UNIVERSIDADE ESTADUAL DE CAMPINAS
Instituto de Computação

Busca Informada por Abordagem Semiótica em Redes Sociais Inclusivas *Online*

Este exemplar corresponde à redação final da Dissertação devidamente corrigida e defendida por Julio Cesar dos Reis e aprovada pela Banca Examinadora.

Campinas, 15 de Abril de 2011

Prof.ª Dr.ª Maria Cecília Calani Baranauskas
(Orientadora)

Dr. Rodrigo Bonacin
(Co-orientador)

Dissertação apresentada ao Instituto de Computação, UNICAMP, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação.

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO IMECC DA UNICAMP
Bibliotecária: Maria Fabiana Bezerra Müller – CRB8/6162

Reis, Julio Cesar dos

R277b Busca informada por abordagem semiótica em redes sociais inclusivas online / Julio Cesar dos Reis-- Campinas, [S.P. : s.n.], 2011.

Orientador : Maria Cecília Calani Baranauskas; Rodrigo Bonacin.

Dissertação (mestrado) - Universidade Estadual de Campinas, Instituto de Computação.

1.Redes sociais online. 2.Web semântica. 3.Ontologia. 4.Semiótica organizacional. 5.Interação humano-computador. I. Baranauskas, Maria Cecília Calani. II.Bonacin, Rodrigo. III. Universidade Estadual de Campinas. Instituto de Computação. IV. Título.

Título em inglês: Search informed by a semiotic approach in inclusive social network services

Palavras-chave em inglês (Keywords): 1.Online social networks. 2.Semantic Web. 3.Ontology. 4.Organizational semiotics. 5.Human-Computer interaction.

Área de concentração: Metodologias e Técnicas para a Computação

Titulação: Mestre em Ciência da Computação

Banca examinadora: Profa. Dra. Maria Cecília Calani Baranauskas (IC – UNICAMP)
Dra. Juliana Salles (MICROSOFT RESEARCH)
Prof. Dr. André Santanché (IC - UNICAMP)

Data da defesa: 30/03/2011

Programa de Pós-Graduação: Mestrado em Ciência da Computação

TERMO DE APROVAÇÃO

Dissertação Defendida e Aprovada em 30 de março de 2011, pela Banca examinadora composta pelos Professores Doutores:

Drª. Juliana Pereira Salles
Microsoft Research

Prof. Dr. André Santanchè
IC / UNICAMP

Profª. Drª. Maria Cecília Calani Baranauskas
IC / UNICAMP

Instituto de Computação
Universidade Estadual de Campinas

Busca Informada por Abordagem Semiótica em Redes Sociais Inclusivas *Online*

Julio Cesar dos Reis
30 de Março de 2011

Banca Examinadora:

- Prof.ª Dr.ª Maria Cecília Calani Baranauskas (Orientadora)
  Instituto de Computação, Universidade Estadual de Campinas (IC/UNICAMP)
- Dr.ª Juliana Salles (Titular)
  Microsoft Research
- Prof. Dr. André Santanché (Titular)
  Instituto de Computação, Universidade Estadual de Campinas (IC/UNICAMP)
- Prof.ª Dr.ª Ariadne Maria Brito Rizzoni Carvalho (Suplente)
  Instituto de Computação, Universidade Estadual de Campinas (IC/UNICAMP)
- Dr.ª Olga Nabuco (Suplente)
  Centro de Tecnologia da Informação Renato Archer (CTI)

Resumo

As Redes Sociais *Online* (RSOs) podem representar uma oportunidade para as pessoas não letradas digitalmente efetivamente familiarizarem-se com tecnologias da informação e comunicação. Para isso, esses sistemas deveriam prover acesso inclusivo, criando situações nas quais a diversidade cultural dos usuários fosse respeitada, e suas dificuldades de acesso minimizadas. RSOs com recursos que promovam acesso a todos, incluindo aqueles às margens da cultura digital, podem ser definidas como Redes Sociais Inclusivas (RSIs). Nestas, as pessoas devem ter a possibilidade de facilmente recuperarem informações corretas, e principalmente que façam sentido a elas.

Nesse contexto, o objetivo deste trabalho é investigar a concepção de mecanismos de busca inclusivos, mais adequados ao conceito de RSI, que possam propiciar resultados de busca mais significativos e personalizados. Embora o uso de mecanismos de busca seja uma das principais alternativas para acessar informação na *Web* atualmente, esses mecanismos ainda são construídos principalmente através de processamento léxico-sintático da informação, resultando em barreiras para muitos usuários. O uso de técnicas de busca semântica impulsionado por tecnologias da *Web* Semântica (WS) apresenta novas possibilidades para o problema em destaque, mas estas ainda são fundamentadas em paradigmas e métodos limitados para a representação do conhecimento em redes sociais.

