

Busca em Sistemas Sócio-Culturais *Online*: Desafios e Recomendações

Júlio Cesar dos Reis¹², M. Cecília C. Baranauskas¹, Rodrigo Bonacin²

¹Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Caixa Postal 6176 – 13.084-971 – Campinas – SP – Brasil

²Centro de Tecnologia da Informação Renato Archer (CTI) – Rodovia Dom Pedro I
km 143.6 – 13.069-901 – Campinas – SP – Brasil

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

Abstract. *Brazil is replete of scenarios characterized by cultural diversity, and also by social and digital inclusion problems. Online socio-cultural micro systems represented by social network services may have a dominant role in influencing the adoption and use of information and communication technologies. They can mean an opportunity for socio-digital inclusion through inclusive social network services (ISN), promoting access to knowledge. This access almost necessarily depends on searching the system for information. This paper aims to show the conception of search engines more appropriate to this context. We show the involved challenges and propose recommendations which prospect a search solution customized for ISN.*

Resumo. *O Brasil está repleto de cenários marcados pela diversidade cultural e também por problemas de inclusão social e digital. Micro sistemas sócio-culturais online representados pelas redes sociais na Web podem ter um papel dominante na influencia da adoção e uso das tecnologias de informação e comunicação. Eles podem significar uma oportunidade para a inclusão sócio-digital através de redes sociais inclusivas (RSI) promovendo o acesso ao conhecimento. Este acesso quase necessariamente passa pela busca de informação no sistema. O objetivo deste artigo é mostrar a concepção de mecanismos de busca mais adequados a este contexto. Mostramos os desafios envolvidos e propomos recomendações que prospectam uma solução de busca personalizada a RSI.*

Palavras-Chave: *Redes Sociais Inclusivas, Busca Semântica, Ontologias*

1. Introdução

Nos últimos anos, com o avanço da tecnologia *Web*, surgiu uma categoria de *software* denominada ‘Software Social’ (*Social Software*). Este *software* introduziu novas oportunidades de interação e troca originando grandes volumes de dados e informação de usuários. Isto possibilitou o surgimento de micro sistemas sócio-culturais *online*. Entende-se que sistemas sócio-culturais emergem de complexas interações e relações

entre atores sociais que compartilham e desenvolvem uma cultura comum, e os micro sistemas são parte deste todo. Estas interações e relações sociais são mediadas também por *software* social como as Redes Sociais na *Web* (Boyd & Ellison, 2008).

As SNS podem ser consideradas exemplos de micro sistemas sócio-culturais que se configuram como sistemas complexos, permitindo aos indivíduos interagirem e compartilharem seus interesses e atividades. As redes formam comunidades que conectam pessoas com objetivos comuns e se apresentam como um espaço virtual propício para a comunicação dos usuários, além de representarem uma oportunidade para compartilhamento de informação, conhecimento e experiências sociais. O acesso a informação e ao conhecimento gerado nas redes sociais pode ser realizado principalmente através dos mecanismos de busca.

A maior parte da informação disponível na *Web* e também nas SNS está em formatos que não podem ser interpretados por máquinas, como textos pouco estruturados ou livres. As ferramentas de busca não têm sido projetadas para interpretar representações do significado dos termos de busca, sendo construídas fundamentalmente para responder com base em comparações de palavras-chave e processamento léxico-sintático, *i.e.*, sem considerar conhecimento semântico. Segundo Mauldin (1991) um dos maiores desafios no desenvolvimento de mecanismos de busca é a capacidade de diferenciação entre resultados relevantes e irrelevantes. Esta diferenciação é prejudicada, principalmente por fatores de polissemia (termos que possuem vários significados) e sinônimos (existência de várias palavras para descrever determinado conceito), além de outros quesitos complexos da linguagem humana.

As dificuldades enfrentadas pelos usuários para obter informação no ambiente *Web*, especialmente aqueles em processo de alfabetização digital, podem ser explicadas principalmente pela sobrecarga de informação apresentada nos sistemas, e também por problemas relacionados com a contextualização do significado dos termos utilizados na busca e no conteúdo. A literatura tem apresentado diversas tentativas para minimizar estes problemas por meio de melhores mecanismos de busca; estas soluções empregam diversas abordagens, incluindo: o uso do conceito de distância social, recomendação, reputação na rede social e também aspectos semânticos. Entretanto, as dificuldades dos usuários juntamente com as deficiências dos mecanismos de busca ainda são barreiras para o acesso a informação em sistemas sócio-culturais *online*.

Segundo um dos grandes desafios atuais de pesquisa em Computação no Brasil para os próximos anos: “Acesso participativo e universal do cidadão brasileiro ao conhecimento” (Baranauskas & Souza, 2006), vivemos em um cenário de vastas diferenças sócio-econômicas, culturais, regionais e de acesso à tecnologia e ao conhecimento; barreiras tecnológicas, educacionais, culturais, sociais e econômicas, têm impedido o acesso e a interação. Inspirado neste desafio, esforços de pesquisa proposta por Baranauskas (2007) visam transformar SNS em um motor para a inclusão digital e a cidadania. Uma rede social com recursos que promovam o acesso de todos, incluindo aqueles à margem da cultura digital, pode ser definida como uma ‘Rede Social Inclusiva’ (RSI) (Neris *et al.*, 2009). Sendo o mecanismo de busca uma das formas mais diretas de prover o acesso a informação nestes sistemas, mostra-se necessário o desenvolvimento de mecanismos mais adequados e especialmente voltados a este fim, que possam contribuir para o acesso participativo e universal através de RSI.

