac{2}{2 + 3 - 2} \cdot \max\left(\frac{2}{3}, \frac{2}{2}\right) = 0,666$$

No capítulo de testes são apresentados mais comparações entre os modelos. Reescrevendo o modelo de similaridade proposto, tem-se:

$$s(C, D) = \frac{|C \cap D|}{|C \cap D| + \alpha * |C - D| + \beta * |D - C|}$$

$$\alpha = \beta = \left(1 - \frac{2 * N3}{N1 + N2 + 2 * N3}\right) + \frac{d(c1, c2)}{\sqrt{n}}$$

5.2 Implementação

Nesta seção serão descritas brevemente as implementações da função de similaridade e a do estudo de caso.

5.2.1 Implementação do modelo de similaridade

Como visto anteriormente, a OWL é baseada em lógica descritiva, além disso, ela é uma linguagem para ontologias no contexto da Web Semântica. A partir disso, a aplicação do mecanismo de similaridade proposto no contexto da OWL, facilita tanto a avaliação com o agente DF, quanto futuros trabalhos relacionados a similaridade no contexto da Web Semântica.

A implementação foi construída com o auxílio de três ferramentas: OWL API, Pellet e Jena. A OWL API é a interface de programação de aplicativos (API) para criar e manipular as ontologias. O Pellet é implementado sobre OWL API, atuando como motor de inferência da ontologia. O Jena é uma API para ontologias, trabalha tanto com OWL quanto com RDFS, e sua utilização não seria necessária, porém ela possui alguns métodos que facilitam a implementação do mecanismo de similaridade.

A instanciação do mecanismo e invocação do método de similaridade é apresentada na figura 09.

```
// endereço da ontologia base
String ontology = getOntology();
// MySimilarityMeasure contém uma instância do Pellet e
// uma árvore com nodos representando conceitos.
MySimilarityMeasure _reasoner = new MySimilarityMeasure(ontology);
...
//conceitos da ontologia para avaliação de similaridade
String concept1 = getConcept1();
String concept2 = getConcept2();

//obtem o valor de similaridade em uma escala de 0 a 1
double sim = _reasoner.getSimilarity(concept1, concept2);
```

FIGURA 09 instanciação do mecanismo de similaridade e invocação do método que calcula similaridade.

O método invocado pelo usuário para obter a similaridade é apresentado na figura 10.

```
public double getSimilarity(String sConcept1, String sConcept2) {

    // MyConcept é uma estrutura que contém todas as características do
    // conceito (super classes, classes equivalentes, classes disjuntas )
    MyConcept    concept1    =    new    MyConcept(
model.getOntClass(sConcept1));
    MyConcept    concept2    =    new
MyConcept(model.getOntClass(sConcept2));
    //retorna a similaridade entre os conceitos segundo o modelo proposto
    return similaridade(concept1, concept2);
}
```

FIGURA 10 método invocado pelo usuário para obter a similaridade.

A implementação do método privado que calcula similaridade é apresentada na figura 11.

```
private double similaridade(MyConcept conceito1, MyConcept conceito2) {
    //obtem o valor do contexto. Determinado pela distância do modelo de
    //Wu e Palmer e pela distância euclidiana normalizada
    _alphaBeta = obterAlphaBeta(conceito1, conceito2);
    double retorno = 0.0;
    int a=0, b=0, ab=0;

    //obtem o número de axiomas exclusivos do conceito1 salvando em a
    a = conceito1.obterNAxiomasExclusivos(conceito2);
    // obtem o número de axiomas exclusivos do conceito1 salvando em b
    b = conceito2.obterNAxiomasExclusivos(conceito1);
    //obtem o número de axiomas que conceito1 e conceito2 compartilham
    ab = conceito1.obterNAxiomasCompartilhados(conceito2);
    // retorna o valor obtido com a função de Tversky.
    return ab/(ab+_alphaBeta*a +_alphaBeta*b);
}
```

FIGURA 11 método que calcula similaridade.

5.2.2 Implementação do estudo de caso

O registro e a busca no facilitador de diretórios utilizando o mecanismo de similaridade serão descritos a seguir. Para registrar um serviço no DF é utilizada uma ontologia de acordo com a OWL-S. Para

o exemplo, e testes, é utilizada a ontologia fornecida pela *OWL-S Coalition*, (MARTIN, 2012) sendo que o grafo da figura 12 representa os conceitos de serviços que poderão ser fornecidos.

