

# HTML5: novidades e contribuições

Daniel Araújo de Lucena<sup>1</sup>; Ana Carolina Costa de Oliveira<sup>2</sup>

<sup>1</sup>Universidade Estadual da Paraíba (UEPB)  
Campus VII – Governador Antônio Mariz – Patos – PB – Brasil

<sup>2</sup>Instituto Federal de Educação, Ciência e Tecnologia da Paraíba (IFPB)  
Núcleo Avançado de Guarabira – Guarabira – PB - Brasil

{daniel\_comput, carolyneolive}@hotmail.com

**Abstract.** This paper aims to confirm some of the contributions of the HTML5 to secure the claims of the Semantic Web. The present paper is justified by countless studies that searched millions of websites, new tags to identify the main parts of a webpage, which makes it easier to the search engines return more relevant results for a search. The paper presents a descriptive and exploratory character. As a result, it can show the contribution that the HTML5 becomes, even in development, a fundamental support for the ambitions of Web 3.0, because it deals directly with the structure of sites.

**Resumo.** Este artigo tem o objetivo de confirmar algumas das contribuições do HTML5 para firmar as pretensões da *Web Semântica*. O presente trabalho justifica-se pelos inúmeros estudos que vasculharam milhões de sites, novas *tags* para identificar as partes principais de uma página eletrônica, o que facilita aos motores de busca retornar resultados mais relevantes durante uma pesquisa. O artigo apresenta caráter descritivo e exploratório. Como resultados, pode-se apresentar a contribuição que o HTML5 se torna, mesmo em desenvolvimento, um apoio fundamental às ambições da *Web 3.0*, pois lida diretamente com a estruturação dos sites.

## 1. Introdução

Na atual concepção de intensa produção e troca de informações entre usuário em que a *Web* está inserida, as dificuldades em retornar dados relevantes numa busca se tornam cada vez mais presentes (ALVARENGA, 2004).

Para ajudar a reverter esse quadro está em desenvolvimento o conceito de *Web Semântica*, ou seja, um ambiente em que a informação estará bem definida e identificada por padrões que possibilitarão aos motores de busca “entenderem” o sentido em que cada informação foi descrita (SOUZA, 2004).

É no intuito de contribuir para que haja essa padronização que o HTML5 foi desenvolvido. A proposta é que essa nova versão da linguagem de Lee consiga estruturar melhor um *site*, permitindo que o motor de busca logo identifique cada área da página eletrônica a partir de suas *tags* específicas.

O presente artigo tem por objetivo confirmar algumas das contribuições do HTML5 para firmar as pretensões da *Web Semântica*. Para assim fazer, o mesmo detém de estudo bibliográfico e pesquisa com desenvolvedores de páginas eletrônicas, os quais subjugarão como relevantes às novidades trazidas pela nova versão do HTML.

## 2. Abordagem teórica

## 2.1 Web Semântica

A partir de 2004 a *Web* começou a se caracterizar como um meio de troca de informações e de colaboração entre os internautas através de *sites* e serviços virtuais. Conhecida como *Web 2.0*, esse período possibilita aos usuários gerar seus conteúdos, criando e compartilhando experiências. É a versão atual da *Web*, exemplificada nas redes sociais como *Facebook* e canais como o *Youtube* (SCHLEMMER, 2009).

Com a possibilidade de o usuário gerar seu próprio conteúdo, a *Web* atual (*Web 2.0*), também conhecida como *Web Sintática* por Breitman (2005), passou a concentrar uma enorme quantidade de informações tornando necessário o desenvolvimento de ferramentas que retornem resultados precisos em meio a informações “irrelevantes”.

Motores de busca como o *Google*, *Yahoo*, entre outros, apresentam limitações quanto à eficiência e eficácia da recuperação da informação, pois até então a busca é baseada em palavras-chave contidas nos textos (ALVARENGA, 2004).

Outro fato é que as tecnologias e linguagens utilizadas para desenvolver os *sites* priorizam os aspectos de exibição e formatação do conteúdo, sendo o conteúdo pobremente descrito impossibilitando que as máquinas conheçam o contexto e o sentido em que a informação foi descrita (SOUZA, 2004).

