

UTILIZAÇÃO DE ONTOLOGIA COMO INTERFACE DE SISTEMAS DE RECUPERAÇÃO DE INFORMAÇÃO: UMA PROPOSTA BASEADA NO MODELO ESPAÇO VETORIAL

USE OF ONTOLOGY AS AN INTERFACE FOR INFORMATION RETRIEVAL SYSTEMS: A PROPOSAL BASED ON THE VECTOR SPACE MODEL

Jorge Janaite Neto^a

Edberto Ferneda^b

RESUMO

Objetivo: Utilizando o Modelo Espaço Vetorial como base teórica, este trabalho propõe um modelo de interface para sistemas de recuperação de informação na qual a estrutura terminológica de uma ontologia é utilizada para auxiliar o usuário na seleção de documentos relevantes dentre aqueles resultantes de uma busca. **Metodologia:** Caracteriza-se como pesquisa de natureza aplicada, exploratória e bibliográfica. A partir das características únicas do Modelo Vetorial, que permite calcular distâncias quantificáveis entre expressões de busca e documentos, e os aspectos linguísticos inerentes ao processo de recuperação de informação, articulou-se uma simbiose entre a estrutura terminológica das ontologias com potenciais recursos visuais desse Modelo. **Conclusões:** A apresentação visual da estrutura terminológica de uma ontologia, juntamente com a representação vetorial dos documentos de um corpus permite o desenvolvimento de interfaces dinâmicas e interativas, proporcionando ao usuário a navegação entre os documentos resultantes da sua busca, tendo por base os termos de uma determinada área de conhecimento.

Descritores: Recuperação de Informação. Modelo Espaço Vetorial. Ontologia. Interface de recuperação de informação. Interface baseada em ontologia.

1 INTRODUÇÃO

A quantidade crescente de informação disponível em meios digitais faz aumentar a importância dos sistemas de recuperação de informação. Porém, a

^a Doutor em Ciência da Informação pela Universidade Estadual Paulista (UNESP), Marília, Brasil. E-mail: jorge.janaite@unesp.br

^b Doutor em Ciências da Comunicação pela Universidade de São Paulo (USP). Docente da Universidade Estadual Paulista (UNESP), Marília, Brasil. E-mail: edberto.ferneda@unesp.br

maioria dos atuais sistemas possuem os mesmos recursos dos sistemas desenvolvidos na década de 1950, baseados na pesquisa por palavras-chave e no resultado em forma de uma listagem sequencial de documentos. Segundo Hobbs, Pfitzner e Powers (2002), as desvantagens desse tipo de sistema são: (1) os usuários devem inspecionar sequencialmente uma lista geralmente longa de documentos; (2) não é apresentado ao usuário a razão pela qual um determinado documento foi recuperado. (3) o usuário não é informado sobre a relação de um documento com a busca; (4) não são apresentadas as semelhanças ou similaridades entre os documentos recuperados, e (5) é exibido apenas um pequeno trecho do documento para que o usuário possa julgar a sua relevância.

Este trabalho tem por objetivo propor novos recursos para sistemas de recuperação de informação a partir de um modelo de interface na qual a terminologia de uma determinada área do conhecimento, apresentada graficamente a partir de uma ontologia OWL, é utilizada como estrutura básica para auxiliar o usuário na busca e na recuperação de documentos relevantes.

2 RECUPERAÇÃO DE INFORMAÇÃO

O processo de recuperação de informação envolve, por um lado, um acervo documental no qual cada documento é representado por um conjunto de termos cujos significados remetem aos assuntos tratados por ele. Por outro lado, os usuários representam 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 ambiente linguístico mediador da comunicação entre um estoque de informação e seus requisitantes. Sua eficiência depende de um controle adequado da linguagem de representação dos itens de informação e das requisições de seus usuários. Essas representações precisam ser descritas em uma estrutura que permita realizar comparações entre elas. A formalização dessa estrutura é denominada “modelo”.

Um modelo de recuperação de informação é a especificação formal de três elementos: a representação dos documentos, a representação da necessidade de informação do usuário e a forma como essas duas representações serão comparadas: a *função de busca* (Ferneda, 2012, p. 20). Os modelos de recuperação de informação estabelecem uma estrutura de representação uniforme para os documentos e para as buscas, o que permite calcular o grau de semelhança (similaridade) entre tais representações. Os documentos resultantes podem assim ser ordenados segundo o grau de similaridade com a expressão de busca.

A grande maioria dos modelos de recuperação de informação é baseada em disciplinas como a lógica, a estatística e a teoria dos conjuntos. Em um estudo sobre os modelos de recuperação de informação, Robertson (1977) justifica esse predomínio pelo fato de que a determinação de um modelo matemático geralmente pressupõe uma cuidadosa análise formal do problema e especificações de hipóteses, além de uma formulação explícita da forma como o modelo depende das hipóteses.

Os chamados “modelos clássicos” de recuperação de informação surgiram nas décadas de 1950 e 1960, quando o crescimento das informações armazenadas eletronicamente levou à necessidade de métodos eficientes para recuperá-las. Durante esse período, pesquisadores propuseram diferentes abordagens para organizar e acessar informações. O Modelo Booleano foi o primeiro a ser formalizado, seguido pelo Modelo Vetorial e, posteriormente, pelo Modelo Probabilístico, cada um deles com aprimoramentos progressivos em relação ao anterior no que diz respeito à precisão e à flexibilidade da recuperação (Baeza-Yates; Ribeiro-Neto, 2011, p. 61).