Em uma rede social, a partir das interações e expressões dos usuários, significados, conceitos e interpretações são construídos, transferidos e compartilhados através do sistema por meio de estruturas complexas, altamente distribuídas, incluindo uma imensa diversidade cultural presente na informação produzida. Novas oportunidades de avanço sobre os mecanismos de busca poderiam ser alcançadas por meio de uma forma mais adequada de modelagem e representação desta informação. Para tal, é necessário criar métodos e ferramentas computacionais para identificar e modelar estas interpretações, expressas através das interações dos usuários com o sistema e também da interação entre usuários. A investigação nesta direção pode gerar diversos avanços entre eles modelos semânticos mais fiéis ao contexto de redes sociais e como consequência buscas semânticas mais precisas. Estas buscas devem levar em consideração os significados utilizados na rede social, e que fazem sentido para os indivíduos daquele contexto.

A descoberta e a representação da semântica empregada em uma rede social configura-se como uma modelagem complexa de um micro sistema sócio-cultural, enquadrando-se no cenário do 2º Grande Desafio de Pesquisa em Computação da Sociedade Brasileira de Computação (SBC)¹ – “Modelagem computacional de sistemas complexos artificiais, naturais e sócio-culturais e da interação homem-natureza”. Neste contexto, verificam-se novas necessidades e esforços que se concentram no entendimento e mapeamento dos dados do ponto de vista semântico, exigindo conhecimentos que vão além do processamento sintático dos dados. A solução para o problema de modelagem neste sentido depende do entendimento da “natureza” da informação e do conhecimento. Compreender como estruturar, modelar, organizar, administrar e promover meios para que a informação disponível nas redes sociais seja mais bem representada computacionalmente, pode levar a formas mais eficientes de acesso a informações através de mecanismos de busca mais adequados.

O objetivo deste artigo é investigar e mostrar a concepção de mecanismos de busca mais adequados as RSI. O artigo mostra os desafios envolvidos neste processo, e discute as principais abordagens para busca em redes sociais presentes na literatura. Fundamentados na análise deste cenário propomos recomendações que prospectam um mecanismo de busca inclusivo. O artigo está organizado da seguinte forma: A Seção 2 apresenta o conceito de ‘redes sociais inclusivas’ visando contextualizar a pesquisa e os desafios de busca atrelados a ela; a Seção 3 apresenta e discute algumas das estratégias de busca possíveis para redes sociais presentes na literatura; A seção 4 apresenta as recomendações e a Seção 5 conclui o trabalho mostrando também os trabalhos futuros.

2. Redes Sociais Inclusivas e Desafios Relacionados à Busca

Segundo Boyd & Ellison (2008) as redes sociais ou as “comunidades de membros” têm grande relevância na *Web*, pois os internautas dedicam a elas grande parte do seu tempo de navegação. Atualmente tem havido um rápido aumento no número de usuários que se

¹ www.sbc.org.br

inscrevem e utilizam ativamente os *sites* de redes sociais como *MySpace*², *FaceBook*³, *Orkut*⁴, *Flickr*⁵ entre outros.

Conforme revelam os dados de Nielsen (2009): (1) O Brasil é o país com o maior número de internautas usando *sites* de relacionamento; (2) 80% dos brasileiros que navegam na Internet estão ligados aos *sites* de "comunidades de membros", *blogs* e redes de relacionamento; (3) Os internautas brasileiros também são os que passam mais tempo neste tipo de *site*, de 1 a cada 4 minutos de navegação na Internet; e (4) Na Espanha 75% dos internautas usam redes de relacionamento, na Itália 73% acessam redes sociais e no Japão 70% dos usuários se comunicam em redes sociais. Ainda segundo Nielsen (2009), as redes sociais são mais populares do que o *e-mail*, com 66,8% de alcance global e estas figuram no quarto lugar entre os recursos mais utilizados na Internet. Adicionalmente, 85,9% dos internautas usam as ferramentas de busca, que é uma das atividades mais realizadas na Internet.

Apesar destas estatísticas otimistas e do sucesso das redes sociais entre pessoas já usuárias da Internet, em contextos sociais como o do Brasil e de outros países em desenvolvimento, ainda há um grande número de pessoas sem acesso à Internet e, conseqüentemente, sem oportunidades para acesso à informação e ao conhecimento. Os indicadores sociais demonstrados pela Pesquisa Nacional por Amostra de Domicílio (PNAD) produzido pelo Instituto Brasileiro de Geografia e Estatística (IBGE)⁶ assinalam que, em 2008, 65% da população brasileira não tiveram acesso à Internet. Além disso, dados do Ministério da Educação no Brasil (MEC)⁷ revelam que cerca de 30 milhões de brasileiros são analfabetos funcionais, definida como a população com mais de 15 anos e menos de 4 anos de escolaridade (21,6% da população).

Dados tão alarmantes, com um alto índice de analfabetismo funcional, ilustram apenas parte do desafio que enfrentamos em termos do projeto de sistemas que deveriam incluir estas pessoas. Soluções para a inclusão neste âmbito, portanto, envolvem a consideração de toda essa diversidade de usuários e contextos no processo de desenvolvimento de sistemas de informação em geral. Há uma necessidade urgente de métodos que se baseiem em um entendimento profundo do conceito de inclusão e das diferenças (Baranauskas, 2007). Desta forma, é extremamente relevante criar espaços e meios para que as pessoas sejam incluídas de maneira natural no acesso e uso da tecnologia. Junto a isso é também importante criar métodos mais sofisticados que possibilitem de maneira efetiva e promovam o acesso e uso da informação veiculada na mídia digital (conteúdo digital) por todas as pessoas, incluindo as pouco letradas digitalmente ou analfabetos funcionais.

