



USO DA ONTOLOGIA SKOS PARA CRIAÇÃO DE UM TESAURO SEMÂNTICO DE PUBLICAÇÕES

Área temática: Gestão do Conhecimento Organizacional

Nilton Freitas Junior

niltonfjunior@gmail.com

Mark Douglas Jacyntho

markjacyntho@gmail.com

***Resumo:** Textos são dados não estruturados, cuja informação é acessível apenas a partir de sua leitura e compreensão. Da mesma forma que a Tecnologia da Informação avança no sentido de divulgar e armazenar documentos textuais, também são desenvolvidas técnicas que procuram tornar mais eficiente a gestão da informação em bases de dados no formato de textos. O uso de um tesauro para relacionar conceitos associados a um texto pode ser considerado uma dessas técnicas. O tesauro estabelece um vocabulário controlado, onde palavras-chave que representam um texto podem ser armazenadas de forma mais completa, quando a elas são relacionadas características semânticas. Este trabalho apresenta a proposta e subsequente construção de um sistema de informação baseado no conceito de tesauro semântico, utilizando a ontologia Simple Knowledge Organization System (SKOS), com dados dispostos no formato semiestruturado Resource Description Framework (RDF), para armazenamento de palavras-chave de textos, a fim de oferecer maior eficiência em consultas e recuperação da informação.*

Palavras-chaves: TESAURO; SKOS; RDF; LINKED DATA WEB SEMÂNTICA



1. INTRODUÇÃO

Publicações textuais tradicionalmente representam conhecimento armazenado, por serem mídia de ideias e fonte de pesquisas, mesmo em tempos anteriores à explosão dos recursos digitais disponíveis no contexto contemporâneo. Um texto, de certa forma, é um banco de dados que segue o mesmo princípio de um sistema que processa seus dados e obtém informação.

A leitura de um texto corresponde ao processamento de dados, e a compreensão adquirida por essa leitura representa a informação nele contida. Mesmo que considere elementos controlados que identificam um texto, como título, resumo e palavras-chave, a certeza do conhecimento semântico associado a esse texto só virá após a sua leitura por alguém que compreenda sua abordagem.

O constante avanço das Tecnologias da Informação (TI) tem possibilitado tanto o crescimento de produções científicas, gerando novos conhecimentos e difundindo informações, quanto o de mecanismos e técnicas capazes de aprimorar a recuperação da informação em bases de dados textuais, através de sistemas de informação capazes de oferecer métodos de controle dos elementos identificadores de um texto (AITCHISON, GILCHRIST e BAWDEN, 2000).

Textos, conforme explicam Prado, Ornelas e Araújo, (2014, p.171-194), são considerados dados não estruturados, pelo fato de não apresentarem uma ordem predeterminada para as entidades que os compõem. Mesmo assim, há a possibilidade de uma linguagem controlada oferecer alguma ordem e até mesmo significado a um texto, como acontece com um e-mail, onde há um assunto (*subject*) a ser definido antes do texto, ou as palavras-chave informadas em um artigo submetido para publicação.

Se considerados grandes repositórios de arquivos em formato de texto, nos quais há a necessidade de buscar informação contida nos mesmos, tais controles são capazes de fornecer apenas uma consulta sintática aos elementos que os representam, o que não necessariamente representará a semântica correta associada ao texto pesquisado.



1.1 Proposta de um vocabulário controlado para recuperação da informação

Com o crescimento da produção científica em ambiente digital, sob o paradigma dos dados textuais não estruturados, gerenciados por sistemas de informação, tornam-se necessários recursos que promovam a indexação dessa grande quantidade de dados armazenados em centros de informação, de modo a apoiar uma melhor gestão do conhecimento (TURBAN e VOLONINO, 2011, p.67; REZENDE, 2005, p.337).

Um vocabulário controlado é conhecido como tesouro: *“um instrumento que reúne termos escolhidos a partir de uma estrutura conceitual previamente estabelecida”* (INEP). Em um tesouro, cada entidade corresponde a um conceito que descreve a si mesmo e o texto ao qual está associado. Estas entidades também estão relacionadas entre si mesmas, através de um sistema de representação de conhecimento.