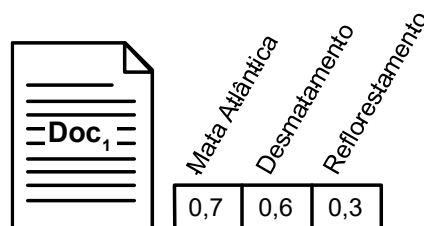
Nessa pesquisa adotou-se o Modelo Vetorial em função de suas características oferecerem uma representação uniforme (dos documentos e da expressão de busca) e baseada em distância, o que é fundamental para o desenvolvimento de interfaces visuais interativas, onde a relação geométrica entre documentos e consultas pode ser explorada e explicitada.

2.1 MODELO ESPAÇO VETORIAL

Segundo Salton (1989, p. 313), os modelos de recuperação de informação podem ser classificados em três categorias: (1) modelos booleanos: baseados na simples comparação entre os termos da busca e os termos de indexação dos documentos; (2) modelos probabilísticos: baseados na estimativa da probabilidade de relevância dos documentos em relação à busca; e (3) modelos espaço-vetoriais: representam tanto os documentos como as buscas por conjuntos de termos e seus respectivos pesos.

No Modelo Espaço Vetorial um documento é representado por um vetor no qual cada elemento determina o peso ou a importância do respectivo termo na representação do conteúdo informacional do documento. Cada elemento do vetor é normalizado de forma a assumir valores entre zero (0) e um (1). Os pesos mais próximos de um (1) indicam termos com maior importância ou mais relevantes para descrever o documento. A Figura 1 apresenta uma ilustração da representação vetorial de um documento Doc1, indexado por 3 termos: “mata Atlântica”, “desmatamento” e “reflorestamento”, com pesos 0,7, 0,6 e 0,3, respectivamente.

Figura 1– Representação Vetorial de um documento



Fonte: Elaborado pelos autores.

De forma semelhante, as expressões de busca também são representadas por um vetor numérico, onde cada valor representa a importância do termo de busca em relação à necessidade de informação do usuário. A utilização de uma mesma representação tanto para os documentos como para as buscas permite calcular o grau de similaridade entre o vetor que representa a busca e cada um dos vetores que representam os documentos do *corpus*.

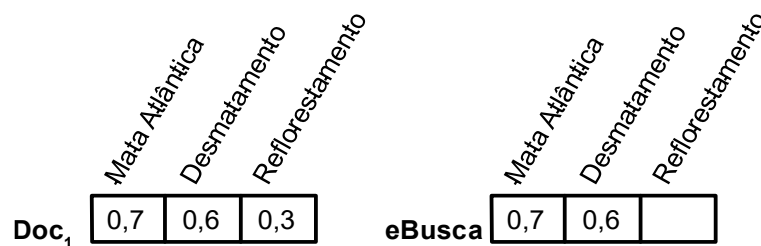
Em um *corpus* contendo **N** documentos, a similaridade (**sim**) entre um documento **d_i** e uma expressão de busca **q** é calculada utilizando a seguinte fórmula:

$$sim(d_i, q) = \frac{\sum_{k=1}^N (w_{k,i} \cdot w_{k,q})}{\sqrt{\sum_{k=1}^N w_{k,i}^2} \cdot \sqrt{\sum_{k=1}^N w_{k,q}^2}}$$

onde $w_{k,i}$ é o peso do *k*-ésimo elemento do vetor que representa o documento **d_i** e $w_{k,q}$ é o peso do *k*-ésimo elemento do vetor da expressão de busca **q**.

Aplicando essa fórmula ao documento **Doc₁** (Figura 1) e na expressão de busca **eBusca** (Figura 2), temos:

Figura 2 – Representação Vetorial da expressão de busca



Fonte: Elaborado pelos autores.

$$sim(Doc_1, eBusca) = \frac{(0,7 \times 0,8) + (0,6 \times 0,5)}{\sqrt{(0,7^2 + 0,6^2 + 0,3^2)} \times \sqrt{(0,8^2 + 0,5^2)}} \cong 0,94$$

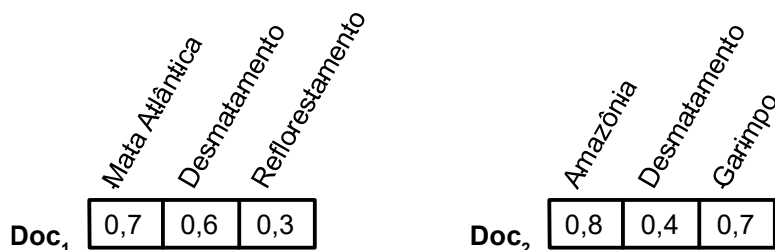
Portanto, a similaridade (**sim**) entre o documento **Doc₁** e a busca **eBusca** é de aproximadamente 0,94.

O valor da similaridade entre a expressão de busca e cada um dos documentos do *corpus* é utilizado no ordenamento dos documentos resultantes de uma busca.