Para desenvolver um mecanismo de busca inclusivo em RSI é preciso "interpretar" a semântica compartilhada pelas pessoas. Nesta direção é necessário utilizar abordagens e métodos que permitam capturar os aspectos sociais e culturais advindos da rede social, como a linguagem coloquial e as relações semânticas usadas. Desta maneira, o mecanismo de busca inclusivo proposto adota uma solução baseada em ontologias, que são construídas com base em uma nova abordagem para o *design* de ontologias *Web*. Esta abordagem é fundamentada em conceitos e métodos da Semiótica Organizacional (SO), combinados com tecnologias da WS. A concepção desta abordagem foi inspirada em resultados de um experimento com usuários reais sobre atividades de busca em um sistema de RSI. A solução proposta pode possibilitar novas estratégias de busca baseadas em ontologia, e conseqüentemente, gerar a possibilidade de promover o acesso participativo e universal ao conhecimento.

Abstract

Social Network Services (SNSs) may represent an opportunity for not digitally literate people to effectively familiarize themselves with information and communication technologies. For that, these systems should provide inclusive access, creating situations where users' cultural diversity is respected and their difficulties minimized. SNSs with resources that promote access for all, including those in the margin of the digital culture can be defined as Inclusive Social Network Services (ISNs). In these networks, ordinary people must have the possibility of easily recovering correct information, which makes sense to them.

In this context, the objective of this work is to investigate the conception of an inclusive search mechanism more adequate to the ISN concept, which may propitiate more meaningful and personalized search results. Although the use of search mechanisms has been the main alternative to access information in the Web, such mechanisms are still built mainly through lexic-syntactical processing of information, resulting in barriers for many users. The use of semantic search techniques driven by Semantic Web (SW) technologies presents new possibilities for the highlighted problem, but it is still grounded in limited paradigms and methods for the representation of the social knowledge.

In order to develop an inclusive search mechanism in ISN it is necessary to "interpret" the meanings shared by people. Approaches and methods that allow capturing the social and cultural aspects from the social network, such as their colloquial language and semantic relations, are necessary for that. Therefore, the inclusive search mechanism proposed in this work adopts an ontology-based solution that is based on a novel approach for the design of Web ontologies. Such approach is grounded on Organizational Semiotics (OS) concepts and methods, combined with SW technologies. The conception of this solution was inspired in experimental results with real users on search activities in an ISN system. This may enable novel ontology-based search strategies, and consequently generates the possibility of promoting the participatory and universal access to knowledge.

Agradecimentos

Há muitas pessoas a agradecer que contribuíram diretamente e indiretamente a esta pesquisa, e na qual desenvolvi relações muito especiais, desde profissionais, e, sobretudo de amizade. A existência deste trabalho depende ontologicamente destas contribuições.

Primeiramente, gostaria de agradecer a Prof.ª Dr.ª M. Cecília C. Baranauskas, que de maneira competente me orientou, dando autonomia, e ao mesmo tempo críticas e sugestões valiosas; sua dedicação e empenho são únicos. Agradeço exclusivamente também ao Dr. Rodrigo Bonacin pelos seus ensinamentos, atenção, incentivo, paciência e a confiança que sempre depositou em meu trabalho. A orientação de vocês, em diferentes dimensões, neste período foi valiosa para o meu crescimento como investigador e ter a oportunidade de trabalhar, estar e aprender com vocês é uma imensa satisfação.

Agradeço não menos, à minha família pelo incentivo e suporte incondicional; a minha querida companheira e noiva Laira, pelo amor, carinho e compreensão em todos os momentos; e também aos amigos de longa data pelo tempo deixado de ser compartilhado devido ao trabalho. Os amigos especiais do IC/UNICAMP e do Grupo de Pesquisa em Interação Humano-Artefato Digital (InterHAD) da UNICAMP, do e-Cidadania, do Centro de Tecnologia da Informação Renato Archer (CTI), e do Núcleo de Informática Aplicada à Educação (NIED), dentre eles: Heiko, Leonelo, Roberto, Elaine, Vânia, Leonardo, Vagner, Fred, Lara, Ciça, Tânia, Olga, Marcos, Gonzaga, Adriana, Aqueo, Walcir e outros não menos especiais; muito obrigado pela ajuda, sugestões, dicas, críticas e o suporte nos diversos momentos.

Agradeço o comprometimento e a colaboração das pessoas da Vila União, e do Telecentro Vila Monte Alegre, de Pedreira, e a todos os participantes dos experimentos. Os revisores anônimos dos artigos, que deram sugestões substâncias para a melhoria do trabalho ao longo da pesquisa, e também aos colegas de estudo da sala 80 do IC/UNICAMP, nas árduas horas de estudos, inclusive aos domingos à noite, para as provas das disciplinas de arquitetura de computadores, e de complexidade de algoritmos. Por fim, agradeço aos membros da Banca Examinadora por aceitarem avaliar este trabalho.