Um sistema de tesouro é mais elaborado do que um simples dicionário. Os termos que compõem um tesouro representam conceitos, estão relacionados entre si e fazem parte de uma estrutura definida de conhecimento. Tal pensamento, em um entendimento focado na natureza de textos publicados e armazenados por meios digitais, seja em ambiente acadêmico ou empresarial, motiva o estudo de sistemas capazes de aprimorar a recuperação deste mesmo conhecimento.

A proposta deste trabalho envolve o desenvolvimento de um sistema de indexação de publicações por meio de um tesouro construído com o emprego de elementos de uma ontologia apropriada, para estabelecer não apenas o relacionamento entre os conceitos associados a um texto, como palavras-chave, mas atribuir valores semânticos a estes conceitos, capazes de oferecer melhores resultados na recuperação da informação.

Objetiva-se, portanto, que o sistema desenvolvido permita que um texto nele incluído possa ter seus elementos de identificação categorizados em um tesouro adequado, proporcionando maiores possibilidade de recuperação da informação por meio de consultas semânticas não ambíguas, realizadas também neste sistema.



O restante do artigo está estruturado da seguinte forma: na seção 2, a ontologia escolhida para a criação do tesauro e os fundamentos da Web Semântica/Dados Ligados são brevemente apresentados; a seção 3 descreve a metodologia usada no desenvolvimento do sistema; a seção 4 aborda os resultados e discussões do trabalho; finalmente, a seção 5 contém as considerações finais e próximos passos.

2. UMA ONTOLOGIA SIMPLES PARA SISTEMAS DE ORGANIZAÇÃO DO CONHECIMENTO

Visando o desenvolvimento de um tesauro controlado, não restrito a uma única área do conhecimento, capaz de se adaptar a diversas utilizações, optou-se por fazer uso da ontologia *Simple Knowledge Organization System* (SKOS) (MILES & BECHHOFER, 2009), pelo fato desta ser apropriada para representar, de forma geral, sistemas de organização do conhecimento.

Breitman (2005) diferencia o conceito filosófico de ontologia daquele aplicado no desenvolvimento de sistemas computacionais. Para esta finalidade, uma ontologia descreve formalmente termos que são utilizados na representação de uma área do conhecimento, através de definições de classes de um domínio, seus relacionamentos com outras entidades definidas na ontologia e as propriedades que estas entidades devem possuir.

A ontologia SKOS é um modelo de dados comum para compartilhamento e a ligação de sistemas de organização do conhecimento. Esta ontologia é considerada um padrão único para sistemas de organização do conhecimento, como: enciclopédias, classificações, títulos de assuntos, taxonomias, tesauros e folksonomias (JACOBS, 2009, tradução nossa).

Muitos desses sistemas de organização do conhecimento compartilham uma estrutura semelhante, permitindo que SKOS concentre grande parte dessas similaridades, permitindo seu intercâmbio entre diversas aplicações, mesmo em ambientes diversificados, como é o caso da Web (MILES & BECHHOFER, 2009). Nesta ontologia, diversos conceitos podem estabelecer relações semânticas entre si, o que a torna ideal para o desenvolvimento de um tesauro semântico.



No ambiente prático do desenvolvimento de sistemas, ontologias requerem um mecanismo padrão para especificar suas instâncias e oferecer meios de interligação com outros dados. Este mecanismo é o *Resource Description Framework* (RDF), um modelo de dados em grafo que permite estabelecer declarações sobre qualquer recurso¹, utilizando classes/propriedades de ontologias (HEATH & BIZER, 2011, p.4). Um grafo RDF é um conjunto de triplas *recurso-propriedade-valor*; conjunto este que forma nós e arcos ao estabelecer relações entre recursos dentro de um mesmo sistema ou com uma fonte de dados externa.

A união entre os critérios da ontologia SKOS e a estrutura de grafos oferecida pelo padrão RDF proporciona a semântica necessária para que a busca por palavras-chave de um texto seja feita de forma que os significados associados às palavras sejam considerados e resultem em respostas mais acuradas. Situações onde palavras polissêmicas representam textos certamente resultam em consultas ambíguas, quando não se tem semântica inteligível por máquina associada às palavras.