No Modelo Vetorial existe uma absoluta homogeneidade nas representações de documentos e buscas. Essa característica permite não só o cálculo de similaridade entre uma expressão de busca e os documentos do

corpus, mas possibilita também calcular a similaridade em pares de documentos. Utilizando a mesma fórmula de similaridade (*sim*) é possível calcular a similaridade entre dois documentos, como ilustrado na Figura 3:

Figura 3 – Representação Vetorial de dois documentos



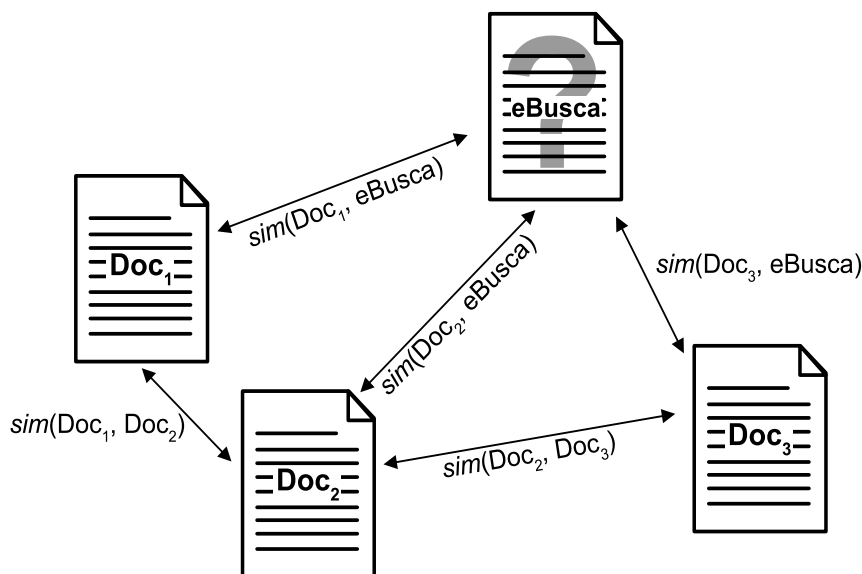
Fonte: Elaborado pelos autores.

$$sim(Doc_1, Doc_2) = \frac{(0,6 \times 0,4)}{\sqrt{(0,7^2 + 0,6^2 + 0,3^2)} \times \sqrt{(0,8^2 + 0,4^2 + 0,7^2)}} \cong 0,21$$

No exemplo acima, a similaridade (*sim*) entre o documento **Doc₁** e o documento **Doc₂** é de aproximadamente 0,21.

A similaridade entre dois documentos ou entre uma expressão de busca e um documento será visualmente representada pelo inverso da distância. Quanto maior a similaridade entre dois elementos, menor a distância entre eles. A Figura 4 ilustra essa característica, apresentando as relações de similaridade (*sim*) entre documentos e busca por meio da distância entre esses objetos.

Figura 4 - Cálculo de Similaridade entre Busca e Documentos



Fonte: Elaborado pelos autores.

A interpretação da distância como o inverso da similaridade entre os elementos envolvidos no processo de recuperação de informação traz diversas possibilidades para o desenvolvimento de visualizações de um acervo documental ou dos documentos resultantes de uma busca.

As características do Modelo Espaço Vetorial permitem uma representação espacial de documentos e buscas no interior de um n -dimensional (estando a magnitude desta dimensionalidade atrelada ao número de termos considerados entre todos os documentos), possibilitando o desenvolvimento de uma grande variedade de modelos de interfaces de recuperação de informação.

3 ONTOLOGIA

O termo ontologia tem sua origem na Filosofia, mas ganhou notoriedade na Ciência da Computação principalmente após o surgimento da Web Semântica. Nesse contexto, Gruber (1995) define ontologia como a especificação formal e explícita de uma conceitualização compartilhada. Formal diz respeito a ser definida seguindo os preceitos da Lógica; explícita indica que os elementos estão claramente definidos sem ambiguidades; conceitualização

refere-se a um modelo abstrato de um fenômeno; e compartilhada significa que os conceitos 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).

A criação de ontologias tem sido beneficiada pelo desenvolvimento de linguagens formais específicas para a descrição ontológica, tais como a *Web Ontology Language* (OWL) / *Web Ontology Language second edition* (OWL2), e de ferramentas especializadas, que oferecem recursos de criação e edição dos diversos elementos que constituem uma ontologia. No entanto, dada a complexidade dos domínios a serem representados, a estrutura de uma ontologia pode conter um conjunto significativo de elementos, tornando sua compreensão uma tarefa bastante árdua.

Com o objetivo de auxiliar na compreensão e no uso de ontologias, surgiram diversas ferramentas para a visualização de ontologias.

3.1 VISUALIZAÇÃO 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 sua compreensão, compartilhamento ou consenso sobre uma determinada área de conhecimento. Assim, as técnicas de visualização de informação podem ser utilizadas para auxiliar na compreensão do domínio representado pelas ontologias.

Katifori *et al.* (2007) apresentam seis diferentes métodos de visualização de ontologias, agrupados de acordo com diferentes características ou tipos de visualização. São eles: (1) *Indented list*; (2) *Node-link and tree*; (3) *Zoomable*; (4) *Space-filling*; (5) *Focus + context or distortion*; (6) *3D Information landscapes*.

