

10. Web semântica e web pragmática: representação e recuperação em acervos digitais

Denysson Axel Ribeiro Mota, Nair Yumiko Kobashi

Estudo comparativo das propostas da *Web Semântica* e *Web Pragmática* com o objetivo propor um modelo que contribua para promover a qualidade da recuperação da informação nesses ambientes. Para isso, abordam-se as origens da *Web Semântica* e *Web Pragmática* e seus conceitos fundamentais, tais como linguagem, representação e recuperação da informação e do conhecimento; termos e terminologia; semântica e pragmática. Dirige-se o foco da abordagem aos problemas da representação e recuperação da informação em acervos bibliográficos digitais, com o uso de Ontologias e *Topic Maps*. Pretende-se, ao final do estudo, apresentar uma modelagem de representação de informação que promovam melhor recuperação da informação nos ambientes da Web.

Acervos Digitais. Pragmática. Recuperação da Informação.
Semântica. Web Pragmática. Web Semântica.

Introdução

O principal objetivo deste trabalho é propor uma abordagem teórico-metodológica de representação de contexto em acervos informacionais digitais, no ambiente da Web

Pragmática, de modo a promover o aprimoramento da recuperação de informação.

Nos estudos de representação e recuperação da informação tem importância fundamental a compreensão da linguagem. Nesses estudos pode-se destacar a perspectiva de Morris (1985), para quem, há três elementos essenciais que definem os significados das palavras: a sintaxe, a semântica e a pragmática.

A Sintaxe estuda a relação dos elementos sígnicos que compõem as palavras individualmente, assim como as estruturas combinatórias entre elas para formar frases e sentenças, desconsiderando-se, nessas análises, a relação entre conceito e sujeito interpretante (MORRIS, 1985).

A Semântica é a disciplina que estuda o significado das palavras e sentenças e a relação dessas palavras, ou signos, com os objetos que eles representam ou significam (MORRIS, 1985). Esses estudos se baseiam nos trabalhos semióticos de Peirce (1977), principalmente na tríade representâmen-objeto-interpretante.

O campo que estuda o uso da língua em contextos é a Pragmática. Esta abordagem leva em conta o ambiente físico, a entonação, as expressões faciais, o histórico e experiências do emissor e receptor, além de outros elementos extralinguísticos utilizados no processo de significação, o que permite analisar não apenas os vocábulos utilizados, mas também a intencionalidade do emissor e as relações entre signos e usuários (MORRIS, 1985).

Na Web tradicional, denominada de Sintática, idealizada por Tim Berners-Lee (BERNERS-LEE; CAILLIAU, 1990), as

páginas são estáticas e o que se tem são documentos interligados por hyperlinks, não havendo nenhum outro dado ou informação extra que identifique ou descreva esses documentos ou seus componentes. Esta Web recorre apenas à sintaxe, como definido por Morris (1985), para apresentação dos dados na web. Os elementos visuais apresentados são uma sequência de símbolos e caracteres, interpretados pelo leitor.

O objetivo da Web Semântica é explicitar as relações entre significante e significado nas páginas. Essas relações, extraídas automaticamente, apresentam documentos formalmente interligados por relações, representados em metadados específicos. Trata-se de uma semântica leve, baseada em enunciados lógicos, em forma de triplas, que permite fazer inferências simples. Com isso, o computador poderá fazer inferências para recuperação mais precisa de informação.

A Web Pragmática surge da percepção de que algo faltaria à proposta da Web Semântica. Diversos autores (SINGH, 2002; DE MOOR; KEELER; RICHMOND, 2002) expressam a perceptível ausência de elementos que complementem o significado dos elementos presentes nos conteúdos na Web, questionando se apenas adicionar elementos semânticos a esses conteúdos será suficiente para atingir o objetivo de criar uma Web significativa. Para superar as insuficiências, as pesquisas sobre a Web Pragmática procuram incluir contextos às representações dos objetos informacionais.

Objetivos

O principal objetivo deste trabalho é propor uma abordagem teórico-metodológica para representação de contexto em acervos informacionais digitais, promovendo a aplicação do conceito de Web Pragmática a este ambiente.

Para alcançar esse objetivo macro, enumeram-se os seguintes objetivos específicos:

- Estudar criticamente os conceitos gerais que guiam a representação e recuperação da informação e do conhecimento;
- Levantar os conceitos fundamentais da Web Semântica e Web Pragmática e discuti-los sob a ótica da Ciência da Informação;
- Definir os possíveis contextos que podem ser representados em acervos informacionais digitais e bases de dados de periódicos científicos;
- Propor uma modelagem que possibilite a adequada representação de contexto em acervos, para uso em Sistemas de Recuperação de Informações.