2.1 Web Semântica e os dados interligados

A construção do tesauro com base em ontologias e no uso do padrão de dados RDF abre possibilidades para que o sistema possa ser caracterizado pelos princípios da Web Semântica, se aplicados mecanismos que permitam a interligação de seus dados com dados de terceiros provenientes de outros sistemas na Web.

A ideia inicial da Web Semântica mantém a publicação e relacionamento de documentos diversos, através de hiperlinks. Porém, a estes documentos são associados metadados de categorização da sua informação, padronizados em RDF, facilitando seu acesso e, principalmente, sua compreensão pela máquina, a partir do significado explícito e estruturado que os mesmos oferecem. Esta padronização de dados se dá através da infraestrutura existente na própria Web tradicional, onde cada recurso é acessado através de um endereço único, conhecido como *Uniform Resource Identifier* (URI) (BREITMAN, 2005, p.5-6).

¹ Recurso é qualquer entidade (pessoa, lugar, produto, ideia, etc.) identificada por um endereço *Web*.



Como subconjunto mais pragmático da *Web Semântica*, em Bernes-Lee (2006) foi proposto o conceito de *Linked Data* (em português, *Dados Interligados*). Assim como a *Web*, a ideia de *Linked Data* é muito simples: *por que não publicar e interligar dados diretamente na Web?* Em outros termos, ao invés de apenas publicar metadados associados a documentos, publicar diretamente estes dados, independentemente de documentos, e interligá-los por meio de *links* RDF (relacionamentos semânticos), criando a chamada *Web de Dados* (em inglês, *Web of Linked Data*). Uma *Web* de dados estruturados, totalmente compreensível por máquinas, tornando, portanto, a busca por informações muito mais precisa e consistente. Um grafo de dados global que não para de crescer (HEATH & BIZER, 2011).

Utilizar um URI para identificar um conceito do tesouro garante que esta identificação seja única em toda a *Web*, evitando qualquer confusão polissêmica. Ao descrever um conceito do tesouro usando RDF, como cada elemento das triplas (recurso-propriedade-valor) pode ser representado por um URI, este pode endereçar um recurso externo ao tesouro, que esteja disponível na *Web*, enriquecendo a descrição do tesouro (*mashup Linked Data*).

Esse cenário contempla os princípios *Linked Data*, que proporcionam condições para explorar novos resultados no sistema, ou até mesmo pela *Web*, a partir de uma pesquisa originalmente feita em um único conceito do tesouro.

3. MÉTODO DE DESENVOLVIMENTO DO TESAURO

No desenvolvimento do sistema foi utilizada a linguagem de programação PHP², que é de código livre, possuindo recursos adequados para a construção da aplicação, além de permitir que esta seja utilizada tanto em plataforma *desktop*, quanto para plataforma *Web*. O sistema conta também com o uso de banco de dados relacional MySQL³, simulando um banco de dados RDF por meio da biblioteca RDF API for PHP (RAP)⁴. A interface gráfica foi

² <http://php.net/>

³ <https://www.mysql.com/>

⁴ <http://wifo5-03.informatik.uni-mannheim.de/bizer/rdffapi/index.html>

implementada com emprego da linguagem HTML5⁵, amparada na utilização do *framework Bootstrap*⁶, englobando com isso inúmeros recursos CSS⁷ e JQuery⁸.

O desenvolvimento seguiu o modelo arquitetural em três camadas, definido por Pressman (2009, p.235-237) como sendo composto por uma camada alta, que atende às operações do usuário; uma camada mais baixa responsável por se comunicar com o sistema operacional; e uma camada de desenvolvimento intermediária, para fornecer serviços e utilitários inerentes à regra de negócio do sistema desenvolvido.

A camada de interface é responsável pela apresentação das telas do sistema e pela captação de dados a serem armazenados. Todo o padrão de interface leva um usuário comum do sistema a não perceber alterações com relação às manipulações de dados em RDF. Este usuário acessa formulários e dados do protótipo como se fossem uma aplicação Web convencional.