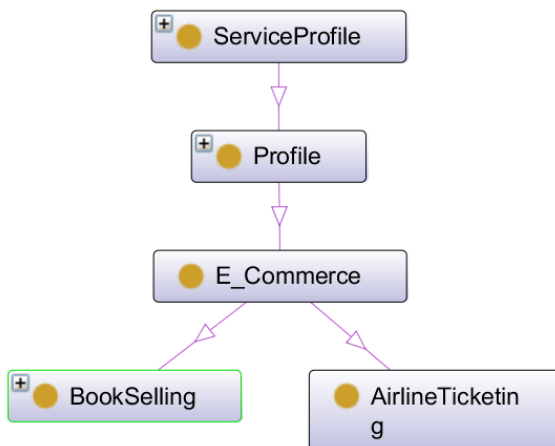


FIGURA 12 Segmento do grafo conceitual de uma ontologia de serviço.

5.2.2.1 Publicando serviços

Para publicar um ou mais serviços, um agente precisa fornecer ao DF o seu AID (identificador do agente), e, possivelmente, uma lista de linguagens e ontologias que outros agentes precisam conhecer para interagir com ele. Para cada serviço publicado, uma descrição é fornecida, contendo nela: o tipo do serviço (deve ser o conceito da ontologia de serviço), o nome do serviço, as linguagens e ontologias necessárias para explorar o serviço.

A figura 13 é apresentada a implementação de um agente que irá registrar um serviço de *BookSelling* no facilitador de diretórios.

```

// criação da descrição de um agente para o catálogo do DF
DFAgentDescription dfd = new DFAgentDescription();
// definindo que a descrição é a do agente que quer registrar o serviço
dfd.setName(getAID());
//criação da descrição de um serviço, tipo de dados do serviço
ServiceDescription sd = new ServiceDescription();
//definição do nome do serviço
sd.setName("My Book Selling Service");
  
```



```

//aqui é definido o tipo, este deve ter o mesmo nome do conceito no Profile da
//OWL-S
sd.setType("BookSelling");
//agentes que quiserem usar esse serviço precisam conhecer a ontologia de
//serviço
sd.addOntologies("/home/thiago/TCC/service.owl");
//agentes que quiserem usar esse serviço precisam "falar" a linguagem FIPA-SL
sd.addLanguages(FIPANames.ContentLanguage.FIPA_SL);
//Adição do serviço na descrição pro DF
dfd.addServices(sd);
// registrando o serviço do agente "this" no DF
DFService.register(this, dfd);

```

FIGURA 13 implementação de um agente que fornece o serviço de BookSelling no DF.

5.2.2.2 Buscando serviços

Para buscar um serviço, um agente precisa fornecer ao DF uma descrição do serviço que está buscando e os dados referentes às restrições para busca, tais como número máximo de resultados e porcentagem mínima de similaridade. O DF busca em sua base pelo serviço, e se for solicitado, por serviços similares utilizando o modelo proposto nesta pesquisa.

Por exemplo, supondo que um agente precise de *E_Commerce* e que serviços com similaridade maior que 60% a *E_Commerce* podem ser aceitos desde que a lista de fornecedores de serviço não seja maior que dez. A figura 14 apresenta uma implementação de um agente que envia para o DF a descrição acima. O resultado obtido é separado em serviços que são *E_Commerce* e serviços similares.

```

// criação da descrição de um agente para busca no catálogo do DF
DFAgentDescription template = new DFAgentDescription();
// criação da descrição do serviço a ser buscado
ServiceDescription templateSd = new ServiceDescription();
// tipo (conceito) a ser buscado
templateSd.setType("E_Commerce");
// adição do serviço na descrição
template.addServices(templateSd);
// classe que modela as restrições de busca
SearchConstraintsWithSimilarity sc = new SearchConstraintsWithSimilarity();
// limita a busca a no máximo 10 resultados
sc.setMaxResults(new Long(10));
//limita o mínimo de similaridade que pode ter (em porcentagem).

```

```

sc.setMinSimilarity(60);
//armazena em "results" os serviços buscados
DFAgentDescription[] results = DFService.search(this, template, sc);
if (results.length > 0) {
    System.out.println("Agent "+getLocalName()+" found the following
services:");
    //percorre todos os serviços encontrados
    for (int i = 0; i < results.length; ++i) {
        //cria uma descrição para cada resultado
        DFAgentDescription dfd = results[i];
        //obtem o identificador do agente que fornece o serviço
        AID provider = dfd.getName();
        //um agente pode fornecer mais de um serviço, e só estamos
        //interessados nos que satisfazem as restrições
        Iterator it = dfd.getAllServices();
        while (it.hasNext()) {
            ServiceDescription sd = (ServiceDescription)
it.next();
            double sim = DFService.getSimilarity(
sd.getType(), "E_Commerce");
            if (sd.getType().equals(" BookSelling"))
                System.out.println("- Service \"\" +
sd.getName() + "\" provided by agent "+provider.getName());
            else if( sim >= sc.getMinSimilarity())
                System.out.println("- Service \"\" +
sd.getName() + "\" provided by agent "+provider.getName()+ "with similarity
of "+ sim);
        }
    }
}

```

FIGURA 14 implementação de um agente que busca serviços no DF.