Rumo a este objetivo, a partir do projeto de pesquisa proposto por Baranauskas (2007) tem se desenvolvido o conceito de 'Redes Sociais Inclusivas' (RSI), constatado que redes sociais desempenham um papel importante na influência, adoção e uso das tecnologias de informação e comunicação (Ramachandran, 2007). Segundo Baranauskas

² www.myspace.com

³ www.facebook.com

⁴ www.orkut.com

⁵ www.flickr.com

⁶ <http://www.ibge.gov.br/home/estatistica/populacao/acessoainternet2008>

⁷ <http://portal.mec.gov.br>

(2007) soluções de inclusão digital devem ser tratadas como instrumento de transformação profunda da sociedade, e dentre as soluções possíveis estão os serviços nas RSI. Trata-se de um “espaço virtual de comunicação”, que seja inclusivo e que permita à comunidade compartilhar conhecimento sobre “o fazer comunitário”. Em geral, esse espaço deve ser generalizado para possibilitar “trocas” (de conhecimento, bens e serviços). Neris *et al.* (2009) mostram a elaboração de um significado socialmente construído para o conceito de RSI. Estas redes tendem a propiciar acesso inclusivo às pessoas, criando situações nas quais a diversidade é respeitada e as dificuldades dos usuários vencidas durante o acesso e uso dos conteúdos digitais.

É importante que os indivíduos tenham a possibilidade de recuperar informação de modo mais preciso, com resultados que façam sentido a eles. O mecanismo de busca, como uma funcionalidade essencial em uma RSI com grande volume de informação, também deve cooperar para o acesso do indivíduo à informação que busca. A melhoria deste recurso em uma RSI pode dar-se principalmente pela naturalidade que o usuário terá para procurar uma informação e pela qualidade de resposta do sistema. O desafio neste aspecto está principalmente relacionado em como propiciar uma experiência harmoniosa entre o mecanismo e o usuário em termos de sua linguagem cotidiana. Em outras palavras a busca deve refletir a realidade semântica do contexto envolvido. Isto está relacionado com esforços de pesquisa atuais para o desenvolvimento de mecanismos de busca adaptativos, baseados principalmente na descoberta do perfil e do comportamento do usuário, além de abordagens que utilizam evolução do domínio (Guelfi *et al.*, 2010).

No contexto deste trabalho, o desafio é desenvolver uma solução computacional para busca situada no significado que emerge no contexto de uso da rede, *i.e.*, os significados que as pessoas trouxeram para a rede, e também os que foram tecidos com o uso do sistema ao longo do tempo (através da interação). A semântica que existe e que emerge na rede não poderá ser representada computacionalmente a menos que os significados utilizados sejam corretamente identificados e modelados. A partir desta modelagem, objetiva-se melhorar a qualidade e precisão na busca da informação. Isso poderá facilitar e prover um acesso mais natural ao conteúdo gerado pelos usuários. Para isso estudos relativos à modelagem e representação semântica da linguagem utilizada em redes sociais necessitam ser desenvolvidos.

Este estudo também vai ao encontro do 2º Grande Desafio de Pesquisa em Computação da SBC, pois uma rede social pode ser entendida como um sistema sócio-cultural complexo. Portanto, criando métodos para uma melhor modelagem de sistemas sócio-culturais, estes possivelmente também poderão contribuir para a geração de mecanismos de busca mais adequados a uma RSI. O desenvolvimento de métodos de extração de requisitos, algoritmos e mecanismos de coleta e processamento de dados que capturem variáveis relativas a interações sociais e sócio-culturais podem influenciar na modelagem dos aspectos semânticos na RSI. A produção de técnicas que lidam com novas maneiras de representação dos significados da informação pode gerar grandes benefícios para ambos os desafios (2 e 4) impactando em um primeiro momento em melhores mecanismos de busca.

3. Estratégias de Busca para Redes Sociais

Um mecanismo de busca em uma SNS permite aos usuários procurar pessoas e informações em uma coleção de documentos e itens (ou outras fontes de informação) através de pesquisas (*query*), em geral formatadas como um conjunto de palavras-chave. Utilizando estas pesquisas, o mecanismo recupera informações que possam ser relevantes para o usuário. Segundo Gürsel & Sen (2009) como poucas pessoas acessam resultados de busca apresentados além das primeiras páginas, a ordenação dos resultados é fator essencial para a satisfação dos usuários. A literatura tem apresentado soluções para mecanismos de busca em SNS baseadas em diversas estratégias, tais como: sistemas de recomendação, confiança e também focada em aspectos semânticos. A maioria dos trabalhos presentes na literatura relacionados à busca especialmente voltada para SNS tem ênfase na procura de pessoas (*social search*) e não exatamente no conteúdo gerado e compartilhado no sistema pelos indivíduos.

3.1. Busca Social

O trabalho pioneiro de Yu & Singh (2003) propõe uma solução para encontrar um participante especialista (sistema de reputação) em uma SNS utilizando um sistema multi-agente. Os agentes artificiais fazem uma busca distribuída através de referências na rede e aprendem o perfil do usuário (preferências e interesses) e seus conhecidos na rede social. Usando destas informações, o agente prioriza as consultas e com base na resposta do seu usuário modifica as taxas de parâmetro de busca de cada agente que forneceu uma resposta, e aqueles que se referem a ele. A busca é feita usando este sistema dinâmico de referências.

Vieira *et al.* (2007) propõem que relacionamentos de amizade são uma evidência importante para a procura de pessoas em uma rede social, e que a posição dos resultados de busca pode ser modelada em função da distância entre os usuários em um grafo, mais especificamente, dos caminhos mais curtos de um grafo de amizade. A mesma idéia é compartilhada por Haynes & Perisic (2009), argumentando que a rede social interpessoal pode ter grande valor para melhorar a ordenação dos resultados de busca de perfil em SNS. O objetivo do estudo de Haynes & Perisic (2009) é verificar se a estrutura de uma rede social interpessoal pode contribuir para melhorar a ordenação dos resultados de busca. A principal conclusão é que a medida de distância social (estrutura da rede social), quando utilizada em conjunto com métodos padrão de busca melhoram a ordenação dos perfis de resultado da busca, e que a estrutura de grafo social deve ser particularmente útil para procurar pessoas em SNS.

