



Rafael Antônio Pinto Pena

**Suporte semântico à publicação de
conteúdo jornalístico na Web**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática do Departamento de Informática do Centro Técnico e Científico da PUC-Rio.

Orientador: Prof. Daniel Schwabe

Rio de Janeiro
Fevereiro de 2012



Rafael Antônio Pinto Pena

**Suporte semântico à publicação de
conteúdo jornalístico na Web**

Dissertação apresentada como requisito parcial para a obtenção do grau de Mestre pelo Programa de Pós-graduação em Informática do Departamento de Informática do Centro Técnico e Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Daniel Schwabe

Orientador e Presidente

Departamento de informática - PUC-Rio

Prof. Edward Hermann Haeusler

Departamento de informática - PUC-Rio

Prof. Giancarlo Guizzardi

Departamento de informática - UFES

Prof. José Eugenio Leal

Coordenador Setorial do Centro

Técnico Científico - PUC-Rio

Rio de Janeiro, 07 de Fevereiro de 2012

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Rafael Antônio Pinto Pena

Graduou-se em Tecnólogo em Processamento e Dados pela Pontifícia Universidade Católica do Rio de Janeiro (PUC-RJ) em Dezembro de 1999. Tem experiência na área de Ciência da Computação, com ênfase em Desenvolvimento de Software. Tem trabalhado em análise de sistemas desde 1997. Mais especificamente com Desenvolvimento de Sistemas para Web, e recentemente, Desenvolvimento de Produtos Jornalísticos para Web.

Ficha Catalográfica

Pena, Rafael Antônio Pinto

Suporte semântico à publicação de conteúdo jornalístico na Web / Rafael Antônio Pinto Pena; Orientador: Daniel Schwabe. - Rio de Janeiro: PUC-Rio, Departamento de Informática, 2012.

v., 105 f.: il. ; 29,7 cm

Dissertação de Mestrado - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática

Referencias bibliográficas incluídas.

Web Semântica; Modelo de Domínio; Conteúdo Jornalístico; Conteúdo Web; Empresas Jornalísticas; Mídia Digital

Agradecimentos

A Deus pela maravilhosa forma de vida que me foi concedida.

A minha esposa Carolyn Maciel pelo apoio, amor, compreensão e por estar ao meu lado em todos os momentos me tranquilizando e fazendo acreditar que seria possível.

Aos meus pais, Fernando Pena e Nina Pena, pela educação, carinho, amor e por me mostrarem a importância do aprendizado e do trabalho.

As minhas irmãs Catarina Pena e Fabiana Pena Ruiz pelo amor, pelos exemplos e ensinamentos.

A minha avó Griselda Schleder, por lutar contra a doença para permanecer viva até ver seus netos formados. Uma inesquecível demonstração de amor e de valorização do estudo.

Ao meu orientador Professor Daniel Schwabe, pelo enorme apoio, confiança e orientação.

Ao Professor Giancarlo Guizzardi pela grande ajuda na modelagem conceitual.

Ao amigo Janilton Maciel pela enorme ajuda no desenvolvimento do protótipo.

Aos amigos da Globo.com, Davidson Fellipe, Franklin Amorim, Claudio Braga, Paulo Ricardo Muniz, Camila Dias, Newton Fleury e Alexandre Moreno por sempre me apoiarem e colaborarem com ideias e melhorias.

Sou grato aos jornalistas que participaram da pesquisa e tanto ajudaram neste trabalho. André Amaral, Gustavo Poli, Bernardo Ferreira, Bernardo Pombo, Eduardo Orgler e Rodrigo Breves.

À Globo.com, pelo financiamento e auxílios concedidos, sem os quais este trabalho não poderia ter sido realizado. Em especial ao Juarez Queiroz por conseguir este benefício para os funcionários e ao Jacques Douglas Varaschim por todo apoio.

Aos professores da Comissão examinadora.