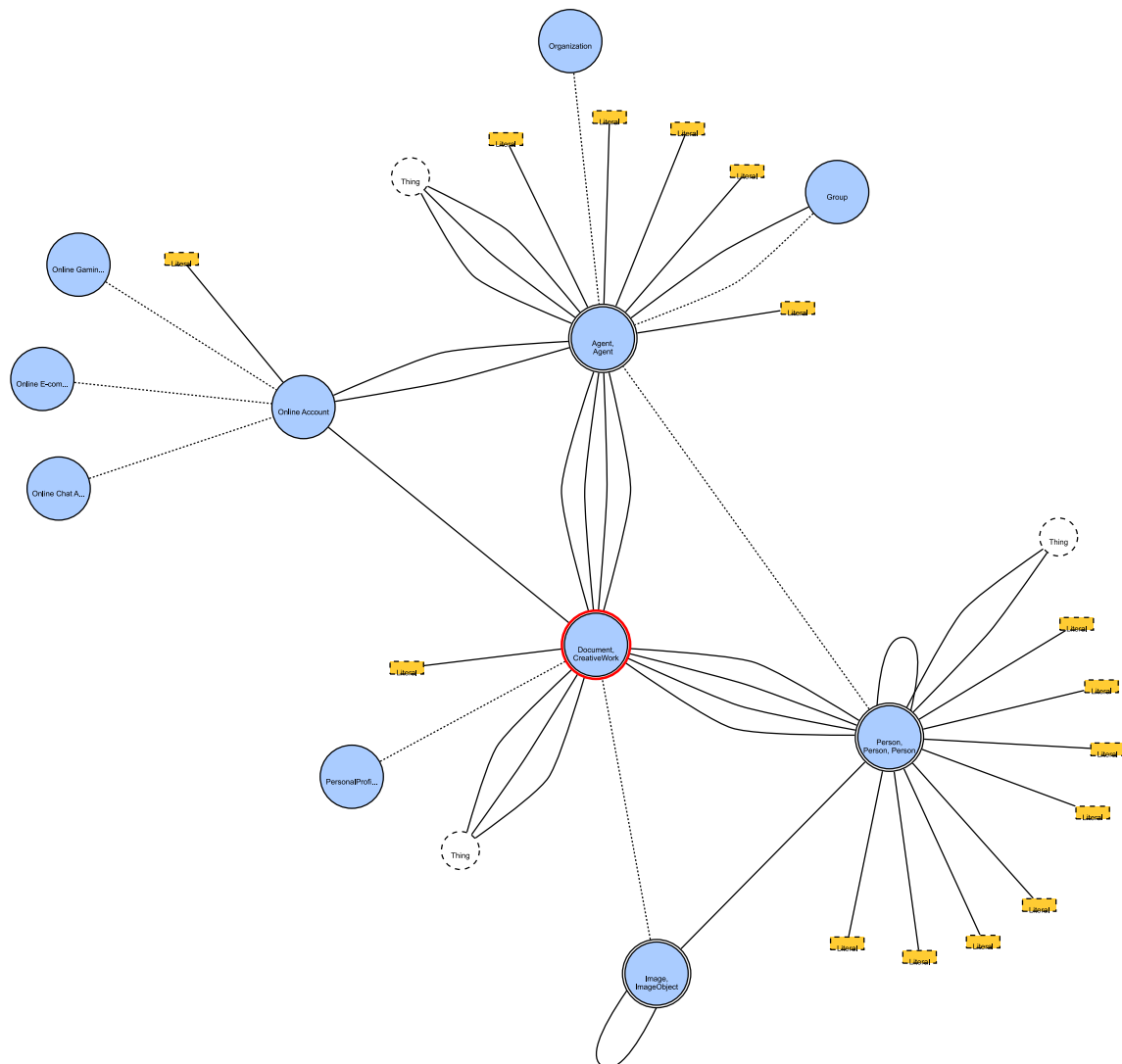
Dudáš *et al.* (2018) apresentam uma classificação dos métodos e ferramentas de visualização de ontologias. As ferramentas são analisadas quanto aos métodos de visualização utilizados, técnicas de interação e tipos de ontologias. Ficou demonstrado que a maioria das ferramentas aplica visualizações bidimensionais *node-link* com foco nas hierarquias de classes. Cor e forma são pouco utilizadas. Segundo os autores, o principal desafio de uma ferramenta de visualização de ontologias está na sua usabilidade, pois necessita fornecer uma visão geral de grandes ontologias e, ao mesmo tempo, mostrar detalhes sob demanda.

No contexto deste trabalho será utilizado o sistema VOWL como base para a proposta de uma interface para sistemas de recuperação de informação.

Segundo Lohmann *et al.* (2016), o VOWL foi inicialmente implementado em duas ferramentas diferentes: o *Protégé* VOWL e o WebVOWL. O *Protégé* VOWL é um plugin Java para este famoso editor de ontologias. O WebVOWL é um aplicativo autônomo baseado em tecnologias Web.

No sistema VOWL (*Visual Notation for OWL Ontologies*) os elementos que compõem uma ontologia OWL são representados por um conjunto de primitivas gráficas e um esquema de cores. A Figura 5 mostra a tela do WebVOWL na qual é visualizada parte da ontologia do vocabulário FOAF (*Friend of a Friend*).

Figura 5 - Visão parcial do vocabulário FOAF apresentado no sistema WebVOWL



Fonte: Adaptado de VOWL (2024)

O VOWL (Figura 5) apresenta a maioria dos elementos que compõem uma ontologia de forma gráfica, interativa e dinâmica. As classes são representadas por círculos rotulados. Classes equivalentes são representadas por círculos com borda dupla. As linhas representam relações entre as classes. Ao selecionar uma determinada classe ela é destacada na cor vermelha. Utilizando o recurso de zoom, os usuários podem explorar em detalhes determinadas partes da ontologia ou analisar a sua estrutura global.

4 PROCEDIMENTOS METODOLÓGICOS

As pesquisas na área de recuperação de informação envolvendo o Modelo Espaço Vetorial (*Vector Space Model*) ganham notoriedade com o trabalho do professor Gerard Salton, com início em 1961 na Universidade de Harvard e posteriormente, após sua mudança em 1965, na Universidade de Cornell. Esse modelo foi aperfeiçoado por décadas e teve um papel significativo no desenvolvimento de toda a área da Recuperação de Informação.

