

Promovendo a Inclusão e o Acesso ao Conhecimento: Um Mecanismo de Busca Informado por Abordagem Semiótica em Redes Sociais

Julio Cesar dos Reis^{1,2}, M. Cecília C. Baranauskas¹ (orientadora) e Rodrigo Bonacin² (co-orientador)

¹Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)

²Centro de Tecnologia da Informação Renato Archer (CTI)

{julio.reis, rodrigo.bonacin}@cti.gov.br, cecilia@ic.unicamp.br

***Abstract.** Due to the context of huge economical differences and cultural diversity in which we live, inclusive-oriented computational solutions and the access for all to the knowledge are essential. Search mechanisms in Social Networks Systems (SNS) could be an instrument for digital inclusion. However, the current search solutions have not yet been taking into account the language, and the semantic understanding shared by people in social networks. This dissertation developed the concept of “inclusive search” in SNSs, aiming at providing search results that make sense to ordinary people. The proposed approach is grounded on Organisational Semiotics, is empirically motivated by scenarios of use, and it was validated with real users. The dissertation is available at the UNICAMP digital library.*

1. Introdução e Motivação

O advento das Redes Sociais *Online* (RSO) tem trazido novas oportunidades para o acesso ao conhecimento. Contudo, existem ainda muitas barreiras que impedem as pessoas de interagirem com as Tecnologias de Informação e Comunicação (TICs) (Baranauskas & Souza, 2006). Propiciar o acesso irrestrito e facilitado ao conhecimento a todas as pessoas, sejam elas letradas (digitalmente) ou não, e independente de suas dificuldades, limitações físicas ou cognitivas, é fundamental no contexto globalizado em que vivemos. No entanto, devido à enorme sobrecarga de informação que os usuários estão expostos na *Web*, adicionado à falta de habilidade ou experiência que muitos possam ter com as TICs, novos mecanismos mais adequados para a procura de informação tornam-se ainda mais urgentes e necessários. Esses mecanismos são essenciais para as pessoas encontrarem conteúdos relevantes e que façam sentido a elas, promovendo seu processo de uso autônomo dos ambientes digitais. Propostas e soluções para a busca de informação na *Web* têm sido largamente investigadas nos últimos anos. Todavia, no geral, o que se percebe nessas propostas é que esses mecanismos ainda são fundamentados em informações não relacionadas com a realidade semântica das pessoas, uma vez que não consideram as diferentes possibilidades de significados e interpretações que os usuários podem ter, ou desenvolver em rede.

Nesse contexto, um dos caminhos de investigação que buscamos envolveu a proposta de um mecanismo de busca que fosse capaz de melhor considerar a semântica compartilhada pelas pessoas em Redes Sociais Inclusivas *Online* (RSI). Tal mecanismo deveria ser fundamentado em métodos que pudessem gerar melhores representações da semântica do contexto sócio-cultural da rede social. RSIs diferem de outras RSOs pela

atenção especial aos aspectos ligados ao acesso universal, onde cada funcionalidade do sistema é desenhada considerando a diversidade e as diferenças de competência das pessoas, *e.g.*, no seu letramento. Mecanismos de busca projetados para RSIs deveriam considerar os aspectos de inclusão e diversidade, tendo em mente que não se pode pressupor usuários familiarizados com os procedimentos e algoritmos de busca da *Web*.

O objetivo desta dissertação foi investigar como desenvolver “mecanismos de busca inclusivos”, *i.e.*, mais adequados à linguagem das pessoas organizadas por meio de RSIs. As principais contribuições desta dissertação envolveram: Proposição de recomendações para o desenvolvimento de mecanismos de busca mais adequados ao conceito de RSI; Proposta de um método para o *design* e construção de ontologias (Gruber, 1993) no contexto de RSI, fundamentado em métodos da Semiótica Organizacional (SO) (Liu, 2000); Estudo e análise de uma estratégia usando ferramentas e técnicas de mineração de texto para a construção de ontologias a partir de dados de RSI; Desenvolvimento de heurísticas e regras de transformação para a modelagem de ontologias fundamentadas nos resultados de métodos e artefatos da SO; Desenvolvimento de um protótipo de busca inclusiva em RSI; Avaliação do mecanismo desenvolvido junto ao público-alvo. Esta investigação está diretamente relacionada à um dos grandes desafios de pesquisa em Computação no Brasil: “Acesso participativo e universal do cidadão brasileiro ao conhecimento” (Baranauskas & Souza, 2006).