Justificativa

A criação de significado envolve elementos sintáticos, semânticos e pragmáticos. Apesar dos avanços alcançados nas pesquisas sobre a recuperação de informações, as questões pragmáticas ainda não estão incorporadas aos ambientes informacionais. Os acervos informacionais digitais, nas bases de recursos informacionais de texto completo ou de representações

desses recursos (dados bibliográficos, resumos e palavras-chave), se enquadram nas características da Web Semântica. É possível potencializar a capacidade de recuperação de informação, nesses espaços, com a adoção de critérios pragmáticos de representação de informação, A definição de perfis de usuários de acervos de especialidade, assim como a publicação de conteúdos, realizada em um domínio científico delimitado e supervisionado, possibilita criar e avaliar um modelo de SRI aderente aos princípios da Web Pragmática.

Procedimentos metodológicos

Considerando os objetivos apresentados acima, enumeram-se os procedimentos metodológicos traçados para alcançar cada um deles:

- Estudar os conceitos gerais que guiam a representação e recuperação da informação e do conhecimento presentes na bibliografia especializada, contrapondo ideias de modo a esclarecer questões aparentemente conflitantes.
- Sistematizar os conceitos fundamentais da Web Semântica e Web Pragmática, e discuti-los sob a ótica da Ciência da Informação. Nesta etapa serão levantados os elementos que fundamentam e compõem as propostas da Web Semântica e Web Pragmática, elencando os conceitos utilizados para desenvolver as tecnologias web. Serão abordados dois formatos de representação informacional para a web: RDF e Topic Maps, com a comparação de pontos fortes e fracos de cada um deles. Também serão discutidos os conceitos de Web

Pragmática, Pragmática Virtual, Web 2.0 e Web 3.0, verificando semelhanças e diferenças nas abordagens de cada um deles.

- Definir os possíveis contextos que podem ser representados para os acervos informacionais digitais e bases de periódicos científicos. Este objetivo será alcançado mediante levantamento na literatura dos possíveis contextos que interferem na interpretação da informação. Serão assim avaliados os possíveis contextos de interpretação e suas possíveis representações. Já existem alguns elementos contextuais passíveis de representação com as atuais tecnologias, como a ontologia de citações CiTO ¹, mas é possível, e necessário, incluir outros contextos.
- Estudar uma modelagem que possibilite a adequada representação de contexto em acervos, para posterior uso na recuperação das informações. Percebe-se a necessidade de criação de um modelo para representar adequadamente os diferentes contextos, sob o escopo da Web Pragmática, utilizando as tecnologias existentes. O ponto principal é estudar um modelo informacional que permita a interação entre as ontologias já existentes e os contextos previamente representados, com visão na possibilidade de identificação das informações que se adequarão ao perfil de usuários de domínios específicos de conhecimentos.

¹ <http://www.essepuntato.it/lode/http://purl.org/spar/cito>

- Testar e avaliar a modelagem proposta em um acervo concreto. Para eliminar a possível limitação de acesso aos artigos, e considerando o alto número de páginas disponíveis na *web*, será utilizada a Base de Dados de Periódicos em Ciência da Informação - BRAPCI¹. Por ser um repositório que contém metadados de quase todos os artigos publicados em periódicos de Ciência da Informação no Brasil, esta base mostra-se como ideal para realizar os testes para verificação das potencialidades da modelagem, em um ambiente real.

Fundamentação

Como exposto anteriormente, sob os estudos linguístico-semióticos de Morris (1985) há três elementos que influenciam o significado das palavras: a sintaxe, a semântica e a pragmática. Os estudos de Morris, e de outros pesquisadores da área, se baseiam nos trabalhos semióticos de Peirce (1977), principalmente na tríade representâmen-objeto-interpretante. Para Morris (1985) a sintaxe é apenas a verificação da composição sígnica das palavras, a junção de letras que indicam o significado, que permite diferenciar “pula” de “mula”, por exemplo.

A semântica adiciona à composição da palavra sua posição em uma frase, no aspecto lógico-linguístico e sua relação com seu denotatum, no aspecto semiótico-filosófico (TAMBA-MECZ, 2006; MORRIS, 1985). Para Morris (1985, p. 55, tradução nossa), a semântica é o estudo da “relação dos signos

com seus designata e, com isso, com os objetos que podem denotar ou que, de fato, denotam”. Esta visão é puramente semântica e relaciona cada palavra ou termo com a imagem mental do objeto ideal que este designa, independentemente de outros elementos que podem influenciar a designação.

O campo que estuda o uso da língua dentro de contextos é a pragmática. Na pragmática, elementos extralinguísticos são utilizados para complementar o significado das palavras, de modo a explicitar a intencionalidade do emissor e do receptor, dando maior importância à relação dos signos com seus usuários (MORRIS, 1985).

Para Armengaud (2006), Morris é o primeiro autor a usar especificamente o termo pragmática, embora os estudos deste campo já ocorressem desde a década de 40, com Bar-Hillel, com especial destaque para o ano de 1950, quando foi publicado o artigo de Strawson *On Referring*, que abordava o uso de expressões dentro de contextos. Armengaud (2006) afirma que não há um único fundador da Pragmática, mas há vários, com diferentes influências: Peirce e Morris, como fundadores diretos; Frege e Wittgenstein, como fundadores indiretos; Carnap e Bar-Hillel, como fundadores intermediários.

