

**UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE FILOSOFIA E CIÊNCIAS, CAMPUS DE MARÍLIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

TATIANE DOS SANTOS DE FREITAS LOPES

**ONTOLOGIA COMO INTERFACE DE APRESENTAÇÃO DE
RESULTADOS DE BUSCA:
uma proposta baseada no modelo espaço vetorial**

MARÍLIA - SP
2017

**UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE FILOSOFIA E CIÊNCIAS, CAMPUS DE MARÍLIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO**

TATIANE DOS SANTOS DE FREITAS LOPES

**ONTOLOGIA COMO INTERFACE DE APRESENTAÇÃO DE
RESULTADOS DE BUSCA:
uma proposta baseada no modelo espaço vetorial**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências - Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP, campus de Marília, como requisito parcial para obtenção do título de Mestre em Ciência da Informação.

ÁREA DE CONCENTRAÇÃO: Informação, Tecnologia e Conhecimento.

LINHA DE PESQUISA: Informação e Tecnologia.

ORIENTADOR: PROF. DR. EDBERTO FERNEDA

MARÍLIA – SP
2017

Lopes, Tatiane dos Santos de Freitas.

L864o Ontologia como interface de apresentação de resultados de busca: uma proposta baseada no modelo espaço vetorial / Tatiane dos Santos de Freitas Lopes. – Marília, 2017.
73 f. ; 30 cm.

Orientador: Edberto Ferneda.

Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista (Unesp), Faculdade de Filosofia e Ciências, 2017.

Bibliografia: f. 71-73

1. Recuperação da Informação. 2. Ontologias (Recuperação da informação). 3. Interfaces (Computador).
I. Título.

CDD 029.7

TATIANE DOS SANTOS DE FREITAS LOPES

**ONTOLOGIA COMO INTERFACE DE APRESENTAÇÃO DE
RESULTADOS DE BUSCA:
uma proposta baseada no modelo espaço vetorial**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Faculdade de Filosofia e Ciências - Universidade Estadual Paulista “Júlio de Mesquita Filho” – UNESP, campus de Marília, como requisito parcial para obtenção do título de Mestre em Ciência da Informação.

ÁREA DE CONCENTRAÇÃO: Informação, Tecnologia e Conhecimento.

LINHA DE PESQUISA: Informação e Tecnologia.

ORIENTADOR: PROF. DR. EDBERTO FERNEDA

Prof. Dr. Edberto Ferneda (orientador)
Faculdade de Filosofia e Ciências, UNESP-Marília

Profa. Dra. Maria José Vicentini Jorente
Faculdade de Filosofia e Ciências, UNESP-Marília

Profa. Dra. Luciana Maria Vieira Pöttker
Instituto Federal do Paraná – IFPR

Marília, 31 de agosto de 2017

*A minha filha Maria Fernanda, meu namorado e
minha família, "sobrinho(a)", pelo apoio e
dedicação na minha formação.*

Agradecimentos

Agradeço a todos que de alguma forma contribuíram para a realização desta pesquisa.



A Deus pela força para realização dos meus sonhos.



Aos meus pais pelo apoio, carinho e incentivo de sempre.



À minha filha pela paciência e compreender minha ausência no período de estudo.



À minha irmã pelo apoio e companheirismo em todos os momentos que preciso e por nunca me deixar faltar nada e ao meu sobrinho(a) que está a caminho.



Ao Fernando pela compreensão e incentivo para continuar sempre.



Ao meu querido orientador, professor Edberto Ferneda pelos ensinamentos, pela confiança e pelas palavras de incentivo e apoio, dizendo que tudo vai dar certo.



À professora Sueli Andruccioli Félix e ao professor Ricardo José Sabatine que incentivou e acreditou em mim na realização do Mestrado e pelo apoio profissional e pessoal.



Aos amigos e à família que de alguma forma compreenderam a minha ausência em muitos momentos.

“Cada sonho que você deixa para trás, é um pedaço
do seu futuro que deixa de existir”

Steve Jobs

Resumo

Um sistema de recuperação de informação é um elemento mediador entre um acervo documental e os usuários que buscam por documentos relevantes. Nesse contexto, as interfaces desempenham uma função importante: em um primeiro momento, auxiliando o usuário na tarefa de expressar a sua necessidade de informação por meio de uma expressão de busca e, em um segundo momento, fornecendo recursos para ajudá-lo a selecionar documentos relevantes dentre os resultados obtidos. A recuperação de informação é um processo linguístico cuja eficiência depende de coincidências terminológicas entre a expressão de busca do usuário e a representação dos documentos. Este trabalho propõe um modelo de interface na qual a estrutura terminológica de uma ontologia é utilizada para auxiliar o usuário na seleção de documentos relevantes dentre aqueles resultantes de sua busca. Caracteriza-se como uma pesquisa de natureza aplicada, e exploratória e bibliográfica quanto aos procedimentos. Conclui-se que a apresentação visual de uma ontologia permite o desenvolvimento de interfaces dinâmicas e interativas, proporcionando ao usuário uma navegação estimulante e prazerosa por entre os documentos resultantes de sua busca, tendo por base os termos de uma determinada área de conhecimento.

Palavras-chave: Recuperação de Informação; Interfaces de Recuperação de Informação; Visualização de Ontologia; Representação Visual de Ontologias

Abstract

An information retrieval system is a mediating element between a document collection and the users who looking for relevant documents. In this context, interfaces play an important role: firstly, assisting the user to expressing their information need by means of a search expression, and secondly by providing resources to help selecting relevant documents from the obtained results. The information retrieval is a linguistic process whose efficiency depends on terminological coincidences between the user's query and the representation of documents. This work proposes an interface model in which the terminological structure of an ontology is used to assist the user in the selection of relevant documents among those resulting from their search. It is characterized as an applied, exploratory and bibliographic research. It is concluded that the visual presentation of ontology allows the development of dynamic and interactive interfaces, providing the user with stimulating and pleasant navigation among the documents resulting from their search, based on the terms of a certain knowledge area.

Keywords: Information retrieval; Information Retrieval Interfaces; Ontology Visualization; Visual Representation of Ontologies

Lista de Figuras

Figura 1 - Representação do processo de recuperação de informação	19
Figura 2 - Página de resultados de pesquisa do Yahoo! Utilizando o termo "arco íris".	25
Figura 3 - Página de resultados do Google para uma busca sobre "foguetes".	26
Figura 4 - Exemplo de declaração dos namespaces em uma ontologia OWL	36
Figura 5 – Exemplo de declaração do cabeçalho de uma ontologia OWL.....	37
Figura 6 – Exemplo de definição de uma classe	38
Figura 7 – Exemplo de definição de uma classe usando blocos OWL.....	39
Figura 8 – Exemplo de definição de uma Hierarquia de classes em OWL	40
Figura 9 – Exemplo de definição de uma classe equivalente	41
Figura 10 – Exemplo de definição de uma classe disjunta.....	41
Figura 11 – Exemplo de definição de um indivíduo	42
Figura 12 – Exemplo de uma forma alternativa para a definição de um indivíduo	42
Figura 13 - Exemplo de propriedade de objetos.....	43
Figura 14 – Exemplo de propriedade de dados	44
Figura 15 – Exemplo de definição de uma hierarquia de propriedades	45
Figura 16 – Representação gráfica de classes e propriedades	45
Figura 17 - Visualização de ontologia em interface de lista indentada	48
Figura 18 – Visualização de ontologia em interface do tipo nó-link	50
Figura 19 - Visualização de ontologia em interface de zoom	51
Figura 20 - Visualização de ontologia em interface do tipo Space Filling.....	52
Figura 21 - Visualização de ontologia em interface do tipo Context+Focus	53
Figura 22 - Visualização de ontologia em interface do tipo Information Landscape.....	54
Figura 23 – Alguns elementos visuais da especificação VOWL.....	56
Figura 24 – Vocabulário Friend of a Friend (FOAF) visualizado com o VOWL	57
Figura 25 – Imagem do WebVOWL com a visualização do vocabulário FOAF.....	58
Figura 26 – Exemplo genérica de um documento e sua representação vetorial.....	60
Figura 27 – Exemplo da descrição vetorial de um documento.....	60
Figura 28 – Especificação de uma expressão de busca	61
Figura 29 – Ontologia OWL e sua representação gráfica	62
Figura 30 – Exemplo da Representação de Relevância.....	64
Figura 31 – Exemplo da especificação de uma expressão de busca.....	65
Figura 32 – Exemplo de apresentação dos documentos resultantes de uma busca	66
Figura 33 – Exemplo de apresentação dos detalhes de um documento.....	67

Lista de Quadros

Quadro 1 – Primitivas gráficas e esquema de cores do VOWL	55
Quadro 2 – Primitivas gráficas utilizadas no modelo proposto.....	62

Sumário

1. Introdução.....	14
1.1 Objetivos.....	15
1.1.1 Geral	15
1.1.2 Específicos	15
1.2 Metodologia.....	16
1.3 Da terminologia adotada.....	16
1.4 Organização do trabalho	17
2. Recuperação de Informação.....	18
2.1 Modelos de Recuperação de Informação.....	21
2.1.1 Modelo Booleano	21
2.1.2 Modelo Espaço Vetorial.....	22
2.1.3 Modelo Probabilístico	23
2.2 Interfaces de Resultados de Busca.....	24
3. Ontologia.....	29
3.1 Ontologias e a Ciência da Informação.....	32
3.2 Web Ontology Language (OWL)	34
3.2.1 Namespaces.....	36
3.2.2 Cabeçalhos	37
3.2.3 Classes.....	38
3.2.4 Indivíduos.....	42
3.2.5 Propriedades	43
4. Apresentação Gráfica de Ontologias.....	47
4.1 Métodos de Visualização de Ontologias.....	47
4.1.1 Indented List (lista indentada).....	47
4.1.2 Node-Link and Tree (nó-link e árvore).....	49
4.1.3 Zoomable Visualizations (visualizações com zoom).....	51
4.1.4 Space Filling (preenchimento de espaço)	52
4.1.5 Context+Focus and Distortion Techniques (contexto+foco e distorção).....	53
4.1.6 Information Landscapes (paisagens informacional)	53
4.2 VOWL	55

5.	Proposta de Interface de Resultados de Busca	59
5.1	Representação Vetorial de um Documento	59
5.2	Representação da Expressão de Busca	60
5.3	Representação Gráfica de uma Ontologia	61
5.4	Representação de Relevância	62
5.5	Modelo de interface de resultados de busca	64
5.5.1	Expressão de busca	65
5.5.2	Resultado da busca	66
5.5.3	Visualização da representação de um documento	67
6.	Conclusões.....	69
	Referências	71

1. Introdução

A quantidade crescente de informação disponível em meios digitais, principalmente na Web, faz aumentar a importância dos sistemas de recuperação de informação. Esses sistemas adquirem cada vez mais evidência em uma sociedade competente em produzir e disseminar informação, mas com dificuldade em encontrar informações relevantes e confiáveis.

A Recuperação de Informação é uma área que em mais de 50 anos de pesquisas produziu inúmeras ideias, conceitos e técnicas. Mas mesmo com um acelerado avanço tecnológico acontecido durante esse período, a busca por informações relevantes e úteis é ainda uma tarefa bastante difícil.

A recuperação de informação envolve, por um lado, um conjunto de documentos, geralmente representados por palavras ou termos cujos significados remetem aos assuntos tratados por eles. Por outro lado, pessoas (usuários) tentam traduzir as suas necessidades de informação por meio de um conjunto de termos de busca a fim de encontrar documentos que venham a satisfazer tais necessidades.

Um sistema de recuperação de informação é um elemento mediador entre um acervo de documentos e os usuários que buscam por documentos relevantes. O sistema deve oferecer um ambiente que auxilie o usuário na tarefa de especificar a sua busca e que provenha recursos para selecionar documentos relevantes entre os resultados obtidos. Nesse sentido, as interfaces desempenham um importante papel nos sistemas de recuperação de informação.

Considerando todo o processo de recuperação de informação, é possível identificar dois momentos em que o usuário realiza uma efetiva interação com o sistema. Em um primeiro momento, o usuário deve traduzir a sua necessidade de informação em um conjunto

de termos de busca que representem o tema ou assunto de seu interesse. Essa tarefa é fundamental em todo o processo de recuperação, podendo ser facilitada por diversos recursos do sistema e por uma interface eficiente.

Um segundo momento de interação do usuário no processo de recuperação de informação se dá durante a escolha de documentos relevantes dentre aqueles que foram recuperados. Nesse momento, a interface cumpre também papel importante, guiando o usuário na escolha dos documentos que realmente lhe serão úteis. Este trabalho restringe-se a esse segundo momento de interação, e propõe uma interface na qual a estrutura terminológica de uma ontologia de domínio é utilizada como base para a apresentação dos documentos resultantes de uma busca do usuário de um sistema de recuperação de informação.

Neste trabalho propõe-se uma inversão da lógica inerente às atuais ferramentas de recuperação de informação, nas quais os documentos são os elementos principais de suas interfaces de apresentação de resultados. Propõe-se nessa dissertação que a terminologia de uma determinada área do conhecimento tenha um papel preponderante em tais interfaces, auxiliando o usuário a identificar similaridades entre os documentos resultantes de uma busca por meio de suas relações conceituais, a partir de uma interface interativa, dinâmica.

1.1 Objetivos

1.1.1 Geral

Propor um modelo de interface para sistemas de recuperação de informação no qual uma ontologia de domínio é utilizada como estrutura básica de visualização dos resultados de busca.

1.1.2 Específicos

- Apresentar a área de Recuperação de Informação (*Information Retrieval*), seus principais modelos;
- Fazer um estudo sobre as principais propostas de interface para sistema de recuperação de informação, com ênfase nas interfaces de resultados de busca;
- Apresentar o conceito de ontologia e a forma como ele está sendo apropriação pela Ciência da Informação e Ciência da Computação;
- Realizar um levantamento sobre as formas de apresentação gráfica da estrutura terminológica de uma ontologia;

- Considerando o modelo vetorial de recuperação de informação, apresentar um modelo de interface para apresentação dos resultados de busca em um sistema de recuperação de informação.

1.2 Metodologia

Este trabalho caracteriza-se como uma pesquisa de natureza aplicada, pois objetiva gerar conhecimentos para aplicação prática, dirigidos à solução de problemas específicos. Quanto aos seus objetivos, caracteriza-se como exploratória e bibliográfica em relação aos procedimentos esta pesquisa.

Para Santos (2010, p.), a pesquisa exploratória é caracterizada pela existência de poucos dados disponíveis e tem como objetivo aprofundar e aperfeiçoar ideias. Portanto, a proposta deste trabalho não é desenvolver ferramentas computacionais, mas simular, por meio de exemplos, uma possível utilização da estrutura conceitual oferecida pelas ontologias na construção de interface de busca e recuperação de informação.

Severino (2007, p.122) afirma que pesquisa bibliográfica é aquela que se realiza a partir dos registros disponíveis, decorrentes de pesquisas anteriores publicadas em documentos impressos (livros, artigos, teses, entre outros). Os dados utilizados são aqueles já trabalhados e registrados por outros pesquisadores. Desse modo, o pesquisador trabalha a partir das contribuições dos autores de estudos analíticos constantes dos textos.

Nesse sentido, cabe notar que, para a realização deste trabalho, além da literatura pertinente à Recuperação de Informação, Ontologias e Interfaces, também coube avaliar, ainda que superficialmente, alguns sistemas de edição e apresentação de ontologias a fim subsidiar o desenvolvimento desta pesquisa

Embora não consista em uma pesquisa experimental, foi necessário o aprendizado de uma ferramenta de construção de ontologias (Protégé) e o conhecimento básico da linguagem de marcação OWL para conseguirmos descrever esses artefatos.

1.3 Da terminologia adotada

O tema principal deste trabalho, Recuperação de Informação, envolve pelo menos dois campos científicos: a Ciência da Informação e a Ciência da Computação. Surge daí problemas terminológicos decorrentes das diferentes nomenclaturas utilizadas para referenciar um mesmo conceito, além da tradução de termos geralmente originados na língua inglesa.

Principalmente na literatura nacional da área, nos deparamos com uma grande variedade de termos para aparentemente representar um mesmo conceito. Sem entrar em preciosismos terminológicos, neste trabalho utilizaremos os termos “motor de busca”, “mecanismo de busca”, “ferramenta de busca”, “buscador”, “sistema de recuperação de informação Web” ou qualquer outra variação, como sinônimos. Alguns termos em inglês foram traduzidos para o termo mais comum na área da Ciência da Informação, em português. Por exemplo, o termo “*search*” e o termo “*query*” são ambos traduzidos para o termo “busca”.

Considerando que este trabalho parte de interesses de investigação em Ciência da Informação, a terminologia utilizada será preferencialmente dessa área. Porém, alguns termos provenientes da Ciência da Computação já estão consolidados e são amplamente utilizados em diversos domínios científicos. Nesse caso a preferência recai sobre o termo mais comumente utilizado.

1.4 Organização do trabalho

Esta dissertação está organizada em sete capítulos. Neste capítulo (**Capítulo 1**) é apresentado uma breve contextualização do problema de pesquisa e os objetivos deste trabalho.

No **Capítulo 2** define-se “Recuperação de Informação” e são apresentados os elementos envolvidos no processo de recuperação de informação e os chamados modelos “clássicos”. Ao final desse capítulo é apresentada uma breve descrição das atuais interfaces dos principais mecanismos de busca.

O **Capítulo 3** apresenta o termo “ontologia”, conceituando seu significado para a Ciência da Informação e Ciência da Computação. Apresenta a linguagem para construção de ontologias OWL, explicando suas particularidades e elementos básicos.

O **Capítulo 4** apresenta uma breve apresentação e métodos e ferramentas de apresentação gráfica de ontologias, dando ênfase à VOWL, sistema no qual será baseado o modelo de interface objeto deste trabalho.

No **Capítulo 5** descreve-se o modelo proposto neste trabalho, seus elementos básicos e a forma de funcionamento da interação com o usuário.