O mestrado no qual esse trabalho foi conduzido terminou em 24 meses e diversas publicações científicas foram geradas: um congresso nacional QUALIS B (SEMISH 2010) (Reis *et al.*, 2010a); uma conferência Ibero-Americana, IADIS Ibero-Americana *WWW/Internet 2010* (CIAWI 2010) QUALIS B2 (Reis *et al.*, 2010b). O artigo publicado nessa conferência foi convidado pelo periódico nacional QUALIS B5 (Revista Fonte) para ser publicado em uma versão estendida (Reis *et al.*, 2011e). Seis artigos em quatro congressos internacionais: dois artigos na *International Conference on Enterprise Information Systems* QUALIS B2 (ICEIS 2010, 2011) (Reis *et al.*, 2010c, 2011c). O artigo publicado em 2010 foi convidado para um capítulo de livro internacional na edição *Lecture Notes in Business Information Processing* (LNBIP) pela Springer (Reis *et al.*, 2011a); dois artigos na *International Conference on Informatics and Semiotics in Organisations* (ICISO 2010, 2011) (Reis *et al.*, 2010d, 2011b). O ICISO é o principal evento científico da área de SO; um artigo no *Workshop Web2Touch* publicado na *IEEE Computer Society Press* (Reis *et al.*, 2010e). Esse artigo foi convidado para ser estendido e foi publicado no *International Journal of Web Portals* (IJWP), *Special Issue Web2Touch Workshop* (Reis *et al.*, 2011f); tivemos também um artigo na *14th International Conference on Human-Computer Interaction*, publicado na edição *Lecture Notes in Computer Science* (LNCS) pela Springer (Reis *et al.*, 2011d). Como resultado do estudo de caso final, um artigo no periódico internacional “*International Journal of Universal Access in the Information Society*” QUALIS B3 foi aceito para publicação (Reis *et al.*, 2012). Este trabalho está organizado da seguinte maneira: A Seção 2 apresenta a problemática da investigação; A Seção 3 descreve a abordagem proposta para o problema; A Seção 4 finaliza com as considerações finais.

2. Problemática

Em RSOs, as soluções de busca atualmente quase necessariamente focam em algoritmos para diferentes abordagens de busca de pessoas na rede, mas não consideram os

conteúdos criados e compartilhados pelas pessoas. Uma vez que os aspectos semânticos podem fazer a diferença para os usuários alcançarem informação, a busca deveria considerar a diversidade de linguagem. A necessidade de recuperar informação semanticamente enriquecida aumentou o interesse em ontologias (Gruber, 1993). Para prover mecanismos de busca com características semânticas é necessário representar e entender um modelo de conhecimento sobre o domínio. Ontologias representam a semântica através de relacionamentos e representações dos conceitos e termos do domínio; logo, buscas fundamentadas em ontologias podem retornar resultados mais relevantes. Na *Web*, ontologias são usualmente descritas por linguagens de ontologias computacionais, e.g., *Web Ontology Language* (OWL) ¹.

Embora muito se tenha evoluído em soluções semânticas, a literatura tem apontado deficiências para as abordagens de modelagem usuais da *Web Semântica* (WebSem), o que tem demandado estudos mais aprofundados sobre a representação do conhecimento na *Web*. Além disso, ainda que a literatura tenha relatado e vislumbrado nos últimos anos novas perspectivas e possibilidades que visam considerar os aspectos semânticos da informação em sistemas computacionais na *Web*, as abordagens para os mecanismos de busca ainda são principalmente fundamentadas em dados sintáticos e vocabulários padrão. Esses vocabulários muitas vezes não fazem parte da linguagem cotidiana das pessoas que usam esses serviços, e.g., pessoas utilizam “postinho” no lugar de “unidade básica de saúde”. Ainda há grandes limitações nas soluções de busca para tratar problemas do tipo polissemia. O problema se torna ainda mais crítico quando se consideram usuários em fase de letramento digital e/ou analfabetos funcionais. Soluções que exigem muito desses usuários como, percorrer os conceitos representados na ontologia através de seus relacionamentos, podem não ter sucesso. Assim, soluções de interação que não considerem apenas os vocabulários formais e padrões (e.g., dicionários com a linguagem culta), mas também a linguagem coloquial, naturalmente compartilhada pelos usuários e mediada pela interação com os sistemas computacionais no dia-a-dia é altamente relevante.