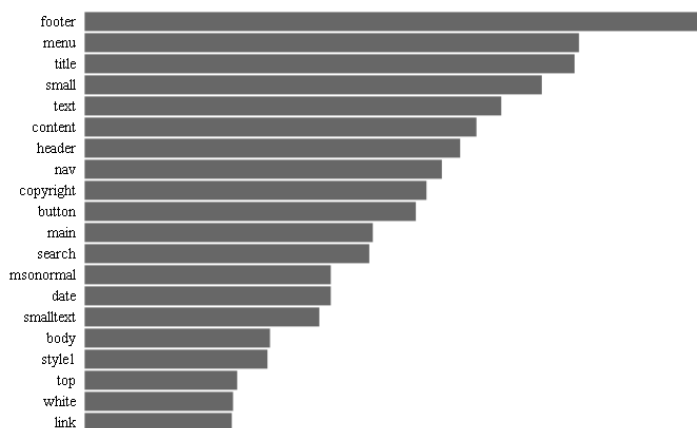
Dessa forma, ao se realizar uma busca na atual *Web* (*Web Sintática*), inúmeros sites serão indexados, porém caberá ao ser humano a tarefa de avaliar, classificar e selecionar as informações que realmente são de interesse (BREITMAN, 2005).

Contrapondo as ideias frustrantes da *Web Sintática*, Tim Berners-Lee propôs em 2001, a chamada *Web Semântica*, também chamada de *Web 3.0* que visa melhorar a satisfação do usuário no momento da busca, retornando-lhe as informações adequadas às suas necessidades (BERNERS-LEE, 2001).

Para atingir os propósitos da *Web Semântica*, é necessária a existência de uma infra-estrutura tecnológica comum da *Internet*, no qual o primeiro passo é a criação de padrões para a descrição de dados e de uma linguagem que permita a construção e codificação de significados compartilhados (SANTOS NETO, 2009).

## 2.2 O HTML5 e suas contribuições para a Web Semântica

Para atingir um padrão que atenda essas exigências, em 2004 a *Google* vasculhou mais de um bilhão de *sites*, para verificar como a *Web* estava estruturada. Como resultados desse estudo foram coletados as vinte classes do HTML mais utilizadas, as quais estão listadas na Figura 1 (SOUZA, 2010).



**Figura 1: Tags HTML mais utilizadas**

A Opera *Software* realizou uma nova pesquisa em 2009, vasculhando cerca de dois bilhões de páginas e não analisando somente as classes utilizadas, mas também os *id's* mais usados. Os resultados dessa pesquisa são apresentados nas Tabelas 1 e 2.

**Tabela 1: 20 classes mais utilizadas**

Popularidade	Valor	Frequência
1	Footer	179,528
2	Menu	146,673
3	style1	138,308
4	Msonormal	123,374
5	Text	122,911
6	Content	113,951
7	Title	91,957
8	style2	89,851
9	Header	89,274
10	Copyright	86,979
11	Button	81,503
12	Main	69,620
13	style3	69,349
14	Small	68,995
15	Nav	68,634
16	Clear	68,571
17	Search	59,802
18	style4	56,032
19	Logo	48,831
20	Body	48,052

**Tabela 2: 20 id's mais utilizadas**

Popularidade	Valor	Frequência
1	Footer	288,061
2	Content	228,661
3	Header	223,726
4	Logo	121,352
5	Container	119,877
6	Main	106,327
7	table1	101,677
8	Menu	96,161
9	layer1	93,920
10	autonumber1	77,350
11	Search	74,887
12	Nav	72,057
13	Wrapper	66,730
14	Top	66,615
15	table2	57,934
16	layer2	56,823
17	Slideber	52,416
18	image1	48,922
19	Banner	44,592
20	Navigation	43,664

Um fato relevante é que os *sites* até então eram desenvolvidos na versão 4.1 da linguagem de Berners-Lee, ou seja, a estrutura padrão das páginas *Web* era basicamente como demonstrada na Figura 2.



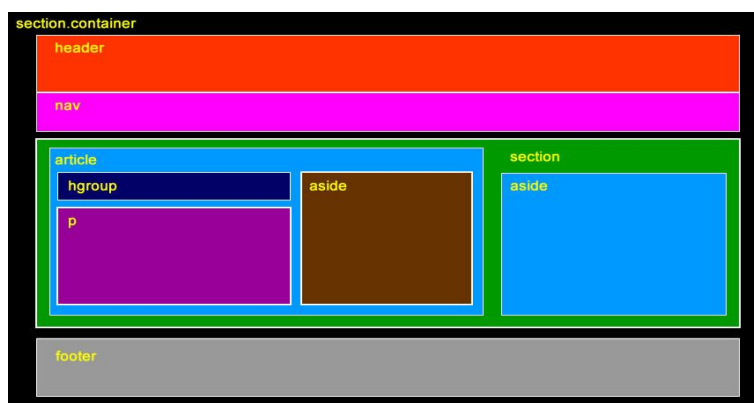
**Figura 2: Estrutura básica de uma página Web em HTML 4.1**