Assumindo que o único agente que registrou serviços foi o *BookSelling*. Quando o agente for buscar por *E_Commerce*, ele não encontrará o serviço exato. Então utilizará o mecanismo de similaridade e buscará por serviços similares. Ao verificar *BookSelling*, o mecanismo encontrará similaridade maior que a restrição, e assim, adicionará a lista de retorno *BookSelling*. Após percorrer todos os registros, retornará para o agente que está buscando serviços, a lista com o serviço *BookSelling*.

6 TESTES

6.1 Testes do modelo de similaridade

Para avaliar o modelo, será utilizado o TBox abaixo, o mesmo é utilizado no modelo de (D'AMATO, et al. 2009). Com isso será possível comparar as avaliações dos dois modelos. Para simplificar, só será utilizado como referência, o sobrenome do autor principal. As comparações feitas serão entre os conceitos: Pai e Mãe; Pai e Mulher; FilhoOuFilha e Mulher; FilhoOuFilha e AvôOuAvó; Mulher e Feminino; PaiOuMãe e Pai; PaiOuMãe e AvôOuAvó, e o feito no exemplo da seção 5.1.3 Pai e AvôOuAvó.

```
TBox  $\mathcal{T} = \{$ 
  Mulher  $\equiv$  Humano  $\sqcap$  Feminino
  Homem  $\equiv$  Humano  $\sqcap$  Masculino
  PaiOuMãe  $\equiv$  Humano  $\sqcap$   $\exists$ TemFilhoOuFilha.Humano
  Mãe  $\equiv$  Mulher  $\sqcap$  PaiOuMãe
  Pai  $\equiv$  Homem  $\sqcap$  PaiOuMãe
  FilhoOuFilha  $\equiv$  Humano  $\sqcap$   $\exists$ TemPaiOuMãe.PaiOuMãe
  AvôOuAvó  $\equiv$  PaiOuMãe  $\sqcap$   $\exists$ TemFilhoOuFilha.
  ( $\exists$ TemFilhoOuFilha.Humano)
  IrmãoOuIrmã  $\equiv$  FilhoOuFilha  $\sqcap$   $\exists$ TemPaiOuMãe.
  ( $\exists$ TemFilhoOuFilha  $\geq 2$ )
  Sobrinha  $\equiv$  Humano  $\sqcap$   $\exists$ TemAvôOuAvó.PaiOuMãe  $\sqcup$ 
   $\exists$ TemTio.Tio
  PrimoOuPrima  $\equiv$  Sobrinha  $\sqcap$   $\exists$ TemTio.
  ( $\exists$ TemFilhoOuFilha.Humano)
 $\}$ 
```

Comparação entre Pai e Mãe:

Pai \equiv Humano \sqcap Masculino \sqcap \exists TemFilhoOuFilha.Humano

Mãe \equiv Humano \sqcap Feminino \sqcap \exists TemFilhoOuFilha.Humano

| Pai \sqcap Mãe | = 2

| Pai - Mãe | = 1

| Mãe - Pai | = 1

Modelo Proposto:

$$\alpha = \beta = \left(1 - \frac{2*2}{1+1+2*2}\right) + \frac{\sqrt{2}}{\sqrt{4}} \cong 0,33 + 0,707 \cong 1,04$$

$$s(\text{Pai}, \text{Mãe}) = \frac{2}{2+1*1,04+1*1,04} \cong 0,49$$

Modelo de D'AMATO:

$$s(\text{Pai}, \text{Mãe}) = \frac{2}{3+3-2} \cdot \max\left(\frac{2}{3}, \frac{2}{3}\right) \cong 0,33$$

Comparação entre Pai e Mulher:

Pai \equiv Humano \sqcap Masculino $\sqcap \exists$ TemFilhoOuFilha.Humano
 Mulher \equiv Humano \sqcap Feminino

$$| \text{Pai} \sqcap \text{Mulher} | = 1$$

$$| \text{Pai} - \text{Mulher} | = 2$$

$$| \text{Mulher} - \text{Pai} | = 1$$

Modelo Proposto:

$$\alpha = \beta = \left(1 - \frac{2*1}{1+2+2*1}\right) + \frac{\sqrt{3}}{\sqrt{4}} \cong +1,466$$

$$s(\text{Pai}, \text{Mulher}) = \frac{1}{1+2*1,466+1*1,466} \cong 0,185$$

Modelo de D'AMATO:

$$s(\text{Pai}, \text{Mulher}) = \frac{1}{3+2-1} \cdot \max\left(\frac{1}{3}, \frac{1}{2}\right) \cong 0,125$$