A gênese dessa pesquisa foi a observação de que as características do Modelo Vetorial permitiriam o desenvolvimento de interfaces visuais para sistemas de recuperação de informação. Embora não fosse possível à época de seu desenvolvimento, o Modelo Vetorial traz recursos que possibilitam a representação espacial dos elementos envolvidos no processo de recuperação de informação.

Porém, embora o Modelo Vetorial traga em si recursos importantes, uma interface visual necessita de uma arquitetura, um alicerce no qual os elementos possam ser ancorados.

Por meio de pesquisas bibliográficas sobre interfaces para sistemas de recuperação de informação, foi encontrado e selecionado o artigo de Lohmann *et al.* (2016), que apresenta uma interface interativa para a visualização de ontologias. Esse seria o alicerce que faltava para o desenvolvimento da interface proposta.

Para os exemplos apresentados nesse trabalho, foram feitas pesquisas em diversos repositórios de Ontologias. Procurávamos uma ontologia relativamente pequena (como poucos conceitos) e com uma representação simples, facilmente legível e compreensível. No repositório de ontologias na área biomédica intitulado *BioPortal*¹ selecionamos a ontologia denominada *PedTerm*, que apresenta conceitos relacionados à saúde e ao desenvolvimento infantil. Para efeitos ilustrativos, os conceitos dessa ontologia, originalmente no idioma inglês, foram traduzidos para o português.

¹ Disponível em: <https://biportal.bioontology.org>

Embora ainda não tenham sido implementadas em um sistema real, acreditamos que as ideias apresentadas a seguir possam contribuir para desenvolvimentos futuros.

5 PROPOSTA DE INTERFACE PARA 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 um sistema. Em um primeiro momento o usuário deve traduzir a sua necessidade de informação em um conjunto de termos que represente o tema ou assunto de seu interesse. Essa tarefa é fundamental no processo de recuperação, podendo ser facilitada por diversos recursos do sistema e por uma interface eficiente. Um segundo momento de interação se dá durante a escolha de documentos relevantes dentre os que foram recuperados. Nesse momento, a interface cumpre também papel importante, guiando o usuário na escolha dos documentos que realmente lhe serão úteis.

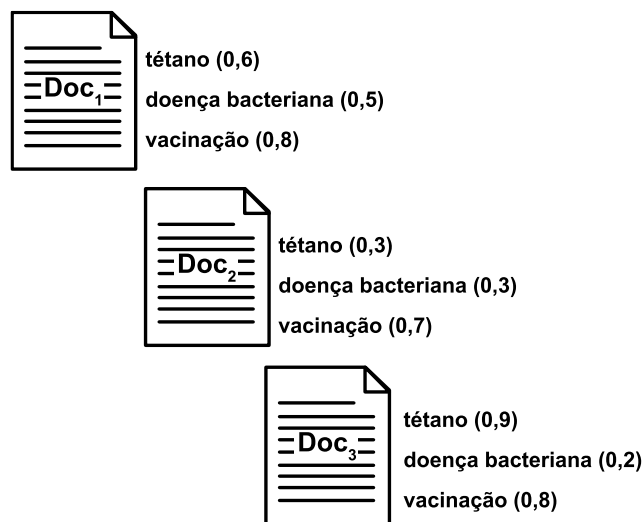
Nas seções seguintes serão apresentados os elementos fundamentais do modelo de interface proposto neste trabalho. Os documentos são representados utilizando a estrutura do Modelo Espaço Vetorial. As buscas são definidas por meio da especificação do tema ou assunto de interesse do usuário e dos termos de busca. Os documentos são apresentados atrelados à ontologia do tema de interesse do usuário.

Os exemplos ilustrativos serão baseados na ontologia *PedTerm*, que apresenta a terminologia relacionada à saúde e ao desenvolvimento infantil e é de acesso gratuito, disponível no *BioPortal*. Como os conceitos dessa ontologia estão originalmente no idioma inglês, os exemplos apresentarão uma tradução desses conceitos, porém conservando a sua estrutura original da ontologia.

5.1 REPRESENTAÇÃO VETORIAL DOS DOCUMENTOS

No modelo vetorial um documento é representado por um conjunto de termos e seus respectivos pesos. A Figura 6 ilustra graficamente os documentos que serão utilizados para exemplificar o modelo de interface aqui proposto.

Figura 6 - Ilustração dos documentos utilizados na apresentação do modelo de interface



Fonte: Elaborado pelos autores

Na ilustração da Figura 6, o documento **doc₁** é indexado pelos termos “tétano”, “doença bacteriana” e “vacinação” com os respectivos pesos: 0.6, 0.5 e 0.8. O documento **doc₂** é indexado pelos termos “tétano” (0.3), “doença bacteriana” (0.3) e “doença infecciosa” (0.7). O documento **doc₃** é representado pelos termos “tétano” (0.9), “doença infecciosa” (0.2) e “*clostridium tetani*” (0.8).

O processo de representação dos documentos não é do escopo deste trabalho. Portanto, assume-se um corpus documental já previamente indexado.

5.2 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 termos (palavras-chave) escolhidas livremente pelo usuário com o objetivo de expressar linguisticamente a sua necessidade de informação.

No modelo aqui proposto, é necessário que o usuário contextualize 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 apresentado na Figura 7, 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. Cada tema está vinculado a uma ontologia. Portanto, é necessário que o sistema tenha previamente cadastrado e disponível um conjunto de ontologias OWL. A definição do assunto permite restringir a terminologia envolvida na busca, diminuindo problemas linguísticos inerentes ao processo de recuperação de informação como ambiguidades e polissemias.

