

# NavCon - uma arquitetura para navegação conceitual em documentos Web

José Renato Villela Dantas, José Bezerra da Silva Filho, and Pedro Porfírio Muniz Farias

Universidade de Fortaleza, Av. Washington Soares, 1321, Fortaleza, Brasil,  
`jose.renato@edu.unifor.br`, `{bezerra,porfirio}@unifor.br`

**Abstract.** This article presents conceptual navigation and NavCon, an architecture that implements this navigation in World-Wide Web pages. NavCon architecture makes use of ontology to contextualize user's search for information. Conceptual links, based on ontologies, are automatically inserted in Web pages. These links permit the users to browse in a graph representing concepts and their relationships to reach documents associated with a desired ontology concept.

## 1 Introdução

A World-Wide Web (ou simplesmente Web) consiste basicamente numa rede de documentos cujos conteúdos são compreensíveis apenas por seres humanos. A Web Semântica é uma evolução da Web proposta por Berners-Lee [2] na qual metadados sejam associados ao conteúdo dos documentos Web de forma que este seja compreendido também por agentes de software que possam fazer inferências, deduzir e buscar informações na rede. A idéia da Web Semântica não se encontra amplamente implementada[18].

Tais metadados podem estar associados a ontologias. Ontologias são uma tecnologia chave para tornar a Web Semântica possível [8]. Uma ontologia descreve formalmente um domínio consistindo de uma lista de termos e dos relacionamentos entre eles. Os termos representam conceitos do domínio [1].

Para efeitos deste trabalho, considera-se que uma ontologia é capaz de capturar e representar adequadamente um contexto. Neste caso, o contexto pode ser entendido como o conjunto de informações relevantes ao usuário para a execução de uma tarefa.

Este artigo introduz o conceito de navegação conceitual e apresenta o NavCon, uma arquitetura que o implementa. A navegação conceitual faz uso de uma rede de conceitos, previamente organizada no formato de ontologias, para proporcionar mecanismos que otimizem a navegação do usuário pela Web. A utilização de um contexto visa reduzir a recuperação de informações desnecessárias auxiliando o usuário a manter o foco em sua pesquisa.

Este artigo encontra-se estruturado da seguinte maneira: na seção 2, apresenta-se o conceito do uso de ontologias para caracterização de um contexto. Na seção 3, explica-se o uso de metadados para recuperação de documentos. Na seção

4, apresenta-se a definição de navegação conceitual. Na seção 5, descreve-se a arquitetura proposta para navegação conceitual e mostra-se um exemplo do protótipo construído a partir desta arquitetura. Por fim, na seção 6, apresentam-se a conclusão e propostas para trabalhos futuros.

## 2 O uso de ontologia para caracterizar um contexto

Com frequência, ontologias são usadas para caracterizar um contexto [19, 6]. Este contexto pode servir como filtro para as informações existentes na Internet que são de interesse para o usuário.

Dey [9] define contexto como “toda informação que pode caracterizar a situação de uma entidade. Uma entidade é uma pessoa, lugar ou objeto considerado relevante para a interação entre o usuário e uma aplicação, incluindo o próprio usuário e a própria aplicação”. O autor ainda define que “um sistema é baseado em contexto se ele usa o contexto para fornecer informação e/ou serviço relevante ao usuário onde a relevância depende da tarefa do usuário”.

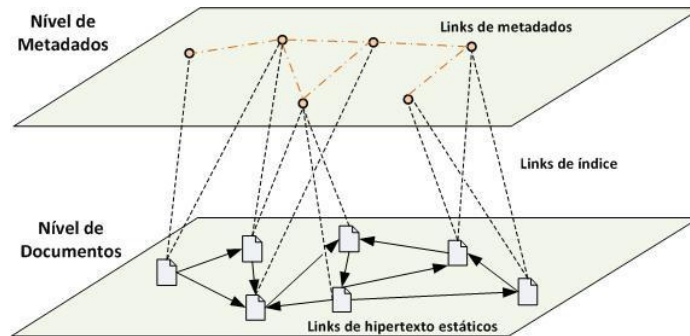
Quando se trata de navegação pela Web para busca de informação, pode-se considerar que o contexto da navegação está associado ao assunto a que o usuário pesquisador se refere. A navegação inserida em um contexto deve exibir para o usuário informação relevante relacionada à tarefa que ele executa. Por exemplo, Ao navegar dentro do contexto de seu ambiente de trabalho, um usuário buscará documentos relacionados aos assuntos deste ambiente. O contexto ajuda a filtrar a informação que será apresentada ou a direcionar a informação para pontos mais relevantes às necessidades do usuário no momento da consulta.