Comparação entre FilhoOuFilha e Mulher:

FilhoOuFilha \equiv Humano $\sqcap \exists$ TemPaiOuMãe.PaiOuMãe
 Mulher \equiv Humano \sqcap Feminino

$$| \text{FilhoOuFilha} \sqcap \text{Mulher} | = 1$$

$$| \text{FilhoOuFilha} - \text{Mulher} | = 1$$

| Mulher - FilhoOuFilha | = 1

Modelo Proposto:

$$\alpha = \beta = \left(1 - \frac{2*1}{1+1+2*1}\right) + \frac{\sqrt{2}}{\sqrt{3}} \cong 0,5 + 0,816 \cong 1,316$$

$$s(\text{FilhoOuFilha}, \text{Mulher}) = \frac{1}{1+1*1,316 + 1*1,316} \cong 0,275$$

Modelo de D'AMATO:

$$s(\text{FilhoOuFilha}, \text{Mulher}) = \frac{1}{2 + 2 - 1} \cdot \max\left(\frac{1}{2}, \frac{1}{2}\right) \cong 0,166$$

Comparação entre FilhoOuFilha e AvôOuAvó:

FilhoOuFilha \equiv Humano \sqcap \exists TemPaiOuMãe.PaiOuMãe
 AvôOuAvó \equiv Humano \sqcap \exists TemFilhoOuFilha.Humano

| FilhoOuFilha \sqcap AvôOuAvó | = 1
 | FilhoOuFilha - AvôOuAvó | = 1
 | AvôOuAvó - FilhoOuFilha | = 1

Modelo Proposto:

$$\alpha = \beta = \left(1 - \frac{2*1}{1+2+2*1}\right) + \frac{\sqrt{2}}{\sqrt{3}} \cong 0,6 + 0,816 \cong 1,41$$

$$s(\text{FilhoOuFilha}, \text{AvôOuAvó}) = \frac{1}{1+1*1,41 + 1*1,41} \cong 0,26$$

Modelo de D'AMATO:

$$s(\text{FilhoOuFilha}, \text{AvôOuAvó}) = \frac{1}{2 + 2 - 1} \cdot \max\left(\frac{1}{2}, \frac{1}{2}\right) \cong 0,166$$

Comparação entre Mulher e Feminino:

Feminino
 Mulher \equiv Humano \sqcap Feminino

| Feminino \sqcap Mulher | = 1
 | Feminino - Mulher | = 0

| Mulher - Feminino | = 1

Modelo Proposto:

$$\alpha = \beta = \left(1 - \frac{2*1}{1+0+2*1}\right) + \frac{\sqrt{1}}{\sqrt{2}} \cong 0,33 + 0,7 \cong 1,03$$

$$s(\text{Feminino, Mulher}) = \frac{1}{1+0*1,03+1*1,03} \cong 0,49$$

Modelo de D'AMATO:

$$s(\text{Feminino, Mulher}) = \frac{1}{2+2-1} \cdot \max\left(\frac{1}{2}, \frac{1}{1}\right) = 0,5$$

Comparação entre PaiOuMãe e Pai:

PaiOuMãe \equiv Humano \sqcap \exists TemFilhoOuFilha.Humano

Pai \equiv Humano \sqcap Masculino \sqcap \exists TemFilhoOuFilha.Humano

| PaiOuMãe \sqcap Pai | = 2

| PaiOuMãe - Pai | = 0

| Pai - PaiOuMãe | = 1

Modelo Proposto:

$$\alpha = \beta = \left(1 - \frac{2*2}{1+0+2*2}\right) + \frac{\sqrt{1}}{\sqrt{3}} \cong 0,2 + 0,57 \cong 0,77$$

$$s(\text{PaiOuMãe, Pai}) = \frac{2}{2+0*0,77+1*0,77} \cong 0,722$$

Modelo de D'AMATO:

$$s(\text{PaiOuMãe, Pai}) = \frac{2}{2+3-2} \cdot \max\left(\frac{2}{2}, \frac{2}{3}\right) \cong 0,66$$

Comparação entre PaiOuMãe e AvôOuAvó:

PaiOuMãe \equiv Humano \sqcap \exists TemFilhoOuFilha.Humano

AvôOuAvó \equiv Humano \sqcap \exists TemFilhoOuFilha.Humano

$$\begin{aligned}
 | \text{PaiOuMãe} \sqcap \text{AvôOuAvó} | &= 2 \\
 | \text{PaiOuMãe} - \text{AvôOuAvó} | &= 0 \\
 | \text{AvôOuAvó} - \text{PaiOuMãe} | &= 0
 \end{aligned}$$

Modelo Proposto:

$$s(\text{PaiOuMãe}, \text{Pai}) = \frac{2}{2} = 1$$

Modelo de D'AMATO:

$$s(\text{PaiOuMãe}, \text{Pai}) = \frac{2}{2 + 2 - 2} \cdot \max\left(\frac{2}{2}, \frac{2}{2}\right) = 1$$

O resumo das comparações de similaridade são apresentados na tabela 3.