A camada de domínio do negócio baseou-se em percepções empíricas sobre processos relevantes para a catalogação de palavras-chave (conceitos) que representam um texto publicado. As propriedades selecionadas para as palavras-chave foram definidas para contemplar propriedades adequadas da ontologia SKOS, como apresentado na tabela 01. Além destas propriedades, também foi utilizada a classe *Concept* da ontologia SKOS para explicitar que cada palavra-chave é um conceito na concepção da ontologia SKOS.

Tabela 01: Propriedades da ontologia SKOS para as palavras-chave

Descrição do atributo	Propriedade	Definição de uso pela ontologia SKOS
A palavra chave cadastrada	<i>prelabel</i>	O termo lexical preferido para a palavra chave.
Descrição narrativa da palavra chave	<i>definition</i>	A explicação formal do significado de um conceito

⁵ <http://www.w3.org/TR/html5/>

⁶ *Bootstrap* é um conjunto *open source* de padrões para os elementos HTML, CSS e Javascript mais utilizados, agilizando o desenvolvimento de interface (*front-end*) para páginas Web, podendo ser também personalizado conforme as necessidades do desenvolvimento.

⁷ CSS, sigla de cascading style sheet (folhas de estilo em cascata) é uma linguagem de programação de estilos voltados para interfaces web, interpretados por um browser em tempo de execução do código.

⁸ JQuery é uma biblioteca JavaScript desenvolvida para facilitar a utilização de códigos desta linguagem no desenvolvimento de sistemas client side, interagindo com HTML

Nome alternativo da palavra chave	<i>altlabel</i>	Acrônimos, abreviações, variações de ortografia, e formas de plural/singular irregulares podem ser incluídas entre os termos alternativos.
Conceito mais amplo da palavra chave	<i>broader</i>	Conceitos genéricos são tipicamente representados como PARENTES em uma hierarquia de conceitos
Conceito mais específico da palavra chave	<i>narrower</i>	Conceitos específicos são tipicamente representados como DESCENDENTES em uma hierarquia de conceitos
Conceito relacionado à palavra chave	<i>related</i>	Um conceito com o qual existe um relacionamento semântico associativo

Fonte: autoria própria; SKOS

O modelo de usabilidade é descrito por um diagrama de casos de uso, onde o usuário tem acesso ao sistema do tesouro diretamente pelos registros existentes no banco de dados. A partir desta visualização inicial, as ações de inclusão, alteração e exclusão dos registros são feitas normalmente. Também é oferecido ao usuário acesso à visualização individualizada de um registro, na forma de uma página catalográfica, além de permitir a emissão de um arquivo com os dados em formato RDF. A figura 1 apresenta este diagrama de casos de uso.