Por fim, no **Capítulo 6** são apresentadas algumas considerações e conclusões deste trabalho.

2.

Recuperação de Informação

Em 1951, Calvin Mooers criou o termo “*Information Retrieval*” (Recuperação de Informação) e definiu os problemas a serem abordados por esta nova disciplina.

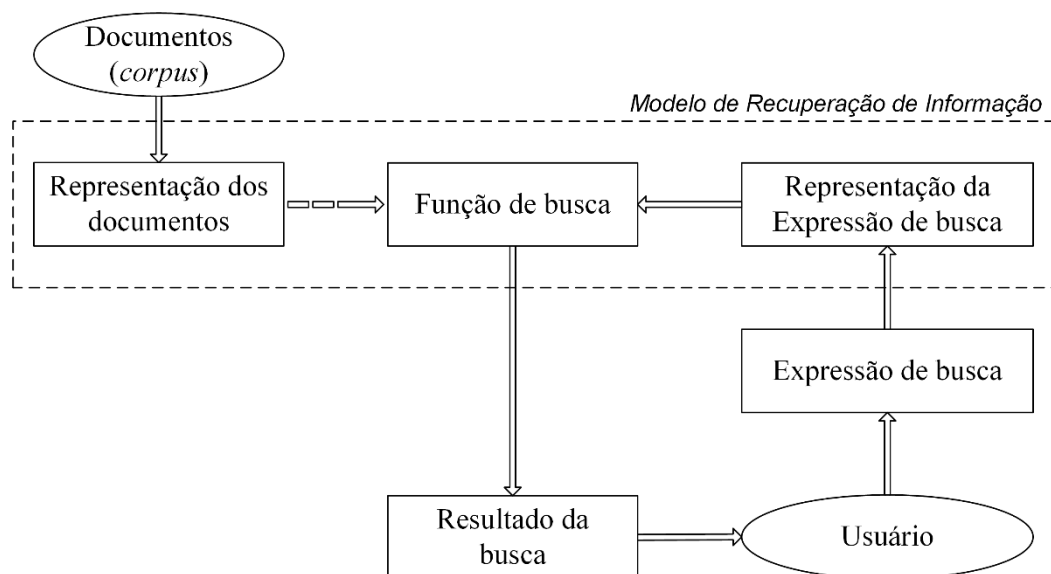
[A Recuperação de Informação] trata dos aspectos intelectuais da descrição da informação e sua especificação para busca, e também de qualquer sistema, técnicas ou máquinas que são empregadas para realizar esta operação (MOOERS, 1951).

Para Manning, Raghavan e Schütze (2009, p.38) o significado do termo “recuperação de informação” pode ser muito amplo:

Recuperação de Informação (RI) é encontrar material (geralmente documentos) de natureza não estruturada (geralmente textos) que satisfaçam as necessidades de informação a partir de grandes coleções (geralmente armazenadas em computadores).

A Figura 1 apresenta uma ilustração simplificada do processo de recuperação de informação.

Figura 1 - Representação do processo de recuperação de informação



Fonte: adaptado de Ferneda, 2012, p.14

Na Figura 1, a elipse superior representa um acervo documental. Para definirmos o que é “documento”, utilizaremos o trabalho de Buckland (1991), que define o conceito de “informação como coisa” e argumenta que os acervos disponíveis nos sistemas de informação seriam registros relacionados a coisas ou objetos. Nesses sistemas informação está vinculada ao objeto que a contém: um *documento*. O termo documento, entendido como coisa informativa, incluiria objetos, artefatos, textos, imagens, sons etc.

Independente dos tipos de documentos gerenciados por um sistema de informação, a eficiência da recuperação é dependente da forma como esses documentos estão representados. A representação de um documento tem por objetivo identificar e descrever resumidamente o seu conteúdo informacional, ao mesmo tempo em que define seus pontos de acesso para a busca em um sistema de informação. A tarefa de representar os documentos é feita em um tempo anterior à execução de qualquer busca. No esquema da Figura 1, a existência de uma seta seccionada tenta mostrar essa assincronia.

Em um sistema de recuperação de informação o usuário expressa sua necessidade de informação por meio de uma expressão de busca, composta geralmente por um conjunto de termos que representa linguisticamente a necessidade de informação do usuário. A principal dificuldade está em prever os termos foram usados para representar os documentos que satisfarão sua necessidade, e ao mesmo tempo evitar a recuperação de documentos não relevantes. Para isso é necessário que o usuário tenha um relativo conhecimento do vocabulário ou da terminologia do domínio de conhecimento de seu interesse.

Para que seja possível uma comparação entre a expressão de busca e cada um dos documentos do *corpus* é necessário que essa seja representada de forma similar à representação dos documentos. Diversos recursos podem ser oferecidos por um sistema a fim de facilitar o usuário na especificação de sua expressão de busca. Porém, essa consulta deve ser representada de uma forma idêntica ou similar à utilizada pelos documentos.

No centro do processo de recuperação de informação está a função de busca, que compara as representações dos documentos com a representação da expressão de busca e recupera os itens que supostamente fornecerão informações relevantes. De forma geral, a função de busca calcula o grau de similaridade entre a expressão de busca e cada um dos documentos do *corpus*. Na maioria dos sistemas, a similaridade é definida por um valor numérico que pretensamente indica o quão relevante é um documento para satisfazer a necessidade de informação do usuário. O resultado de uma busca é geralmente composto por uma lista de referências a documentos, ordenada pelo grau de similaridade calculada pela função de busca.

Um modelo de recuperação de informação é a especificação formal de três elementos: a *representação dos documentos*, a *representação da expressão de busca* e a *função de busca* (FERNEDA, 2012, p.20). De maneira mais formal, Baeza-Yates e Ribeiro-Neto (2011, p.58) definem modelo de recuperação de informação como uma quadrupla:

$$[\mathbf{D}, \mathbf{Q}, F, R(q_i, d_j)]$$

D é um conjunto composto por visões lógicas (representações) dos documentos no *corpus*;

Q é um conjunto composto de visões lógicas das necessidades de informação dos usuários;

F é um *framework* para a modelagem de representações dos documentos, consultas e seus relacionamentos;

$R(q_i, d_j)$ é uma função de ordenamento (*ranking*) que atribui um número real à relação entre uma representação da consulta q_i de **Q** e a representação de um documento d_j de **D**.

Apesar de alguns dos modelos de recuperação de informação terem sido criados entre as décadas de 1960 e 1970, as suas principais ideias ainda estão presentes na maioria dos sistemas de recuperação atuais e nos mecanismos de busca da Web.

2.1 Modelos de Recuperação de Informação

Na Figura 1 há três blocos circundados por uma linha tracejada. O conjunto formado por estes três elementos é chamado por Ferneda (2012) de Modelo de Recuperação de Informação.

Pode-se considerar que a “Representação dos documentos” é o resultado do processo de indexação. A “Representação da Expressão de busca” traduz as necessidades informacionais dos usuários e a “Função de busca” é responsável pela comparação entre as representações dos documentos e a representação de uma determinada expressão de busca.

2.1.1 Modelo Booleano

No Modelo Booleano (FERNEDA, 2012, cap.3) um documento é representado por um conjunto de termos de indexação. As buscas são formuladas por meio de uma expressão booleana composta por termos ligados através dos operadores lógicos AND, OR e NOT, e apresentam como resultado o conjunto de documentos cuja representação satisfaz as restrições lógicas da expressão de busca.

Uma expressão conjuntiva de enunciado t_1 **AND** t_2 recuperará documentos indexados por ambos os termos (t_1 e t_2). Uma expressão disjuntiva t_1 **OR** t_2 recuperará o conjunto dos documentos indexados pelo termo t_1 ou pelo termo t_2 . Uma expressão que utiliza apenas um termo t_1 terá como resultado o conjunto de documentos indexados por esse termo. A expressão **NOT** t_1 recuperará os documentos que não são indexados pelo termo t_1 . As expressões t_1 **NOT** t_2 ou t_1 **AND NOT** t_2 terão como resultado o conjunto dos documentos que são indexados por t_1 e que não são indexados por t_2 .

Os termos e operadores booleanos podem ser combinados para especificar buscas mais detalhadas ou restritivas. Como a ordem de execução das operações lógicas de uma expressão influencia no resultado da busca, muitas vezes é necessário explicitar essa ordem delimitando partes da expressão por meio de parênteses. A definição de expressões complexas exige um conhecimento profundo da lógica booleana. O conhecimento da lógica booleana é importante também para entender e avaliar os resultados obtidos em uma busca.

O modelo booleano possui diversas limitações. Algumas delas, apresentadas por Ferneda (2012, p.27) são:

- Não existe uma forma de atribuir importância relativa (pesos) aos termos de indexação dos documentos nem aos diferentes termos da expressão de busca. Assume-se implicitamente que todos os termos possuem o mesmo peso.
- O resultado de uma busca booleana se caracteriza por uma simples partição do *corpus* em dois subconjuntos: os documentos que atendem à expressão de busca e aqueles que não atendem. Presume-se que todos os documentos recuperados são de igual relevância, não havendo nenhum mecanismo pelo qual os documentos possam ser ordenados;
- Sem um treinamento apropriado, o usuário leigo será capaz de formular somente buscas simples. Para buscas que exijam expressões mais complexas é necessário um conhecimento sólido da lógica booleana.

Muitos sistemas se desenvolveram utilizando o modelo booleano como ponto de partida para o desenvolvimento de novos recursos de recuperação. Neste sentido o modelo booleano pode ser considerado o modelo mais utilizado nos sistemas de recuperação de informação e nos mecanismos de busca da Web.

2.1.2 Modelo Espaço Vetorial

O Modelo Espaço Vetorial (ou simplesmente Modelo Vetorial) propõe uma abordagem onde é possível obter documentos que respondem parcialmente a uma expressão de busca. Isto é possível associando valores de relevância diferentes aos termos tanto da indexação dos documentos quanto da expressão de busca formulada pelo usuário. Com o uso deste recurso é possível calcular a similaridade entre uma dada expressão de busca e cada um dos documentos do *corpus*, permitindo o ordenamento dos documentos resultantes pelo grau de similaridade de cada documento em relação à busca (FERNEDA, 2012, p.31).

Quanto ao funcionamento deste modelo, Kuramoto (2002, p.2) explica que:

Enquanto o Booleano baseia-se na comparação exata entre os termos de uma consulta e aqueles presentes nos documentos, o modelo Vetorial baseia-se na comparação parcial entre a representação dos documentos e a da consulta do usuário. Isto é possível graças à atribuição de pesos aos termos de indexação presentes na consulta e aqueles presentes nos documentos. Essa ponderação, atribuída a cada termo de indexação, permite calcular o grau de similaridade entre um documento e uma consulta. Portanto, a cada documento que atende a uma consulta é atribuído um grau de similaridade, possibilitando, ao sistema de recuperação de informação, apresentar os resultados de uma consulta de forma ordenada, normalmente, em ordem decrescente dos

respectivos graus de similaridade. Trata-se de uma maneira de evidenciar a relevância de cada documento em relação a uma dada consulta. Esses pesos são calculados a partir da frequência de ocorrência do referido termo no documento

No Modelo Vetorial um documento é representado por um vetor numérico onde cada elemento representa o peso, ou relevância do respectivo termo de indexação para a representação do documento. O valor do peso é normalizado com valores entre um (1) e zero (0), sendo os valores mais próximos de um (1) indicam termos com mais importância para a descrição do documento. Da mesma forma, a expressão de busca é também representada por um vetor numérico com os pesos dos termos que a representam a necessidade de informação do usuário. A utilização de uma mesma representação tanto para os documentos como para as expressões de busca permite calcular o grau de similaridade entre o vetor que representa uma determinada expressão de busca e cada um dos vetores que representam os documentos do *corpus*.

2.1.3 Modelo Probabilístico

O Modelo Probabilístico foi proposto inicialmente por Maron e Kuhns (1960) e posteriormente explorado por diversos outros pesquisadores, tais como Robertson e Jones (1976). A ideia é tratar o processo de recuperação de informação como um processo probabilístico, pois é caracterizado por seu grau de incerteza no julgamento de relevância dos documentos em relação a uma expressão de busca. Assim, é mais realista pensar em uma probabilidade de relevância do que em uma pretensa relevância exata, como a utilizada nos modelos booleano e vetorial.

A expressão de busca, composta por um ou mais termos, o usuário expressa sua necessidade de informação e a submete ao sistema. Por meio de cálculos de probabilidade o sistema calcula, para cada documento do *corpus*, um valor numérico (similaridade), que representa a provável relevância do documento para a consulta. Esse valor é utilizado para ordenar os resultados da busca. Tendo um primeiro conjunto de documentos, o usuário pode marcar alguns deles que considera verdadeiramente relevantes para a sua necessidade. O conjunto de documentos marcados pode ser então submetido ao sistema, permitindo fornecer resultados mais precisos. Esse processo, denominado *relevance feedback*, pode ser repetido até que o usuário se sinta satisfeito com os resultados.

Uma virtude do modelo probabilístico está em reconhecer que a atribuição de relevância é uma tarefa do usuário, pois é o único modelo que incorpora explicitamente o processo de *relevance feedback* como base para a sua operacionalização.

Esses três modelos de recuperação de informação (booleano, vetorial e probabilístico) compartilham um mesmo paradigma de representação no qual os termos extraídos dos documentos e das buscas são suficientes para se efetivar o processo de recuperação de informação, considerando-o um sistema fechado no qual o significado dos elementos lexicais é dado pelas suas inter-relações no interior de um *corpus* documental.

2.2 Interfaces de Resultados de Busca

Esta seção apresentará resumidamente a forma de apresentação de resultados de busca dos mecanismos de busca atuais, destacando algumas de suas características. Essas características dos atuais buscadores é a principal motivação e justificativa deste trabalho.

Ao exibir os resultados de uma busca, os documentos devem ser apresentados na íntegra ou por meio de alguma forma de representação do conteúdo desses documentos. Essa representação refere-se a um resumo do documento e é uma parte fundamental do sucesso da interface de recuperação de informação. O design dessas representações e a forma como os resultados de busca são apresentados é uma área de pesquisa e experimentação.

A qualidade das representações dos documentos pode afetar significativamente a relevância percebida na lista de resultados de uma busca. Nos atuais buscadores Web, o título da página geralmente é exibido de forma destacada, juntamente com o URL e algumas vezes outros metadados. O resumo, contendo trecho de texto extraído do documento, também é crítico para avaliação de resultados de recuperação.

Atualmente, o padrão de exibição de resultados padrão é uma lista vertical de resumos textuais. Em alguns casos, os resumos são pequenos trechos extraídos do texto completo no quais aparecem os termos utilizados na busca. Em outros casos, tipos especializados de metadados são mostrados. Por exemplo, uma busca utilizando o termo "arco íris" pode retornar imagens como um elemento do conjunto de resultados (Figura 2).

Figura 2 - Página de resultados de pesquisa do Yahoo! Utilizando o termo "arco íris".

YAHOO! BRASIL

Inicio Mail Notícias Esportes Finanças Celebridades Vida e Estilo Cinema Respostas Flickr Mais

"Arco Iris" Buscar

Web

Imagens

Vídeos

Notícias

Respostas

Shopping

"Arco Iris" - Imagens



Mais "Arco Iris" imagens

Qualquer período

Dia anterior

Semana passada

Mês passado

Em toda web

Somente páginas em português

Arco-íris – Wikipédia, a enciclopédia livre
pt.wikipedia.org/wiki/Arco-íris Em cache
Um arco-íris (também popularmente denominado arco-da-velha [1]) é um fenômeno óptico e meteorológico que separa a luz do sol em seu espectro (aproximadamente ...

Arco Íris Cosméticos
www.arcoiriscosmeticos.com.br Em cache
A Arco Iris Cosméticos surgiu em 2010 com o intuito de satisfazer seus clientes com os melhores produtos, preços e prazos de entrega. Nossa met...

Arco-Íris - Fenômeno Óptico - InfoEscola
www.infoescola.com/meteorologia/arco-iris Em cache
Artigo sobre os Arco-íris, quais são as condições necessárias para a sua formação, qual é a explicação física para o surgimento de arco-íris no céu, etc.

Como se forma um arco-íris? - noticias.terra.com.br
noticias.terra.com.br/...Como+se+forma+um+arcoiris.html Em cache
Um arco-íris aparece quando a luz branca do sol é interceptada por uma gota d'água da atmosfera. Parte da luz é refratada para dentro da gota ...

Significado das Cores do arco-íris - O que significam ...
www.significados.com.br/cores-do-arco-iris Em cache
O que significam as Cores do arco-íris. Conceito e Significado das Cores do arco-íris: O arco-íris é composto por 7 cores, que são apresentadas nesta ordem ...