3. Um Mecanismo de Busca Inclusivo

Como um possível caminho para o problema de busca visando a inclusão, nesta pesquisa articulamos técnicas, métodos e artefatos que possibilitam o entendimento, a discussão e a modelagem de conceitos compartilhados em RSIs. Para alcançar esse objetivo foi proposto e desenvolvido um mecanismo de busca fundamentado em ontologias construídas a partir dos conteúdos da rede social. Essas ontologias também são modeladas a partir de uma nova abordagem para o *design* de ontologias.

Construção das ontologias. O principal desafio endereçado pela presente abordagem foi desenvolver uma solução computacional de busca situada nos significados que as pessoas trouxeram para a rede, e também os que foram tecidos através da interação. Para tanto, foram estudadas possíveis ferramentas e técnicas de mineração de texto em um método semi-automático que inclui a extração de conceitos advindos dos conteúdos do sistema para auxiliar na construção das ontologias a partir do conteúdo da RSI. Um experimento foi realizado com dados reais de uma RSI, o *VilanaRede*² (VnR). Um dos desafios foi lidar com informação em linguagem natural,

¹ www.w3.org/TR/owl-features

² www.vilanarede.org.br

não estruturada, além de conter características de informalidade e abrangência de assuntos. O processo de extração de termos do conteúdo estudado apontou resultados de pesquisa (Reis *et al.*, 2011e) que demonstram a importância da aplicação de métodos apropriados ao contexto de investigação considerado.

Abordagem Semiótica para ontologias. As ontologias são criadas com base nos resultados de extração das ferramentas de mineração de texto. Contudo, para alcançar uma busca inclusiva, desenvolvemos uma nova abordagem para a engenharia das ontologias *Web* para informar a busca inclusiva. Essa proposta inclui o papel do indivíduo que interpreta signos (representados como conceitos na ontologia *Web*). Portanto, os conceitos de “Agentes” e “Dependências Ontológicas”, que vêm do resultado do *Método de Análise Semântica* (SAM) da SO (Liu, 2000) foram incluídos em ontologias *Web* descritas em OWL (Reis *et al.*, 2010d). Com isso, tanto os “Agentes” como os “*Affordances*” (Liu, 2000) (conceitos do SAM) são transformados em classes OWL, e são relacionados uns aos outros seguindo as dependências ontológicas modeladas no Diagrama de Ontologia (DO) do SAM. Logo, na abordagem, ao invés de modelar os significados a partir dos conteúdos da rede usando OWL diretamente, eles são primeiro modelados usando o DO.

Essa abordagem deu origem ao conceito de “Ontologia *Web* Semiótica” (*Semiotic Web Ontology*) SWO (Reis *et al.*, 2010d). SWO é um modelo semântico (ontologia computacionalmente tratável) construído incluindo conceitos do SAM como os conceitos de “Agentes”, “*Affordances*” e “dependências ontológicas”, combinando assim conceitos do SAM com tecnologias da WebSem. Reis *et al.* (2012) propõem um método geral chamado WODAS (*Design de Ontologia Web Suportado por Semiótica*) (Figura 1) para criar SWOs em contexto de RSO. A Figura 1 apresenta uma visão geral dos métodos implementados. Com base nesse método, ontologias são criadas a partir dos conteúdos da rede social, e usadas pelo mecanismo de busca inclusivo.