Resumo

Pena, Rafael Antônio Pinto; Schwabe, Daniel. **Suporte semântico à publicação de conteúdo jornalístico na Web**. Rio de Janeiro, 2012. 105p. Dissertação de Mestrado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

O consumo de conteúdo jornalístico na Web aumenta a cada dia. No entanto, boa parte desse conteúdo ainda é produzido segundo paradigmas da mídia impressa. Paralelamente a isto, a Web Semântica ou Web 3.0, adiciona uma camada de "inteligência" à Web, onde computadores são capazes de extrair significados dos conteúdos acessados na Web, e conseqüentemente, processá-los. Neste trabalho foi definido e testado um modelo de publicação de conteúdo jornalístico apoiado pela Web Semântica. Foi desenvolvida uma ferramenta de apoio para produtores de conteúdo, com o objetivo de melhorar o processo de construção de narrativas jornalísticas para web, através da sugestão de padrões narrativos e de informações objetivas que dão suporte ao tipo de narrativa desejado. Um estudo de caso foi realizado para avaliar em ambiente real o modelo proposto. Um grupo de jornalistas usou a ferramenta desenvolvida, com aceitação unânime.

Palavras-chave

Web Semântica; Modelo de Domínio; Conteúdo Jornalístico; Conteúdo Web; Empresas Jornalísticas; Mídia Digital.

Abstract

Pena, Rafael Antônio Pinto; Schwabe, Daniel (Advisor) . **Semantic Support for the publication of news content on the Web**. Rio de Janeiro, 2012. 105p. MSc Dissertation - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The use of journalistic content on the web increases every day. However, much of the content is still produced according to the paradigms of print media. Parallel to this, the Semantic Web or Web 3.0, adds a layer of "intelligence" to the Web, where computers are able to extract meaning from the visited web content, and consequently, process them. This work defined and tested a model of publishing news content supported by the Semantic Web. It was developed a support tool for content producers, in order to improve the process of construction of news stories for the web, using the suggestion of narrative patterns and objective information that support the desired type of narrative. A case study was conducted to evaluate the proposed model in a real environment. A group of journalists used the developed tool, with unanimous acceptance.

Keywords

Semantic Web; Domain Model; Journalistic Content; Web Content; Media Companies; Digital Media.

Sumário

1	Introdução	12
2	Fundamentos e estado da arte	14
2.1	Fundamentos	14
2.1.1	Web Semântica e Linked Open Data	14
2.1.2	Sistema de frames e slots	20
2.1.3	UFO (<i>Unified Foundational Ontology</i>)	22
2.2	Estado da arte	32
2.2.1	Utilização de dados semânticos em empresas de mídia	32
2.2.2	Modelos de estruturas narrativas	37
3	Modelo de publicação de conteúdo jornalístico na web	39
3.1	Descrição do contexto de uso	39
3.2	Descrição do fluxo desejado	41
4	Modelos de ontologias	45
4.1	Descrição da metodologia de modelagem usada	45
4.1.1	Construção de um modelo de jogo de futebol	46
4.1.2	Definição dos estereótipos e esquemas narrativos	48
4.2	Ontologia do Futebol	51
4.2.1	Fragmento das participações em um jogo	52
4.2.2	Fragmento das etapas de disputa de um jogo	54
4.2.3	Fragmento do lance de disputa comum	55
4.2.4	Fragmento do lance de jogo individual	58
4.2.5	Fragmento dos Lances Complexos	61
4.2.5.1	Lances faltosos	61
4.2.5.2	Lances de jogo	63
4.2.6	Resumo do modelo	66
4.3	Modelo de estereótipos e narrativas	68
4.3.1	Regras de detecção de estereótipos	68

4.3.2	Formação dos esquemas narrativos	75
5	Estudo de caso	80
5.1	Protótipo funcional	80
5.1.1	Interface de autoria de matéria	80
5.1.2	Exemplos de templates de estereótipos	85
5.2	A experiência de uso	91
5.2.1	Indicadores sobre a utilização do protótipo	92
5.2.2	Pesquisa de opinião sobre o uso do protótipo	96
6	Conclusões e próximos passos	98
6.1	Conclusões	98
6.2	Próximos passos	100
6.2.1	Oportunidades relacionadas ao modelo de publicação	100
6.2.2	Oportunidades relacionadas aos esquemas narrativos	101
7	Referências bibliográficas	103