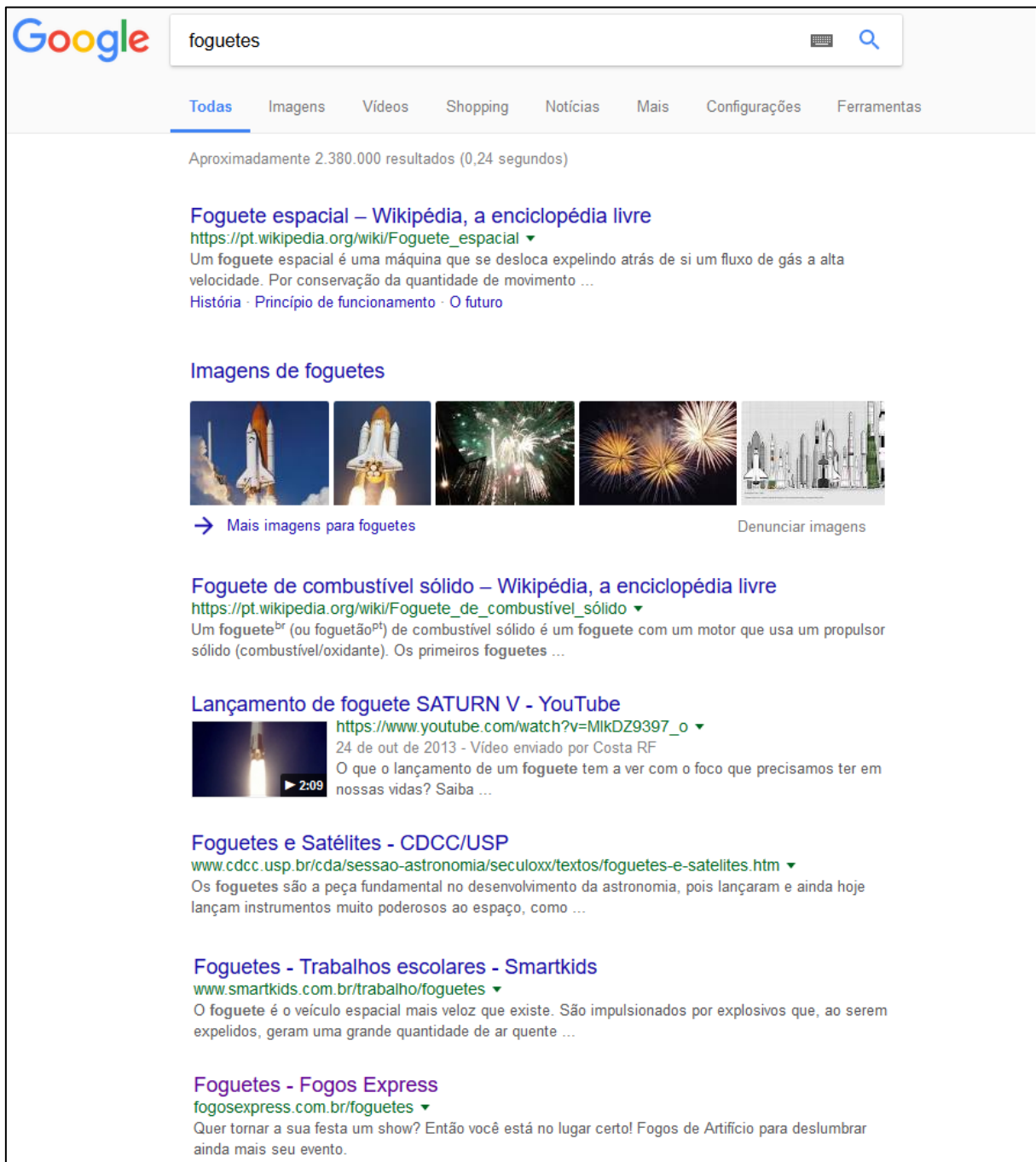
Grupo Arco-íris :: Cidadania LGBT – Desde 1993
www.arco-iris.org.br Em cache
nota oficial do grupo arco-iris sobre a parada do orgulho... saiba mais. nota pública saiba mais. reunião com os voluntários da 21ª parada do orgulho ...

Fonte: Resultado de busca realizada no Yahoo! (<https://br.yahoo.com/>) (29/06/2017)

Na Figura 2, os resultados incluem links para um conjunto de imagens sobre o termo ("arco íris"); um link para um artigo da enciclopédia Wikipédia sobre "arco íris" e como eles se formam, mas também um link para o site de uma empresa de cosméticos e outro link para um grupo LGBT.

Uma busca com o termo "foguetes" pode recuperar informações sobre a construção de foguetes espaciais (Figura 3).

Figura 3 - Página de resultados do Google para uma busca sobre "foguetes".



Fonte: Resultado de busca realizada no Google (<https://www.google.com.br/>) (29/06/2017)

Na Figura 3, diferentes tipos de informação são mostrados para diferentes significados da palavra “foguetes”. O primeiro resultado é um link para a enciclopédia Wikipédia sobre “Foguete Espacial”. Em seguida é apresentada uma série de imagens sobre o termo “foguetes” onde fica claro polissemia do termo, pois mistura fotos do lançamento do ônibus espacial com a apresentação de fogos de artifício. O último resultado apresentado na Figura 3 é um link para uma loja de fogos de artifício.

Observa-se que tanto no Yahoo! quanto no Google são utilizados alguns recursos úteis que auxiliam no julgamento da relevância dos documentos recuperados. Abaixo do título do documento, em destaque, aparece um link para o documento e em seguida um pequeno resumo do documento, um trecho de texto (2 linhas) onde aparece o termo de busca.

Determinar o tamanho do trecho de texto para ser colocado em cada documento do resultado é um problema complexo. Muitas vezes, os artigos mais relevantes são aqueles que contêm todos os termos da consulta próximos uns dos outros, mas para resultados menos relevantes é necessário um processo de escolha de frases contíguas que contenham termos da busca. Alguns resultados sugerem que é melhor mostrar frases completas em vez de cortá-las, mas, por outro lado, frases muito longas geralmente não são desejáveis na lista de resultados (ROSE; ORR; KANTAMNENI, 2007).

Alguns estudos de usabilidade sugerem que mostrar os termos da busca destacados (negrito) no contexto em que aparecem melhora a capacidade do usuário para avaliar a relevância dos resultados (WHITE; JOSE; RUTHVEN, 2003).

Segundo Hobbs, Pfitzner e Powers (200-), a maioria dos atuais mecanismos de busca possui basicamente os mesmos recursos que os sistemas de recuperação de informação propostos na década de 1950. O usuário fornece uma lista de termos que visam traduzir linguisticamente sua necessidade de informação e obtém como resultado uma lista de documentos que contém esses termos. Segundo os autores, as desvantagens desse tipo de interfaces são:

- Os usuários devem inspecionar sequencialmente uma lista geralmente longa de documentos.
- O usuário geralmente não é informado a razão pela qual o documento foi recuperado.
- O usuário não é informado sobre a relação de um determinado documento com a busca.
- Não são representadas as semelhanças entre os documentos.
- É exibido apenas um pequeno trecho do documento para que o usuário possa julgar a relevância.

Este trabalho propõe a solução dessas desvantagens, por meio da utilização do Modelo Espaço Vetorial associado à estrutura terminológica fornecida pelas ontologias.

O Modelo Vetorial mostra-se adequado para o desenvolvimento de interfaces de recuperação de informação, pois incorpora uma representação numérica das buscas e dos documentos, permitindo o cálculo da similaridade entre esses elementos. Tais características possibilitam uma representação espacial de documentos e buscas no interior de um ambiente terminológico.

As ontologias oferecem um conjunto de conceitos (termos) contextualizados em um determinado domínio do conhecimento. Unindo-se as características do Modelo Vetorial, este trabalho propõe um modelo de interface no qual os termos da ontologia, visualmente representados, formam a estrutura básica para busca por documentos relevantes.

No Capítulo 3 será apresentado a definição de ontologia e como as ontologias são vistas e utilizadas na Ciência da Informação. É apresentada também as principais características da linguagem de definição de ontologias, a linguagem OWL.

3.

Ontologia

O termo ontologia possui definições distintas dependendo da área de conhecimento ao qual é citado. Ferneda (2013, p. 48) afirma que na filosofia o termo ontologia possui sua origem na Metafísica, a “Filosofia Primeira” de Aristóteles, que trata do estudo do ser em sua essência.

O dicionário da língua portuguesa Aulete Digital define a palavra ontologia de forma abrangente:

(on.to.lo.gi.a)

sf.

1. Fil. Parte da filosofia que estuda a natureza dos seres, o ser enquanto ser.
2. Fil. Doutrina sobre o ser.
3. Hist. Med. Doutrina segundo a qual os fenômenos patológicos têm existência própria, não tendo relação com fenômenos fisiológicos.
4. Inf. Campo da informática que trata de conceitualizar de forma explícita e formal (portanto processável por máquina e compartilhável) conceitos e restrições relacionados a certo domínio de interesses

[F.: ont(o)- + -logia.]

Dentre os significados apresentados, a definição ligada ao campo da informática é a que mais se aproxima dos propósitos almejados por esta pesquisa, pois permite que utilizemos a ontologia como uma ferramenta de auxílio a organização e representação do conhecimento de forma processável por computadores.

No âmbito da ciência da computação uma definição clássica é a de Gruber (1995), para o qual uma ontologia é uma **especificação formal e explícita de uma conceitualização compartilhada**. *Formal* diz respeito a ser legível por computador. *Explícita*, indica que os

elementos estão claramente definidos; *conceitualização* refere-se a um modelo abstrato de um fenômeno; e *compartilhada* significa que os conceitos presentes representam um conhecimento consensual, aceito por um grupo de pessoas.

Uma ontologia pode ser considerada como um vocabulário de representação, geralmente especializado em um domínio ou assunto, qualificado por conceituações de tipos de objetos e suas relações no mundo. Em outras palavras, é um corpo de conhecimento que descreve algum domínio, utilizando um vocabulário de representação (CHANDRASEKARAN; JOSEPHSON; BENJAMIN, 1999).

Jasper e Uschold (1999, p.11-2) ressaltam a necessidade de explicitação dos relacionados entre os termos pertencentes a uma ontologia:

Uma ontologia pode possuir uma variedade de formas, mas necessariamente incluirá um vocabulário de termos, e alguma especificação de seus significados. Isto inclui definições e uma indicação de como os conceitos estão inter-relacionados, o que impõem uma estrutura no domínio e restringe possíveis interpretações de termos.

Os autores supracitados destacam um ponto do uso de ontologias, a possibilidade de criar restrições, explicitar as relações entre os termos, agregar rótulos em múltiplos idiomas conceituando assim cada termo com expressividade suficiente para que este sejam compreensíveis tanto por humanos quanto por máquinas com o mesmo nível de compreensão.

A utilização de ontologias oferece uma série de vantagens, sendo que no âmbito da ciência da computação destacamos alguns aspectos apontados por Guimarães (2002, p.53):

- Ontologias fornecem um vocabulário para representação do conhecimento que traz uma conceitualização que o sustenta, evitando ambiguidades.
- Ontologias permitem o compartilhamento de conhecimento. Sendo assim, caso exista uma ontologia que modele adequadamente certo domínio de conhecimento, essa pode ser compartilhada e usada por pessoas que desenvolvam aplicações dentro desse domínio.
- Fornecem uma descrição exata do conhecimento. Diferentemente da linguagem natural em que as palavras podem ter semântica totalmente diferente conforme o seu contexto, a ontologia por ser escrita seguindo uma estrutura formal, não deixa espaço para o gap semântico existente na linguagem natural. Por exemplo, a palavra “globo” pode significar um corpo esférico, porém também é o nome de um canal de

televisão brasileiro. A descrição exata de cada termo em uma ontologia evita a interpretação dúbia de termos como este.

Há possibilidade de mapeamento do idioma da ontologia sem que com isso seja alterada a sua conceitualização, ou seja, uma mesma conceitualização pode ser expressa em vários idiomas. É possível estender o uso de uma ontologia genérica de forma a que ela se adéque a um domínio específico.

As vantagens apresentadas por Guimarães são descritas como positivas a área de ciência da computação, porém a ciência da informação e a área de recuperação da informação mais especificamente podem ser fortemente beneficiadas por estes aspectos positivos provindos do uso de ontologias, compreendendo assim a necessidade da aproximação das áreas e o compartilhamento das tecnologias.

Quando ao nível de generalidade Fernald (2013, p.58) afirma que as ontologias podem ser classificadas em três níveis: ontologias genéricas, as que descrevem os conceitos mais gerais, como objetos, ações, eventos, etc. As ontologias de domínio e de tarefa que descrevem respectivamente um domínio como medicina ou direito ou uma tarefa ou atividade como uma venda; por último temos as ontologias de aplicação que descrevem conceitos dependentes de um determinado domínio e tarefa particular.

Uma ontologia é composta por uma série de componentes, uma breve descrição de cada elemento é apresentada por Ramalho (2010, p.38):

- **Classes e Subclasses:** As classes e subclasses de uma ontologia agrupam um conjunto de elementos, “coisas”, do “mundo real”, que são representadas e categorizadas de acordo com suas similaridades, levando-se em consideração um domínio concreto. Os elementos podem representar coisas físicas ou conceituais, desde objetos inanimados até teorias científicas ou correntes teóricas;
- **Propriedades Descritivas:** Descrevem as características, adjetivos e/ou qualidades das classes;
- **Propriedades Relacionais:** Trata-se dos relacionamentos entre classes pertencentes ou não a uma mesma hierarquia, descrevendo e rotulando os tipos de relações existentes no domínio representado;

- **Regras e Axiomas:** Enunciados lógicos que possibilitam impor condições como tipos de valores aceitos, descrevendo formalmente as regras da ontologia e possibilitando a realização de inferências automáticas a partir de informações que não necessariamente foram explicitadas no domínio, mas que podem estar implícitas na estrutura da ontologia;
- **Instâncias:** Indicam os valores das classes e subclasses, constituindo uma representação de objetos ou indivíduos pertencentes ao domínio modelado, de acordo com as características das classes, relacionamentos e restrições definidas;
- **Valores:** Atribuem valores concretos às propriedades descritivas, indicando os formatos e tipos de valores aceitos em cada classe.

A construção de uma ontologia, de acordo com FERNEDA (2013, p.62) pode ser pensada como:

[...] a união de peças que formam uma estrutura completa. Classes e subclasses definem um “esqueleto” na forma de uma hierarquia [...], complementada por propriedades descritivas, propriedades relacionais, regras e axiomas. A sua abrangência (domínio) deve ser previamente definida e estabelece uma área do conhecimento ou uma parte do mundo que se pretende tratar.

Conforme o contexto apresentado nota-se que as ontologias fornecem uma estrutura conceitual que visa representar formalmente os conceitos e suas relações, regras e restrições lógicas de um determinado domínio, cujo objetivo principal de sua construção é a necessidade de um vocabulário compartilhado cujas informações possam ser trocadas e reusadas por usuários ou agentes inteligentes (FERNEDA, 2013; SANTARÉM SEGUNDO, 2010).

3.1 Ontologias e a Ciência da Informação

A Ciência da Informação é uma área ligada às tecnologias de informação desde suas origens. Marcondes e Campos (2008) afirmam que a área de CI vem de uma longa tradição teórica, metodológica e prática, que converge para questões mais atuais ligadas à Web Semântica e para construção e utilização de ontologias viabilizando assim o acesso e a organização de acervos documentais, assim como o processamento semântico de informações em computadores.

Segundo Soergel (1999) e Vickery (1997), o termo ontologia começou a ser utilizado na literatura da Ciência da Informação no final da década de 1990, mas principalmente por pesquisadores da área de Organização do Conhecimento. Nesta época, os instrumentos e métodos de classificação despertavam maior interesse da Computação, devido principalmente a necessidade de desenvolvimento de instrumentos de organização da informação no ambiente Web.

A Organização do Conhecimento vem se consolidando como um importante campo de investigação da Ciência da Informação, a partir da fundação da *International Society for Knowledge Organization* (ISKO), em 1989 (PICKLER, 2014, p.60). A autora ainda explica que:

A Organização do Conhecimento é uma disciplina que estuda uma série de fundamentos teóricos do tratamento e recuperação da informação, avaliando o uso de instrumentos lógico-linguísticos para controlar os processos de representação, classificação, ordenação e armazenamento do conteúdo informativo dos documentos com a finalidade de permitir sua recuperação e disseminação.

A organização do conhecimento visando seu posterior uso é de importância primordial, desde a antiguidade até os dias atuais. Marcondes e Campos (2008, p.109) pontuam que o conhecimento científico, tão importante para a sociedade, vem passando por um desafio com o surgimento da Web. A velocidade em que são transferidos documentos e informações pela internet é indiscutível, um avanço grandioso, o que acabou permitindo um crescimento desordenado dos conteúdos disponíveis na rede, e o mais crítico é que grande parte do volume de materiais na rede são compreensíveis apenas por seres humanos.

As ontologias então se apresentam como:

Uma categoria de instrumentos de representação do conhecimento, possibilitando e favorecendo melhorias nos processos de representação, organização, disseminação e recuperação de conteúdos documentais. Se na Computação uma ontologia pode ser considerada um artefato tecnológico, no contexto da Ciência da Informação caracteriza-se como um instrumento de nível epistemológico, concebido para favorecer a representação formal dos conceitos e dos relacionamentos existentes entre eles em um domínio específico (FERNEDA, 2013, p.64)

Entre os instrumentos de representação tradicionalmente utilizados na área de Ciência da Informação, os tesauros são os mais semelhantes às ontologias. Como destacam Sales e Café (2008), ambos os instrumentos são constituídos por linguagens de estruturas combinatórias, de caráter especializado, representando termos e conceitos organizados a partir

de tipos de relacionamentos, no entanto, estudos têm constatado que, apesar de possuírem características em comum, tais instrumentos constituem-se como diferentes modelos de representação do conhecimento.

Outro fator determinante para a distinção entre ontologias e os demais modelos de representação utilizados na CI é a natureza das relações. Pickler (2014, p.61) destaca que para possibilitar a realização de inferências automáticas é necessário que as relações expressas na ontologia sejam formalmente descritas, rotuladas e categorizadas. Assim as ontologias proporcionam liberdade para representar tipos de relacionamentos que não seriam possíveis em outros modelos de representação.

O potencial de expressividade das ontologias as coloca como um instrumento valioso ao arsenal teórico e prático da Ciência da Informação. Porém, é necessário que os profissionais da informação sejam capazes de colocar em prática.

3.2 Web Ontology Language (OWL)

Atualmente a OWL é recomendada pelo consórcio W3C como a principal linguagem para a construção de ontologias. Ela tem como objetivo principal atender às necessidades de aplicação da Web Semântica e ser efetivamente utilizada por aplicações que necessitem processar o conteúdo de informações, e não somente apresentar a visualização destas informações. (SANTARÉM SEGUNDO, 2010, p.127).