Observa-se que no HTML 4.1, o elemento *div* é bastante utilizado. A desvantagem é que sua função consiste apenas em estruturar o conteúdo, não apresentando significado algum. Assim os motores de busca não conseguem distinguir se determinada área do site é um cabeçalho, menu, artigo, rodapé *etc.*

Em contrapartida, o HTML 5 propõe novos elementos baseados nas pesquisas já mencionadas para estruturar as informações com foco na semântica. A inclusão desses elementos contribui para a otimização dos resultados dos motores de busca.

Essas novas *tags* oferecem uma definição mais relevante à informação, tornando viável sua publicação devidamente segmentada. A partir desses elementos será possível o desenvolvimento de aplicações capazes de manipular a informação com mais eficiência, sem a necessidade de classes e *id's* subjetivas (ANDRADE, 2009).

Como pode ser observada na Figura 3, cada área do site ganha um significado próprio. A descrição de alguns elementos está presente na Tabela 3.



**Figura 3: Estrutura básica de uma página HTML5.**

Padronização, codificação e compartilhamento são algumas das sentenças que norteiam o desenvolvimento do HTML 5, contribuindo para tornar a *Web Semântica* uma realidade, e por sua vez, tornando o processo de recuperação da informação menos insatisfatório (MARCHI, 2009).

Isso será possível porque os motores de busca, conhecendo cada parte integrante do *site*, poderão fazer inferências sob as informações de acordo com o sentido e o contexto, retornando assim, resultados mais relevantes ao usuário.

De acordo com os estudos desenvolvidos pela *Google* e *Opera*, foram propostas novas *tags* para o HTML 5, as quais algumas estão descritas na Tabela 3.

**Tabela 3: tags HTML5 e suas funcionalidades**

<article>	Define um artigo. Um artigo deve ser algo independente, por exemplo, um <i>post</i> de um <i>blog</i> , de um fórum, comentários.
<aside>	Define um conteúdo secundário de uma página, por exemplo, um menu secundário ou um banner de anunciantes.
<audio>	Define um conteúdo de áudio. Usando <code>src = "nome do arquivo"</code> pode ser definido a fonte do áudio.
<footer>	Define um rodapé de uma página.
<header>	Define um cabeçalho de uma página.
<nav>	Define <i>links</i> de navegação, por exemplo, um menu.
<section>	Define uma seção do conteúdo. Pode ser usado para organizar capítulos, cabeçalhos, rodapé, <i>etc.</i>

Vale ressaltar que o HTML5 ainda não está concluído e que o entusiasmo com essa quinta versão deve ser acompanhado com a evolução de outras tecnologias que conjuntamente poderão edificar um novo conceito de *Web* para a todos os usuários, pois sozinha, a linguagem de Tim Berners-Lee é inerte.

### 3. Metodologia

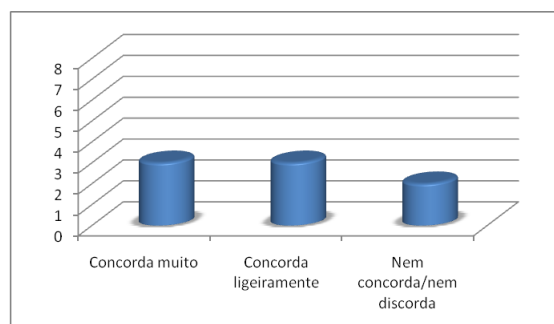
Esta pesquisa é do tipo descritivo-exploratória. Preliminarmente, realizou-se uma pesquisa bibliográfica com a finalidade de obter subsídios teóricos necessários para desenvolver uma contextualização, argumentações e observações, assegurando a qualidade das informações.

A pesquisa foi feita com oito programadores das cidades de Patos-PB, Uiraúna-PB e São Paulo-SP, através de um questionário que, de acordo com Marconi e Lakatos (2007), é um instrumento de coleta de dados constituído por uma série ordenada de perguntas, classificadas em perguntas fechadas, abertas, múltiplas escolhas, entre outras.

Optou-se pela escala de Lickert por ser necessário especificar o nível de concordância sobre variáveis de notória relevância. Essa escala permite elaborar um conjunto de proposições que tenham relação direta com o objetivo do trabalho. Essas proposições foram apresentadas aos programadores que indicaram suas reações.