Figura 7 - Exemplo da especificação de uma busca

Tema/Assunto	
Pediatria (PedTerm)	▼
tétano "doença infecciosa"	

Fonte: Elaborado pelos autores.

Após escolher o tema ou assunto, o usuário define um conjunto de termos que irão compor a sua expressão de busca. No exemplo da Figura 7, o assunto de interesse do usuário é “Pediatria” e a expressão de busca é composta pelo termo simples “tétano” e pelo termo composto "doença infecciosa". Após efetuar a sua busca, o usuário terá como resultado a representação gráfica da ontologia do assunto escolhido, onde serão sinalizados os termos utilizados na sua busca (“tétano” e "doença infecciosa").

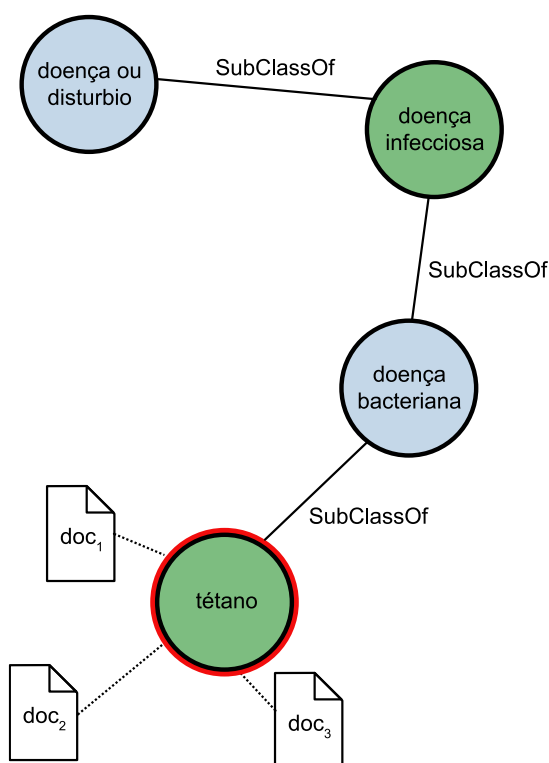
5.3 RESULTADO DA BUSCA

Como resultado da busca, será apresentada ao usuário a representação gráfica da ontologia referente ao tema escolhido, com a sinalização (em verde) dos termos coincidentes aos termos utilizados na expressão de busca. Ao selecionar um determinado termo na ontologia, serão apresentados os documentos indexados por esse termo. Os documentos irão “orbitar” o termo da

ontologia a uma distância inversamente proporcional ao peso do termo de indexação de cada documento.

A busca exemplificada na Figura 7 terá como resultado a apresentação gráfica da ontologia *PedTerm* com os termos “tétano” e “doença infecciosa” em destaque. Ao selecionar o termo “tétano” na ontologia, ícones representando os documentos indexados por esse termo serão dispostos ao seu redor, como apresentado na Figura 8.

Figura 8 - Documentos resultantes de uma busca pelo termo “tétano”

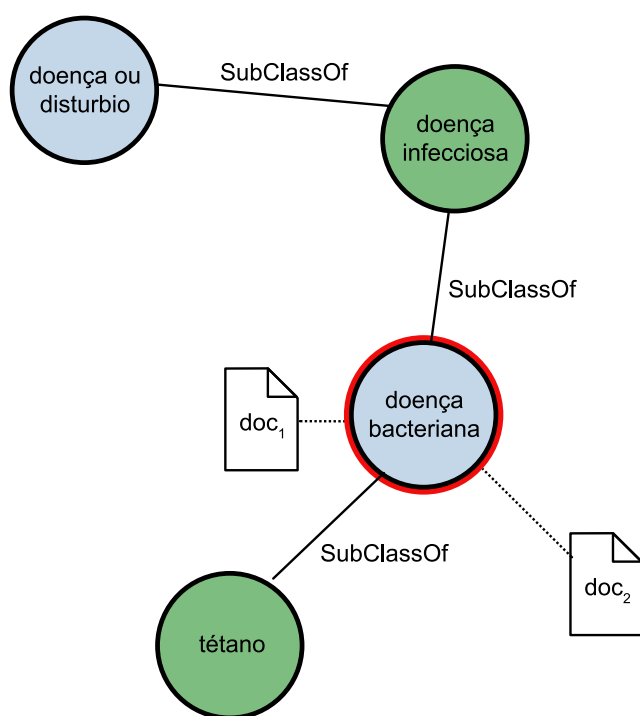


Fonte: Elaborado pelos autores.

Na Figura 8, o usuário selecionou na ontologia o termo “tétano”. Como pode ser visto na Figura 6, os três documentos do *corpus* são indexados por esse termo, com diferentes pesos. Esses pesos são considerados como uma medida da relevância e é representada pela distância do documento em relação ao termo selecionado na ontologia. No documento **doc₃** o termo “tétano” possui peso 0,9 e por essa razão ele é apresentado bem próximo ao termo da ontologia. No documento **doc₂** o termo “tétano” possui peso 0,3 e está posicionado mais distante do termo da ontologia.

Na Figura 9, o termo selecionado é “doença bacteriana”. Embora não seja um termo utilizado na busca, o usuário tem a liberdade de navegar pela ontologia, descobrindo documentos que podem lhe ser úteis. Existem dois documentos indexados pelo termo “doença bacteriana”: **doc₁** e **doc₂**, com pesos 0,5 e 0,3 respectivamente. Assim, **doc₂** aparece mais distante e **doc₁** aparece mais próximo ao termo selecionado na ontologia.