A abordagem de Gürsel & Sen (2009) foca em aspectos de recomendação e confiança para busca de itens em SNS. Eles propõem um sistema de busca de itens baseado em redes sociais (*Social Network based Item Search* - SNIS). Este sistema usa um *framework* baseado em agentes artificiais que “mineram” o grupo de relacionamentos de um usuário na rede, com o intuito de melhorar os resultados de busca utilizando as conexões sociais, facilitando a busca de itens de interesse. SNIS foi implementado no *Flickr* para a busca de fotos. Este mecanismo de busca proposto por Gürsel & Sen (2009) fornece as preferências de um usuário pelo monitoramento indireto de avaliações anteriores deste. Para classificar estas preferências, o sistema identifica tópicos de preferências específicos da pessoa para os itens de seus amigos na rede. Em

outras palavras isto permite que um usuário tenha preferências diferentes para os itens correspondentes a diferentes tópicos destacados por um dado amigo. O objetivo é classificar os resultados de busca destacando os itens de preferência recentemente postados pelos amigos na rede social, que será de especial interesse para uma pessoa. Segundo Gürsel & Sen (2009) a abordagem gera resultados de busca que podem melhorar a precisão de busca. Estas propostas tentam melhorar a busca em sistemas de redes sociais, contudo ainda utilizam abordagens de buscas sintáticas tradicionais e limitadas.

3.2. Busca Semântica

De acordo com Kassim & Rahmany (2009) devido ao grande crescimento do volume de informação na *Web* e de sua complexidade, os motores de busca tradicionais já não são mais capazes de prover resultados precisos. A deficiência destes mecanismos de busca é que eles são baseados apenas em palavras-chave e também não são capazes de lidar com aspectos de polissemia e sinônimos, e por muitas vezes, retornam resultados que não satisfazem às necessidades dos usuários. Busca semântica tem se tornado uma alternativa para superar as deficiências dos mecanismos tradicionais. Diferentes destes, os mecanismos de busca semântica tentam analisar e compreender o que o usuário deseja na pesquisa em um contexto através de ‘raciocínio lógico’, possibilitando melhores resultados. Não temos o objetivo neste artigo de fazer uma revisão exaustiva da literatura sobre o tópico de busca semântica, pois já existem trabalhos com este propósito na literatura, *e.g.* Mangold (2007) e Wei *et al.* (2007). Contudo, nesta seção elucidamos principalmente os trabalhos mais significativos e também os mais atuais, dando uma visão geral sobre os mesmos, e um panorama com tendências sobre o tópico que podem trazer contribuições para buscas mais apropriadas em contextos de RSI.

Nos últimos anos, esforços consideráveis em pesquisas têm sido dedicados a aplicação de tecnologias da *Web Semântica* (Berners-Lee *et al.*, 2001) no desenvolvimento de mecanismos de busca semântica. Segundo Berkan (2007) estes mecanismos trazem diversas vantagens como: (1) tornam mais fáceis a localização de informações relevantes para o assunto de interesse do usuário, poupando tempo de leitura em resultados não relacionados; (2) tratam consultas mais bem elaboradas e (3) através destas consultas, que podem ser analisadas por algoritmos semânticos em tempo de execução, possibilitam resultados de pesquisa com contextos mais precisos.

Os mecanismos semânticos podem superar as limitações das buscas tradicionais, pois eles possibilitam o uso de uma ontologia para ‘inferir’ informações sobre conceitos. Em computação, segundo Gruber (1993) ontologia é uma especificação formal e explícita de uma conceitualização compartilhada, que fornece descrições sobre conhecimento. Para Studer *et al.* (1998), ontologia é um entendimento comum e compartilhado de algum domínio que pode ser comunicado entre pessoas e computadores. Em outras palavras é uma especificação formal que deve ser capaz de ser lida e entendida por máquinas. Trabalhos como o de Heflin & Hendler (2000), Guha *et al.* (2003), além de Bonino *et al.* (2004) e Fang *et al.* (2005) introduzem a idéia do uso de ontologias em mecanismos de busca semântica. A adição de semântica explícita pode melhorar as buscas. As ontologias tentam representar os significados em um modelo semântico para melhor contextualizar os significados dos termos no processo de busca,

permitindo que um sistema de busca semântico identifique corretamente os conceitos e suas relações.

Segundo Mangold (2007), a busca tradicional de documento é estabelecida principalmente sobre a ocorrência de palavras em documentos (sintático); já a busca semântica é um processo de recuperação de documentos que explora o conhecimento no domínio, que pode ser formalizado por meio da ontologia. Bonino *et al.* (2004) dizem que o ponto chave para o processo de refinamento de uma busca semântica está na disponibilidade de uma ontologia de domínio, e na capacidade de compreender as relações semânticas entre os conceitos ontológicos. Isto é importante, pois as buscas são bem dependentes de contextos devido aos vários significados de uma mesma palavra *e.g.*, a palavra macaco pode significar tanto um animal, quanto um artefato que auxilia o humano a trocar o pneu do carro, dependendo do contexto aplicado. Uma busca que não leve em consideração o significado pode retornar, portanto, resultados imprecisos, desconexos ou incompletos.

Grande parte das abordagens para busca semântica encontradas na literatura propõe soluções arquiteturais genéricas e distribuídas, voltadas principalmente para a *Web* (Semântica). Outras são voltadas para sistemas de informação em contextos mais específicos, mas poucas são especialmente voltadas para SNS. Há uma série de propostas e projetos-piloto implementados e avaliados em vários domínios de aplicação, *e.g.* a máquina de busca semântica do projeto TRUST⁸ descrita em Amaral *et al.* (2004). Estes autores apresentam um motor semântico de perguntas e respostas, assim como sua arquitetura, as ferramentas e mecanismos de processamento implementados para o módulo da língua Portuguesa. O objetivo final do referido projeto é desenvolver um motor de busca semântico e multi-línguas capaz de processamento de linguagem natural.