A documentação oficial disponibilizada pelo W3C afirma que a OWL é uma linguagem de web semântica projetada para representar o conhecimento em toda sua riqueza e complexidade, além de grupos de coisas e as relações entre as coisas. Ela é baseada em lógica computacional onde o conhecimento é expresso de forma que o computador possa explorá-lo e verificá-lo sistematicamente, além de ter acesso a informações explícitas ou implícitas por meio de inferências. A OWL é parte da “pilha” de tecnologias propostas pelo W3C, que inclui ainda RDF, RDFS, SPARQL, entre outras.

A linguagem OWL é subdividida em três sub-linguagens, projetadas para uso de implementadores e comunidades específicas: OWL Lite, OWL DL e OWL Full, seus detalhes e características são apresentadas na sequência:

- OWL Lite: dá suporte à criação de hierarquias simplificadas, com restrições simples. Devido a um menor conjunto de funcionalidades ela se torna mais simples

e de fácil utilização, sendo um de seus principais objetivos fornecer um rápido caminho de migração para tesauros e outras taxonomias.

- OWL DL: possui o mesmo vocabulário da linguagem OWL Full e dá suporte aos usuários que desejam o máximo de expressividade, sem perder a completude computacional (todas as conclusões são garantidas de serem computadas) e capacidade de decisão (todas as computações serão finalizadas em um tempo finito) dos sistemas de raciocínio. O OWL DL inclui todos os construtores da linguagem. OWL com restrições, como separação entre tipos (uma classe não pode ser ao mesmo tempo um indivíduo ou tipo, e uma propriedade não pode ser ao mesmo tempo um indivíduo ou uma classe). OWL DL tem expressividade menor que o OWL Full, mas conta com melhor eficiência, pois garante que todas as conclusões sejam computáveis e resolvidas em tempo finito. Ela tem este nome devido a sua correspondência à Lógica de Descrição ou Description Logic.
- OWL Full é a versão com o máximo de expressividade e liberdade sintática do RDF, sem nenhuma garantia computacional. A linguagem OWL Full não conta com as restrições da OWL DL, e justamente por isso pode ser bem mais adaptada a situações onde o ponto mais importante é a expressividade. Ela suporta o mesmo subconjunto da OWL DL porém tem restrições diferentes, como, por exemplo, permitir que uma classe seja ao mesmo tempo uma classe e um indivíduo.

A escolha de uma ou outra especificação da linguagem está diretamente ligada a análise prévia do domínio e do tipo de ontologia que se pretende criar (SANTARÉM SEGUNDO, 2010, p 129).

Os desenvolvedores de ontologias devem escolher a versão de acordo com suas necessidades. Basicamente a escolha consiste em balancear o nível de expressividade com decidibilidade e eficiência na inferência de informações, isto porque, quanto maior o nível de expressividade utilizado na descrição de uma ontologia, maior será a complexidade dos algoritmos envolvidos em processos de busca de informações (FILHO; LÓSCIO, 2009).

Um arquivo OWL é formado por alguns “blocos” bem definidos, sendo que cada um deles tem um papel muito claro junto a estrutura da ontologia. Agora faremos uma breve descrição de cada um destes blocos que compõem a estrutura de um arquivo OWL bem formado.

3.2.1 Namespaces

O início de um arquivo OWL é formado pela declaração de *namespaces*, que são os vocabulários que serão empregados dentro da ontologia. Para que os termos dentro de uma ontologia possam ser descritos de forma a serem interpretados sem ambiguidades é de suma importância declarar corretamente os namespaces.

O componente padrão inicial de uma ontologia inclui um conjunto de declarações em XML dos namespaces incluídas dentro de uma tag `rdf:RDF`. Veja na Figura 4 uma típica declaração dos namespaces em uma ontologia sobre vinhos.

Figura 4 - Exemplo de declaração dos namespaces em uma ontologia OWL

```
1 <rdf:RDF
2   xmlns      ="http://www.w3.org/TR/2004/REC-owl-guide-20040210/wine#"
3   xmlns:vin  ="http://www.w3.org/TR/2004/REC-owl-guide-20040210/wine#"
4   xml:base   ="http://www.w3.org/TR/2004/REC-owl-guide-20040210/wine#"
5   xmlns:food ="http://www.w3.org/TR/2004/REC-owl-guide-20040210/food#"
6   xmlns:owl  ="http://www.w3.org/2002/07/owl#"
7   xmlns:rdf  ="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
8   xmlns:rdfs ="http://www.w3.org/2000/01/rdf-schema#"
9   xmlns:xsd  ="http://www.w3.org/2001/XMLSchema#">
```

Fonte: Adaptado de PICKLER, 2014, p.78

Na Figura 4 é possível observar que a linha 2 não possui prefixo indicando que nessa ontologia todos os nomes sem prefixo terão como domínio a própria ontologia. A linha 3 define o prefixo “vin” para referenciar uma ontologia de vinhos pré-definida. Essa ontologia sobre vinhos é bastante utilizada nos documentos disponibilizados pelo W3C para exemplificar conceitos da Web Semântica e nos diversos manuais sobre a linguagem OWL.

A linha 4 define o prefixo base, indicando que essa ontologia se utilizará de uma ontologia já construída anteriormente como base para a construção da nova ontologia. Verificase assim a capacidade de reuso da informação existente, pois esta é uma forma de utilizar ontologias prontas para criar novas, mais específicas ou que atendam as novas necessidades.

A linha 5 define o prefixo food que se refere uma segunda ontologia já pronta, é desta forma que podem ser utilizadas várias ontologias (e vocabulários) para se representar o conhecimento.

As últimas linhas, 6 até 9 são os prefixos “owl”, “rdf”, “rdfs” e “xsd”, estes são obrigatórios para o uso da linguagem OWL, pois indicam o uso do léxico e da sintaxe referenciados na URLs.

A utilização destes prefixos tem como principal objetivo facilitar o desenvolvimento da ontologia, abreviando a declaração completa das definições, resultando assim em uma maior clareza no código. É possível também utilizar as URLs completas diretamente ao invés dos prefixos, porém isso pode gerar mais dificuldade na leitura e desenvolvimento das ontologias.

3.2.2 Cabeçalhos

Após a definição dos *namespaces*, o próximo passo na criação de um arquivo OWL é a definição de um cabeçalho, que é responsável por indicar um conjunto de informações a respeito da ontologia que está sendo desenvolvida.

É neste momento que devem ser inseridas as informações que dão suporte a tarefas cruciais do desenvolvimento de uma ontologia, como comentários, versão da ontologia, importação de código pré-existente, além da caracterização dos metadados referentes a ontologia que está sendo desenvolvida.

Estas informações devem ser agrupadas dentro da tag *owl:Ontology*, como pode ser visto na Figura 5.

Figura 5 – Exemplo de declaração do cabeçalho de uma ontologia OWL

```
1 <owl:Ontology rdf:about="">
2   <rdfs:label>
3     Ontologia sobre Vinhos
4   </rdfs:label>
5   <rdfs:comment>
6     Um exemplo de ontologia OWL
7   </rdfs:comment>
8   <owl:versionInfo>
9     06/05/2006 15:45:00
10  </owl:versionInfo>
11  <owl:priorVersion rdf:resource="http://www.w3.org/TR/2003/PR-owl-guide-20031215/wine"/>
12  <owl:imports rdf:resource="http://www.w3.org/TR/2004/REC-owl-guide-20040210/food"/>
```

Fonte: Adaptado de PICKLER, 2014, p.79

A tag inicial *owl:Ontology* na linha 1 indica que neste bloco estão descritos os metadados da ontologia que está sendo desenvolvida.

A tag *rdfs:label* na linha 2 tem a função de nomear a ontologia. O atributo *rdfs:comment* na linha 5 permite a definição de comentários sobre a ontologia. O atributo *owl:versionInfo* da linha 8 permite inserir um texto contendo informações sobre a versão da ontologia. O atributo *owl:priorVersion* na linha 11 faz uma referência explícita à versão anterior a essa ontologia. O atributo *owl:imports* na linha 12 permite a inserção de dados de outros arquivos dentro do documento que está sendo desenvolvido.

É importante ressaltar a diferença entre indicar o *namespace* para uma ontologia e a importação da mesma usando a tag *owl:imports*. A indicação do *namespace* ocorre quando se deseja utilizar parte da estrutura de outro documento, como definição de classes ou atributos, por exemplo. Já a utilização da importação com a tag *owl:imports* indica que o conteúdo completo da outra ontologia será inserido no documento.

Após descrever um pouco sobre as declarações iniciais, do cabeçalho e dos *namespaces* seguimos abordando os elementos básicos da construção de uma ontologia OWL, são eles as classes, as instâncias destas classes (indivíduos) e as propriedades (relacionamentos) entre essas instâncias.

3.2.3 Classes

As classes proporcionam um mecanismo de abstração para agrupar recursos com características similares, ou seja, uma classe define um grupo de indivíduos que compartilham algumas propriedades (CARVALHO; LIMA, 2005).

As classes são utilizadas para definir um conceito de um determinado domínio e pessoas, automóveis, ou qualquer outra entidade concreta ou abstrata que se deseja representar. As classes e seus relacionamentos hierárquicos são os elementos básicos para a criação da taxonomia que serve de alicerce para a construção de qualquer ontologia. Na linguagem OWL, a classe principal é expressa como *owl:Thing*, sendo que essa classe predefinida é a classe mais genérica (superclasse) da qual derivam todas as demais classes criadas pelo usuário.

Uma classe OWL é definida utilizando a tag *owl:Class* seguida de um atributo identificador *rdf:ID*. A Figura 6 mostra a criação de uma classe.

Figura 6 – Exemplo de definição de uma classe

```
1  
2 <owl:Class rdf:ID="Computador" />  
3
```

Fonte: Adaptado de PICKLER, 2014, p.80

A sintaxe que pode ser vista na Figura 6 é uma das formas mais simplistas de criação de uma classe. Esta classe é referenciada no corpo da ontologia utilizando o caractere “#” seguido do nome da classe, tal como outras linguagens de programação. Assim, para se referir a esta classe podemos utilizar a seguinte declaração: #Computador. Isso pode ser usado por

exemplo ao se referir a um recurso, usando o código `rdf:resource="#Computador"` ou ainda utilizando a cláusula `about`, como em `rdf:about="#Computador"`.

Uma forma mais comum de se criar classe é criando um bloco delimitado por abertura e fechamento da tag `owl:Class`, como podemos ver na Figura 7. Nesta segunda forma podemos agregar dentro da classe propriedades e relacionamentos desta classe.

Figura 7 – Exemplo de definição de uma classe usando blocos OWL

```
1
2 <owl:Class rdf:ID="Fabricante">
3   <rdf:label>Fabricante</rdf:label>
4   ...
5 </owl:Class>
6
```

Fonte: Adaptado de PICKLER, 2014, p.81

A Figura 7 mostra a definição de um rótulo usando a propriedade *rdf:label* para a classe Fabricante. Essa é apenas uma das propriedades que pode ser definidas para uma classe OWL.

Como já foi dito uma ontologia deve se apoiar em uma estrutura taxonômica, na qual as classes se organizam em uma forma hierárquica. O código para se declarar isso em uma ontologia OWL é baseado na tag *rdfs:subClassOf*, como podemos ver na Figura 8.

Figura 8 – Exemplo de definição de uma Hierarquia de classes em OWL

```
1
2 <owl:Class rdf:ID="Computador"/>
3   <rdfs:label>Computador</rdfs:label>
4 </owl:Class>
5
6 <owl:Class rdf:ID="Desktop">
7   <rdfs:label>Desktop</rdfs:label>
8   <rdfs:subClassOf rdf:resource="#Computador"/>
9 </owl:Class>
10
11 <owl:Class rdf:ID="Notebook">
12   <rdfs:label>Notebook</rdfs:label>
13   <rdfs:subClassOf rdf:resource="#Computador"/>
14 </owl:Class>
15
16 <owl:Class rdf:ID="AllInOne">
17   <rdfs:label>All in One</rdfs:label>
18   <rdfs:subClassOf rdf:resource="#Desktop"/>
19 </owl:Class>
20
21 <owl:Class rdf:ID="Netbook">
22   <rdfs:label>Netbook</rdfs:label>
23   <rdfs:subClassOf rdf:resource="#Notebook"/>
24 </owl:Class>
25
```

Fonte: Adaptado de PICKLER, 2014, p.82

O código apresentado na Figura 8 mostra que as classes “Desktop” e “Notebook” são subclasses da classe “Computador”. Assim, o conjunto de indivíduos da classe Notebook e os indivíduos da classe Desktop devem ser um subconjunto de indivíduos da classe Computador, isto é, a classe Computador possui características (atributos) que são comuns à classe Notebook e à classe Desktop.

Utilizando a *tag* *rdfs:subClassOf* é possível definir na ontologia os relacionamentos de especificidade e generalização, considerando o caminho que vai de uma classe para uma subclasse (especificidade) ou de uma subclasse à sua correspondente classe superior (generalização).

Este recurso abre valiosas possibilidades para o processo de recuperação da informação, visto que com essa possibilidade de “navegação” em uma ontologia é bastante útil, por exemplo, é possível a partir de do termo “AllInOne” o sistema de recuperação compreenda que AllInOne é um tipo de Desktop, que por sua vez é um tipo de computador, “compreendendo” assim o contexto da busca de um usuário.

Um outro recurso da linguagem OWL é a definição de classes equivalentes ou sinônimas, para isso é utilizada a *tag* `owl:equivalentClass`. Na Figura 9 pode-se observar a definição de uma classe equivalente.

Figura 9 – Exemplo de definição de uma classe equivalente

```
1
2 <owl:Class rdf:ID="Laptop"/>
3   <rdfs:label>Laptop</rdfs:label>
4   <owl:equivalentClass rdf:resource="#Notebook"/>
5 </owl:Class>
6
```

Fonte: Adaptado de PICKLER, 2014, p.83

Na Figura 9 uma classe “Laptop” foi definida como uma classe equivalente à classe “Notebook”, desta forma é possível reduzir ambiguidades e agregar conceitos por seus significados verdadeiros, não apenas pela grafia do termo.

Este recurso também pode ser utilizado para promover a interoperabilidade entre duas ontologias diferentes, declarando algumas classes equivalentes é possível estabelecer um consenso entre termos com grafias diferentes e significados idênticos.

Ainda falando sobre a definição das relações em uma hierarquia de classes temos um recurso que permite especificar quando um indivíduo não pode ser membro de duas classes, este recurso é conhecido como disjunção. Na Figura 10 é apresentado o código que explicita que a classe “Desktop” é disjunta à classe “Notebook”, isto é, não pode existir um computador que seja ao mesmo tempo um Desktop e um Notebook.

Figura 10 – Exemplo de definição de uma classe disjunta

```
1
2 <owl:Class rdf:ID="Desktop"/>
3   <rdfs:label>Desktop</rdfs:label>
4   <owl:disjointWith rdf:resource="#Notebook"/>
5 </owl:Class>
6
```

Fonte: Adaptado de PICKLER, 2014, p.84

Filho e Lóscio (2009) destacam que a linguagem OWL permite ainda a criação de novas classes a partir de classes previamente existentes, sendo possível especificar que:

- *unionOf*: é possível criar uma classe como sendo a união de duas ou mais classes. Exemplo: a classe de acadêmicos pode ser definida como a união das classes docentes e discentes.

- *intersectionOf*: uma classe também pode ser originada da intersecção de duas classes existentes. Por exemplo: pode-se criar uma classe “Alunos professores” através da intersecção das classes discentes e docentes.

3.2.4 Indivíduos

Os indivíduos são definidos como objetos do mundo que sempre são ligados às classes previamente definidas, ou seja, são membros das classes. Em uma ontologia todo indivíduo deve pertencer a uma classe existente, que foi criada previamente.

Para inserir um indivíduo em uma ontologia OWL devemos apresentá-lo como membro de uma classe, para isso utiliza-se o código mostrado na Figura 11.

Figura 11 – Exemplo de definição de um indivíduo

```
1  
2 <Notebook rdf:ID="MeuComputador" />  
3
```

Fonte: Adaptado de PICKLER, 2014, p.87

A Figura 11 indica que o indivíduo “MeuComputador” é uma instância da classe Notebook, definida na Figura 8. Esta construção declara um fato na ontologia que permite ao motor de inferência chegar à seguinte conclusão “MeuComputador é um Notebook”.

Outra forma de criar um indivíduo é defini-lo inicialmente como uma instância da classe geral *owl:Thing* e em seguida defini-lo com um tipo *rdf:type* da classe à qual esse indivíduo está associado. Nesta segunda abordagem a leitura e compreensão da declaração do indivíduo e de suas características ficam mais claras. Veja um exemplo desta forma alternativa para declaração de um indivíduo na Figura 12.

Figura 12 – Exemplo de uma forma alternativa para a definição de um indivíduo

```
1  
2 <owl:Thing rdf:ID="MeuComputador" />  
3  
4 <owl:Thing rdf:about="#MeuComputador">  
5   <rdfs:label>Meu Computador</rdfs:label>  
6   <rdf:type resource="#Notebook" />  
7 </owl:Thing>  
8
```

Fonte: Adaptado de PICKLER, 2014, p.87

3.2.5 Propriedades

Propriedades são recursos da linguagem OWL que têm o propósito de descrever fatos em geral, elas são usadas para estabelecer relacionamentos entre dois indivíduos ou entre indivíduos e valores de dados. Através das propriedades, pode-se fazer referência a todos os membros de uma classe, ou seja, afirmar fatos gerais sobre todos os membros de uma classe ou apenas sobre um indivíduo específico.

A linguagem OWL dispõem de duas categorias principais de propriedades:

- Propriedades de objetos: estabelecem relação entre classes ou indivíduos;
- Propriedades de dados: indicam a relação entre indivíduos e valores de dados, estes valores devem ser expressos em RDF utilizando os tipos definidos no XML Schema.

Qualquer propriedade OWL é subclasse da classe *rdf:Property*. Propriedades de objetos são definidas como instância da classe *owl:ObjectProperty*, já as propriedades de dados são definidas como instâncias da classe *owl:DatatypeProperty*. Na Figura 13 pode-se ver a definição de uma propriedade de objeto denominada “Marca” que está associada à classe “Computador”.