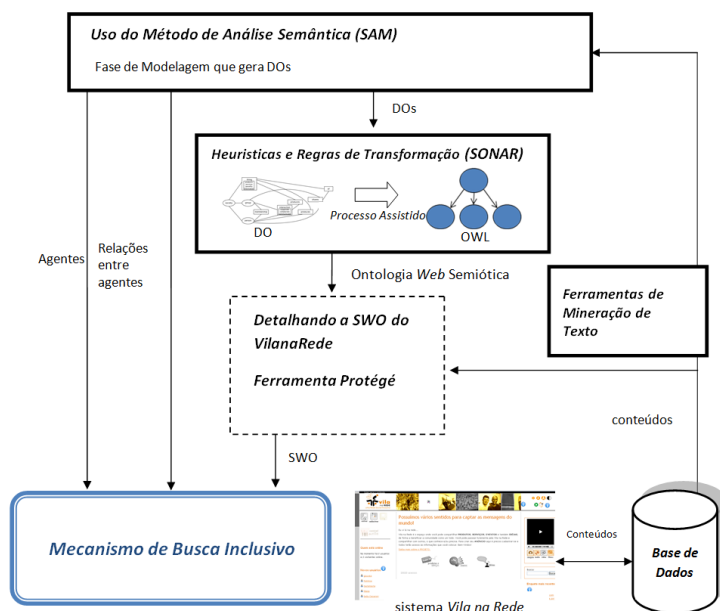


Figura 1. Métodos envolvidos na proposta (adaptado de Reis *et al.*, 2012)

A fim de alcançar uma ontologia *Web* computacionalmente tratável descrita em OWL a partir do DO resultante do SAM, heurísticas (Reis *et al.*, 2011c) e regras de transformação (Reis *et al.*, 2011b) são aplicadas para produzir SWOs. Heurísticas

objetivam mostrar a relação de transformação explícita entre o artefato (DO) e o código OWL. As heurísticas tornam explícitas as relações entre os modelos, e também tornam possível a construção de um diagrama a partir do outro. Essas heurísticas foram implementadas como regras de transformação em um processo semi-automático.

O Mecanismo de Busca Inclusivo. Esse mecanismo age fundamentado nas ontologias construídas usando o WODAS, o que difere de métodos tradicionais. Ao modelar as ontologias, os termos informais encontrados a partir do conteúdo da RSI são ligados a possíveis conceitos formais. Portanto, quando um usuário procura por alguma informação utilizando sua descrição informal ou regional, conteúdos que reflitam o conceito formal são candidatos a serem recuperados. O ponto chave para a estratégia de personalização proposta na busca inclusiva é a utilização de informações sobre os agentes sociais modelados na ontologia como um contexto para a busca. Os possíveis significados de uma palavra estão ligados a agentes em ambas as situações: durante o processamento do índice de busca do conteúdo gerado pelos usuários da RSI e durante o tempo de busca. Assim, a desambiguação é determinada pelos agentes.

Implementação. Com base na arquitetura do VnR, o mecanismo de busca inclusivo foi desenvolvido como serviços *Web*. Os Serviços *Web* da busca gerenciam todas as informações necessárias e fornecem todos os métodos para lidar com a solicitação de busca, o processamento do índice, a adição de agentes e relacionamentos entre agentes e usuários do VnR, bem como com a gestão de ontologias para obter significados e expansão das consultas de busca. Os serviços de busca incluem também o algoritmo de busca que fornece os resultados da busca organizada por listas.

Validação. A partir do protótipo foi realizado um experimento a fim de investigar o comportamento da interação de usuários representativos no contexto do VnR. O experimento avaliou o uso da busca inclusiva no VnR por meio de cenários de busca real e teve como meta avaliar o mecanismo de busca implementado no VnR com 25 usuários finais. De maneira geral, os resultados indicaram que os participantes apreciaram os anúncios (conteúdo principal do VnR) resultantes da busca fornecida pelo mecanismo proposto, bem como a forma como ele os organiza. Uma vez que a RSI retorna resultados mais completos, significativos e melhor organizados, essa pode fornecer meios melhores para os usuários encontrarem informação. Detalhes dessa validação e uma discussão sobre trabalhos relacionados estão em Reis *et al.* (2012).

4. Considerações Finais

O contexto único de multiplicidade de riquezas culturais e de problemas sociais presente na sociedade contemporânea, principalmente em países em desenvolvimento como o Brasil, exige que a criação de aplicações computacionais seja abordada de maneira socio-técnica e inclusiva. Considerando esse cenário, a pesquisa descrita nesta dissertação abordou o desenvolvimento de mecanismos de busca mais adequados ao conceito de RSIs. Nesta investigação buscaram-se respostas para a criação de métodos em direção a mecanismos inclusivos que pudessem gerar resultados de busca com mais sentido para os usuários. Com base na abordagem proposta protótipos foram construídos e a solução foi validada em um estudo de caso. Diversos artigos em eventos, capítulos de livros e periódicos nacionais e internacionais foram publicados.

Agradecimentos: A dissertação foi realizada no contexto do projeto *e-Cidadania* financiado pela *Microsoft Research – FAPESP* Instituto (processo nro. 2007/54564-1).