Celino *et al.* (2006) apresenta o *Squiggle*: um *framework* extensível de busca semântica para o desenvolvimento de aplicações de busca semântica de domínio específico, que leva em consideração particularidades necessárias na busca de conteúdos multimídia. Seu modelo de representação do conhecimento baseia-se no vocabulário SKOS⁹, que permite ao sistema sugerir significados de consultas por um processo de inferência simples, *e.g.*, sugerem rótulos alternativos ou sinônimos para uma imagem. Há uma aplicação deste *framework* para o contexto de música, chamado *Squille Music*¹⁰. Segundo Lei *et al.* (2006) as propostas de busca semântica não prestam atenção especial aos usuários finais comuns, que não necessariamente estão familiarizados com domínios específicos de dados semânticos, ontologias (navegação para modificação de consultas manualmente), ou linguagens computacionais de consulta. Desta maneira, o principal objetivo do trabalho de Lei *et al.* (2006) foi propor vários meios para “esconder” a complexidade da busca semântica, tentando torná-la mais eficiente e fácil de usar aos usuários finais. Eles focam não necessariamente em aspectos de interface da busca, contudo nos aspectos da formação da consulta (*query*). Argumentam que superam a limitação dos motores de busca semântica baseados

⁸ <http://www.trustsemantics.tip.net.pl/>

⁹ <http://www.w3.org/2004/02/skos/>

¹⁰ <http://squiggle.cefriel.it/music/>

somente em uma palavra-chave, suportando consultas semânticas complexas em termos de várias palavras-chave em uma interface de busca simples como a do *Google*¹¹.

Mangold (2007) faz uma revisão e classificação (*survey*) das principais propostas de busca semântica presentes na literatura, assim como os pontos em aberto e os esforços necessários de investigação. O mesmo autor apresenta resultados sobre soluções de busca; 10 trabalhos pesquisados foram classificados e analisados, mostrando os trabalhos mais efetivos do ponto de vista semântico em cada categoria de classificação, descritas como: arquitetura, acoplamento, transparência, contexto do usuário, modificação de consultas, estrutura e tecnologia de ontologias. O trabalho de Wei *et al.* (2007) também faz uma breve revisão sobre as principais propostas de busca semântica na literatura observando aspectos de semelhanças, objetivos, aplicações, metodologias além de tecnologias envolvidas entre as propostas, e apresenta o IRIS juntamente com sua arquitetura e componentes. Este é um protótipo de sistema de busca semântica que ajuda pesquisadores a procurar e explorar coleções de grande número de publicações científicas.

Em específico sobre busca semântica em SNS há a proposta de Choudhari *et al.* (2008). Estes autores descrevem uma arquitetura para desenvolver uma busca semântica com uma abordagem a partir de agrupamento de termos relacionados semanticamente. Na proposta de Choudhari *et al.* (2008), se o usuário pretender fazer uma busca a partir de uma palavra-chave, esta busca poderia ser generalizada para todos os termos do agrupamento. Por exemplo, supondo que “nadar”, “correr” e “esporte” estejam em um agrupamento (de acordo com os autores, os termos destes agrupamentos são criados segundo informações dos perfis dos usuários), ao fazer a busca semântica, seria equivalente utilizar qualquer uma destas palavras-chave. O sistema automaticamente usaria todas as palavras-chave neste agrupamento para efetuar a busca.

3.3. Discussão e Análise com foco em RSI

Esta subseção tem o propósito de analisar e discutir as soluções já desenvolvidas na literatura para busca em SNS e busca semântica. Os trabalhos de Guha *et al.* (2003), Bonino *et al.* (2004) e Fang *et al.* (2005) são pioneiros no uso de ontologias em mecanismos de busca, mas ainda envolvem modelos de recuperação de informação clássicos, como o modelo de espaço vetorial. Bonino *et al.* (2004) atentam-se sobre a navegação na ontologia para o refinamento de consultas. Contudo, argumentamos que são necessários processos de refinamento e expansão de pesquisa (*query*) que vão além da navegação em ontologias que consideram apenas simples relações de generalização e especialização, conforme proposto por Bonino *et al.* (2004).

Para melhorar a experiência de busca dos usuários, Bonino *et al.* (2004) alegam que uma grande melhoria na relevância dos resultados poderia ser alcançada sabendo-se “exatamente” o que o usuário quer dizer quando especifica um termo de busca, e tendo a descrição do conteúdo da informação. Propostas como de Guha *et al.* (2003), propõem que os usuários escolham a denotação de um termo ambíguo, e exploram o histórico de consultas do usuário para resolver problemas de ambigüidade. No contexto de RSI, devido à inexperiência dos usuários, tal abordagem não é factível, pois demanda mais habilidades de interação. Lei *et al.* (2006) argumentam que a solução de busca proposta

¹¹ www.google.com

por eles interpreta a consulta do usuário, achando o significado semântico explícito das palavras-chave, mas não esclarecem como e também não citam situações de polissemia, nas quais se torna difícil encontrar o significado correto dos termos. Sobre a ordenação dos resultados de busca, a solução de Lei *et al.* (2006) é feita conforme o grau de satisfação do usuário pelos resultados, mas também em um contexto de polissemia, não explicam como garantir que os resultados satisfaçam o usuário.

A proposta de Choudhari *et al.* (2008) não utiliza nem menciona o uso de ontologias para efetuar a busca semântica. Os autores não tratam em nenhum momento de questões de polissemia, que é um grande e tradicional problema em busca, ou articulam idéias para soluções mais sofisticadas de busca semântica. O objetivo do trabalho seria desenvolver uma aplicação (*plug in*) para o *FaceBook* e testar o sistema nesta SNS. O trabalho está focado na busca semântica de pessoas (usuários da rede). Até onde é de nosso conhecimento, esta proposta é a única atualmente a lidar especificamente com busca semântica em SNS.

