

AILTON FEITOSA

ORGANIZAÇÃO DA INFORMAÇÃO NA WEB:  
DAS TAGS À WEB SEMÂNTICA

Estudos Avançados em Ciência da Informação - V. 2



# Sumário

Lista de abreviaturas e siglas .....	9
1 Introdução .....	11
2 Elementos de Organização da Informação .....	17
2.1 Documentos .....	17
2.2 Ciclo Documentário .....	18
2.3 Tratamento intelectual da informação .....	21
2.3.1 Indexação .....	22
2.3.1.1 Índices e Resumos .....	23
2.3.1.2 Indexação automática .....	24
2.3.2 Recuperação da Informação .....	28
2.3.2.1 Modelos de Recuperação da Informação .....	30
2.4 Resumo do capítulo .....	31
3 Serviços de Indexação e Recuperação na <i>web</i> .....	33
3.1 Breve histórico dos serviços de busca na Web .....	33
3.2 Meta Tags: as primeiras técnicas de otimização da web para os mecanismos de busca .....	47
3.3 Padrões de meta dados: tendências da organização na web .....	50
3.4 Resumo do capítulo .....	59
4 <i>Web Semântica</i> .....	63
4.1 Caracterização da <i>web semântica</i> .....	63
4.2 Ontologias .....	70
4.3 Infra-estrutura da <i>Web Semântica</i> .....	78
4.4 As gerações da <i>web</i> .....	83

4.4.1 A 1ª geração: HTML E XHTML .....	85
4.4.2 A 2ª geração: XML, DTD e XML Schema .....	86
4.4.2.1 XML .....	88
4.4.2.2 XML-Schema .....	93
4.4.3 A 3ª geração: padrões e linguagens de especificação de ontologias (RDF, RDF Schema, DAML+OIL, OWL) .....	98
4.4.3.1 RDF.....	101
4.4.3.2 RDFS .....	118
4.4.3.3 OWL .....	113
4.5 Resumo do capítulo .....	120
5 Considerações finais .....	123
Bibliografia .....	125

# 1 Introdução

O advento do serviço *World Wide Web* na *Internet* proporcionou grande crescimento na quantidade de informações, de usuários e de computadores ligados à rede. De acordo com dados das instituições de pesquisa *NUA Internet Surveys*<sup>1</sup> e *Nielsen NetRatings*<sup>2</sup>, em todo o mundo, dados do ano de 2003 apontavam para cerca de 605,60 milhões de pessoas com acesso à *Internet* (tabela 1) e, no Brasil, em julho de 2003, cerca de 14,3 milhões. Segundo estimativas da NUA, somente nos Estados Unidos, até o final de 2003, foram realizados negócios entre empresas na rede atingindo a soma de 1,3 trilhão de dólares (ilustração 1). Dados do Comitê Gestor da *Internet* no Brasil (2004), indicavam a existência, em janeiro de 2004, de mais de 223 milhões de servidores de *Internet* em todo o mundo, distribuídos entre trinta países. A tabela 2 apresenta as dez primeiras posições nessa classificação, destacando o Brasil em 8º lugar. Ao lado dessas estatísticas, o Serviço de Domínios para a *Internet* no Brasil – Registro.br (<http://registro.br>) contabilizava, em dezembro de 2003, mais de 509 mil domínios em todo o país, número que está hoje (2006), na ordem de mais de 859 mil (ilustração 2). Em todo o mundo são quase 57 milhões de *sites* publicados na *web* (SERVER WATCH, 2004). Atualizando as informações para o ano de 2005, pesquisas do site *ClickZnetwork*<sup>3</sup> apontam para um número de 1,08 bilhão de usuários, sendo 22,32 milhões o número global de brasileiros com acesso à rede, dos quais cerca de 12,52 milhões são usuários ativos. A previsão para o ano de 2010 é de que o número de usuários da *Internet* em todo o mundo salte para 1,8 bilhão.

---

<sup>1</sup> [http://www.nua.com/surveys/how\\_many\\_online/index.html](http://www.nua.com/surveys/how_many_online/index.html)

<sup>2</sup> [http://www.nielsennetratings.com/news.jsp?section=dat\\_to&country=br](http://www.nielsennetratings.com/news.jsp?section=dat_to&country=br)

<sup>3</sup> [http://www.clickz.com/stats/web\\_worldwide/](http://www.clickz.com/stats/web_worldwide/)