Figura 13 - Exemplo de propriedade de objetos

```
1
2 <owl:ObjectProperty rdf:ID="Marca">
3   <rdfs:label>Marca</rdfs:label>
4   <rdfs:domain rdf:resource="#Computador"/>
5   <rdfs:range rdf:resource="#Fabricante"/>
6 </owl:ObjectProperty>
7
```

Fonte: Adaptado de PICKLER, 2014, p.84

Na Figura 13 temos um código que apresenta a definição de uma propriedade de objetos, onde a linha 4 especifica que a classe “Computador” possui uma propriedade denominada “Marca” que deve ser obrigatoriamente preenchida com valores da classe “Fabricante” (linha 5), sendo que esta deve estar previamente definida na ontologia.

O uso de propriedades de dados é demonstrado na Figura 14, onde duas propriedades são definidas, vejamos:

Figura 14 – Exemplo de propriedade de dados

```
1
2 <owl:DatatypeProperty rdf:ID="MemoriaRam">
3   <rdfs:label>Memória RAM</rdfs:label>
4   <rdfs:domain rdf:resource="#Computador"/>
5   <rdfs:range rdf:resource="&xsd;positiveInteger"/>
6 </owl:DatatypeProperty>
7
8 <owl:DatatypeProperty rdf:ID="Nome">
9   <rdfs:label>Nome</rdfs:label>
10  <rdfs:domain rdf:resource="#Fabricante"/>
11  <rdfs:range rdf:resource="&xsd:string"/>
12 </owl:DatatypeProperty>
13
```

Fonte: Adaptado de PICKLER, 2014, p.85

A primeira propriedade apresentada na Figura 14 é uma propriedade da classe Computador, isso é indicado na linha 4, o nome desta propriedade é “MemoriaRam” como pode ser visto na linha 2, ela ainda possui um rótulo em formato mais legível na linha 3 e os valores aceitáveis desta propriedade são inteiros positivos, tal restrição foi especificada na linha 5, por meio da referência ao léxico do XML Schema Datatypes. É possível usar os tipos pertencentes ao XML Schema devido a declaração do namespace correspondente (xsd) na Figura 2.

A segunda propriedade da Figura 14 define um nome do tipo string para a classe “Fabricante”, seguindo a mesma lógica da propriedade “MemoriaRam”. Além da definição de propriedades como estas supracitadas é possível também criar hierarquias de propriedades, assim como as hierarquias de classes.

Para definir uma hierarquia de propriedades utilizamos a declaração *rdfs:subPropertyOf*, desta forma é possível afirmar que uma determinada propriedade é uma sub-propriedade de alguma outra. Veja na Figura 15 algumas sub-propriedades para a classe Fabricante.

Figura 15 – Exemplo de definição de uma hierarquia de propriedades

```

1
2 <owl:DatatypeProperty rdf:ID="Endereco">
3   <rdfs:label>Endereço</rdfs:label>
4   <rdfs:domain rdf:resource="#Fabricante"/>
5 </owl:DatatypeProperty>
6
7 <owl:DatatypeProperty rdf:ID="Logradouro">
8   <rdfs:label>Logradouro</rdfs:label>
9   <rdfs:subPropertyOf rdf:resource="#Endereco"/>
10  <rdfs:domain rdf:resource="#Fabricante"/>
11  <rdfs:range rdf:resource="#xsd:string"/>
12 </owl:DatatypeProperty>
13
14 <owl:DatatypeProperty rdf:ID="Numero">
15   <rdfs:label>Número</rdfs:label>
16   <rdfs:subPropertyOf rdf:resource="#Endereco"/>
17   <rdfs:domain rdf:resource="#Fabricante"/>
18   <rdfs:range rdf:resource="#xsd:positiveInteger"/>
19 </owl:DatatypeProperty>
20
21 <owl:DatatypeProperty rdf:ID="Complemento">
22   <rdfs:label>Complemento</rdfs:label>
23   <rdfs:subPropertyOf rdf:resource="#Endereco"/>
24   <rdfs:domain rdf:resource="#Fabricante"/>
25   <rdfs:range rdf:resource="#xsd:string"/>
26 </owl:DatatypeProperty>
27

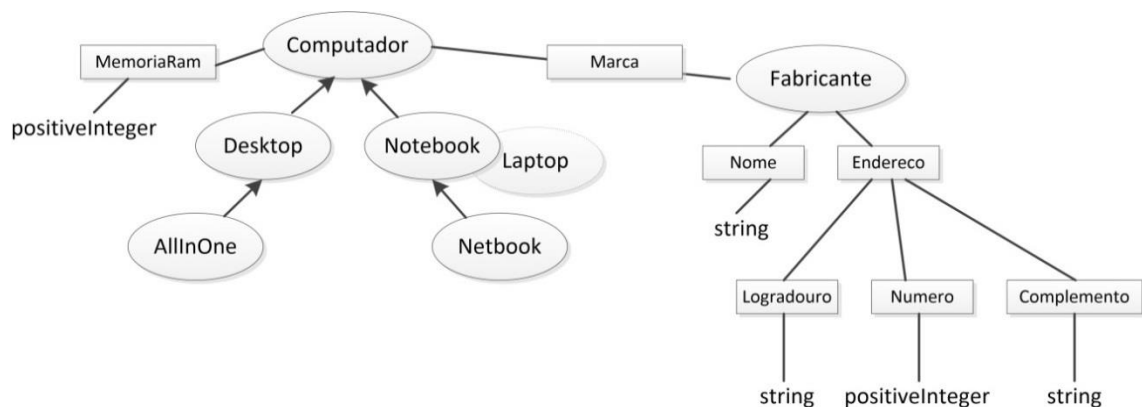
```

Fonte: Adaptado de PICKLER, 2014, p.86

A Figura 15 apresenta uma propriedade “Endereco” com suas sub-propriedades “Logradouro”, “Numero” e “Complemento”, sendo a propriedade “Endereco” relacionada à classe “Fabricante”.

Uma representação gráfica da hierarquia de classes, junto com as propriedades definidas nos exemplos anteriores é mostrada na Figura 16.

Figura 16 – Representação gráfica de classes e propriedades



Fonte: Adaptado de PICKLER, 2014, p.86

Há ainda uma série de características e especializações que podem ser aplicadas às propriedades que não são relevantes para no contexto deste trabalho e, portanto, não serão detalhadas.

A eficiência do processo de recuperação de informação é dependente da correta interpretação e representação do conteúdo informacional dos documentos e da representação eficiente da necessidade de informação do usuário em um conjunto de termos de busca. As ontologias podem contribuir no processo de recuperação de informação provendo um maior nível semântico dessas duas instâncias de representação.

As características da linguagem OWL, permitem a construção de software capazes de transformar as descrições textuais de seus elementos em representações visuais de tais elementos. No Capítulo 4 serão apresentadas algumas formas de representação visual de ontologias.

4.

Apresentação Gráfica de Ontologias

Segundo Campos (2010), a conceitualização proporcionada pelas ontologias pode ser entendida como uma abstração, uma visão simplificada do mundo, com o propósito de auxiliar na compreensão, compartilhamento ou consenso sobre uma área de conhecimento. Assim, as técnicas de visualização de informação podem ser utilizadas em ferramentas para auxiliar na compreensão do domínio representado pelas ontologias. No entanto, Kriglstein e Wallner (2011) afirmam que a escolha por uma técnica adequada para representar uma ontologia é uma tarefa bastante complexa, pois não existe uma técnica que possa ser empregada a qualquer tipo de ontologia nem tampouco para qualquer tipo de usuário ou necessidade de informação.

4.1 Métodos de Visualização de Ontologias

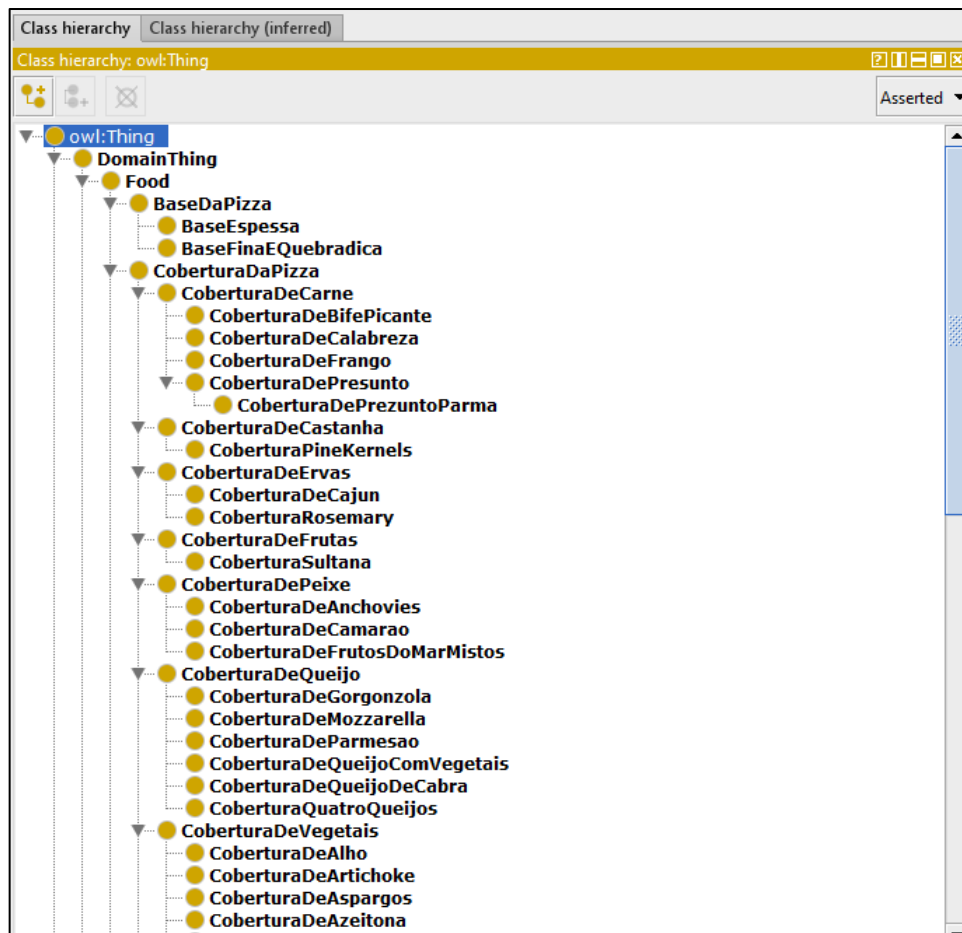
Katifori et al (2007) apresentam seis métodos de visualização de ontologias, são eles: *Indented List*; *Node-Link and Tree*; *Zoomable Visualizations*; *Space Filling*; *Context+Focus and Distortion Techniques*; *Information Landscapes*. A seguir serão apresentados detalhadamente esses métodos.

4.1.1 *Indented List* (lista indentada)

A maioria dos sistemas de visualização de ontologia oferece uma estrutura hierárquica semelhante ao Windows Explorer. Neste tipo de visualização, a hierarquia de classes da ontologia é representada como uma árvore, tal como utilizada na ferramenta Protégé (Figura

17). Essa interface, do tipo lista indentada, permite que os usuários cliquem sobre os itens de uma ontologia e tenham uma visão dos seus detalhes, conforme necessidade.

Figura 17 - Visualização de ontologia em interface de lista indentada



Fonte: Elaborado pela autora com o uso da ferramenta Protégé.

A principal vantagem da utilização de lista indentada é a simplicidade de implementação e a sua familiaridade. Esse método de visualização tem uma clara vantagem em comparação com quase todas as outras técnicas: não há sobreposição de rótulos e não é necessário mover o mouse sobre um item para visualizar o seu rótulo. O recurso de retração e expansão de nós é uma característica útil para se concentrar em partes específicas da hierarquia, especialmente em hierarquias grandes. Além disso, a simplicidade da interface facilita a navegação rápida. A tarefa de localizar uma classe ou uma instância específica é mais fácil do que na maioria das outras visualizações, pois o layout *top-down* permite uma exploração sistemática de toda a ontologia.

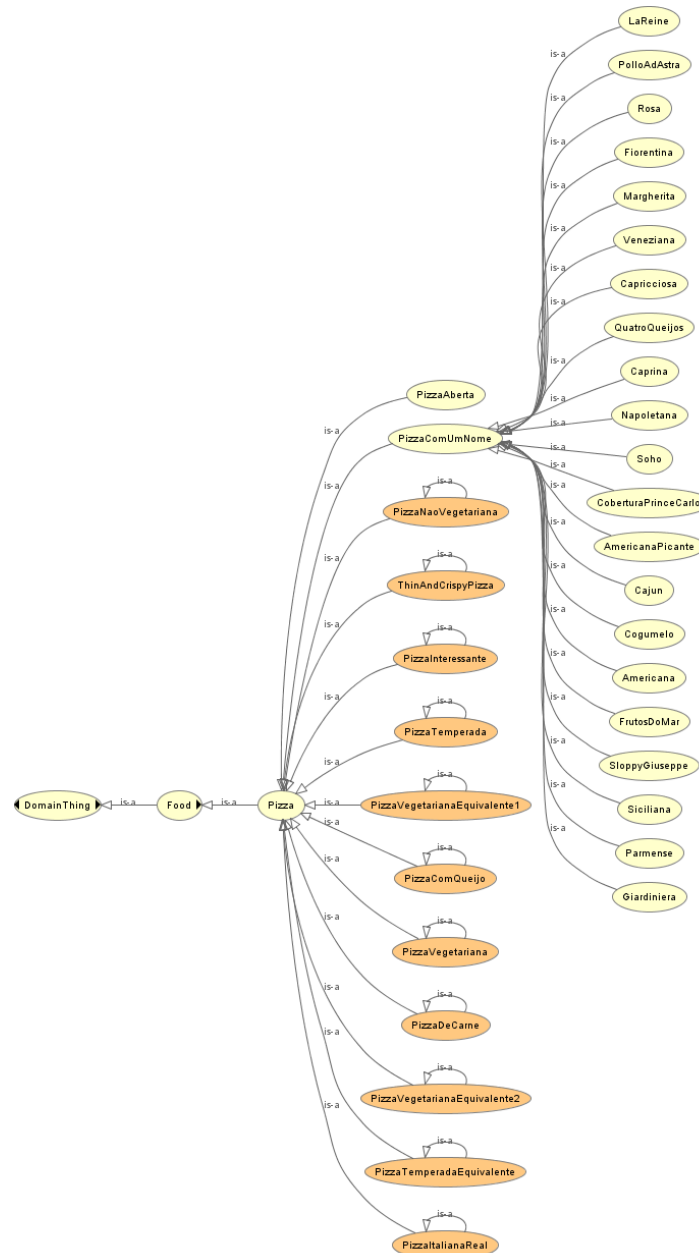
Um problema dessa técnica de visualização está na sua principal característica que é a representação em árvore e não na forma de grafo. Sendo assim, apenas as relações

taxonômicas são exibidas, deixando de lado os demais tipos de relações. Outra limitação é que em grandes ontologias apenas uma pequena parte da ontologia pode ser visível de uma só vez.

4.1.2 *Node-Link and Tree* (nó-link e árvore)

Nesta categoria as ontologias são representadas como um conjunto de nós interligados, apresentando a taxonomia com um layout *top-down* ou *left-to-right*. Normalmente, o usuário pode expandir e retrair os nós e suas sub-árvores a fim de ajustar o detalhamento das informações mostradas e evitar sobrecarga de informações na tela. A Figura 18 apresenta um exemplo utilizando a ferramenta Protégé.

Figura 18 – Visualização de ontologia em interface do tipo nó-link



Fonte: Elaborado pela autora com o uso da ferramenta Protégé.

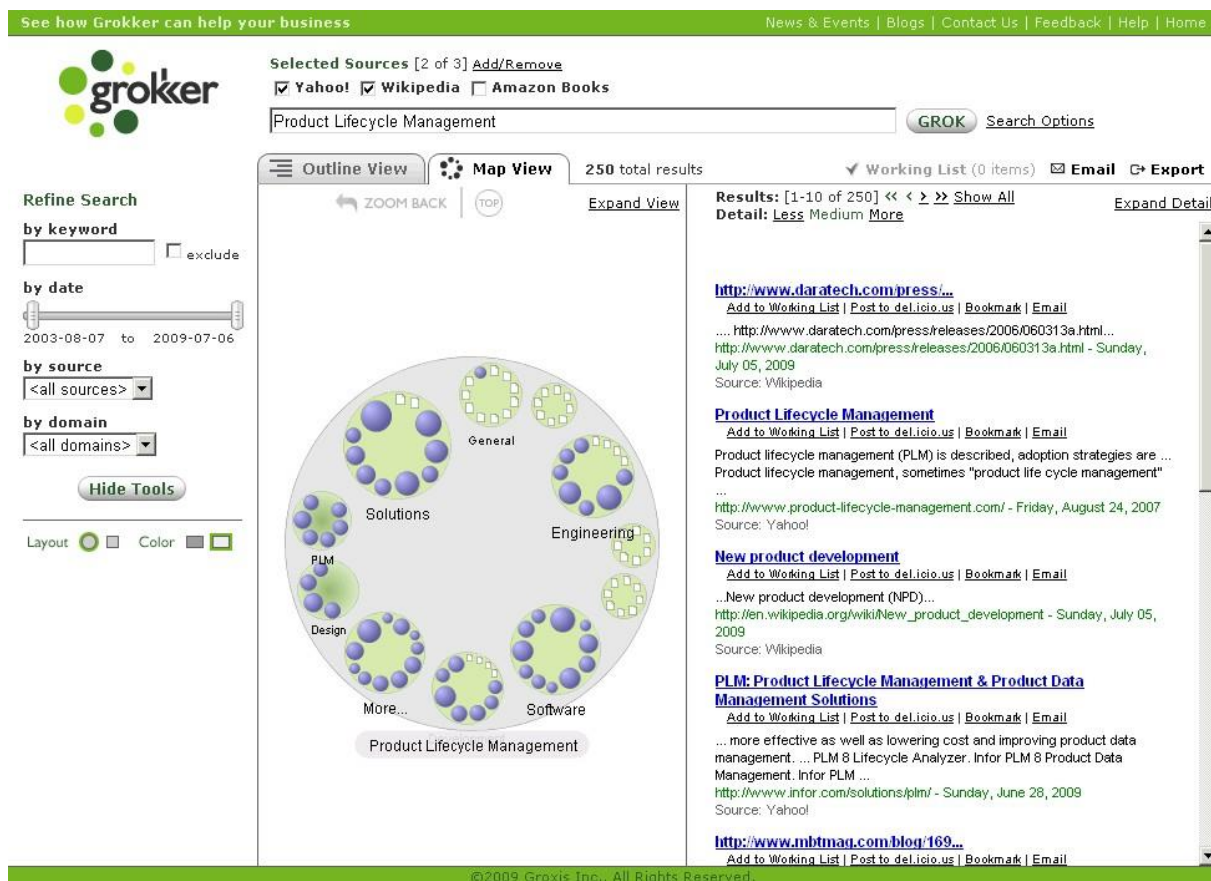
Os diagramas nó-link em formato de árvore são formas relativamente intuitivas de representar hierarquias, possibilitando uma visão geral de sua estrutura e a fácil identificação de algumas de suas características, tal como profundidade e largura. Por outro lado, os métodos nó-link e árvore geralmente fazem uso ineficiente do espaço da tela, deixando grandes espaços no lado da raiz da árvore enquanto o lado oposto fica sobrecarregado. Além disso, mesmo árvores com poucos nós muitas vezes precisam de múltiplas telas para serem completamente exibidas, pois apenas parte do diagrama é visível em um mesmo momento.