Lista de imagens

Figura 1 – Quadrado de Aristóteles. Figura extraída de [Guizzardi et al, 2009]	23
Figura 2 - Fragmento da UFO-B. Objetos e Eventos. Figura extraída de [Guizzardi et al, 2008]	26
Figura 3 - Fragmento da UFO-B: Relações de Allen. Figura extraída de [Guizzardi et al, 2008]	27
Figura 4 - Fragmento da Ontologia de Processo de Software de ODE. Extraída de [Guizzardi et al, 2008]	29
Figura 5 - Fragmento Remodelado da Ontologia de Processo de Software (Substanciais). Extraída de [Guizzardi et al, 2008]	29
Figura 6 - Fragmento Remodelado da Ontologia de Processo de Software (Eventos). Extraída de [Guizzardi et al, 2008]	31
Figura 7 - Extrato da ontologia de fundamentação. Extraída de [Almeida et al, 2008]	32
Figura 9 - Esquema simplificado do processo de publicação de conteúdo web existente	39
Figura 10 - Esquema simplificado do processo de publicação de conteúdo web proposto no trabalho	43
Figura 11 - Fragmento das participações em um jogo	52
Figura 12 - Fragmento das participações em um jogo alinhado com a UFO	53
Figura 13 - Fragmento das etapas de disputa de um jogo	54
Figura 14 - Fragmento das etapas de disputa de um jogo alinhado com a UFO	55
Figura 15 - Fragmento do lance de disputa comum	56
Figura 16 - Fragmento do lance de disputa comum alinhado com a UFO	57

Figura 17 - Fragmento do lance de jogo individual	58
Figura 18 - Fragmento do lance de jogo individual alinhado com a UFO	61
Figura 19 – Fragmento do lances faltosos	62
Figura 20 – Fragmento dos lances faltosos alinhado com a UFO	63
Figura 21 – Fragmento do lances válidos	64
Figura 22 – Fragmento dos lances válidos alinhado com a UFO	66
Figura 23 - Matriz de estereótipos	68
Figura 24 - Interface de autoria de matérias	82
Figura 25 - Interface para seleção de jogo para crônica	83
Figura 26 - Interface de sugestão de estereótipos	84
Figura 27 - Interface de autoria de matéria com template de estereótipo selecionado	85
Figura 28 - Template narrativo para estereótipo de jogo movimentado	86
Figura 29 - Template narrativo para estereótipo de jogo morno	86
Figura 30 - Template narrativo para estereótipo de goleada	87
Figura 31 - Template narrativo para estereótipo de placar duvidoso	88
Figura 32 - Trecho de matéria feita com template Goleada	88
Figura 33 - Trecho de matéria feita com template Placar Duvidoso	89
Figura 34 - Trecho de matéria feita com template Artilheiro	90
Figura 35 - Gráfico comparativo entre quantidade de matérias produzidas e quantidade de matérias produzidas com templates narrativos do protótipo	92
Figura 36 - Gráfico comparativo entre quantidade de jogos e a quantidade de jogos para os quais pelo menos uma matéria foi produzida com template narrativo do protótipo	94
Figura 37 – Gráfico com as quantidades de uso dos templates narrativos do protótipo	95

Lista de Tabelas

Tabela 1 - Estereótipo Jogo Movimentado	69
Tabela 2 - Estereótipo Jogo Morno	70
Tabela 3 - Estereótipo Jogo Truncado	71
Tabela 4 - Estereótipo Goleada	71
Tabela 5 - Estereótipo Jogo Recorde no campeonato	72
Tabela 6 - Estereótipo Placar Duvidoso	73
Tabela 7 - Estereótipo Melhor Performande do Jogo	75
Tabela 8 - Template narrativo Jogo Movimentado / Jogo Morno	75
Tabela 9 - Template narrativo Goleada	76
Tabela 10 - Template narrativo Jogo Truncado	77
Tabela 11 - Template narrativo Placar Duvidoso	77
Tabela 12 - Template narrativo Primeiro no ranking do campeonato	78
Tabela 13 - Template narrativo Atleta em Destaque	79
Tabela 14 - Resultado da pesquisa de opinião feitas com os jornalistas envolvidos	97