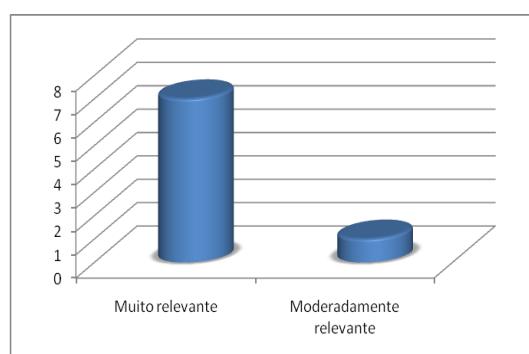
#### 4. Estudo de caso

De acordo com os resultados, seis programadores concordam que o HTML5 contribui para que os motores de busca identifiquem a informação com maior relevância, como mostra a Figura 4.



**Figura 4: HTML5 na identificação da informação pelos motores de busca**

No que diz respeito às ambições da *Web 3.0*, todos os programadores afirmam que as contribuições do HTML5 são relevantes, como mostra a Figura 5.



**Figura 5: Contribuições do HTML5 para a Web Semântica**

Diante do estudo bibliográfico realizado, verificou-se que, mesmo ainda em fase de desenvolvimento, a nova versão da linguagem HTML contribui para tornar possíveis os ideais da *Web 3.0*, devido ao acréscimo de novas *tags*.

Fato se torna notório quando 87,5% dos programadores entrevistados consideram que o HTML5, através de suas novas *tags*, são muito relevante para fazer da *Web 3.0* uma realidade cada vez mais presente no mundo das relações virtuais.

Assim, a forma do usuário se relacionar com a informação disponível na rede mundial de computadores ganha um novo e importante membro: o computador (ou outro dispositivo), pois eliminará da responsabilidade dos usuários a árdua tarefa de filtrar os resultados de uma pesquisa.

## 5. Considerações finais

Os programadores entrevistados concordam que a automatização da capacidade de interpretar os dados espalhados pela *Internet* está sendo acelerada também com o desenvolvimento do HTML5. Ou seja, marcações com modelagens mais adequadas de identificadores permitirão que computadores façam isso.

Assim apoiado na revisão bibliográfica e debruçado nos resultados da pesquisa, verifica-se que o HTML5 se torna, mesmo em desenvolvimento, um apoio fundamental às ambições da *Web 3.0*, pois lida diretamente com a estruturação dos sites.

## Referências

- ANDRADE, Diogo Araújo de; MARCHI, Késsia Rita da Costa. HTML5 - Novos padrões para a *Web Semântica*. Paranaíba: Universidade Paranaense, 2009.
- BERNERS-LEE, Tim; HENDLER, James; LASSILA, Ora. The Semantic *Web*: A new form of *Web* content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*, New York, p.1-12, 19 maio 2001.
- BREITMAN, Karin. *Web Semântica: a Internet do futuro*. Rio de Janeiro: LTC, 2005
- FERREIRA, Elcio; EIS, Diego. Curso W3C Escritório Brasil. São Paulo. 2010. 106 p.
- MARCONI, Marina de Andrade; LAKATOS, Eva Maria. *Técnicas de Pesquisa*. 6. ed. São Paulo: Atlas, 2007. 279 p.
- ROBERTO, Rafael Liberato. Uma abordagem para identificação de interesses dos usuários durante a navegação em *Websites* semânticos. 2006. 113 f. Dissertação (Mestrado) - Universidade Estadual de Maringá, Maringá, 2006.
- SANTOS NETO, Gilberto Martins dos. Anotação semântica de recursos *Web* baseada em ontologias. 2009. 83 f. Dissertação (Mestrado) - Universidade Federal do Amazonas, Manaus, 2009.
- SOUZA, David William Rosa de; CARDOSO, Rafael Cunha. O HTML5 e suas novas API. Pelotas: Tecnologias de Sistemas para a *Internet* Instituto Federal de Educação, Ciência e Tecnologia Sul-Rio-Grandense. 2010.
- SOUZA, Renato Rocha; ALVARENGA, Lídia. A *Web Semântica* e suas contribuições para a ciência da informação. *Ciência da Informação*, Brasília, p.132-141, 03 jan. 2004.
- TREIN, Daiana; SCHLEMMER, Eliane D. R.. Projetos de aprendizagem baseados em problema no contexto da *Web 2.0*: possibilidades para a prática pedagógica. *Revista E-curriculum*, São Paulo, 02 jun. 2009.