4.1.3 Zoomable Visualizations (visualizações com zoom)

Esta categoria engloba os métodos que apresentam os nós de níveis mais baixos da hierarquia aninhados no interior de seus nós-pai. Essas técnicas permitem que o usuário faça zoom nos nós-filhos para ampliá-los. Um exemplo de utilização dessa técnica é o Grokker (Figura 19).

O Grokker (KOSHMAN, 2006) foi uma ferramenta interativa que oferecia uma representação gráfica dos resultados de um mecanismo de busca na Web ou outro repositório de informação. O mecanismo de agrupamento apresenta os documentos como uma série de círculos aninhados. Os usuários podem navegar na hierarquia clicando em um círculo, que é então ampliado, tornando o conteúdo visível.

Figura 19 - Visualização de ontologia em interface de zoom



Fonte: <http://blogs.3ds.com/perspectives/visual-search-engines/>

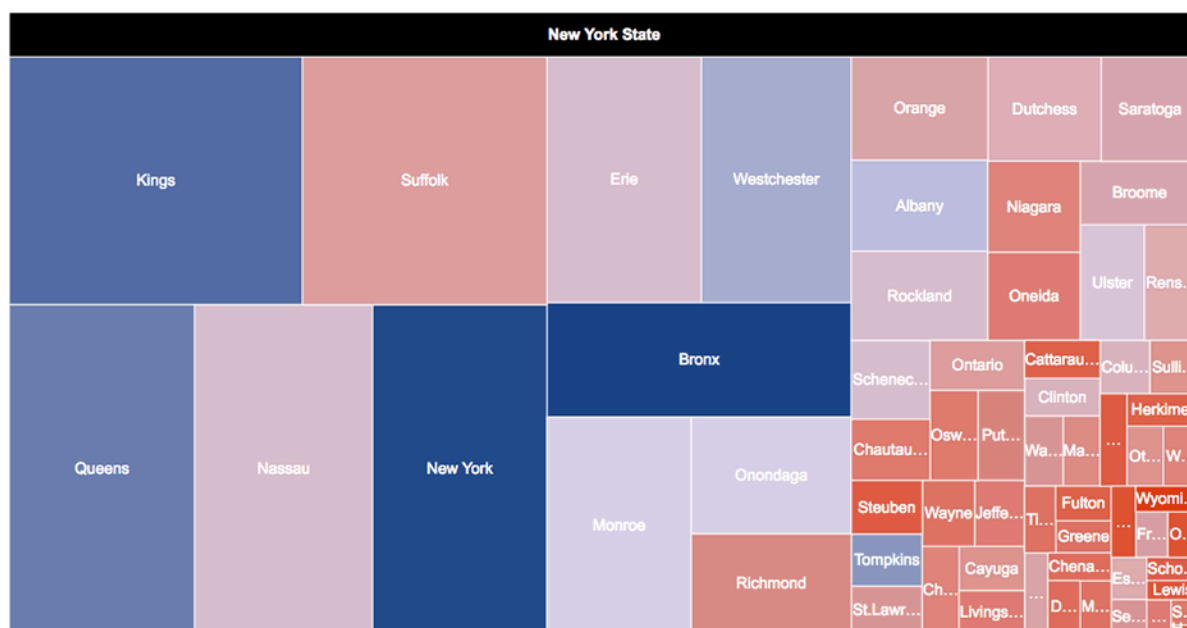
As interfaces que utilizam esse tipo de técnica são, em geral, eficientes na navegar e localização de nós específicos. No entanto, eles não oferecem uma visão geral efetiva da estrutura hierárquica e não auxiliam o usuário na formação de uma imagem mental da hierarquia.

Rivadeneira e Bederson (2003) sugerem que tais interfaces poderiam ser melhoradas com a apresentação de pistas de navegação que informariam os usuários quais elementos já foram visitados e pistas hierárquicas que poderiam informar aos usuários em que nível eles estão e quão profunda é a estrutura.

4.1.4 *Space Filling* (preenchimento de espaço)

As técnicas *Space Filling* buscam utilizar todo o espaço da tela, subdividindo-a entre os nós de uma hierarquia. O espaço atribuído a um nó-pai é subdividido entre os seus nós-filhos. O tamanho e a cor de cada subdivisão correspondem a uma propriedade da informação que se deseja representar. Um exemplo de utilização de técnicas de *Space Filling* é apresentado na Figura 20.

Figura 20 - Visualização de ontologia em interface do tipo *Space Filling*



Fonte <http://alloveralbany.com/archive/2016/11/09/new-york-state-presidential-election-2016-county-r>

A Figura 20 mostra uma representação do resultado das eleições presidenciais americanas de 2016 no estado de Nova York. O tamanho de cada quadrado representa a porcentagem de votos totais de cada condado. Os tons de azul representam a quantidade de votos de Hillary Clinton e os tons de vermelho representam os votos a Donald Trump.

As técnicas de *Space Filling* são particularmente adequadas para a representação visual de certas propriedades de uma ontologia ao nível de instância. São eficazes quando o usuário se preocupa principalmente com as propriedades dos nós de nível hierárquico mais

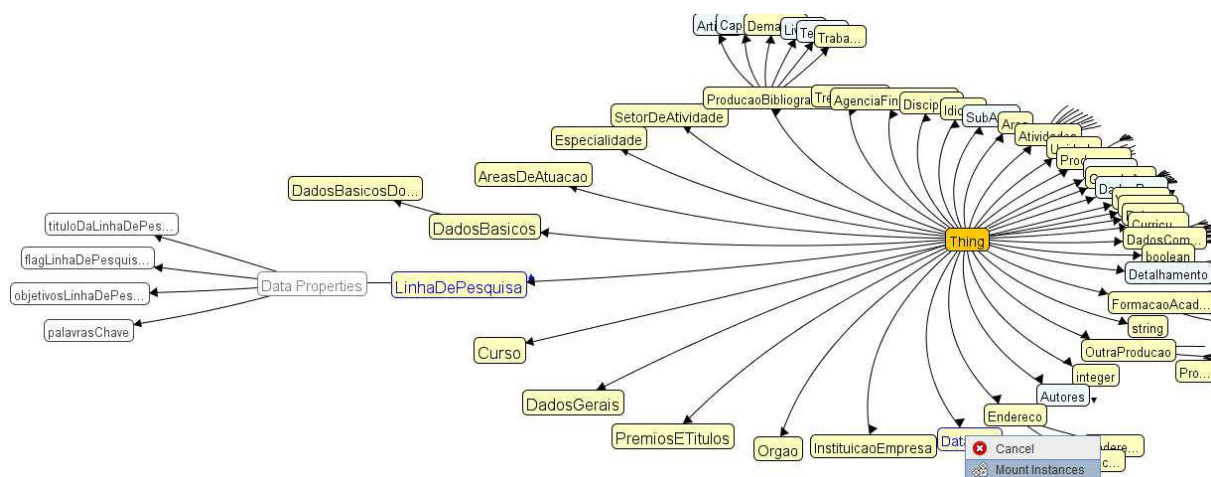
baixo, sem se ater à estrutura como um todo. Assim, essas técnicas não são eficazes para representar a estrutura de uma ontologia, tornando difícil reconstruir sua hierárquica.

4.1.5 *Context+Focus and Distortion Techniques (contexto+foco e distorção)*

Este grupo de técnicas baseia-se na noção de distorção da visão do gráfico apresentado para combinar contexto e foco. Geralmente o nó central é apresentado em destaque e os outros nós em torno dele são reduzidos ou distorcidos. A técnica mais utilizada nesse grupo é a árvore hiperbólica (Figura 21).

Em uma árvore hiperbólica geralmente a raiz da árvore é colocada inicialmente no centro de uma área circular com os seus nós-filho ao redor. Um nó de interesse pode ser movido para o centro da árvore para ser exibido com mais detalhes, mantendo o contexto dos nós relacionados à ele.

Figura 21 - Visualização de ontologia em interface do tipo *Context+Focus*



Fonte: PIZZINATO; VIEIRA; RIGO, 2010

As técnicas *focus+context* permitem uma visualização geral da ontologia e exibem vários nós ao mesmo tempo. Permitem ainda a visualização detalhada de um determinado nó, sem que se perca o contexto onde ele está inserido na hierarquia. Permite ainda uma navegação e localização rápida de classes ou instâncias específicas. No entanto, eles não oferecem uma representação muito óbvia de toda a sua estrutura hierárquica.

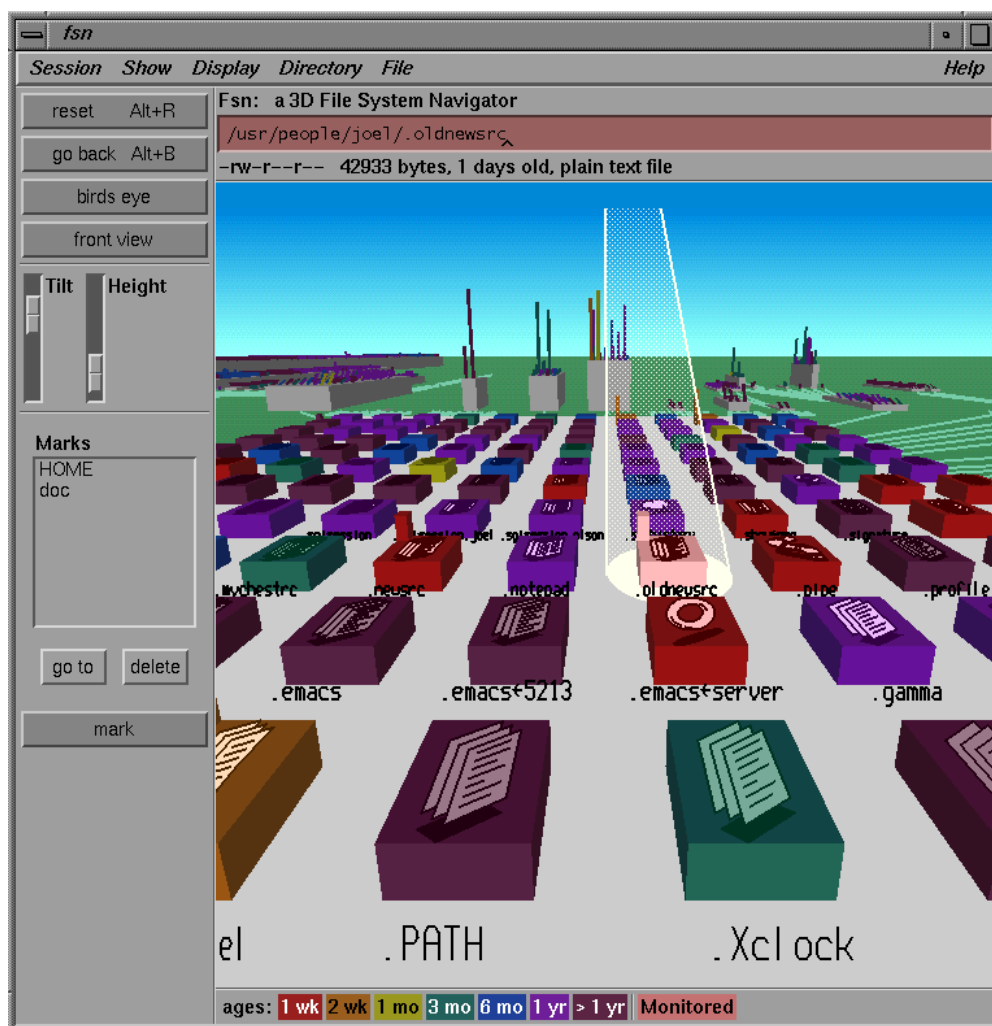
4.1.6 *Information Landscapes (paisagens informacional)*

Information Landscape é uma metáfora muito utilizada em ambientes de Realidade Virtual para gerenciamento de documentos. Os documentos, representados por objetos tridimensionais, são dispostos em um plano em perspectiva horizontal, simulando uma

planície. Em aplicações de visualização de ontologias, a tridimensionalidade pode ser utilizada para fornecer uma dimensão extra onde as propriedades dos nós poderiam ser codificadas e as relações apresentadas.

Uma ferramenta que utiliza esta técnica é o *File System Navigator* (FSN) (Figura 22). O FSN foi criado como um explorador de arquivos 3D para sistemas UNIX¹. A altura dos nós representa o número de arquivos contidos em uma pasta. Olhando de cima, os nós formam uma árvore bidimensional, que representa a hierarquia.

Figura 22 - Visualização de ontologia em interface do tipo *Information Landscape*



Fonte: HERMAN; MELANÇON; MARSHALL, 2000

Embora as interfaces do tipo *Information Landscape* sejam atrativas e úteis em diversos contextos, não é muito claro se podem ser úteis para representar ontologias. Poucos

¹ **UNIX** é um sistema operacional multitarefa e multiusuário criado por Ken Thompson e Dennis Ritchie na década de 1970.

experimentos práticos foram feitos, sendo, portanto, escassas as avaliações extensivas sobre a visualização de ontologias utilizando estas técnicas.

4.2 VOWL

Visual Notation for OWL Ontologies (VOWL) é um sistema baseada em apenas um conjunto de primitivas gráficas que formam o alfabeto de um idioma visual: as classes são representadas como círculos que são conectados por linhas e setas que representam as relações. Além desses elementos, a VOWL permite a representação gráfica da maioria dos elementos de uma ontologia.

Os blocos básicos de construção são um conjunto de primitivas gráficas e um esquema de cores para uma série de conceitos OWL. Os rótulos de propriedades e os tipos de dados são mostrados em retângulos. As informações sobre indivíduos e valores de dados são exibidas na própria visualização ou em outra parte da interface do usuário. Um segundo ingrediente do VOWL é um esquema de cores que complementa as primitivas gráficas. As cores permitem uma distinção fácil entre diferentes tipos de classes e propriedades.

O alfabeto da linguagem visual é formado por apenas um conjunto de primitivas e recursos gráficos. O Quadro 1(a) apresenta a lista dessas primitivas e os elementos da ontologia aos quais são aplicados.

Quadro 1 – Primitivas gráficas e esquema de cores do VOWL

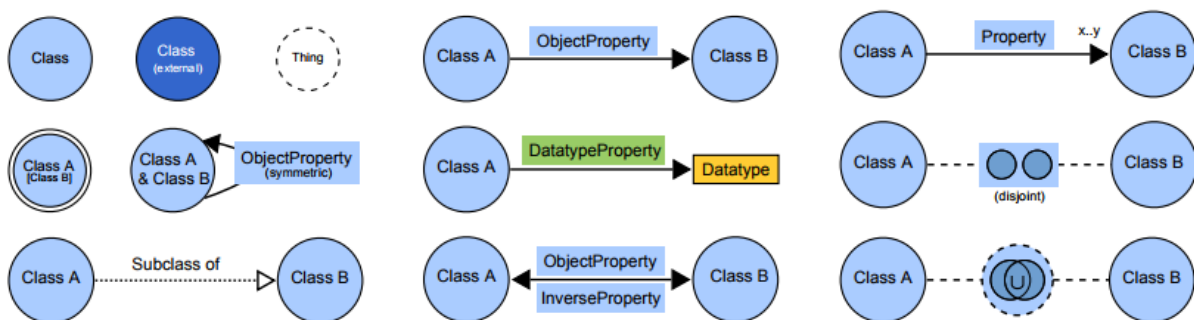
(a) Primitivas gráficas		(b) Esquema de cores	
Primitiva	Aplicação	Cor	Aplicação
	Classes		Classes, propriedade de objetos, disjunção
	Propriedades.		Elementos RDF e RDF Schema
	Direção das propriedades		Classes e propriedades externas
	Tipos de dados, rótulo de propriedades		Classes e propriedades obsoletas
 	Classes e propriedades especiais		Tipos de dados, literais
texto número símbolo	Rótulos e cardinalidade		Propriedades de tipo de dados
			Elementos em destaque (selecionado)

Fonte: Adaptado de LOHMANN *et al.*, 2016, p.5

Estudos mostraram que as cores são muito úteis na identificação dos diferentes elementos de uma ontologia. Por essa razão o esquema de cores define claramente o primeiro plano e fundo para todos os elementos. O Quadro 1(b) apresenta a relação de cores e os respectivos elementos da ontologia aos quais elas são aplicadas.