## 3 Utilização de metadados para recuperação de documentos

Vista como um sistema de hipertexto, a Web é uma rede de links entre documentos na qual o usuário navega usando marcações visuais. Estes links são estáticos e são definidos pelo autor do documento no momento de sua criação. Segundo Carr et alii [5], o formato de links estáticos inseridos em páginas HTML pode ser considerado um sistema de hipermídia fechado. Em sistemas de hipermídia abertos (OHS), os links são metadados armazenados e mantidos em separado dos dados de lay-out da página.

Um dos principais problemas relacionados aos links estáticos da Web é sua constante desatualização. Como os tradicionais links estáticos estão definidos dentro dos documentos HTML, eles ficam inconsistentes sempre que as páginas às quais se referenciam mudam de endereço. O fato dos links estáticos conterem, em si mesmos, a direção a ser navegada é apontado como outro problema deste tipo de link. A direção de navegação que os links irão seguir é definida pelo autor e não pode ser alterada pelo consumidor da página [7, 8].

A figura 1 mostra o modelo de uso de metadados para recuperação de documentos. O modelo apresenta dois níveis de ligações: nível de metadados e nível de



**Fig. 1.** Modelo de recuperação de documentos usando metadados.

documentos. No nível de documentos, os documentos possuem abstração entre si feitas através de links estáticos definidos nos documentos.

Usualmente, podem-se usar texto, tags ou conceitos para representar os metadados. Os metadados podem possuir relacionamentos entre si. Os documentos podem se ligar a outros documentos através do nível de metadados. Os metadados ligam-se aos documentos por links de índice.

O uso de texto como metadados para recuperação de documentos pode ser encontrado em ferramentas de busca como Google [12] e Yahoo! [26]. O texto é subdividido em palavras que podem ser consideradas como metadados uma vez que são utilizadas para a criação de índices invertidos que registram as ocorrências de cada palavra em um conjunto de documentos.

Tags são usadas em ferramentas para anotação social [25]. Neste tipo de ferramenta, sugere-se que o próprio usuário crie a camada de metadados definindo quais tags se aplicam aos documentos. Tipicamente os documentos estão ligados a menos tags que palavras. Por outro lado, como cada tag pode ser composta pela combinação de várias palavras, o número de tags pode resultar significativamente superior ao número de palavras de uma linguagem.

Os sistemas Twine [23], Quintura [17] e Semantic Turkey [13] são exemplos de ferramentas que usam tags como metadados para recuperação de documentos. O Twine é uma ferramenta de anotação social que permite criar, compartilhar e organizar as tags em grupos de interesses específicos. O Quintura é uma ferramenta de busca que adiciona recursos para navegação por tags relacionadas à pesquisa realizada pelo usuário. O Semantic Turkey é uma ferramenta de anotação para páginas Web que permite inserir anotações nas páginas visitadas pelo usuário para recuperação posterior destas páginas.

Recursos RDF e conceitos presentes em ontologias também podem ser usadas como metadados. Assim, na camada de metadados, pode-se encontrar conceitos que são semanticamente relacionados entre si e, a partir de um índice, relacionados com os documentos. O uso de conceitos na camada de metadados é vantajoso pois pode ser utilizado para definir um contexto onde o usuário irá navegar.

Dentre as ferramentas que usam recursos RDF como metadados pode-se citar o Tabulator [3] e o Disco [10]. Estes dois sistemas são browsers Web que permitem navegar por uma rede de recursos RDF e acessar documentos relacionados a estes recursos.