Em um contexto de RSI, ao analisar o trabalho de Mangold (2007), observa-se que em cada uma das categorias descritas tem-se a necessidade de se repensar as soluções, devido a peculiaridades do contexto. Este trabalho é um importante ponto de partida sobre as decisões arquiteturais para o desenvolvimento de uma busca semântica voltada a redes sociais utilizando ontologias. Os trabalhos descritos e analisados em sua pesquisa são importantes, pois apresentam diversas arquiteturas e soluções para a recuperação semântica de documentos. No entanto, é necessário, ainda, desenvolver uma estratégia arquitetural que melhor se adéque às necessidades e restrições do contexto de RSI, pois nenhum destes trabalhos endereça busca semântica para redes sociais. O trabalho de Mangold (2007) também faz considerações interessantes sobre a categoria de transparência. Esta categoria refere-se às capacidades semânticas do sistema serem transparentes ou não para o usuário em sua interação com a aplicação. Segundo Mangold (2007) a solução de busca semanticamente mais eficaz proporciona tanto transparência para os usuários inexperientes como comportamento interativo para usuários mais experientes. Lei *et al.* (2006) prospectam um mecanismo de busca semântico dito o mais simples possível para o usuário. Para isso escolheram busca baseada em palavras-chave ao invés de pergunta e resposta em linguagem natural. A abordagem de Amaral *et al.* (2004) usa sistema de perguntas e respostas, que pode não ser o mais apropriado para o contexto de RSI.

Além deste aspecto, das propostas investigadas, poucas se preocupam e investigam a questão de ordenação dos resultados de busca de forma mais apropriada no sentido semântico. Diversas melhorias, com soluções mais adequadas ainda são necessárias em contextos de SNS, endereçando os aspectos semânticos de maneira melhor elaborada, tratando problemas de representação e polissemia.

4. Recomendações

As soluções de busca voltadas para SNS atualmente são principalmente direcionadas para busca de pessoas com base em seus perfis. O único trabalho relativo à busca semântica em SNS, também está voltado a este mesmo fim. Examinamos a necessidade do desenvolvimento de soluções de buscas mais adequadas a SNS que possam endereçar outros tipos de conteúdo produzidos pelos usuários da rede. Adicionalmente, não foram

encontrados na literatura até o momento pesquisas que tenham foco especificamente em busca semântica em redes sociais considerando aspectos de acessibilidade e inclusão. O desenvolvimento de um mecanismo de busca em uma RSI deveria comportar estes novos desafios.

As recomendações que propomos podem ser aplicadas para SNS em geral, mas são especialmente descritas com um forte apelo voltado para busca em RSI. Resultados práticos (Reis *et al.*, 2010) com usuários reais de uma RSI, mostram que a linguagem coloquial dos usuários deve ser considerada durante o desenvolvimento de mecanismos de busca em RSI. As pessoas em uma rede social podem criar o seu próprio vocabulário, compartilhando significados em uma comunidade. Os resultados nos mostraram que é necessário construir modelos computacionalmente tratáveis do ponto de vista semântico para lidar com estes aspectos. Conseqüentemente, é argumentado que um sistema de rede social necessita de um mecanismo de busca semântico, pois as dificuldades dos usuários podem ser minimizadas com o uso deste. Também foi observada a necessidade de projetar um mecanismo que retorne resultados personalizados ao usuário. Para isso é preciso uma arquitetura de solução que inclua componentes de: indexação, expansão de consulta e ordenação dos resultados especialmente pensados para este contexto.

Devido à natureza dos sistemas de redes sociais, é essencial levar em consideração a cultura da rede, incluindo aspectos de linguagem e aplicação dos significados utilizados. Estes aspectos vão muito além de apenas considerar a distância social entre as pessoas. Os usuários na maioria das vezes utilizam termos não cultos ao se expressarem através do sistema, *e.g.*: podem utilizar o termo “postinho” no lugar de “Unidade Básica de Saúde”, pois é o que faz sentido a eles, contudo na realidade, semanticamente significam a mesma coisa. Portanto, ao tentarem recuperar a informação, estes fatores devem ser levados em consideração pelo sistema computacional. O mecanismo de busca deve considerar os significados criados e utilizados na rede social, podendo assim propiciar buscas mais precisas. Isto inclui dar ênfase a linguagem local (regionalismo) e cotidiana das pessoas que utilizam a rede, e ao mesmo tempo ligar a linguagem informal a resultados descritos na linguagem culta, dando a oportunidade aos usuários de terem acesso a novos conteúdos e aprender com eles. Para isso, precisamos de técnicas e ferramentas que permitam efetuar extração e mineração dos dados do sistema para tentar capturar a semântica compartilhada.

Além da captura, é imprescindível uma modelagem adequada da semântica da linguagem utilizada na rede. Neste contexto, carecemos de métodos que modelem os significados utilizados na rede com base em aspectos sociais e cognitivos. Para isso, novas abordagens para a representação devem ser investigadas e propostas com o intuito de prover melhor capacidade semântica ao mecanismo de busca. Salientamos a importância do modelo semântico utilizado de base para a busca semântica, como um dos fatores mais decisivos para uma busca semântica mais efetiva; *i.e.*, para que uma busca seja de qualidade e propicie melhores resultados, não podemos negligenciar o processo de desenvolvimento do modelo semântico (modelagem), e sua estratégia de utilização. Este modelo, que pode ser uma ontologia, deve ser bem projetado e desenvolvido. A estrutura de uma ontologia é um critério importante, uma vez que caracteriza a flexibilidade do mecanismo de busca. Em termos tecnológicos

recomendamos o uso da *Web Ontology Language* (OWL)¹² para a descrição formal das ontologias, principalmente por ser um padrão da *World Wide Web Consortium* (W3C)¹³, possuindo suporte de *software* editores, e.g. *Protégé*¹⁴.