A especificação VOWL inclui representações gráficas para os elementos OWL mais comuns com base nas primitivas e cores mencionadas acima. Alguns desses elementos são mostrados na Figura 23.

Figura 23 – Alguns elementos visuais da especificação VOWL



Fonte: LOHMANN et al., 2016, p.6

Os elementos visuais são combinados compondo um grafo representando toda a ontologia. Os grafos VOWL são visualizados usando um grafo baseado em força (*force-directed graph*)².

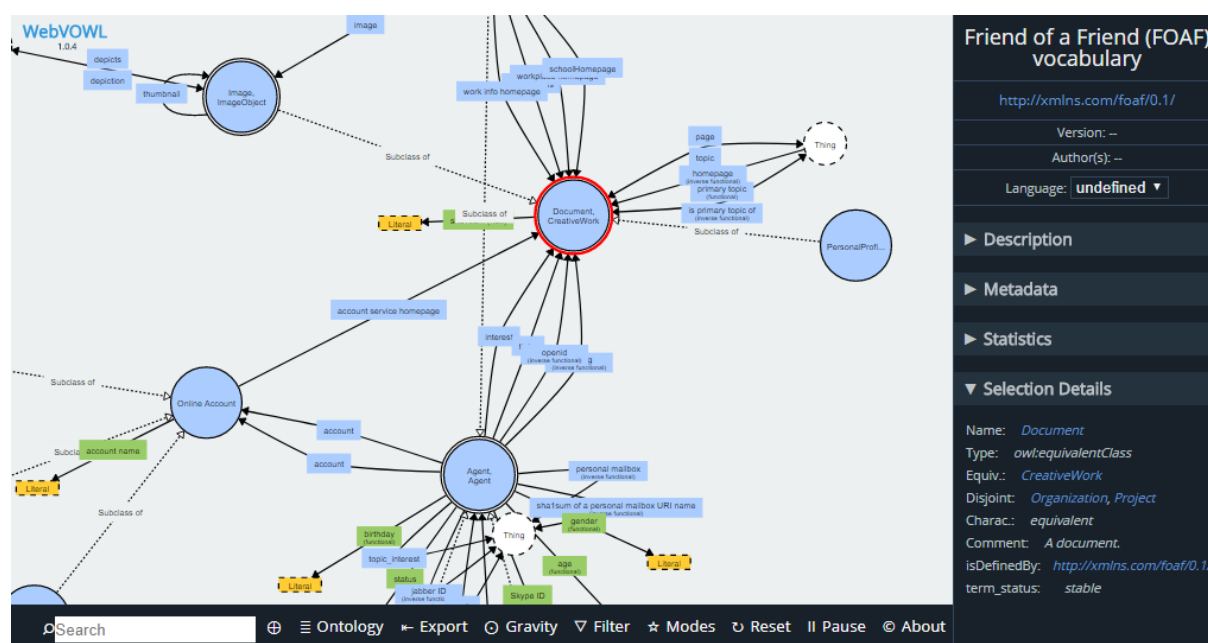
² Algoritmos baseados em força constituem uma das famílias clássicas de algoritmos de desenho de grafos. Sendo uma subclasse dos layouts baseados em otimização, esses algoritmos tratam o grafo como um sistema físico, atribuindo forças aos nodos e arestas e minimizando a energia até que um layout estável seja alcançado.

selecionados. O *layout* baseado em força pode ser adaptado, pois as forças de atração entre nós podem ser modificadas e o *layout* automático pode ser desativado em favor de um reposicionamento manual de nós.

As interfaces de usuário de ambas as ferramentas consistem-se de três partes: um visualizador na área maior, uma barra lateral apresentando detalhes sobre o elemento selecionado no momento e os controles que permitem ajustar o *layout* do grafo.

Na Figura 25 apresenta o WebVOWL com a visualização de parte do vocabulário FOAF. WebVOWL é a primeira ferramenta para visualização de ontologias completamente baseada em padrões abertos.

Figura 25 – Imagem do WebVOWL com a visualização do vocabulário FOAF



Fonte: Elaborado pela autora com o uso da ferramenta WebVOWL

Com base em estudo de usuários realizado com a primeira versão do VOWL, a versão atual chegou-se à notação aqui demonstrada. A VOWL 2 é uma linguagem visual que pode ser utilizada por usuários leigos e experientes em ontologias.

A VOWL2 será utilizada como base para a proposta de interface apresentada no Capítulo 5, porém com um conjunto de primitivas gráficas restrito a apenas 2 elementos.

5.

Proposta de Interface de Resultados de Busca

Nesse capítulo será apresentado um modelo de interface de recuperação de informação no qual a estrutura terminológica de uma ontologia é utilizada como base para a apresentação dos resultados de busca. Considera-se um acervo onde cada documento é representado por um vetor, tal como no Modelo Vetorial.

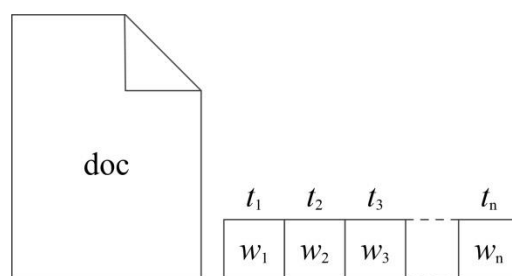
Nos exemplos apresentados neste capítulo será utilizado uma ontologia de termos de pediatria (PedTerm) que apresenta uma terminologia relacionada à saúde e ao desenvolvimento infantil desde o pré-natal até os 21 anos de idade. Essa ontologia é de acesso gratuito e está disponível no BioPortal⁴, um grande repositório de ontologias na área biomédica. Como os conceitos dessa ontologia estão no idioma inglês, os exemplos apresentarão uma tradução desses conceitos, porém a estrutura da ontologia não será alterada.

5.1 Representação Vetorial de um Documento

No modelo vetorial, um documento é representado por um vetor onde cada elemento representa o peso, ou a relevância, do respectivo termo de indexação para a representação do documento. Cada elemento do vetor é normalizado de forma a assumir valores entre zero e um. Os pesos mais próximos de um (1) indicam termos com maior importância para a descrição do conteúdo informacional do documento. A Figura 26 apresenta uma ilustração genérica de um documento que será utilizada no transcorrer desse capítulo.

⁴<https://bioportal.bioontology.org/ontologies/PEDTERM>

Figura 26 – Exemplo genérica de um documento e sua representação vetorial

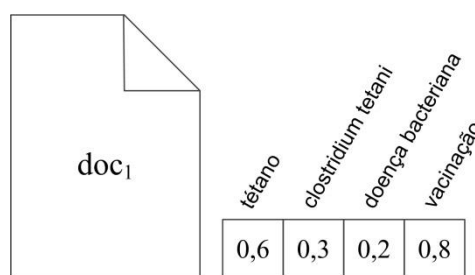


Fonte: desenvolvido pela autora

Na Figura 26, o documento **doc** é representado (indexado) pelos termos $t_1, t_2, t_3, \dots t_n$. A cada termo está associado um valor numérico entre zero e um ($w_1, w_2, w_3 \dots w_n$) que representa a importância ou relevância relativa do termo na representação do documento.

Na Figura 27 é apresentado um exemplo literal de um documento (**doc₁**).

Figura 27 – Exemplo da descrição vetorial de um documento



Fonte: desenvolvido pela autora

No exemplo da Figura 27 é possível presumir que **doc₁** é um documento que trata sobre **vacinação (0,8)** de **tétano (0,6)**, que é uma **doença bacteriana (0,2)** causada pela toxina **clostridium tetani (0,3)**.

O processo de indexação, no qual os vetores representativos de cada documento serão criados, não está no escopo deste trabalho. Assim, assume-se um *corpus* documental já previamente indexado.

5.2 Representação da Expressão de Busca

A expressão de busca é o meio que o usuário emprega para comunicar a sua necessidade para o sistema de recuperação de informação. É geralmente composta por um conjunto de palavras (ou termos) escolhidas livremente pelo usuário com o objetivo de expressar linguisticamente a sua necessidade de informação.

A principal dificuldade do usuário nesse processo está em prever quais termos foram usados na representação dos documentos que satisfarão sua necessidade e ao mesmo tempo evitar termos que possam recuperação de documentos não relevantes.

Na interface proposta a busca deverá ser expressa em duas etapas, utilizando dois campos a ser preenchido pelo usuário. Inicialmente o usuário terá que definir o assunto ou tema de seu interesse a partir de uma lista previamente definida. Cada item dessa lista refere-se a uma ontologia armazenada no sistema. Ao definir o tema, o usuário estará especificando a ontologia que será utilizada para auxiliá-lo em sua busca. Em seguida, o usuário especifica a sua expressão de busca, composta por um conjunto de palavras ou termos que irão representar a sua necessidade de informação. Uma imagem ilustrativa da interface de busca proposta é apresentada na Figura 28.

Figura 28 – Especificação de uma expressão de busca

A interface de busca é composta por dois campos principais. O primeiro campo, rotulado 'Tema/Assunto', é uma lista suspensa com o item 'Pediatria (PedTerm)' selecionado. O segundo campo é uma barra de texto contendo 'vacinação tétano' e um botão de busca com um ícone de lupa.

Fonte: desenvolvido pela autora

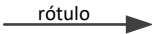
A Figura 28 apresenta um exemplo de uma busca hipotética na qual o usuário selecionou “Pediatria (PedTerm)” de uma lista de temas (ontologias) armazenados e registrados no sistema. Em seguida o usuário digitou as palavras (termos de busca) “vacinação” e “tétano” para representar a sua necessidade de informação.

5.3 Representação Gráfica de uma Ontologia

A ontologia definida na busca será apresentada graficamente ao usuário a fim de auxiliá-lo na navegação pelos conceitos do domínio da ontologia, permitindo recuperar documentos relacionados a esses conceitos. Para esse propósito, é necessário definir um uma notação gráfica para cada elemento de uma ontologia OWL. A notação utilizada nos exemplos apresentados a seguir é uma adaptação simplificada da notação adotada pelo sistema VOWL (LOHMANN et al, 2016).

Os elementos básicos (primitivas) para a representação visual de uma ontologia OWL são apresentados no Quadro 2.

Quadro 2 – Primitivas gráficas utilizadas no modelo proposto

Elemento	Primitiva gráfica
Classes	
Propriedades e relacionamentos	

Fonte: adaptado de LOHMANN et al, 2016

Neste trabalho serão utilizados apenas dois elementos para representar visualmente uma ontologia OWL. Uma classe será representada por um círculo rotulado com a sua identificação (ClassID). Propriedades e relacionamentos serão representados por uma seta rotulada com o nome da propriedade ou relacionamento. Na Figura 29 é apresentada uma hierarquia de classes em OWL e a sua correspondente representação gráfica.

Figura 29 – Ontologia OWL e sua representação gráfica

```

<owl:Class rdf:ID="Doenca_ou_disturbio">

</owl:Class>

<owl:Class rdf:ID="Doenca_infecciosa">

  <rdfs:subClassOf rdf:resource="#Doenca_ou_disturbio"/>

</owl:Class>

<owl:Class rdf:ID="Doenca_bacteriana">

  <rdfs:subClassOf rdf:resource="#Doenca_infecciosa"/>

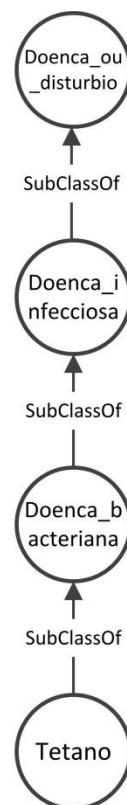
</owl:Class>

<owl:Class rdf:ID="Tetano">

  <rdfs:subClassOf rdf:resource="#Doenca_bacteriana"/>

</owl:Class>

```



Fonte: desenvolvido pela autora

5.4 Representação de Relevância

O conceito de relevância é crucial na área de Recuperação de Informação, sendo muitas vezes utilizado na própria enunciação dos objetivos dessa área (COOPER, 1971; MIZZARO, 1998). Nas definições do termo “relevância” disponíveis nos diversos dicionários *on-line*, esse termo é tratado, na maioria das vezes, como sinônimo de “importância”.

O iDicionário Aulete⁵ traz a seguinte definição:

(re.le.vân.ci:a)

sf.

1. Qualidade ou característica do que é relevante; IMPORTÂNCIA: *temas da maior relevância*. [Antôn.: irrelevância.]

[F.: relevar+ -ância.]

No dicionário Michaelis⁶ tem-se:

re·le·vân·ci·a

sf

1 Qualidade do que é relevante.

2 Aquilo que tem importância ou relevo num contexto determinado; pertinência: “*Os temas transversais constituíram um conjunto de temas de grande relevância para uma educação que visa a formação de sujeitos realmente capazes de conviver em harmonia*” (NG).

3 Parte saliente de uma superfície plana.

No dicionário Aurélio⁷ *on-line* tem-se ainda as seguintes definições de relevância:

Relevo. Importância; lado vantajoso de alguma coisa.

O necessário, o indispensável. Que releva. Que sobressai. Importante.

Embora não esteja declarado explicitamente em suas definições de dicionário, é possível inferir que o conceito de relevância é inerentemente subjetivo e relativo a um indivíduo, a uma pessoa que julgue a importância de algo para si.

Na proposta de interface apresentada neste trabalho, toma-se uma ontologia como base terminológica na qual os documentos resultantes de uma busca serão apresentados. Os documentos “orbitarão” os conceitos da ontologia que coincidem com os termos de busca. A relevância de cada documento é representada pela sua distância em relação ao conceito da ontologia.

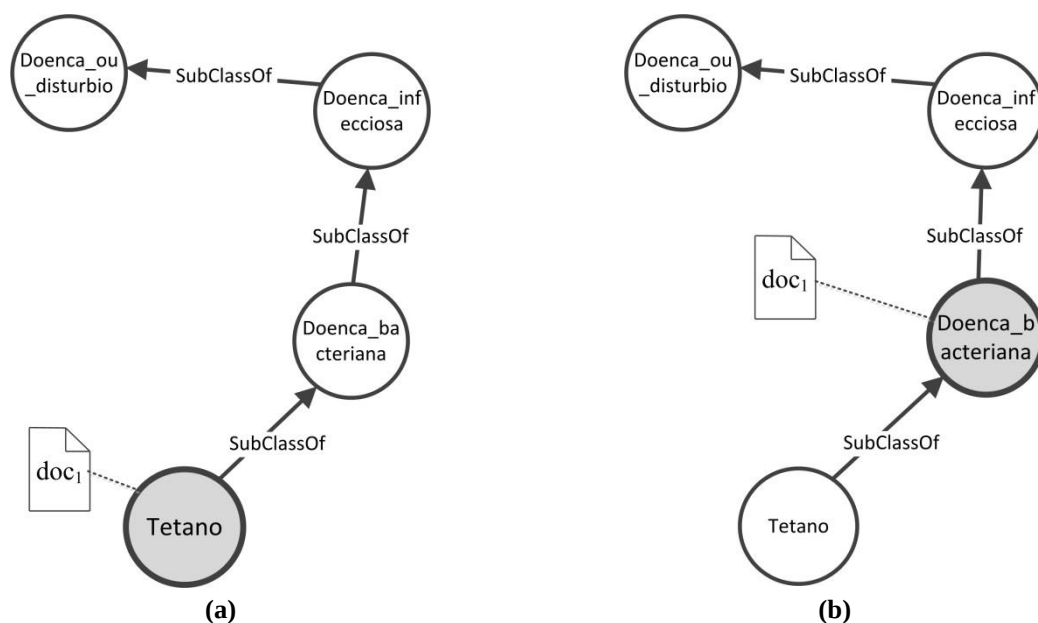
Dado um termo (conceito) da ontologia, os documentos que são indexados por esse termo serão posicionados ao seu redor, a uma distância inversamente proporcional ao valor do peso do correspondente termo de indexação. A Figura 30 apresenta um exemplo utilizando o **doc₁** (Figura 27), juntamente com a representação gráfica do trecho da ontologia da Figura 29.

⁵ <http://aulete.uol.com.br/ontologia>.

⁶ <http://michaelis.uol.com.br/>

⁷ <https://dicionariodoaurelio.com/>

Figura 30 – Exemplo da Representação de Relevância



Fonte: desenvolvido pela autora

O conceito de relevância é representado pela distância de um documento em relação a um determinado termo da ontologia. Na Figura 30a o termo de referência (selecionado) é “Tetano”. O documento **doc₁** é indexado por esse termo com peso igual a 0,6. Por meio desse peso é calculada a sua distância em relação ao termo da ontologia.

Na Figura 30b o termo de referência é “Doenca_bacteriana”. O documento **doc₁** é indexado por esse termo com peso igual a 0,2 e por isso aparece “orbitando” o termo da ontologia a uma distância maior. A distância do documento em relação ao termo da ontologia é **inversamente proporcional** ao peso do termo de indexação. Quanto maior o peso de um termo de indexação, menor será a distância do documento em relação ao conceito da ontologia.

Para não sobrecarregar o exemplo da Figura 30, foi utilizado um único documento (**doc₁**) e a seleção de um único termo na ontologia. Na próxima seção será apresentado um exemplo completo do modelo de interface proposto nesse trabalho.

5.5 Modelo de interface de resultados de busca

Nesta sessão serão exemplificados os passos do processo de recuperação de informação e da interface proposta nesse trabalho.

A fim de exemplificar o modelo de interface aqui proposto, será considerado o seguinte *corpus* documental, apresentado na Tabela 1

Tabela 1 – Documentos utilizados nos exemplos do modelo proposto

documento	Termos de Indexação				
	tétano	clostridium tetani	doença bacteriana	vacinação	doença infecciosa
doc ₁	0,6	0,3	0,5	0,8	-
doc ₂	0,3	0,8	0,3	-	0,3
doc ₃	0,9	0,2	-	0,8	-
doc ₄	0,2	-	0,6	-	0,8

Nos exemplos apresentados a seguir serão considerados esses quatro documentos representados por seus respectivos vetores.