Conceito 1	Conceito 2	Similaridade Modelo Proposto	Similaridade Modelo D'AMATO
Pai	Mãe	0,492	0,333
Pai	PaiOuMãe	0,722	0,666
Pai	AvôOuAvó	0,616	0,666
Pai	Mulher	0,185	0,125
FilhoOuFilha	Mulher	0,275	0,166
FilhoOuFilha	AvôOuAvó	0,261	0,166
Mulher	Feminino	0,492	0,5
AvôOuAvó	PaiOuMãe	1	1

TABELA 03 Comparação entre avaliações de similaridade do modelo proposto com o modelo de D'AMATO.

Como o modelo de D'AMATO avalia similaridade de forma não estruturada, para duas avaliações de similaridade que possuem conceitos correspondentes, o valor de similaridade será o mesmo. Por exemplo, os conceitos FilhoOuFilha, Mulher e AvôOuAvó podem ser reescritos como:

$$\begin{aligned}
 \text{FilhoOuFilha} &\equiv \text{Humano} \sqcap \exists \text{TemPaiOuMãe.PaiOuMãe} \\
 \text{Mulher} &\equiv \text{Humano} \sqcap \text{Feminino} \\
 \text{AvôOuAvó} &\equiv \text{Humano} \sqcap \exists \text{TemFilhoOuFilha.Humano}
 \end{aligned}$$

Assim, a avaliação de similaridade entre FilhoOuFilha e Mulher e entre FilhoOuFilha e AvôOuAvó, resultarão no mesmo valor (0,166). Já no modelo proposto, ao determinar os parâmetros α e β é considerada a estrutura da representação, com isso, a similaridade entre duas avaliações correspondentes, não necessariamente, resultarão no mesmo valor. Para o caso entre FilhoOuFilha e Mulher e entre FilhoOuFilha e AvôOuAvó, as similaridades são, respectivamente, 0,275 e 0,261.

Uma comparação conceitual interessante é entre AvôOuAvó e PaiOuMãe. Para ambos os modelos, o valor de similaridade é igual a 1, uma vez que as características dos conceitos são as mesmas, o que muda é o domínio para o qual a regra TemFilhoOuFilha é aplicada.

6.2 Testes com OWL

Os testes com OWL foram divididos em dois, primeiro são apresentados testes só com ontologias OWL e em seguida com agentes utilizando ontologias.

6.2.1 Testes com ontologias

Para os testes com OWL, foram utilizados alguns conceitos de duas ontologias. A tabela 04 apresenta a similaridade entre alguns conceitos da ontologia wine.owl.

Conceito 1	Conceito 2	Similaridade
WineBody	WineTaste	0,72
WineBody	WineFlavor	0,49
WineFlavor	Region	0
WineDescriptor	WineTaste	0,49
WineDescriptor	WineFlavor	0,27

TABELA 04 similaridade entre conceitos OWL na ontologia wine com modelo proposto

Na figura 15 é apresentada a estrutura envolvendo os conceitos comparados. Em relação aos dois primeiros testes WineBody é mais similar a sua generalização WineTaste, do que a seu irmão, WineFlavor. WineFlavor e Region não possuem características em comum, de modo que α e β obtiveram o valor máximo (2), além disso como a interseção de característica é vazia, a similaridade deu 0. WineDescriptor é mais similar a seu primeiro descendente, WineTaste, do que a seu segundo descendente WineFlavor.

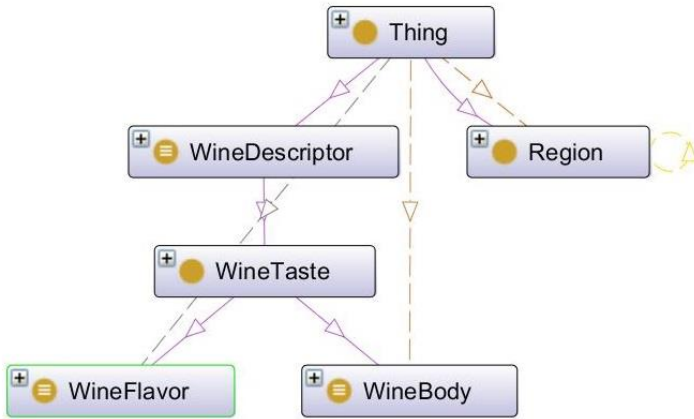


FIGURA 15 grafo parcial da ontologia *wine*.

