

FEITOSA, Ailton. **Organização da informação na web**: das tags à web semântica. Brasília: Thesaurus, 2006. 132 p.

OS CAMINHOS DA WEB

Ailton Feitosa é Doutor em Ciência da Informação pela Universidade de Brasília e Professor da mesma universidade. É consultor em Tecnologias de Informação e Comunicação e um dos pioneiros em pesquisas que focam as interfaces entre a Ciência da Informação e a web semântica. Em **Organização da Informação na Web: das tags à web semântica**, Feitosa discorre cronologicamente a evolução da *World Wide Web* impulsionando pesquisadores interessados no assunto a percorrem e a refletirem os caminhos, em via de desenvolvimento e evolução, da web semântica.

O livro de Feitosa vem corroborar a atual efervescência da web semântica em uma introdução, informativa, três capítulos principais que se revelam como a espinha dorsal de sua obra e uma conclusão que sintetiza seu conteúdo. Para o autor a web semântica é a evolução da *web* atual rumo a uma *Internet* inteligente constituída de sistemas capazes de serem inteligíveis por máquinas e pessoas.

Diferente daquilo que é comumente encontrado em introduções de livros, Feitosa inicia sua obra com uma carga significativa de dados e informações referentes ao uso da *Internet* e o que este uso acarreta para o universo da informação. Em vez de introduzir com frases de efeito que apresentem a estrutura e o conteúdo do livro de forma genérica, o autor opta por informar seu leitor, logo de início, a respeito dos usuários conectados à *Internet* em todo mundo, a posição dos países pela quantidade de servidores de *Internet* e de domínios registrados no Brasil. Esta realidade revela que embora a América Latina supere apenas a África e o Oriente Médio em número de usuários, o Brasil é o oitavo país do mundo em números de servidores de *Internet*. O aumento vertiginoso do consumo de informações via *Internet*, é enfatizado como mola propulsora para o progresso dos serviços ofertados pela *web*.

Debruçando-se em conceitos da Ciência da Informação, Feitosa apresenta no segundo capítulo uma visão geral do ciclo documentário discorrendo a respeito do tratamento intelectual da informação. Destina maior atenção ao processo de

indexação de conteúdos informacionais, sobretudo à contribuição que a indexação automática ou semi-automática, baseada em processos lingüísticos e estatísticos, pode fornecer à recuperação de conteúdos na *web*. Nesse capítulo ao abordar o tema 'precisão' e 'revocação' na busca automática de informação, Feitosa explica claramente seus fundamentos, mas, quando opta por ilustrar sua explicação por meio de fórmulas matemáticas, lança mão de siglas que se confundem, principalmente no que diz respeito ao princípio de que precisão e revocação são grandezas inversamente proporcionais. Porém, tal inadequação se limita à esfera da codificação das fórmulas, fato que não impede a compreensão do texto, apenas dá vazão a possíveis questionamentos.

Os serviços de indexação e recuperação na *web* são abordados no terceiro capítulo, que inicia com um breve histórico acerca dos buscadores, meta-buscadores e diretórios de busca disponíveis na *web*. Explica como se dá o processo de comunicação e localização no universo digital por meio das meta tags, que são as primeiras técnicas de otimização para mecanismos de busca automática. Em seguida percorre com mais detalhes a tendência dos padrões de metadados na organização da informação na *web*. Apresenta as metodologias de recuperação que surgiram e que estão sendo gradativamente mais utilizadas para a organização e recuperação de conteúdos *web*, metodologias essas baseadas em conhecimento da Inteligência Artificial, Ciências Cognitivas e Matemática, além de lançar mão também das tecnologias que se fundamentam na descrição de documentos e controle terminológico de termos semanticamente relacionados. O abuso na utilização de tags (marcas) que servem para identificar metas e *links* exigiu dos atores da *web* o desenvolvimento de padrões de metadados, tais como os originários da Ciência da Informação *Dublin Core*, AACR2, MARC etc.

O quarto capítulo é o mais descritivo e extenso da obra de Feitosa. Nele o autor trata dos aspectos teóricos da web semântica. Descreve detalhadamente sua trajetória, desde sua concep-

ção em 1989 até seu estágio atual. Foca atenção nos elementos que constituem todo o funcionamento da *web*, as ontologias, as linguagens de marcação e os esquemas construídos com base nessas linguagens. Explícita como a junção de ontologias e esquemas de linguagens de marcação podem formar a base para uma *web* inteligível por máquinas e programas concebendo um entendimento comum e compartilhado em um determinado assunto. O autor descreve as três gerações da *web* até o momento. A primeira é pautada na linguagem *Hypertext Markup Language* - HTML, e possibilitou a exibição de documentos de qualquer localização física; a segunda (embasada pela linguagem *eXtensible Markup Language* - XML) possibilitou diferentes formas de apresentação para uma mesma estrutura documental e; a terceira, que separou os significados das estruturas, estendendo a noção de informação, viabiliza (ou pretende viabilizar) uma semântica reconhecível por máquina. Em realidade a *web* semântica, como bem afirma o autor, ainda está por vir, é a evolução da *web* rumo a uma *Internet* inteligente constituída de sistemas capazes de decodificar significados concebidos por humanos.

A obra é destinada principalmente àqueles que atuam com representação e gestão do conhecimento, recuperação da informação e inteligência artificial. A clareza e a concisão do texto permitem também que qualquer leitor interessado em compreender o universo da *web* e os fundamentos da *web* semântica apreenda seu conteúdo. A contribuição da respectiva obra para a Ciência da Informação está na reunião de informações concernentes à *web* semântica que correntemente se encontram espalhadas na literatura interdisciplinar da Ciência da Computação, Engenharia da Computação e Ciência da Informação. Feitosa consegue reunir em um único livro o conteúdo de textos de diversas áreas.

Rodrigo de Sales

Mestrando em Ciência da Informação -
UFSC. Bacharel em Biblioteconomia - UFSC. E-mail: rodrigo.sales.s@gmail.com