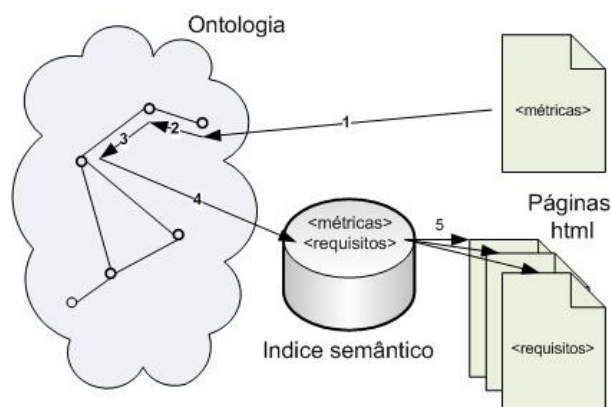
Outro sistema baseado no uso de ontologias como metadados e na teoria de sistemas de hipermídia aberto, é o sistema COHSE [4, 27] - Conceptual Open Hypermedia Service. O COHSE combina o uso de sistemas de hipermídia aberta com ontologias para criar links em documentos Web baseados em ontologias. O sistema gera links que disparam consultas, sobre os termos marcados com estes links, em ferramentas de busca tais como o Google e outros.

#### 4 Navegação conceitual

A navegação conceitual é uma forma de se navegar em sítios Web inserido dentro de um contexto. Tal contexto pode ser representado por uma ontologia que fornece suporte à navegação.

Um navegador conceitual navega por relacionamentos em uma rede de conceitos para atingir documentos associados a estes conceitos.

A figura 2 mostra uma seqüência de passos durante uma navegação conceitual. No passo 1, a partir do link conceitual <métricas> tem-se acesso ao respectivo conceito de <métricas> na ontologia. Nos passos 2 e 3 navega-se através da rede de conceitos da ontologia até o conceito de <requisitos>. A partir daí, consultando, no passo 4, o índice semântico, obtém-se, no passo 5, o conjunto de páginas relacionadas ao conceito <requisitos>.



**Fig. 2.** Exemplo de navegação conceitual.

A navegação conceitual supõe a existência de um índice semântico que associa conceitos de uma ontologia a páginas Web onde estes conceitos ocorrem. Um

índice semântico pode ser produzido por um sistema de indexação semântico. Tudhope [22] define que um sistema de indexação semântica é aquele que possui um vocabulário controlado, geralmente definido manualmente, com os termos do índice representando conceitos únicos que se inter-relacionam. Fonsêca [11] sugere que a indexação semântica pode tornar mais precisa a busca do usuário por informações textuais, especialmente em páginas Web.

Ao utilizar uma ontologia para conduzir a navegação, tem-se como vantagem permitir a visualização de uma estrutura de conhecimento específica sobre um determinado assunto dando oportunidade aos usuários de um melhor entendimento desse conhecimento. A estrutura do conhecimento é representada pela ontologia. Os usuários terão acesso à estrutura de conhecimento ao navegar pelos conceitos da ontologia e nos documentos relacionados a partir dos conceitos. Além disto, como o índice semântico é continuamente atualizado, os problemas de links para documentos inexistentes deixam de ocorrer.

A navegação conceitual utiliza recursos de navegação por uma ontologia para acessar os conceitos representados nesta ontologia. A navegação na ontologia pode ser feita usando recursos gráficos para facilitar a identificação dos conceitos e seus relacionamentos. O endereço apontado pelo link conceitual exibe um gráfico mostrando os conceitos da ontologia e os relacionamentos entre os conceitos.

## 5 Uma arquitetura para navegação conceitual

A arquitetura proposta neste artigo, denominada NavCon, pode ser dividida em dois módulos. O primeiro módulo manipula o conteúdo de páginas Web inserindo links conceituais. O segundo módulo, a partir destes links conceituais, conduz a navegação pela ontologia até os documentos desejados.