Para a determinação da similaridade entre os conceitos da tabela 05, foi utilizada a ontologia *food.owl*. Como o número de características e a distância estrutural é a mesma na comparação entre Pasta e Shellfish e Pasta e Fish, o valor de similaridade foi o mesmo. O mesmo ocorre entre PastaWithWhiteSauce e Seafood. A figura 16 exibe a estrutura dos conceitos comparados da tabela 05.

Conceito 1	Conceito 2	Similaridade
Pasta	Shellfish	0,28
Pasta	Fish	0,28
PastaWithWhiteSauce	Seafood	0,28
Fish	Shellfish	0,51
PastaWithWhiteSauce	PastaWithRedSauce	0,51

TABELA 05 similaridade entre conceitos OWL na ontologia *food* com modelo proposto.

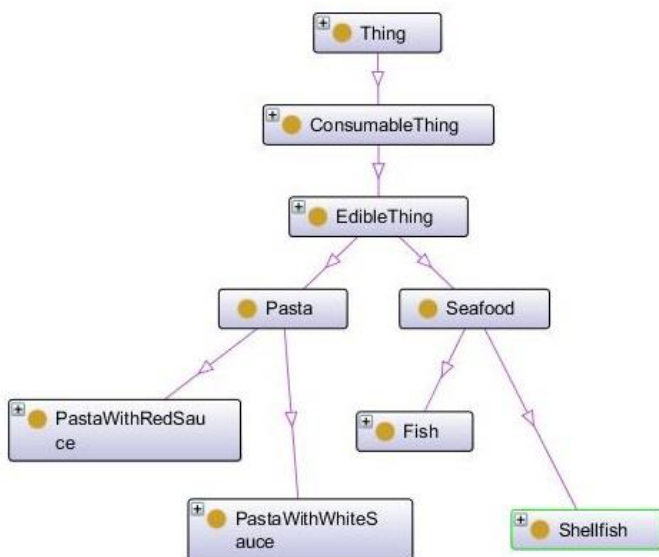


FIGURA 16 grafo parcial da ontologia *food*.

6.2.2 Testes com agentes

Os testes com agentes foram feitos a partir da plataforma JADE, uma vez que ela fornece todas as abstrações para trabalhar com o padrão FIPA. Além do facilitador de diretórios, foram adicionados na plataforma dois agentes para registrar serviços e um para buscar.

Os dois serviços registrados foram aqueles contidos no exemplo da OWL-S *Coalition*: venda *online* de passagens aéreas (*AirlineTicketing*) e venda *online* de livros (*BookSelling*). Sendo que cada um dos dois agentes cadastrados na plataforma para fornecer serviços registrou um.

O número de fornecedores de serviço que o DF retorna para o agente que está buscando depende do valor definido para similaridade. A tabela 06 apresenta a similaridade entre os dois serviços, e entre cada serviço com o seu ancestral em comum (*E_Commerce*).

Serviço 1	Serviço 2	Similaridade
BookSelling	E_Commerce	0,659
AirlineTicketing	E_Commerce	0,659
BookSelling	AirlineTicketing	0,427

TABELA 06 Similaridade entre serviços (conceitos)

Com isso, várias situações podem ocorrer. Por exemplo:

Se o agente estiver buscando pelo serviço AirlineTicketing ou no mínimo 70% similares a ele, só encontrará o serviço de AirlineTicketing. Caso a busca seja por E_Commerce ou pelo menos 60% similares a ele, retornarão BookSelling e AirlineTicketing.

Se busca for por BookSelling e serviços com similaridade superior a 60%, não retornará nenhum serviço, uma vez que, embora E_Commerce seja 65% similar, não existe nenhum serviço registrado como E_Commerce.

6 CONSIDERAÇÕES FINAIS

Um conceito em lógica descritiva pode ser representado a partir da combinação entre conceitos e restrições do domínio e operadores da linguagem. A expressividade da representação varia de acordo com a LD utilizada. Para representar conhecimento, foi utilizada nessa pesquisa a OWL, uma família de linguagens construídas a partir de LD's. Por isso, o ferramental disponível para LD, tais como o modelo de similaridade proposto, também é válido para OWL.

Sabendo que similaridade é obtida através do julgamento humano, torna-se difícil a definição de um modelo único e universalmente aceito para avaliação de similaridade. Para desenvolver o modelo de similaridade, essa pesquisa analisou diversas maneiras de medição de similaridade. Dentre os trabalhos analisados, foram encontrados modelos que trabalham com: estrutura da representação, distância semântica, características comuns e distintas, métricas, além de modelos híbridos, como por exemplo, o modelo proposto.