Referências

- Baranauskas, M.C.C. & Souza, C.S. (2006) “Desafio 4 Acesso Participativo e Universal do Cidadão Brasileiro ao Conhecimento”. *Computação Brasil*, ano VII, n. 23, p.7.
- Gruber, T. R., (1993) “A translation approach to portable ontologies” *Knowledge Acquisition*. Vol.5, n.2.
- Liu, K. (2000) “Semiotics in information systems engineering”. Cambridge Univ. Press.
- Reis, J. C.; Baranauskas, M. C.; Bonacin, R. (2010a) “Busca em Sistemas Sócio-Culturais Online: Desafios e Recomendações”. In: *Anais do XXX Congresso da Sociedade Brasileira de Computação*. XXXVII Seminário Integrado de Software e Hardware (SEMISH). Belo Horizonte, Brasil, 2010, p. 380-394.
- Reis, J. C., Bonacin, R., Baranauskas, M.C.C. (2010b). “Ferramentas e Técnicas de Mineração de Texto em Redes Sociais Inclusivas: Promovendo Meios para o Acesso Universal ao Conhecimento”. In: *Proceedings of IADIS Ibero-Americana WWW/Internet 2010*. Carvoeiro, Portugal, p. 1-10.
- Reis, J. C.; Bonacin, R.; Baranauskas, M. C. C. (2010c) “New Perspectives for Search in Social Networks: A Challenge for Inclusion”. In *Proceedings of the 12th International Conference on Enterprise Information Systems*. p. 53-62.
- Reis, J.C., Bonacin, R. e Baranauskas, M.C.C., (2010d) “A Semiotic-based Approach to the design of Web Ontologies”. In: *Proceedings of 12th International Conference on Informatics and Semiotics in Organisations (ICISO 2010)*. Reading, UK, p. 60-67.
- Reis, J. C.; Bonacin, R.; Baranauskas, M. C. C. (2010e). “Search Informed by a Semiotic Approach in Social Network Services”. In *Proceedings of the. Workshop Web2Touch - living experience Through Web*. Tozeur, Tunisia, p. 321-326.
- Reis, J.C., Bonacin, R. e Baranauskas, M.C.C. (2011a) “Prospecting an Inclusive Search Mechanism for Social Network Services”. In: *ICEIS 2010 - Revised Selected Papers*. In J. Filipe and J. Cordeiro (Eds.). *Lecture Notes in Business Information Processing (LNBIP)*. Vol. 73, p. 555-570.
- Reis, J.C., Bonacin, R. e Baranauskas, M.C.C. (2011b) “An Assisted Process for Building Semiotic Web Ontology”. In: *Proceedings of the 13th International Conference on Informatics and Semiotics in Organisations (ICISO 2011)*, Leeuwarden, Holland, vol. 1, p. 167-174.
- Reis, J.C., Bonacin, R. e Baranauskas, M.C.C. (2011c) “Constructing Web Ontologies Informed by Semantic Analysis Method: Heuristics and Transformation Rules”. In: *Proceedings of the 13th International Conference on Enterprise Information Systems (ICEIS 2011)*. Beijing, China, p. 203-206.
- Reis, J. C.; Bonacin, R.; Baranauskas, M. C. C. (2011d) “Beyond the Social Search: Personalizing the Semantic Search in Social Networks”. In: *Online Communities and Social Computing*. Ozok, A. and Zaphiris, Panayiotis (Eds.), *Lecture Notes in Computer Science (LNCS)*. Springer Berlin Heidelberg. Vol. 6778, p. 345-354.
- Reis, J.C., Bonacin, R. e Baranauskas, M.C.C., (2011e) “O Uso de Ferramentas de Mineração de Texto para Modelar Significados em Redes Sociais Inclusivas”. In: *Revista Fonte*. Ano 8, num 11, p. 59-69.
- Reis, J. C.; Bonacin, R.; Baranauskas, M. C. C. (2011f). “A Semiotic-Based Approach for Search Mechanism in Social Network Services”. In: *International Journal of Web Portals (IJWP)*. Vol. 3, Issue 3, p. 27-40.
- Reis, J.C., Bonacin, R. e Baranauskas, M.C.C. (2012) “Addressing Universal Access in Social Networks: An Inclusive Search Mechanism”. In: *International Journal of Universal Access in the Information Society* (aceito para publicação)