Figura 9 - Documentos indexados pelo termo “doença bacteriana”



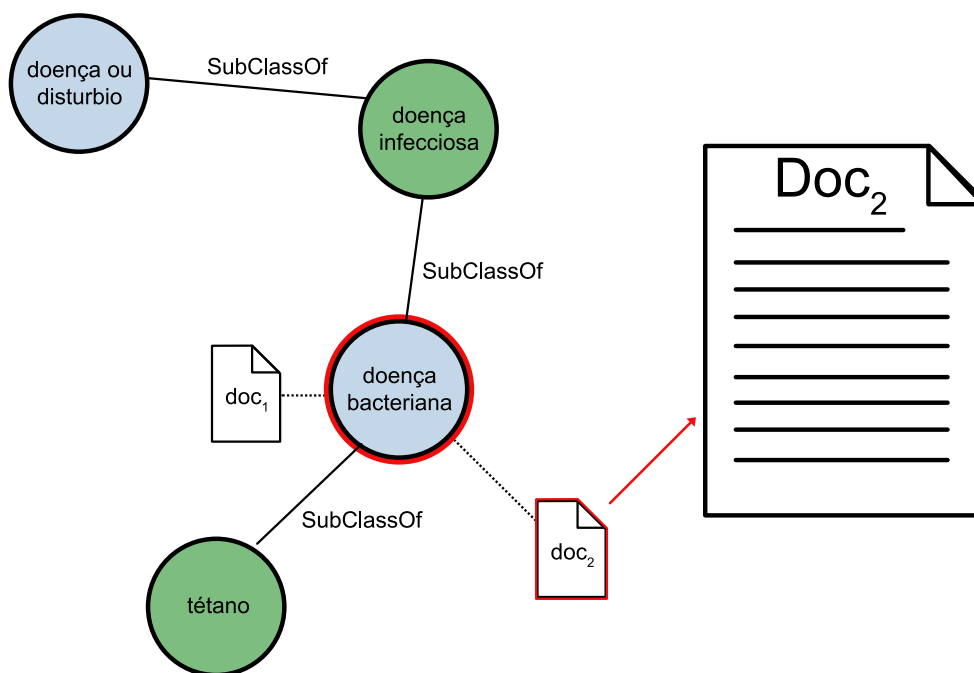
Fonte: Elaborado pelos autores.

Para evitar uma sobrecarga da interface, ao selecionar um determinado conceito da ontologia, somente os documentos relacionados a esse conceito serão apresentados. Os documentos que orbitavam outro conceito não serão mais apresentados. Na Figura 9, ao ser selecionado o conceito “doença bacteriana” serão apresentados apenas os documentos indexados por esse termo; os documentos que orbitavam o conceito “tétano” não serão mais apresentados.

5.4 VISUALIZAÇÃO DE DOCUMENTOS

O julgamento da relevância de um determinado documento só é possível se o usuário tiver conhecimento de seu conteúdo. A Figura 10 apresenta um exemplo da visualização dos detalhes de um documento.

Figura 10 - Apresentação do conteúdo de um documento



Fonte: Elaborado pelos autores.

Por meio de um comando via teclado ou mouse, é apresentado um trecho do documento em uma janela do tipo *pop-up* ou algum outro recurso semelhante, dependendo do formato do arquivo. No exemplo da Figura 10, considerando que o documento **doc₂** está em formato HTML, ele poderia ser apresentado no navegador Web.

A apresentação de parte ou de todo o conteúdo dos documentos resultantes de uma busca é de fundamental importância, pois permitirá que o usuário julgue se o documento é ou não relevante para a sua necessidade de informação.

6 CONSIDERAÇÕES FINAIS

A imensa quantidade de informação disponível em ambientes digitais, juntamente com o seu aumento progressivo contínuo traz para os sistemas de indexação automática e recuperação de informação a responsabilidade de auxiliar na organização e busca por informações relevantes e úteis para o cotidiano das pessoas em meio a esta infinitude informacional crescente. Nesse contexto, as interfaces dos sistemas de recuperação devem 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 grande parte das atividades humanas, o processo de recuperação de informação é uma tarefa que envolve linguagem. Por um lado, cada documento é representado por expressões linguísticas que representam o seu conteúdo informacional. Por outro lado, os usuários expressam linguisticamente a sua necessidade de informação.

O modelo de interface proposto neste trabalho reconhece o caráter linguístico do processo de recuperação de informação e apresenta a estrutura terminológica de uma ontologia como alicerce sobre o qual se assenta a visualização dos documentos resultantes de uma busca. Altera-se, assim, a lógica presente nos atuais sistemas de recuperação de informação e nos mecanismos de busca Web, onde os elementos principais são os documentos, apresentados geralmente na forma de uma longa e estática lista. A navegação por meio da estrutura terminológica de uma ontologia, possibilita ainda que o usuário se familiarize com a terminologia adotada em uma determinada área do conhecimento de seu interesse.

Como sugestões para trabalhos futuros, vemos a necessidade de implementação de um protótipo deste sistema que venha a ser um campo de provas para verificar a viabilidade tecnológica dessa proposta, assim como estudar com mais profundidade as suas vantagens e desvantagens.