### 5.1 Módulo de inserção de links conceituais em página Web

O primeiro módulo tem como finalidade inserir links conceituais em páginas Web. O processo de inserção dos links conceituais é ilustrado na figura 3. Um usuário faz, do modo usual, uma requisição de uma página Web qualquer. Um proxy captura a página solicitada pelo usuário e a percorre buscando identificar termos presentes que sejam associados a conceitos da ontologia. Os conceitos encontrados são marcados com links conceituais. No final, a página alterada é enviada para o cliente.

A ontologia que será usada, representada em linguagem OWL, é configurada no proxy e define o contexto no qual o usuário está inserido. O serviço de marcação implementado oferece um recurso para selecionar as ontologias que definem o contexto.

O contexto definido será comum a todos os usuários uma vez que a seleção das ontologias não é feita de maneira individual mas coletiva. Pretende-se, futuramente, incluir um recurso de seleção das ontologias para permitir a seleção

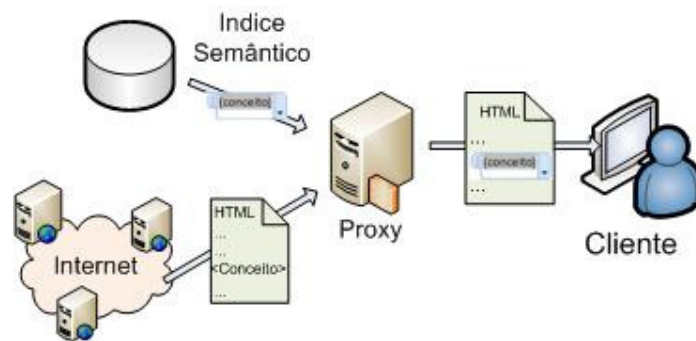


Fig. 3. Inclusão de link conceitual em uma página Web.

por usuário. As duas formas de seleção são importantes se considerarmos os ambientes diferentes de uso dos recursos da Web. Por exemplo, em uma intranet dentro de uma empresa, o contexto seria definido por um gestor ou grupo de gestores que ficariam responsáveis por configurar as ontologias no proxy. Já em um ambiente mais aberto, como a própria Internet, um recurso para o usuário configurar seu perfil passa a ser mais vantajoso. O usuário pode informar quais ontologias prefere usar definindo assim o contexto no qual deseja se inserir.

O sistema permite indicar mais de uma ontologia para a definição do contexto. Sendo assim, pode ocorrer de várias ontologias serem utilizadas ao mesmo tempo. O sistema identifica, pelos conceitos encontrados na página Web, à qual ontologia a página está relacionada.

Para a implementação do primeiro módulo, utilizou-se o servidor proxy WBI 4.5 [24], da IBM. O WBI é um proxy configurável, usando uma API Java, apropriado para a criação de aplicações intermediárias na Web. Segundo o WBI, aplicações intermediárias são entidades computacionais que podem ser colocadas em qualquer posição no caminho do fluxo da informação. As aplicações intermediárias são usadas para adicionar funcionalidades a um sistema quando nem a fonte produtora de dados nem o consumidor podem ser modificados.

No WBI foi implementado um serviço que recebe uma página Web, varre o conteúdo desta página identificando conceitos, busca estes conceitos em ontologias pré-definidas e faz a marcação no código HTML da página inserindo um link conceitual. O link conceitual diferencia-se dos demais links pois aponta para o endereço que aciona o applet de navegação na ontologia. Um código Javascript também é adicionado ao conteúdo da página para exibir um ícone próximo ao link conceitual. Este ícone permite, visualmente, identificar quais links da página são links conceituais distinguindo-os dos links originais da página.

O ícone de link conceitual é exibido apenas quando o ponteiro do mouse encontra-se sobre o link conceitual. A intenção é evitar ao máximo que a inserção de links conceituais faça alterações no lay-out original da página.

No momento, o sistema que faz a marcação não sobrescreve ou substitui os links já existentes na página. Se um conceito identificado pelo sistema já possui um link criado pelo construtor da página, este conceito não é marcado com o link conceitual.