5.5.1 Expressão de busca

Embora não seja assunto desse trabalho, é importante enfatizar a necessidade do usuário em contextualizar a sua busca por meio da escolha do tema ou assunto de interesse. O tema escolhido estará vinculado a uma ontologia que será o elemento principal da interface de apresentação e seleção dos documentos resultantes da sua busca.

Como mencionado anteriormente, a especificação da busca é feita por meio de dois campos. No primeiro campo é realizada a escolha do tema ou assunto de interesse do usuário do sistema. Cada um desses assuntos está ligado diretamente a uma ontologia do domínio do referido assunto. A ideia, portanto, é que o sistema de recuperação tenha previamente cadastrado um conjunto de ontologias. A definição do assunto/tema permite restringir a terminologia envolvida na busca, diminuindo problemas linguísticos inerentes ao processo de recuperação de informação, tal como a ambiguidade e a polissemia.

Figura 31 – Exemplo da especificação de uma expressão de busca

Tema/Assunto

Pediatria (PedTerm)

▼

tétano “doença infecciosa”



Fonte: desenvolvido pela autora

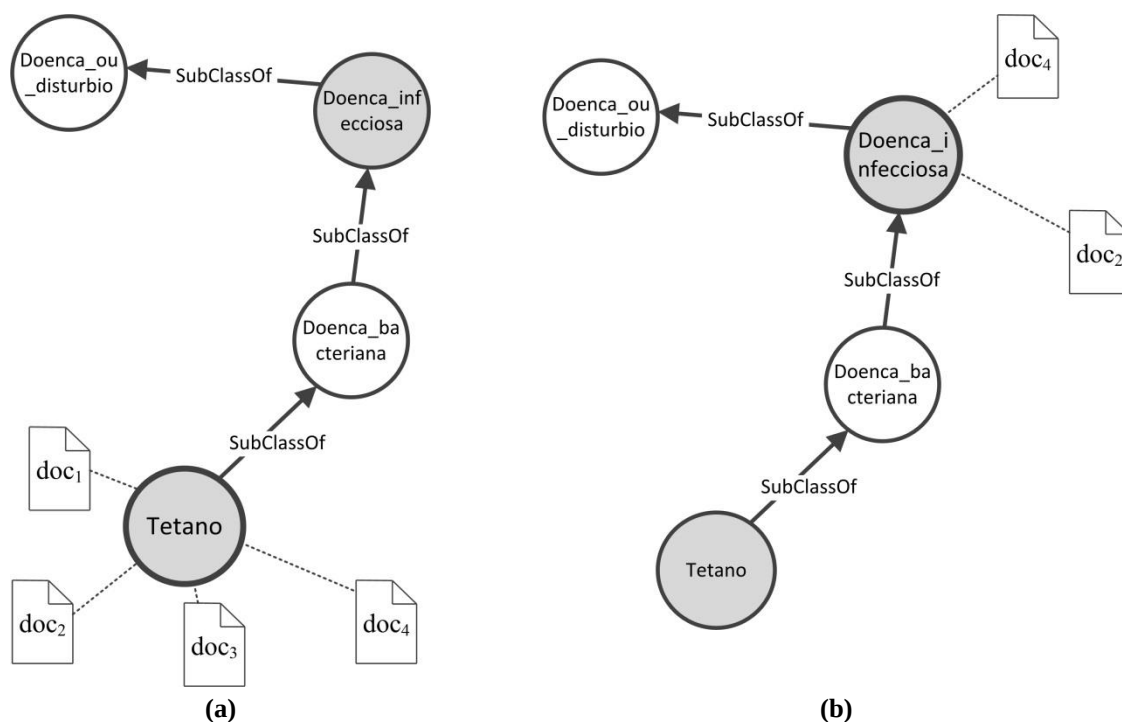
Após escolher o tema ou assunto, o usuário define um conjunto de termos que irão compor a sua expressão de busca. No exemplo acima, o assunto de interesse do usuário é “Pediatria” e a expressão de busca é composta pelos termos “tétano” e “doença infecciosa”. Após efetuar a sua busca, o usuário terá como resultado a representação gráfica e dinâmica da

ontologia do assunto escolhido com algum tipo de sinalização dos termos utilizados na sua busca.

5.5.2 Resultado da busca

Como resultado de uma busca, será apresentada ao usuário a representação gráfica da ontologia referente ao assunto/tema escolhido, com algum tipo de sinalização dos termos coincidentes aos termos utilizados na expressão de busca. Ao selecionar (clicar) um determinado termo sinalizado na ontologia, serão apresentados os documentos indexados pelo termo. Os documentos irão “orbitar” o termo da ontologia a uma distância inversamente proporcional ao peso do termo no vetor que representa cada documento.

Figura 32 – Exemplo de apresentação dos documentos resultantes de uma busca



Fonte: desenvolvido pela autora

Na Figura 32a o usuário selecionou na ontologia o termo “Tetano”. No *corpus* existem quatro documentos indexados por esse termo, com diferentes pesos. Esses pesos são utilizados na representação da relevância de cada documento para esse termo. A relevância, por sua vez, é representada pela distância do documento em relação ao termo selecionado. No documento **doc3** o termo “tétano” possui peso 0,9, por essa razão ele é representado bem próximo ao termo da ontologia. No documento **doc4** o termo “tétano” tem peso 0,2 e está posicionado mais distante do termo selecionado (“Tetano”).

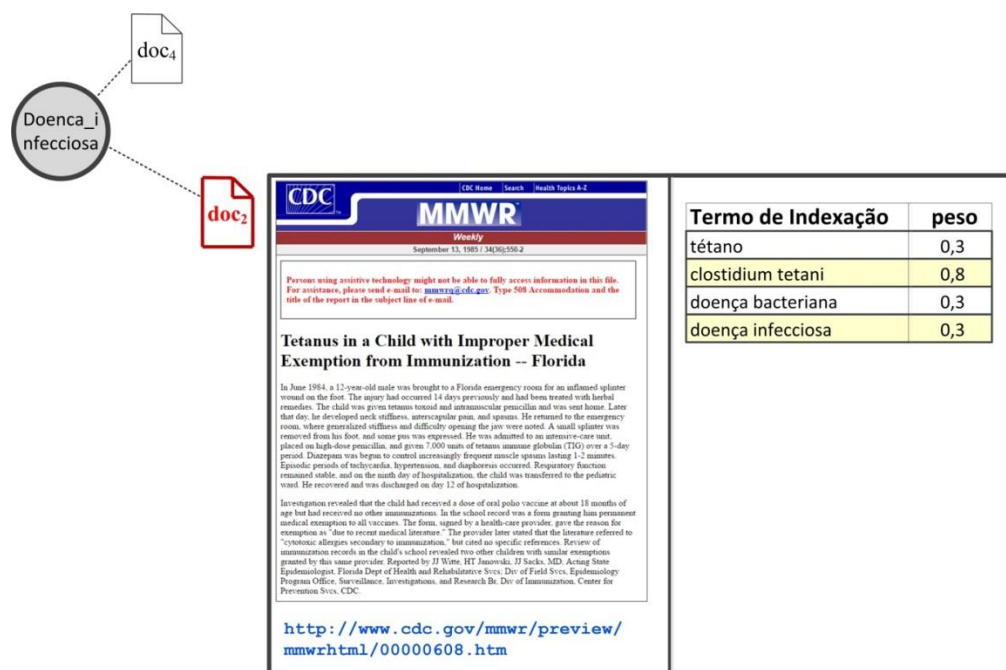
Na Figura 32b o termo selecionado é “Doença_infecciosa”. Existem apenas dois documentos indexados por esse termo: **doc2** e **doc4**, com pesos 0,3 e 0,8 respectivamente. Assim, **doc2** aparece mais distante e **doc4** aparece mais próximo ao termo selecionado na ontologia.

Ao selecionar um determinado conceito da ontologia, somente os documentos relacionados a esse conceito serão apresentados; os documentos que orbitam outro conceito não serão mais apresentados. Na Figura 32b, ao ser selecionado o conceito “Doença_infecciosa” serão apresentados apenas os documentos indexados por esse termo, os documentos que “orbitam” o conceito “Tetano” não serão mais apresentados ao usuário, evitando assim uma sobrecarga de informação na interface.

5.5.3 Visualiação da representação de um documento

O julgamento da relevância de um determinado documento só é possível se o usuário tiver conhecimento de seu conteúdo e/ou da forma como ele está representado (indexado). A Figura 33 apresenta um exemplo de apresentação dos detalhes de um documento.

Figura 33 – Exemplo de apresentação dos detalhes de um documento



Fonte: desenvolvido pela autora

A Figura 33 utiliza um trecho da Figura 32b na qual o documento **doc2** está selecionado. Por meio de um comando via teclado ou botão do mouse, é apresentado um trecho ou todo o documento em uma janela *pop-up* ou algum outro recurso semelhante. No

exemplo, é apresentado à esquerda um trecho do documento, no caso um documento HTML e seu endereço (URL). À direita é apresentado o vetor representativo do documento, com seus os termos de indexação e os respectivos pesos.

Independente do recurso a ser implementado em um sistema de recuperação de informação, a apresentação de parte ou de todo o conteúdo dos documentos resultantes de uma busca é de fundamental importância, pois é o usuário quem irá julgar se um documento é ou não relevante para os seus interesses. É importante também que o usuário tenha acesso à forma como um determinado documento foi indexado: seus termos de indexação com seus respectivos termos. Dessa forma ele terá clareza das razões pelas quais um determinado documento foi recuperado pelo sistema.

6.

Conclusões

A quantidade crescente de informação disponível na Web traz para os sistemas de recuperação de informação a responsabilidade de filtrar desse imenso repositório, informações relevantes e úteis para uma ampla gama de interesses e variados tipos de usuário.

Nesse contexto, as interfaces de busca carregam a responsabilidade de proporcionar ao usuário um meio de expressar adequadamente a sua necessidade de informação e prover um ambiente eficiente para selecionar os documentos realmente relevantes para seus interesses.

Assim como a maioria das atividades humanas, o processo de recuperação de informação é uma tarefa que envolve linguagem. Por um lado, os documentos são representados por expressões linguísticas (termos de indexação) que representam o conteúdo informacional dos documentos. Por outro lado, os usuários expressam linguisticamente a sua necessidade de informação por meio de um conjunto de termos de busca. A eficiência da recuperação de informação relaciona-se, portanto, com as coincidências terminológicas entre documentos e busca.

Esse trabalho apresentou um modelo de interface para sistemas de recuperação de informação na qual a estrutura terminológica (conceitos) de uma ontologia é utilizada como “alicerce” sobre o qual se assenta a visualização dos documentos resultantes de uma busca. Altera-se assim a lógica presente nos atuais na maioria dos sistemas de recuperação de informação e das ferramentas de busca na Web, onde os documentos são destaque.

No modelo proposto a ênfase recai na terminologia da área de interesse do usuário, apresentada de forma visual e dinâmica, por meio da qual ele irá selecionar os documentos que considerar relevantes.

Embora este trabalho não tivesse a pretensão do desenvolvido de uma aplicação, acreditamos que o modelo proposto aqui possa futuramente servir de base para a construção de um sistema computacional completo.

Referências

- BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern Information Retrieval**. 2^a ed. Addison-Wesley, 2011.
- BUCKLAND, M.K. Information as thing. **Journal of the American Society of Information Science**, v.42, n.5, 1991. p.351-360.
- CAMPOS, M.L.A. O papel das definições na pesquisa em ontologia. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 15, n. 1, p.220-238, jan./abr. 2010
- CHANDRASEKARAN, B.; JOSEPHSON, J.R.; BENJAMINS, V. R. What are ontologies, and why do we need them? **IEEE Intelligent Systems**, v.14, n.1, 1999.
- COOPER, W.S. A Definition of Relevance for Information Retrieval. **Information Storage and Retrieval**, v.7, pp.19-37, 1971.
- FERNEDA, E. **Introdução Modelos Computacionais de Recuperação de Informação**. Rio de Janeiro: Ciência Moderna, 2012.
- FERNEDA, E. **Ontologia como recurso de padronização terminológica em um Sistema de Recuperação de Informação**. 2013. 109 f. Relatório (Estágio Pós-Doutoral), Ciência da Informação, Universidade Federal da Paraíba, João Pessoa. 2013.
- FILHO, F.W.; LÓSCIO, B.F. Web Semântica: Conceitos e Tecnologias. In: Pedro de Alcantara. (Org.). **II Escola Regional de Computação - Ceará, Maranhão e Piauí - ERCEMAPI 2009**. : SBC, 2009.
- GRUBER, T. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. **International Journal Human-Computer Studies**, v.43, n.5-6, 1995.
- GUIMARÃES, F.J.Z. **Utilização de ontologias no domínio B2C**. 2002. 195 f. Dissertação (Mestrado em Informática), Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2002.
- HERMAN, I.; MELANÇON, G.; MARSHALL, M.S. Graph visualization and navigation in information visualization: A survey. **IEEE Transactions on Visualization and Computer Graphics**, v.6, n.1. p.24-43

- HOBBS V.; PFITZNER, D.; POWERS, D. **A Survey of Information Retrieval Interfaces.** (200-) Disponível em
<https://www.academia.edu/18544734/A_Survey_of_Information_Retrieval_Interfaces>,
Acessado em 30/07/2017.
- JASPER, R.; USCHOLD, M. A Framework for understanding and classifying ontology applications. In: **KRR5-99**, Stockholm. 1999.
- KATIFORI, A.; HALATSIS, C.; LEPOURAS, C.; VASSILAKIS, C; GIANNOPOULOU, E, et al. Ontology visualization methods: a survey. **ACM Computing Surveys (CSUR)**, v. 39, n. 4, p. 10-43, mar. 2007.
- KRIGLSTEIN, S.; WALLNER, G. Development process and evaluation of the ontology visualization tool knoocks: a case study. In: **International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP/IVAPP)**, 2011, Faro. Preceedings. Faro: SciTePress, 2011. p. 187-197.
- KOSHMAN, S. Visualization-based information retrieval on the Web. **Library & Information Science Research**. 28. p.192-207, 2006.
- KURAMOTO, Hélio. Sintagmas nominais: uma nova proposta para a recuperação de informação. **DataGramaZero - Revista de Ciência da Informação**, Rio de Janeiro, v.3, n.1, fev. 2002.
- LIMA, J.C.; CARVALHO, C.L. **Ontologias-owl (web ontology language)**. Technical report, Universidade Federal de Goiás, 2005.
- LOHMANN, S.; NEGRU, S.; HAAG, F.; ERTL, T. Visualizing Ontologies with VOWL. **Semantic Web Journal**, v.7, n.4, pp. 399-419, 2016
- MANNING, C.D.; RAGHAVAN, P.; SCHÜTZE, H. **An Introduction to Information Retrieval**. Cambridge, England: Cambridge University Press, 2009. 581 p.
- MARCONDES, C., ALMEIDA CAMPOS, M. Ontologia e Web semântica: o espaço da pesquisa em Ciência da Informação. **Ponto de Acesso**, Salvador, v.2, n.1, p.107-136, 2008.
- MARON, M.E.; KUHNS, J.L. On relevance, probabilistic indexing and information retrieval. **Journal of the ACM**, v. 7, n. 3, p.216-244, 1960.
- MIZZARO, S. How many relevances in Information Retrieval? **Interacting with Computers**, v.10, n.3, p.303–320, 1998.
- MOOERS, C. N. Zatocoding applied to mechanical organization of knowledge. **American Documentation**, USA, v.2, p.20-32, 1951.

PICKLER, Maria Elisa Valentim. **Diretrizes para utilização de ontologias na indexação automática**. 2014. 101f. Dissertação (Mestrado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2014.

PIZZINATO, P.M.; RIGO, S.J.; VIEIRA, R.. Um Visualizador Online de Ontologias Utilizando Árvores Hiperbólicas. In: **Ontobras 2010** - Seminário de Pesquisa em Ontologia no Brasil, 2010.

RAMALHO, R.A.S. **Desenvolvimento e utilização de ontologias em bibliotecas digitais: uma proposta de aplicação**. 2010. 145 f. Tese (Doutorado) - Universidade Estadual Paulista, Faculdade de Filosofia e Ciências, 2010.

RIVADENEIRA, W.; BEDERSON, B. B. **A Study of Search Result Clustering Interfaces: Comparing Textual and Zoomable Interfaces**, University of Maryland HCIL Tech. Rep. HCIL-2003-36, October, 2003.

ROBERTSON, S.E.; JONES, K.S. Relevance weighting of search terms. **Journal of the American Society for Information Science**, v. 27, n. 3, p.129-146, 1976.

ROSE, D.E.; ORR, D.; KANTAMNENI, R.G.P. Summary attributes and perceived search quality. **Proceedings of the 16th international conference on World Wide Web (WWW'07)**, p.1201–1202, 2007.

SALES, R.; CAFÉ, L. **Semelhanças e Diferenças entre Tesouros e Ontologias**. DataGramaZero, Rio de Janeiro, v.9, n.4, ago. 2008.

SANTAREM SEGUNDO, J E. **Representação Iterativa: um modelo para repositórios digitais**. 2010. 224f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010.

SANTOS, I. E. dos. **Manual de métodos e técnicas de pesquisa científica**. 7.ed. Niterói: Impetus, 2010.

SEVERINO, Antônio Joaquim. **Metodologia do trabalho científico**. 23. ed. rev. atual. São Paulo: Cortez, 2007.

SOERGEL, D. The rise of ontologies or the reinvention of classification. **Journal of the American Society for Information Science**. v. 50, n. 12, 1999.

VICKERY, B. C. Ontologies. **Journal of Information Science**. v.23, n.4, 1997.

WHITE, R.W.; JOSE, J.; RUTHVEN, I. A task-oriented study on the influencing effects of query-biased summarisation in web searching. **Information Processing and Management**, v.39, n.5, 2003, p.707–733.