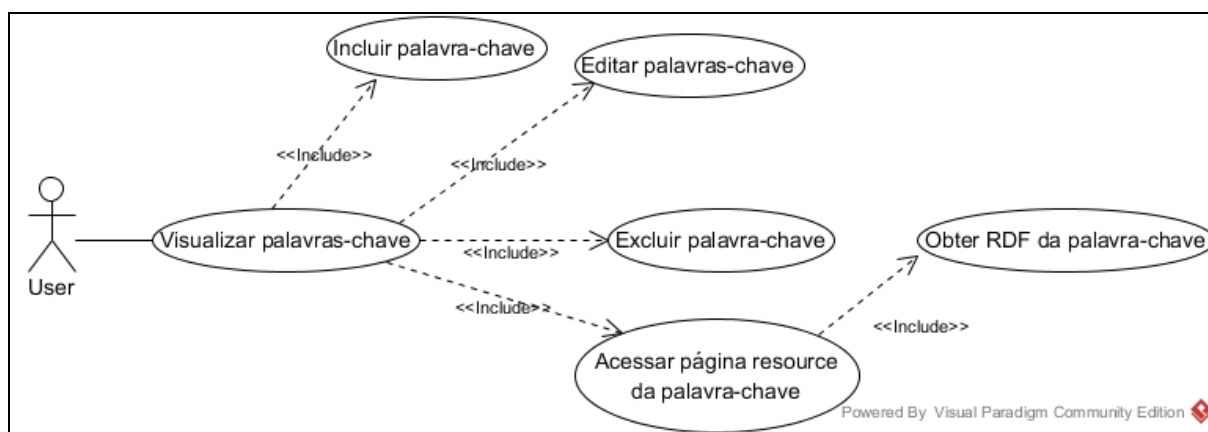


FIGURA 1: Diagrama de casos de uso do tesouro. Autoria própria; Visual Paradigm⁹

Cabe aos usuários do sistema informar dados adequados para cada uma dessas propriedades, em cada registro de palavra-chave inserido (figura 2). Esta característica de construção colaborativa do conhecimento, ou folksonomia, é apresentada por Fuks (2011) como um mecanismo eficaz para circulação e recuperação de informações, onde os próprios usuários de um sistema escolhem livremente associações relevantes para um determinado recurso.

⁹ Visual Paradigm Community Edition. <https://www.visual-paradigm.com/download/community.jsp>



Cadastro de Palavras-chave

Editar registro

Palavra-chave:

Definição:

OFEREÇA MAIS DETALHES PARA ESTA PALAVRA-CHAVE:

Nome alternativo:

Conceito mais amplo:

Conceito mais específico:

Conceito relacionado:

FIGURA 2: Interface para cadastro de palavras-chave no tesauro. Autoria própria.

A inserção de cada palavra-chave no sistema gera registros RDF com as propriedades descritas na tabela 1. O exemplo a seguir ilustra o registro RDF da palavra polissêmica “cabo”:

```
<http://www.niltonjunior.com.br/pws/resource/keyword/Cabo> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>  
<http://www.w3.org/2004/02/skos/core#Concept> .
```

```
<http://www.niltonjunior.com.br/pws/resource/keyword/Cabo> <http://www.w3.org/2004/02/skos/core#prefLabel> "Cabo"@PT .
```

```
<http://www.niltonjunior.com.br/pws/resource/keyword/Cabo> <http://www.w3.org/2004/02/skos/core#definition> "Posto militar superior  
ao soldado e inferior ao sargento"@PT .
```

```
<http://www.niltonjunior.com.br/pws/resource/keyword/Cabo> <http://www.w3.org/2004/02/skos/core#altLabel> "Cabo Militar"@PT .
```

```
<http://www.niltonjunior.com.br/pws/resource/keyword/Cabo> <http://www.w3.org/2004/02/skos/core#broader> "Hierarquia Militar"@PT.
```

```
<http://www.niltonjunior.com.br/pws/resource/keyword/Cabo> <http://www.w3.org/2004/02/skos/core#narrower> "Patente Militar"@PT .
```

```
<http://www.niltonjunior.com.br/pws/resource/keyword/Cabo> <http://www.w3.org/2004/02/skos/core#related> "Exército"@PT .
```

Em sistemas convencionais, uma busca sintática pela palavra-chave “cabo” resultaria em registros diversos, contemplando todas as possibilidades que porventura existam dentro do banco de dados (figura 3). Uma busca por registros no sistema de tesauro semântico considera as propriedades estabelecidas pela ontologia SKOS para oferecer buscas mais precisas, baseadas em significados. Dessa forma, o tesauro proporciona meios para buscas mais precisas, favorecendo a recuperação da informação. Seus registros, por estarem dispostos em formato RDF, podem fazer relações entre si, o que contempla características de um tesauro.



ID	Palavra-chave	URI
56	Cabo	http://www.niltonjunior.com.br/pws/resource/keyword/Cabo
64	Cabo Frio	http://www.niltonjunior.com.br/pws/resource/keyword/Cabo_Frio
65	Arraial do Cabo	http://www.niltonjunior.com.br/pws/resource/keyword/Arraial_do_Cabo

Exibindo 1 até 3 de 3 registro(s) (Encontrados em 50 registros)

FIGURA 3: Busca sintática no tesouro pela palavra-chave cabo. Autoria própria.

Simulando um possível cenário de ambiguidades dentro do sistema, onde existam publicações relacionadas a palavras-chaves distintas, porém com o termo “cabo” em seus nomes, uma consulta sintática retorna vários resultados, porém não aqueles que o usuário possa realmente estar procurando. Entretanto, basta que este usuário informe algum termo semântico em sua pesquisa para que o resultado seja especificamente aquele que procura. Considerando o RDF apresentado anteriormente, a figura 4 apresenta este contexto.