A arquitetura NavCon apresenta uma proposta semelhante ao COHSE, descrito na seção 3, pelo fato de inserir links conceituais, baseados em ontologias, em páginas Web. A principal diferença é que o COHSE aciona ferramentas de busca textuais, como o Google, a partir de um conceito marcado em uma página Web enquanto o NavCon permite a navegação através de um grafo de conceitos e utiliza um índice semântico para fornecer acesso a outros documentos.

## 5.2 Módulo de navegação em conceitos de uma ontologia

O segundo módulo realiza, a partir do acionamento do link conceitual, a navegação pela ontologia e a recuperação, utilizando, o índice semântico, dos documentos HTML relacionados ao conceito desejado.

A ferramenta de navegação acionada pelo link conceitual é uma adaptação do TGLinkBrowser [20] e é basicamente constituída por um applet que exhibe um grafo contendo os conceitos da ontologia e seus relacionamentos. O usuário pode navegar por estes conceitos. Uma vez selecionado um conceito, é possível acessar documentos externos a ele relacionados.

O TGLinkBrowser foi desenvolvido usando o componente TouchGraph [21] que é uma biblioteca Java para criação e navegação de gráficos de rede interativos. O TouchGraph oferece recursos de visualização tais como cores, tamanhos e formas distintas para indicar categorias, relacionamentos e importância relativa dos itens. Além disso, oferece visualização panorâmica e recursos de zoom, mudanças de perspectivas através de mecanismo de segurar e arrastar itens mudando assim a perspectiva de visão da rede, ocultar e exibir nós e seus respectivos relacionamentos e possibilidade de visualização de recursos externos à rede visualizada.

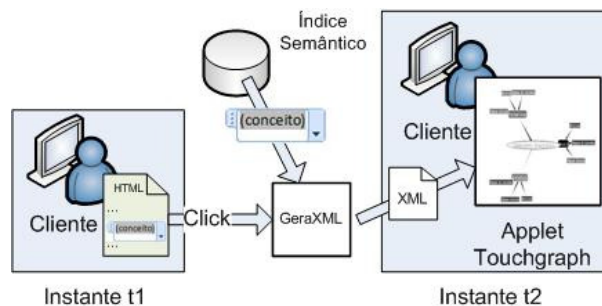


Fig. 4. Modelo de navegação gráfica sobre conceitos.

A figura 4 ilustra a implementação do NavCon com o TGLinkBrowser. No instante t1, aciona-se o link conceitual. Após este acionamento, representado pela seta “Click”, é executado o procedimento GeraXML. Este procedimento consulta a ontologia e o índice semântico e gera um arquivo XML que descreve o grafo a ser apresentado pelo Applet Touchgraph ao usuário no instante t2. O GeraXML foi implementado usando bibliotecas do Jena 2 [14].

O NavCon exibe uma rede de relacionamentos construída a partir das classes de uma ontologia onde os nós indicam as classes e as arestas indicam os relacionamentos entre as classes.

### 5.3 Exemplo de uso da arquitetura

Apresenta-se a seguir um exemplo de navegação conceitual com a utilização do NavCon. Neste exemplo, utilizou-se a ontologia Pizza [15] encontrada no tutorial do programa de edição de ontologias Protégé-OWL [16].