REFERÊNCIAS

- BAEZA-YATES, R.; RIBEIRO NETO, B. **Modern Information Retrieval**. 2. ed. Boston: Addison-Wesley, 2011.
- 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, 2010. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/22422>. Acesso em: 14 maio 2025.
- CHANDRASEKARAN, B.; JOSEPHSON, J. R.; BENJAMINS, V. R. What are ontologies, and why do we need them? **IEEE Intelligent Systems**, [S. l.], v. 14, n. 1, p. 20-16, 1999. Disponível em: https://www.researchgate.net/publication/3420451_What_Are_Ontologies_and_Why_Do_We_Need_Them. Acesso em: 14 maio 2025.
- DUDÁŠ, M.; LOHMANN, S.; SVÁTEK, V.; PAVLOV, D. Ontology visualization methods and tools: a survey of the state of the art. **The Knowledge Engineering Review**, [S. l.], v. 33, n. 10, p. 1-39, 2018. Disponível em: <https://www.cambridge.org/core/journals/knowledge-engineering-review/article/abs/ontology-visualization-methods-and-tools-a-survey-of-the-state-of-the-art/5EA8C64D7DF60A84F6D2B7B9A09B6E6A>. Acesso em: 14 maio 2025.
- FERNEDA, E. **Introdução Modelos Computacionais de Recuperação de Informação**. Rio de Janeiro: Ciência Moderna, 2012.
- GRUBER, T. Toward Principles for the Design of Ontologies Used for Knowledge Sharing. **International Journal Human-Computer Studies**, [S. l.], v. 43, p. 907-928, 1995. Disponível em: https://www.researchgate.net/publication/2626138_Toward_Principles_for_the_Design_of_Ontologies_Used_for_Knowledge_Sharing. Acesso em: 14 maio 2025.
- HOBBS, V.; PFITZNER, D.; POWERS, D. A survey of information retrieval interfaces, Flinders Unviserity, 2002. Disponível em: <http://www.academia.edu/download/40120742/200204-drft-IRS.pdf>. Acesso em 05 fev. 2023.
- KATIFORI, A.; HALATSIS, C.; LEPOURAS, C.; VASSILAKIS, C.; GIANNOPOULOU, E. Ontology visualization methods: a survey. **ACM Computing Surveys**, [S. l.], v. 39, n. 4, 2007. Disponível em: <http://www.dit.unin.it/~p2p/RelatedWork/Matching/a10-katifori.pdf>. Acesso em: 14 maio 2025.

LOHMANN, S.; NEGRU, S.; HAAG, F.; ERTL, T. Visualizing Ontologies with VOWL. **Semantic Web Journal**, [S. l.], v. 7, n. 4, p. 399-419, 2016. Disponível em: <https://dl.acm.org/doi/abs/10.3233/SW-150200>. Acesso em: 14 maio 2025.

ROBERTSON, S. E. Theories and models in information retrieval. **Journal of Documentation**, [S. l.], n. 33, p. 126-148, 1977. Disponível em: <https://www.semanticscholar.org/paper/THEORIES-AND-MODELS-IN-INFORMATION-RETRIEVAL-Robertson/2a9abc1aaca1ecbe15b252148cdd97544075151e>. Acesso em: 14 maio 2025.

SALTON, G.; WONG, A.; YANG, C. S. A Vector Space Model for Automatic Indexing. **Communications of the ACM**, [S. l.], v. 18, n. 11, 613-620, 1975. Disponível em: <https://dl.acm.org/doi/10.1145/361219.361220>. Acesso em: 14 maio 2025.

USE OF ONTOLOGY AS AN INTERFACE FOR INFORMATION RETRIEVAL SYSTEMS: A PROPOSAL BASED ON THE VECTOR SPACE MODEL

ABSTRACT

Objective: Using the Vector Space Model as a theoretical basis, 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.
Methodology: This is characterized as applied, exploratory and bibliographic research. Based on the unique characteristics of the Vector Model, which allows calculating quantifiable distances between queries and documents, and the linguistic aspects inherent to the information retrieval process, a symbiosis was created between the terminological structure of the ontologies and the potential visual resources of that Model.
Conclusions: It is concluded that the visual presentation of an ontology allows the development of dynamic and interactive interfaces, providing the user with the navigation between the documents resulting from their search, based on the terms of a certain knowledge area.

Descriptors: Information Retrieval. Vector Space Model. Ontology. Information Retrieval Interface. Ontology-based interface.

USO DE LA ONTOLOGÍA COMO INTERFAZ PARA SISTEMAS DE RECUPERACIÓN DE INFORMACIÓN: UNA PROPUESTA BASADA EN EL MODELO DE ESPACIO VECTORIAL

RESUMEN

Objetivo: Utilizando como base teórica el Modelo de Espacio Vectorial, este trabajo propone un modelo de interfaz para sistemas de recuperación de información en el que

se utiliza la estructura terminológica de una ontología para asistir al usuario en la selección de documentos relevantes entre los resultantes de una búsqueda.

Metodología: Se caracteriza por ser una investigación aplicada, exploratoria y bibliográfica. A partir de las características propias del Modelo Vectorial, que permite calcular distancias cuantificables entre consultas y documentos, y de los aspectos lingüísticos inherentes al proceso de recuperación de información, se creó una simbiosis entre la estructura terminológica de las ontologías y los potenciales recursos visuales de dicho Modelo. **Conclusiones:** La presentación visual de una ontología permite el desarrollo de interfaces dinámicas e interactivas, facilitando al usuario la navegación entre los documentos resultantes de su búsqueda, en base a los términos de una determinada área de conocimiento.

Descriptores: Recuperación de Información. Modelo espacio vectorial. Ontología. Interfaz de recuperación de información. Interfaz basada en ontología.

Recebido em: 13.02.2023

Aceito em: 09.10.2024