Pesquisa semântica de palavras-chave utilizadas em publicações

Termo para pesquisa:

OFEREÇA MAIS OPÇÕES PARA ENRIQUECER A PESQUISA:

Conceito mais amplo:

Conceito mais específico:

Conceito alternativo:

Conceito relacionado:

Pesquisa semântica de palavras-chave utilizadas em publicações

RESULTADOS:

- Cabo <<http://www.niltonjunior.com.br/pws/resource/keyword/Cabo>>

FIGURA 4: Busca semântica no tesouro pela palavra-chave cabo. Autoria própria.



Outro benefício da utilização do padrão RDF para os registros do tesouro é a possibilidade dos *mashups* semânticos, capazes de oferecer mais informações aos resultados obtidos no sistema. Um bom exemplo de dados ligados é o projeto *DBpedia*¹⁰, que desenvolve “um esforço comunitário com a finalidade de extrair informações estruturadas da *Wikipedia*¹¹ e tornar essa informação disponível na *Web* em RDF” (DBpedia, tradução nossa). O registro RDF da palavra-chave “Cabo Frio”, apresentado a seguir, mostra a utilização deste recurso:

```
<http://www.niltonjunior.com.br/pws/resource/keyword/Cabo_Frio> <http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
<http://www.w3.org/2004/02/skos/core#Concept> .
<http://www.niltonjunior.com.br/pws/resource/keyword/Cabo_Frio> <http://www.w3.org/2004/02/skos/core#prefLabel> "Cabo Frio"@PT .
<http://www.niltonjunior.com.br/pws/resource/keyword/Cabo_Frio> <http://www.w3.org/2004/02/skos/core#definition> "Município do
Estado do Rio de Janeiro"@PT .
<http://www.niltonjunior.com.br/pws/resource/keyword/Cabo_Frio> <http://www.w3.org/2004/02/skos/core#altLabel> "Cidade de Cabo
Frio"@PT .
<http://www.niltonjunior.com.br/pws/resource/keyword/Cabo_Frio> <http://www.w3.org/2004/02/skos/core#broader> "Cidades
Fluminenses"@PT .
<http://www.niltonjunior.com.br/pws/resource/keyword/Cabo_Frio> <http://www.w3.org/2004/02/skos/core#narrower> "Cidade
Turística"@PT .
<http://www.niltonjunior.com.br/pws/resource/keyword/Cabo_Frio> <http://www.w3.org/2004/02/skos/core#related> "Região dos
Lagos"@PT .
<http://www.niltonjunior.com.br/pws/resource/keyword/Cabo_Frio> <http://www.w3.org/2000/01/rdf-schema#seeAlso>
<http://dbpedia.org/resource/Cabo_Frio> .
```

A propriedade *seeAlso*, de RDF-Schema, indica um URI da Dbpedia, externo ao sistema, que será acesso quando a palavra-chave consultada estiver associada a uma publicação, como mostra a figura 5.

O ALINHAMENTO DE ROCHAS ALCALINAS POÇOS DE CALDAS-CABO FRIO (RJ) E SUA CONTINUIDADE NA CADEIA VITÓRIA-TRINDADE

[artigo]

<http://www.niltonjunior.com.br/pws/resource/publication/10>

RESUMO:

Com base em algumas evidências geológicas, geomorfológicas e geocronológicas, propõe-se que o Alinhamento de Rochas Alcalinas Poços de Caldas-Cabo Frio, que se estende na direção WNW-ESE, no Estado do Rio de Janeiro, tenha se implantado quando da passagem da placa litosférica Sul-Americana por sobre um hot spot. Durante o Eoceno, esse hot spot teria chegado ao limite leste do continente, ou seja, ao litoral do Estado do Rio de Janeiro, e foi acompanhado por importantes eventos tectônicos e magmáticos. Associado a esses eventos, teria ocorrido o deslocamento relativo do hot spot no sentido do alinhamento de rochas vulcânicas da Cadeia Vitória-Trindade. Como prováveis consequências desse tectonismo, dada a contemporaneidade, citam-se o intenso vulcanismo eocênico da Plataforma de Cabo Frio e do Arquipélago de Abrolhos; o início da formação das bacias tafrogênicas do sudeste brasileiro (Taubaté, Resende, Volta Redonda e outras), o que ocorreu há pouco mais de 40 Ma; e as pronunciadas discordâncias sedimentares eocênicas e/ou o magmatismo contemporâneo observado nas sequências sedimentares das bacias marginais brasileiras de Pelotas, Santos, Campos, Espírito Santo, Cumuruxatiba e Jequitinhonha. Duas importantes feições morfológicas podem ser observadas no sentido de reforçar essa proposição. Uma delas diz respeito à pronunciada inflexão para sudoeste do limite oeste da Cadeia Vitória-Trindade, contrastando com sua direção preferencial leste-oeste, ao longo da zona oceânica. A outra, refere-se à tendência do alinhamento no sentido nordeste dos campos de petróleo da Bacia de Campos, a mais importante bacia produtora de petróleo do Brasil, coincidente com a direção proposta para o "caminhamento" do hot spot.

PALAVRAS-CHAVE:

- Cabo Frio, Rochas Alcalinas, Tectonismo

Conteúdo relacionado a CABO FRIO



Cabo Frio é um município brasileiro do estado do Rio de Janeiro. Localiza-se a 22°52'46" de latitude sul e 42°01'07" de longitude oeste, a uma altitude de quatro metros acima do nível do mar. Faz divisa com Armação dos Búzios a leste, Arraial do Cabo a sul, Araruama e São Pedro da Aldeia a oeste, e Casimiro de Abreu e Silva Jardim a norte. É o sétimo município mais antigo do Brasil e o principal da Região dos Lagos.

Veja mais informações em: http://dbpedia.org/resource/Cabo_Frio

RDF PDF RDF

FIGURA 5: Palavra-chave com *mashup* semântico. Thomaz Filho e Rodrigues (1999); Autoria própria.

4. RESULTADOS E DISCUSSÕES

Embora a alimentação dos dados no tesauro possa seguir uma concepção rígida, dentro um ambiente onde o mesmo possa ser implantado, é interessante também a ideia da utilização do sistema com participação do usuário, no que diz respeito ao princípio da folksonomia. Um tesauro de grandes proporções em seus termos pode ser beneficiado pelo número de pessoas que ofereçam contribuições para sua composição.

4.1 Possibilidades para Web Semântica

Para uma completa utilização do sistema de tesauro semântico aqui apresentado, o mesmo precisa fazer parte de um sistema maior, que gerencie também as publicações a serem classificados e cuja informação deva ser recuperada. Para explorar todo o potencial previsto na Web Semântica, esse sistema deve oferecer mecanismos de criação e associação de URI's únicas a cada registro nele inserido.

Estando o tesauro associado a um sistema maior e ambos preparados para funcionamento e utilização na Web, as possibilidades de ligação a dados externos são grandes, podendo existir *mashups Linked Data* para as entidades do tesauro, proporcionando resultados mais complexos para identificação de recursos nele relacionados.

5. CONSIDERAÇÕES FINAIS

Este trabalho apresentou o desenvolvimento de um sistema de tesauro para indexação de publicações, baseado em SKOS - uma ontologia específica para representação de conhecimento, a fim de proporcionar maior teor semântico à categorização das publicações, viabilizando, assim, a construção de sistemas nos quais a consulta por tais publicações possa ser realizada de forma precisa e consistente pela máquina para nos auxiliar.



No seu desenvolvimento, o sistema de tesouro mostrou-se mais adequado como um componente de software do que como um aplicativo independente. Tanto que o sistema já se encontra incorporado a um protótipo de maior abrangência, que reúne outras tecnologias capazes de ampliar ainda mais seus recursos, como por exemplo a Mineração de Textos.

Tal sistema faria uso do tesouro em um ambiente acadêmico, bibliotecas ou outros repositórios de recursos textuais, até mesmo ambientes empresariais e cartoriais, onde dados não estruturados são tão presentes. Favorece a utilização do sistema o fato deste oferecer buscas precisas, baseadas em semântica, sem ambiguidade de polissemia, além da ligação com dados externos, utilizando informações além daquelas que foram cadastradas no sistema.