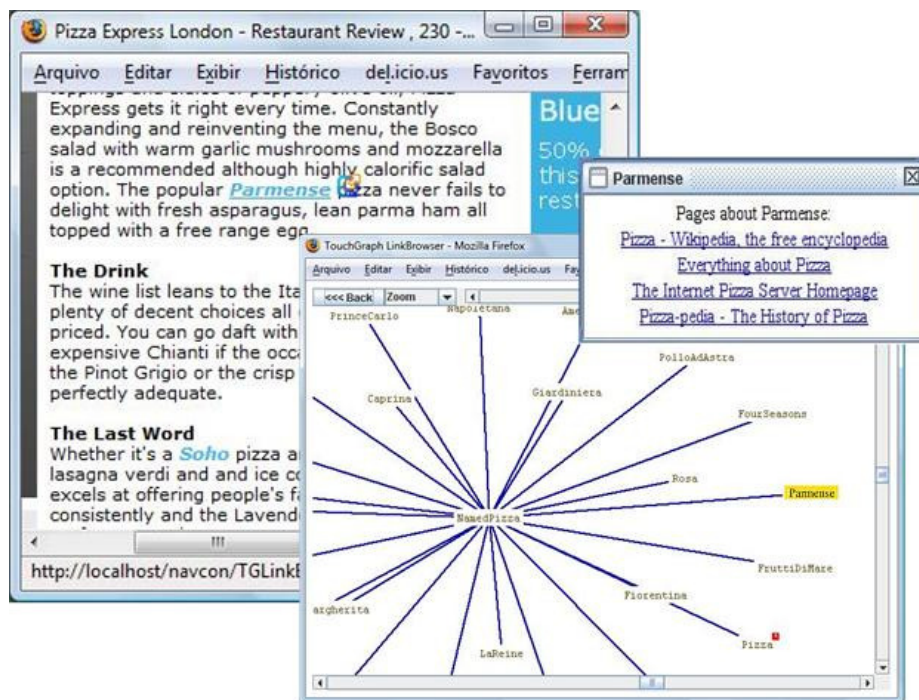


Fig. 5. Página com marcações conceituais.

A figura 5 exibe três janelas. A primeira janela mostra um exemplo de página Web com links conceituais introduzidos pelo NavCon. Ao carregar a página



solicitada pelo usuário, o NavCon identificou os termos ⟨Parmense⟩ e ⟨Soho⟩ e fez a marcação com links conceituais. Pode-se perceber, logo acima do link, o ícone indicando que trata-se de um link conceitual.

Ao serem acionados, os links conceituais disparam o applet, apresentado na segunda janela, que permite visualizar os conceitos da ontologia relacionados ao termo marcado. O conceito realçado é o que foi acionado na página Web exibida na primeira janela, no caso, o termo ⟨Parmense⟩. O grafo exibe os conceitos que possuem relacionamento direto com o conceito selecionado. A partir deste grafo inicial, pode-se navegar pelos demais conceitos que inicialmente não são exibidos. O nome do relacionamento é exibido quando a aresta que liga dois conceitos é acionada.

A terceira janela mostra uma lista de links relacionados ao conceito ⟨Parmense⟩. Esta lista foi obtida ao passar o ponteiro do mouse sobre o conceito ⟨Parmense⟩ exibido na segunda janela. Através dos links exibidos na terceira janela, tem-se acesso aos documentos para os quais eles apontam.

## 6 Conclusão

Neste artigo, apresentamos o conceito de navegação conceitual. A navegação conceitual utiliza recursos de web semântica e técnicas de open linked systems para propor uma forma do usuário navegar na Web inserido em um contexto. Este contexto é determinado pela ontologia utilizada. Assim, navega-se não apenas em documentos, mas através dos relacionamentos entre conceitos encontrados nos documentos.

Em seguida, descrevemos a arquitetura do NavCon para navegação conceitual. Esta arquitetura realiza a inserção de links conceituais em páginas web através de um proxy. O NavCon oferece recursos à navegação pela rede de conceitos das ontologias, consulta ao índice semântico a partir de um conceito da ontologia e acessa os documentos relacionados a este conceito através do índice semântico.

Como trabalhos futuros, pretende-se evoluir algumas das funcionalidades já encontradas na ferramenta. O algoritmo de identificação de conceitos na página web precisa ser trabalhado para melhorar a precisão na identificação de conceitos com mais de um termo. A definição de contextos com mais de uma ontologia precisa ser desenvolvida de forma que o NavCon possa identificar corretamente à qual ontologia o conceito pertence. É interessante elaborar uma forma de apresentar opções ao usuário quando o mesmo conceito for localizado em mais de uma ontologia. Por fim, pretende-se permitir que um usuário possa selecionar as ontologias e criar seu contexto individual sem que essa definição fique a cargo de um administrador do sistema.