No processo de busca, vislumbramos a necessidade de levar em consideração o papel do indivíduo na rede (informações do perfil) como uma maneira de personalizar a busca. No contexto de SNS, é particularmente crítica a personalização do processo de busca. Segundo Fang *et al.* (2005) para melhorar o desempenho de um mecanismo, palavras-chave do usuário devem ser devidamente interpretadas de acordo com seus significados em um domínio específico. Argumentamos que além do domínio específico, em uma rede social, o perfil dos usuários tem papel fundamental para encontrar o significado dos termos de busca e do conteúdo. Logo, uma recomendação importante é tentar utilizar o perfil do usuário como uma estratégia para o mecanismo de busca, designando os significados dos termos de acordo com usuários específicos, *i.e.*, considerar o significado dos termos (busca e conteúdo) conforme quem produz e consome a informação na rede social. Ao utilizar no processo de busca o perfil junto aos padrões de comportamento do usuário, poderemos tratar melhor aspectos de polissemia e sinônimos utilizados na rede.

Considerando que a maioria dos trabalhos para SNS focam na busca de pessoas da rede, e a maior parte dos trabalhos sobre busca semântica é para o contexto distribuído da *Web*, precisamos definitivamente de uma abordagem de busca semântica voltada para SNS incluindo a busca de conteúdo gerado nas mesmas, além das pessoas da rede. A Tabela 1 faz uma síntese das principais recomendações propostas para esse fim.

Tabela 1. Síntese das principais recomendações propostas

| Funcionalidade | Recomendações |
|--|--|
| <i>Estratégia de busca</i> | -Semântica |
| <i>Conhecimento do domínio</i> | -Extraído e baseado em atividades na rede social |
| <i>Modelagem</i> | -Feita através de ontologias, que devem ser desenvolvidas por meio de métodos e processos específicos fundamentados em aspectos humanos |
| <i>Tecnologia de representação computacional</i> | -OWL |
| <i>Processo de busca</i> | -Personalizado: leva em consideração o perfil do indivíduo na rede pela busca - Considera o significado dos termos conforme quem produz e consome a informação - Trata melhor aspectos de polissemia e sinônimos utilizados na rede social - Liga termos informais e coloquiais a sinônimos na linguagem culta a fim de propiciar o acesso a informação e oportunidades de aprendizagem |
| <i>Foco da Busca</i> | -No conteúdo produzido na rede, além de pessoas |
| <i>Entrada de dados</i> | -Interface simples através de palavras-chave |
| <i>Ordenação dos resultados</i> | -Apropriada ao contexto de RSI e baseada nos aspectos semânticos modelados |

¹² <http://www.w3.org/TR/owl-features>

¹³ <http://www.w3.org/2001/sw/Activity>

¹⁴ <http://protege.stanford.edu/>

Além dos aspectos de *back-end* do mecanismo, recomendamos uma solução minimalista na camada de interface para a interação com o mecanismo de busca semântico. Soluções de interface que propõem navegação em ontologias e modelos de interação complexos estão fora da realidade da maioria dos usuários que objetivamos incluir. É recomendável a utilização de uma solução que faça uso de palavras-chave para a busca. Também na camada de interface, com relação às estratégias de ordenação dos resultados de busca, são necessárias soluções apropriadas ao contexto de RSI, que sejam fundamentadas nos aspectos semânticos relativos à pessoa que efetua a busca e melhor organizadas em pacotes semanticamente relacionados, *e.g.* um pacote de resultado específico que agrupe apenas conteúdos que contenham sinônimos do termo buscado. Estas recomendações podem gerar oportunidades de desenvolvimento de buscas mais adequadas e naturais aos usuários de uma RSI, uma vez que cobrem os aspectos essenciais em um mecanismo de busca semântica, incluindo: a modelagem, estratégia de busca, interface além da ordenação dos resultados.

5. Conclusão

Nos dias atuais o mundo contemporâneo é caracterizado por diversos problemas sociais e ambientais. Soluções tecnológicas que cooperem na minimização destes incluem o acesso e uso de maneira facilitada à informação contida na mídia digital, podendo ser uma oportunidade de inclusão sócio-digital e de formação da consciência ambiental das pessoas. Para isso necessitamos desenvolver aplicações que não introduzam barreiras para o acesso, que contribuam para a formação de uma cultura digital respeitando a diversidade. As Redes Sociais Inclusivas na *Web* (RSI) têm sido desenvolvidas com este foco e configuram-se como micro sistemas sócio-culturais complexos (Desafio 2 da SBC). Desenvolver mecanismos de busca mais adequados a estes sistemas pode definitivamente ajudar no acesso participativo e universal do cidadão brasileiro ao conhecimento (Desafio 4 da SBC).

Este artigo contribuiu discutindo propostas da literatura para busca semântica, elucidando os desafios atuais relacionados aos mecanismos de busca e propondo recomendações com o objetivo de prospectar uma solução mais apropriada a RSI. Fundamentados na análise e discussão desenvolvida, concluímos que uma RSI necessita de um mecanismo de busca de cunho semântico especialmente projetado a ela. As soluções na literatura contribuem, mas ainda são insuficientes para tal objetivo. Apresentar os desafios neste contexto junto às recomendações propostas representa um passo inicial rumo a uma solução. Em continuidade a este trabalho temos investigado técnicas para capturar a semântica utilizada na RSI, assim como metodologias apropriadas para a sua modelagem. Trabalhos futuros também envolvem uma implementação piloto do mecanismo de busca em uma RSI (em produção) baseado nas recomendações apresentadas.

Agradecimentos

Este trabalho é financiado pela Microsoft *Research* – FAPESP Instituto para Pesquisas em Tecnologia da Informação (processo nro. 2007/54564-1) e também pelo CNPq/CTI (680.041/2006-0).