Para de Moor, Keeler e Richmond (2002), a Web Pragmática é uma Web em que processos pragmáticos essenciais são definidos e automatizados. A pragmática é considerada como o estudo da relação do signo com seus intérpretes, intérpretes estes inseridos em um contexto, que pode ser tanto situacional, como individual ou social, segundo Morris (1985) e Peirce (1977), amplamente reconhecidos como os precursores conceituais da Web Pragmática (DE MOOR, KEELER e

¹ <http://www.brapci.inf.br/>

RICHMOND, 2002). Dessa forma, é possível ver a Web Pragmática como a possível representação de contexto para a Web, que poderá melhorar os processos de recuperação da informação.

Resultados esperados

A necessidade de se considerar o contexto na recuperação da informação não propriamente uma novidade. Os estudos de relevância na recuperação, por meio da avaliação da relevância (ou não) dos resultados, para os usuários já levavam em consideração o contexto. Ingwersen e Järvelin (2005) ressaltam que a recuperação da informação é um processo de busca que acontece dentro de um contexto determinado por diversos elementos além da tarefa do indivíduo: seu contexto social, organizacional e cultural, assim como contextos sistêmicos. Para esses autores, toda informação é construída com base em dois elementos contextuais: o modelo de mundo do agente (seu contexto social e cognitivo) e a mensagem consumida em um contexto. Porém a mensagem também foi construída dentro de um contexto, por um autor, com um modelo de mundo próprio. Quanto menos informação o agente tem sobre o contexto de produção e consumo da mensagem, maiores são as possibilidades de interpretação. Isso ocorre de forma semelhante quando considerada a interpretação não apenas de uma fonte informacional, mas também de diferentes informações cujas fontes contextuais não estão acessíveis ao agente.

A operacionalização de conceitos pragmáticos, para a adequada representação de contexto no ambiente web deverá

trazer melhorias aos resultados dos processos de recuperação da informação, porque permitirá avaliar a relevância de forma contextual. Independente da tecnologia utilizada – RDF, RDF com reificação ou Topic Maps em XML, conhecido como XTM – a representação da informação nesse ambiente, e disponibilização para toda a comunidade acadêmica e profissional, trará benefícios porque, ao integrar e relacionar informações, de forma contextualizada, poderá reduzir a ambiguidade e polissemia no processo de representação e recuperação de informações.

Considerações finais

Nos ambientes informacionais digitais, principalmente no escopo da Web Semântica e/ou Web Pragmática, a representação das informações deve estar adequadamente traduzida para os padrões de disponibilização de dados definidos para esses ambientes. No caso das representações contextuais, essa representação pode ocorrer de diferentes formas: RDF simples, RDF reificado e XTM.

Dentre as possíveis representações contextuais, este trabalho visa explorar as seguintes opções: Citações, Domínio-base do Documento, Área de Formação do Indivíduo, Domínio da Palavra-Chave, Área de Publicação da Revista, Áreas de Publicação do Indivíduo e Interesses do Indivíduo.

Estes são apenas alguns exemplos de contextos passíveis de representação, e das tecnologias que podem ser usadas para modelar e representar contextos no processo de representação da informação de um documento em um ambiente informacional

digital. Outros contextos de interpretação e outras formas de representação podem ser avaliadas para obter a forma mais adequada e simples de contextualizar informações atualmente armazenadas em repositórios digitais, que poderão ser disponibilizadas abertamente para a Web, no escopo da Web Semântica, transformando esta última em uma Web Pragmática.

Principais referências

ARMENGAUD, F. **A Pragmática**. São Paulo: Parábola Editorial, 2006.

BERNERS-LEE, T.; CAILLIAU, R. **WorldWideWeb**: Proposal for a HyperText Project. 1990. Disponível em: <
<http://www.w3.org/Proposal.html> >. Acesso em: 13 out. 2014.

DE MOOR, A.; KEELER, M.; RICHMOND, G. Towards a pragmatic web, In: UTA, Priss et al. Conceptual Structures: Integration and Interfaces. **Lecture Notes in Computer Science**, v. 2393, p. 235-249, 2002. Disponível em <
<http://www.cspeirce.com/menu/library/aboutcsp/richmond/web.pdf> >. Acesso em 23 mar. 2014.

INGWERSEN, P.; JÄRVELIN, K. **The turn**: integration of information seeking and retrieval in context. Dordrecht, The Netherlands: Springer, 2005.

LEVINSON, S. **Pragmática**. São Paulo: Martins Fontes, 2007.

LYONS, J. **Semântica** - I. Lisboa, Ed. Presença/Martins Fontes, 1980.

MOESCHLER, J.; REBOUL, A. **Diccionario enciclopédico de Pragmática**. Madrid, Ed. Arrecife, 1999.

MORRIS, C. **Fundamentos de la teoría de los signos**. Barcelona: Paidós, 1985.

PEIRCE, C. S. **Semiótica**. São Paulo: Perspectiva, 1977.

SINGH, M. P. The Pragmatic Web. **IEEE Internet Computing**, v. 6, n. 3, May/June, p. 4-5, 2002a.

TAMBA-MECZ, I. **A Semântica**. São Paulo: Parábola Editorial, 2006.