## References

1. Antoniou, G., van Harmelen, F.: A Semantic Web Primer. The MIT Press, USA (2004)

2. Berners-Lee, T., Hendler, J., Lassila, O.: The Semantic Web. *Scientific American*, May (2001)
3. Berners-Lee, T., Chen, Y., Chilton, L., Connolly, D., Dhanaraj, R., Hollenbach, J., Lerer, A., Sheets, D.: Tabulator: Exploring and analyzing linked data on the semantic web. In: *Proceedings of the 3rd International Semantic Web User Interaction Workshop*, Atenas (2007)
4. Carr, L., Hall, W., Bechhofer, S., Goble, C.: Conceptual Linking: Ontology-based Open Hypermedia. In: *Proceedings of the 10th international conference on World Wide Web*, pp. 334–342. ACM, Hong Kong (2001)
5. Carr, L., De Roure, D., Hall, W., Hill, G.: Implementing an Open Link Service for the World-Wide Web. *World Wide Web* 1(2), 61–71 (1998)
6. Benta, K., Rarău, A., Cremene, M.: Ontology Based Affective Context Representation. In: *Proceedings of the 2007 Euro American conference on Telematics and information systems*, pp. 1–9, ACM, Faro (2007)
7. Crampes, M., Ranwez, S.: Ontology-Supported and Ontology-Driven Conceptual Navigation on the World-Wide Web. In: *Proceedings of the eleventh ACM on Hypertext and hypermedia*, pp. 191–199. ACM, San Antonio (2000)
8. Davies, J., Fensel, D., van Harmelen, F.: *Towards the semantic web: ontology-driven knowledge management*. John Wiley & Sons, England (2003)
9. Dey, A.: Understanding and Using Context. *Personal Ubiquitous Comput.* 5(1), 4–7 (2001)
10. Disco: Hyperdata Browser. <http://sites.wiwiss.fu-berlin.de/suhl/bizer/ng4j/disco/>
11. Fõnseca, J.: Um modelo de indexação semântica para intranet. Unifor, Brasil (2002)
12. Google <http://www.google.com>
13. Griesi, D., Pazienza, M., Stellato, A.: Semantic Turkey: a Semantic Bookmarking Tool. In: *Proceedings of the 4th European Semantic Web Conference*, pp. 779–788, Springer, Innsbruck (2007)
14. Jena 2 <http://jena.sourceforge.net/>
15. Pizza Ontology <http://www.co-ode.org/ontologies/>
16. Protégé <http://protege.stanford.edu/>
17. Quintura <http://company.quintura.com/quintura/>
18. Shadbolt, N., Berners-Lee, T., Hall, W.: The Semantic Web Revisited. *IEEE Intelligent Systems* 21(3), 96–101 (2006)
19. Strang, T., Linnhoff-Popien, C., Frank, K.: CoOL: A Context Ontology Language to enable Contextual Interoperability. In: *Proceedings of 4th IFIP WG 6.1 International Conference on Distributed Applications and Interoperable Systems*, pp. 236–247, Springer Verlag, Paris (2003)
20. TGLinkBrowser [http://sourceforge.net/project/showfiles.php?group\\_id=30469](http://sourceforge.net/project/showfiles.php?group_id=30469)
21. Touchgraph <http://www.touchgraph.com/>
22. Tudhope, D., Cunliffe, D.: Semantically indexed hypermedia: linking information disciplines. *ACM Comput. Surv.* 31(4), 4 (1999)
23. Twine <http://www.twine.com>
24. WBI <http://www.almaden.ibm.com/cs/wbi/>
25. Wu, X., Zhang, L., Yu, Y.: Exploring social annotations for the semantic web. In: *Proceedings of the 15th international Conference on World Wide Web*, pp. 417–426, ACM, Edinburgh (2006)
26. Yahoo! <http://www.yahoo.com>
27. Yesilada, Y., Bechhofer, S., Horan, B.: COHSE:Dynamic Linking of Web Resources. <http://research.sun.com/techrep/2007/abstract-167.html> (2007)