Referências

- Amaral, C.; Laurent, D.; Martins, A.; Mendes, A.; Pinto, C. (2004): Design and implementation of a semantic search engine for Portuguese. In Proc. of 4th International Conference on Language Resources and Evaluation, LREC 2004, Lisbon, Portugal, Vol. I, pp. 247–250.
- Baranauskas, M.C.C. & Souza, C.S. (2006): Desafio 4 Acesso Participativo e Universal do Cidadão Brasileiro ao Conhecimento. Computação Brasil, ano VII, núm. 23, p.7
- Baranauskas, M.C.C. (2007): e-Cidadania: Systems and Methods for the Constitution of a Culture mediated by Information and Communication Technology. Research Proposal for the Microsoft Research-FAPESP Institute.
- Berkan, C. R. (2007): Semantic Search: An Antidote for Poor Relevancy. <http://www.readwriteweb.com/archives/semantic_search_antidote_for_poor_relevancy.php> Acessado em Fevereiro de 2009.
- Berners-Lee, T.; Hendler, J.; Lassila, O. (2001): The Semantic Web, Scientific American, May 2001 issue.
- Bonino, D.; Corno, F.; Farinetti, L.; Bosca, A. (2004): Ontology Driven Semantic Search. WSEAS Transaction on Information Science and Application, Issue 6, Vol. 1, pp. 1597-1605.
- Boyd, D. M. & Ellison, N. B. (2008): Social Network Sites: Definition, History, and Scholarship. Journal of Computer-Mediated Communication. Num. 13, pp. 210-230.
- Celino, I; Valle, E, D; Cerzza, D; Turati, A. (2006): Squiggle: A semantic search engine for indexing and retrieval of multimedia content. Proc. of SAMT 2006, pp. 20–34.
- Choudhari, Ajay; Jain, Mohit; Sinharoy, Avik; Zhang, Min. (2008): Semantic Search in Social Networks. <http://www.cc.gatech.edu/projects/disl/courses/8803/2008/project/project_deliverables/group22/proposal.pdf>. Acessado em Janeiro de 2010.
- Fang, Wei-Dong; Zhang, Ling; Wang, Yan-Xuan; Dong, Shou-Bin. (2005): Toward a Semantic Search Engine based on Ontologies. In Proc. of the 4th International Conference on Machine Learning and Cybernetics. pp. 1913-1918.
- Gruber, T. R. (1993): A translation approach to portable ontologies. Knowledge Acquisition, Vol. 5, Num. 2, pp. 199-220.
- Guelfi, Nicolas; Pruski, Cédric; Reynaud, Chantal (2010): Experimental Assessment of the TARGET Adaptive Ontology-based Web Search Framework. In Workshop Web2Touch in conjunction with 10th International Conference on New Technologies of Distributed Systems. Tozeur – Tunisia (to be published)
- Guha, R.; McCool, R.; Miller, E. (2003): Semantic Search. Proceedings of the 12th international conference on World Wide Web. pp. 700-709. Budapest, Hungary
- Gürsel, Anil & Sen, Sandip. (2009): Improving search in social networks by agent based mining. In Proc. of the 21st international joint conference on Artificial intelligence table of contents. Pasadena, California, USA. pp. 2034-2039

- Haynes, J. & Perisic, I. (2009): Mapping Search Relevance to Social Networks. International Conference on Knowledge Discovery and Data Mining. In Proc. of the 3rd Workshop on Social Network Mining and Analysis. Paris, France. Article N° 2.
- Heflin, J. & Hendler, J. (2000): Searching the web with SHOE. Artificial Intelligence for Web Search. In AAAI Workshop, AAAI Press, Menlo Park, CA, pp. 35–40.
- Kassim, J. M. & Rahmany, M. (2009): Introduction to Semantic Search Engine. In International Conference on Electrical Engineering and Informatics. Selangor, Malaysia. Vol. 02, pp. 380-386.
- Lei, Y.; Uren, V.S.; Motta, E. (2006): SemSearch: a search engine for the Semantic Web. In Proc. of EKAW 2006, Managing Knowledge in a World of Networks. POdebrady, Czech Republic. pp. 238-245
- Mangold, C. (2007): A survey and classification of semantic search approaches. Int. J. Metadata, Semantics and Ontology, Vol. 2, No. 1, pp.23–34
- Mauldin, M. L. (1991): Conceptual Information Retrieval: A Case Study in Adaptive Partial Parsing. Kluwer Academic Publishers Norwell, MA, USA.
- Neris, V. P. A.; Almeida, L. D. ; Miranda, L. C. ; Hayashi, E. ; Baranauskas, M. C. C. (2009): Towards a Socially-constructed Meaning for Inclusive Social Network Systems. In: International Conference on Informatics and Semiotics in Organisations. Beijing. pp. 247-254.
- Nielsen (2009): Global Faces and Networked Places. The Nielsen Company USA. <http://blog.nielsen.com/nielsenwire/wp-content/uploads/2009/03/nielsen_globalfaces_mar09.pdf> Acessado em Julho de 2009.
- Ramachandran, Divya; Kam, Matthew; Chiu, Jane; Canny, John; Franke, James L. (2007): Social Dynamics of Early Stage Co-Design in Developing Regions. In Proc. of the SIGCHI Conference on Human Factors in Computing Systems. California, USA. pp. 1087-1096.
- Reis, J.C.; Bonacin, R.; Baranauskas, M. C. C. (2010): Analysing Search in Inclusive Social Network Services. IC-10-02, www.ic.unicamp.br/~reltech
- Studer, R.; Benjamins, V. R; Fensel, D. (1998): Knowledge engineering: principles and methods. Data & Knowledge Engineering, v.25, n.1/2, pp. 161-197.
- Vieira, M. V.; Fonseca, B. M.; Damazio, R.; Golgher, P. B.; Reis, D. C.; Neto, B.R. (2007): Efficient Search Ranking in Social Networks. In Proc. of the 16th ACM Conference on information and knowledge management. Lisbon, Portugal. pp. 563-572.
- Wei, W.; Barnaghi, P. M.; Bargiela, A. (2007): The Anatomy and Design of A Semantic Search Engine, Tech. rep., School of Computer Science, University of Nottingham Malaysia Campus.
- Yu, Bin & Singh, Munindar P. (2003): Searching social networks. In Proc. of the 2th International joint Conference on Autonomous agents and multiagent systems. New York, NY, USA. ACM Press. pp. 65–72