Desenvolvimentos mais aprofundados para o sistema de tesouro podem seguir na utilização de mais propriedades da ontologia SKOS, além de adicionar novas ontologias candidatas, que oferecem mais recursos a cada registro. Possibilidades de *mashup Linked Data* externos ao sistema, também tendem a tornar resultados de termos mais completos.

A usabilidade e a aderência do sistema podem passar por testes mais específicos, de acordo com sua utilização, enquanto componente de outros sistemas maiores. Neste momento, o tesouro encontra-se disponível como parte de um protótipo de catalogação semântica de publicações acadêmicas, disponível em <http://www.niltonfjunior.com.br/pws>. Interessados em verificar o protótipo podem fazer acesso ao sistema, utilizando o termo *admin* para ambos os campos nome de usuário e senha.

REFERÊNCIAS

AITCHISON, Jean; GILCHRIST, Alan; BAWDEN, David. Thesaurus construction and use: a practical manual. Psychology Press, 2000.

BERNES-LEE, T. Linked Data. Design Issues about Web architecture, June 2006. Disponível em <<http://www.w3.org/DesignIssues/LinkedData.html>>. Acesso em: 28 abr. 2016.

BREITMAN, Karin Koogan. *Web semântica: a internet do futuro*. Rio de Janeiro: LTC, 2005.

DBPEDIA. Disponível em <<http://wiki.dbpedia.org/about>>. Acesso em 28 abr. 2016.

FUKS, Hugo. *Sistemas colaborativos*. Elsevier Brasil, 2011.

HEATH, Tom; BIZER, Christian. *Linked Data: Evolving the Web into a Global Data Space*. First Edition. Ebook ISBN: 9781608454310, 2011.

INEP - Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. O que é o Thesaurus. O que é thesaurus. Disponível em <<http://portal.inep.gov.br/pesquisa-thesaurus>>. Acesso em 21 abr. 2016.

JACOBS, Ian. SKOS Connects Diverse Knowledge Organization Systems to Linked Data. 2009. Disponível em <<https://www.w3.org/2009/07/skos-pr>>. Acesso em 21 abr. 2016.

MILES, A.; BECHHOFFER, S. SKOS Simple Knowledge Organization System Namespace Document. 2009. Disponível em <<http://www.w3.org/2009/08/skos-reference/skos.html>>. Acesso em 21 abr. 2016.

MySQL. Disponível em: <www.mysql.com/>. Acesso em 23 abr. 2016.

PHP.NET. O que é PHP? PHP.NET, 2015. Disponível em <http://php.net/manual/pt_BR/intro-what-is.php>. Acesso em 23 abr. 2016.

PRADO, Edmir; ORNELLAS, Regina; ARAÚJO, Luciano. Fundamentos de Sistemas de Informação. Elsevier Brasil, 2014.

PRESSMAN, Roger S. Engenharia de Software, Sétima Edição. Editora McGrawHill: Porto Alegre, 2011.

RAP - RDF API for PHP V0.9.6. Disponível em <<http://wifo5-03.informatik.uni-mannheim.de/bizer/rdfapi/index.html>>. Acesso em 22 abr. 2016.

REZENDE, S.O. Sistemas Inteligentes: fundamentos e aplicações. Barueri, SP: Manole, 2005.

THOMAZ FILHO, Antônio; RODRIGUES, Ana Lúcia. O alinhamento de rochas alcalinas Poços de Caldas-Cabo Frio (RJ) e sua continuidade na cadeia Vitória-Trindade. Brazilian Journal of Geology, v. 29, n. 2, p. 189-194, 1999.

TURBAN, E, VOLONINO, L. Tecnologia da Informação para Gestão: em busca do melhor desempenho estratégico e operacional. 8.ed. Porto Alegre, RS: Bookman Editora, 2011.

Visual Paradigm Community Edition. FREE for non-commercial use only. Disponível em <<https://www.visual-paradigm.com/download/community.jsp>>. Acesso em 23 abr. 2016.