O modelo de similaridade definido nessa pesquisa foi construído de forma gradativa. Inicialmente, tomou-se como base a função relacional de Tversky, por ser base em outras pesquisas e, porque os parâmetros α e β poderiam ser determinados a partir de testes para se encaixar no contexto de recomendação de serviços do DF. Num segundo momento, surgiu a ideia de utilizar distância semântica para determinar os parâmetros α e β , uma vez que, quanto mais distantes os conceitos forem, provavelmente menos características em comum eles terão. A distância entre os conceitos foi determinada através do balanceamento da distância estrutural e não estrutural. A especificação dos parâmetros α e β foi aprimorada baseando-se em ontologias em que as versões anteriores tinham resultados insatisfatórios.

Ao comparar o modelo proposto com o de D'AMATO, observou-se que o definido nesta pesquisa obteve em 6 dos 8 casos similaridade maior ou igual ao proposto por D'AMATO. Em um dos dois casos em que avaliou como menos similar, a diferença é inferior a 1%. No outro caso, comparando P_{ai} e $AvôOuAvó$, constatou-se que, para o modelo de D'AMATO, a similaridade entre estes conceitos é a igual a similaridade entre P_{ai} e $P_{ai}OuMãe$, já no modelo proposto, P_{ai} é mais similar a $P_{ai}OuMãe$ do que a $AvôOuAvó$. Isso ocorre pelo fato de que o modelo proposto considera a estrutura da representação.

A implementação do mecanismo de avaliação de similaridade utilizou o motor de inferência de código aberto Pellet, e algumas classes

da API JENA. A implementação da plataforma de agentes segundo o modelo FIPA e com a equação de similaridade foi feita a partir do mediador JADE. Existem alguns problemas relacionados à implementação, porém a recomendação de serviços similares pôde ser avaliada.

O padrão do facilitador de diretórios é buscar os serviços que satisfazem exatamente as condições definidas na consulta. O mecanismo de similaridade desenvolvido é uma ferramenta para o facilitador de diretórios. Assim, o DF proposto, além de fornecer a lista padrão, pode fornecer serviços similares. Ou fornecer apenas estes serviços, quando o serviço buscado não existir. Por exemplo, quando foi buscado pelo serviço não prestado E_Commerce, retornaram os serviços de BookSelling e AirlineTicketing por respeitarem a restrição de similaridade e por existir um agente que registrou o serviço.

6.1 Trabalhos futuros

Devido à abrangência deste trabalho, itens interessantes não foram aprofundados:

- Avaliar o mecanismo de similaridade proposto em outros contextos, como por exemplo, a WordNet;

- Transformar o mecanismo de similaridade em um operador de linguagem;

 - Avaliar o comportamento do modelo ao trabalhar com o ABox;

 - Considerar as especializações dos conceitos envolvidos no cálculo de similaridade;

 - Avaliar similaridade assimétrica;

 - Avaliar comportamento em relação ao modelo de Redes.

REFERÊNCIAS

- BAADER, Franz; MCGUINNESS; Deborah L.; NARDI, Daniele; PATEL-SCHNEIDER, Peter F., **The Description Logic Handbook - Theory, Implementation and Applications**. 2003.
- BELLIFEMINE, Fabio Luigi; CAIRE, Giovanni; GREENWOOD, Dominic. **developing multi-agent systems with JADE**. John Wiley & Sons, Ltd. 2007.
- BERNERS LEE, Tim; HENDLER, James; LASSILA, Ora. **The Semantic Web**. Scientific American, [S.1.], v.284, n.5,p. 35-43, 2001.
- BRACHMAN, Ronald J.; LEVESQUE, Hector J.; **Knowledge Representation And Reasoning**. 2003. Disponível em: <<http://rair.cogsci.rpi.edu/pai/library/brachmanbook7-17-03.pdf>>. Acesso Dez/13.
- CHEN, Chaur-Chin; CHU, Hsueh-Ting; The 29th Annual International Computer Software and Applications Conference (COMPSAC'05), 2005 IEEE.
- DAVIS, Randall; SHROBE, Howard; SZOLOVITS, Peter. **What Is a Knowledge Representation?** AI Magazine. Volume 14 Number 1. 1993 (©AAAI)
- DIETZE, Stefan; GUGLIOTTA, A. ; DOMINGUE, J. **Exploiting Metrics for Similarity-based Semantic Web Service Discovery**. Web Services, 2009. ICWS 2009.
- DONG, Hai; HUSSAIN, Farookh Khadeer; CHANG, Elizabeth. **A Hybrid Concept Similarity Measure Model for Ontology Environment**. Springer Berlin Heidelberg: 2009.
- D'AMATO, Claudia; FANIZZI, Nicola; ESPOSITO, Floriana. **A Semantic Similarity Measure for Expressive Description Logics**. CoRR Nov 2009.
- D'AMATO, Claudia; STAAB, Steffen; FANIZZI, Nicola. **On the Influence of Description Logics Ontologies on Conceptual Similarity**. 2008.

D'OTTAVIANO, Itala M. Loffredo. **Non-Classical Logics and Applications**. 04-2009

GERSTING, Judith L., Tradução: Valéria de Magalhães Iório – **Fundamentos Matemáticos Para a Ciência da Computação**. 2004.
GOODMAN, Nelson; **Problems and Projects**. The Bobbs-Merrill Company, INC. Indianapolis and New York:1972.

GRUBER, Thomas R. **A translation approach to portable ontology specifications**. 1993.

IEEE Foundation for Intelligent Physical Agents, 2013. Disponível em: <<http://www.fipa.org/specs/fipa00023/XC00023H.html>>. Acesso Jan/13

IEEE Foundation for Intelligent Physical Agents, 2013. Disponível em: <<http://www.fipa.org/>>. Acesso Jan/13

JANOWICZ, Krzysztof; WILKES, Marc. **SIM-DLA: A Novel Semantic Similarity Measure for Description Logics Reducing Inter-concept to Inter-instance Similarity**. Springer Berlin Heidelberg: 2009.

JENNINGS, Nicholas R.; SYCARA, Katia; WOOLDRIDGE, Michael; **A Roadmap of Agent Research and Development: Autonomous Agents and Multi-Agent Systems**, Ed.: 1, (1998): Kluwer Academic Publishers, Boston. Manufactured in The Netherlands. Pag: 7–38.

JIAN; SANTINI; **Similarity Measures**. IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, Vol. 21, N°. 9, Setembro/1999.

KLUSCHA, Matthias; FRIESB, Benedikt; SYCARAC, Katia. **OWLS-MX: A hybrid Semantic Web service matchmaker for OWL-S services**. Web Semantics: Science, Services and Agents on the World Wide Web Volume 7, Issue 2, April 2009, Pages 121–133

LIU, Ling, ÖZSU, M. Tamer. **In the Encyclopedia of Database Systems**, Springer-Verlag, 2009. Disponível em:

<<http://tomgruber.org/writing/ontology-definition-2007.htm>>. Acesso Nov/12.

NORVIG, P.; RUSSELL, S., **Artificial Intelligence (Second Edition)**, 2004.

Ontologia Wine. Disponível em <<http://protege.cim3.net/file/pub/ontologies/wine/wine.owl>>. Acesso em Dez/13.

Ontologia Food. Disponível em <<http://www.w3.org/TR/2003/CR-owl-guide-20030818/food.owl>>. Acesso em Dez/13.

OWL Web Ontology Language Guide: W3C Recommendation: 10 February, 2004. Disponível em: <<http://www.w3.org/TR/owl-guide/>>. Acesso Jul/13.

OWL 2 Web Ontology Language Document Overview: W3C Recommendation 27 October, 2009. Disponível em: <<http://www.w3.org/TR/2009/REC-owl2-overview-20091027/>>. Acesso em Jul/13.

OWL 2 Web Ontology Language Document Overview (Second Edition): W3C Recommendation, 11 December, 2012. Disponível em: <<http://www.w3.org/TR/owl2-overview/#Semantics>>. Acesso em Jul/13.

QUILLIAN, Ross M.; **Word concepts: A theory and simulation of some basic semantic capabilities.** 1967.

RICHARDSON, Roberto Jarry. **Pesquisa Social Métodos e Técnicas.** São Paulo, SP: Atlas, 1985.

SÁNCHEZ, David; BATET, Montserrat; ISERN, David; VALLS, Aida. **Ontology-based semantic similarity: A new feature-based approach.** Expert Systems with Applications 39, 2012.

SCHMIDT-SCHAUB, Manfred; SMOLKA, Gert; **Attributive concept descriptions with complements.** Artificial Inteligente, Elsevier, 1991.

SCHWERING Angela. **Approaches to Semantic Similarity Measurement for Geo-Spatial Data: A Survey**. Blackwell Publishing, 2008.

SOWA, John F.; **Knowledge Representation: Logical, Philosophical, and Computational Foundations**. Pacific Grove, CA: Brooks/Cole, 2000.

SOWA, John F. **Semantic Networks**. Disponível em: <<http://www.jfsowa.com/pubs/semnet.htm>>. Acesso em outubro/13.

TVERSKY, Amos; **Features of Similarity**: Hebrew University, Jerusalém, Israel:1977.

WOOLDRIDGE, Michael; JENNINGS, Nicholas R., **Intelligent Agents: Theory and Practice**. Cambridge University Press, 1995.

WU, Z., PALMER, M.. **Verb semantics and lexical selection**. In 32nd annual Meeting of the Association for Computational Linguistics (pp. 133–138). Las Cruces, New Mexico: Association for Computational Linguistics 1994.