"*O Acesso é poder e o poder é informação!*"

Fernando Anitelli

Sumário

Lista de Tabelas

Lista de Figuras

Lista de Abreviaturas e Siglas

CTI	Centro de Tecnologia da Informação Renato Archer
DO	Diagrama de Ontologia
HCI	*Human-Computer Interaction*
IBGE	Instituto Brasileiro de Geografia e Estatística
IC	Instituto de Computação
INAF	Indicador Nacional de Alfabetismo Funcional
InterHAD	Grupo de Pesquisa em Interação Humano-Artefato Digital
IC	Instituto de Computação
IPEA	Instituto de Pesquisa Econômica Aplicada
IPM	Instituto Paulo Montenegro
ISN	*Inclusive Social Network*
MEASUR	*Methods for Eliciting, Analyzing and Specifying Users' Requirements*
NIED	Núcleo de Informática Aplicada a Educação
OC	*Ontology Chart*
OS	*Organizacional Semiotics*
OWL	*Web Ontology Language*
PNAD	Pesquisa Nacional por Amostra de Domicílios
RDF	*Resource Description Framework*
RDFS	*Resource Description Framework Schema*
RSO	Rede Social *Online*
RSI	Rede Social Inclusiva
SAM	*Semantic Analysis Method*
SBC	Sociedade Brasileira de Computação
SNS	*Social Network Service*
SO	Semiótica Organizacional
SW	*Semantic Web*
SWRL	*Semantic Web Rule Language*
TIC	Tecnologia de Informação e Comunicação
UML	*Unified Modeling Language*

UNICAMP Universidade Estadual de Campinas
W3C *World Wide Web Consortium*
XML *eXtensible Markup Language*
WS *Web* Semântica

Capítulo 1

Introdução

Ao longo dos últimos anos, a *Web* tem gerado oportunidades para as pessoas estarem constantemente conectadas e propensas a criar e a disponibilizar novos conteúdos digitais. Recentemente, com as chamadas Redes Sociais *Online* (RSOs) [9], pessoas tem se organizado em Comunidades *Online,* compartilhando e gerando ainda mais informações, agora de maneira menos isolada, mais articulada e potencialmente significativa para a rede.

Tão importante quanto gerar novos conteúdos é propiciar o acesso irrestrito e facilitado a todas as pessoas, sejam elas letradas (digitalmente) ou não, e independente de suas dificuldades, ou limitações físicas ou cognitivas. No entanto, devido à enorme sobrecarga de informação na qual usuários estão expostos no ambiente *Web*, adicionado à falta de habilidade que muitos possam ter com as Tecnologias da Informação e Comunicação (TICs), novos mecanismos e abordagens mais adequadas para a procura de informação tornam-se ainda mais urgentes e necessários. Estes são altamente importantes para se desenvolverem meios propícios para as pessoas encontrarem conteúdos relevantes e que façam sentido a elas.

Propostas e soluções para a busca de informação na *Web* (*e.g.* mecanismos de busca) têm sido largamente desenvolvidas e investigadas nos últimos anos. Todavia, no geral, o que se percebe nessas é que esses mecanismos ainda são profundamente e estritamente fundamentados em informações não relacionadas com a realidade semântica das pessoas organizadas em rede, uma vez que não consideram as diferentes possibilidades de significados e interpretações que os usuários daquele contexto podem ter, ou desenvolver. Nesse contexto, o objetivo deste trabalho é investigar como desenvolver mecanismos de busca inclusivos, *i.e.,* mais adequados à linguagem das pessoas

organizadas por meio de RSO. Um dos caminhos que se apresenta é propor um mecanismo de busca que seja capaz de melhor considerar a semântica compartilhada pelas pessoas em Redes Sociais Inclusivas *Online* (RSIs) [45], sendo fundamentado em métodos que possam gerar melhores representações da semântica do contexto sócio-cultural da rede. Denominamos RSIs às RSOs com recursos que promovam o acesso de todos, incluindo aqueles à margem da cultura digital. Em RSI deve-se criar situações nas quais as dificuldades dos usuários sejam vencidas durante o acesso dos conteúdos digitais.

Esta dissertação, apresentada na forma de uma coletânea de artigos, reconstrói o caminho tomado desde os fundamentos teóricos e metodológicos, a prospecção, desenvolvimento e a avaliação de um mecanismo de busca mais adequado para RSI. Em especial, este primeiro capítulo apresenta de forma sintética na seção 1.1 o contexto, a motivação e a problemática tratados nesta dissertação. Já na seção 1.2 são especificados os objetivos e a metodologia de pesquisa utilizada, e na seção 1.3 são apresentados os capítulos que compõem esta dissertação.

1.1 Contexto, Motivação e Problemática

O Brasil, assim como outros países em desenvolvimento, está repleto de cenários marcados pela diversidade cultural, e também por problemas sociais e de inclusão digital. A diversidade de habilidades e competências, assim como a realidade social do público-alvo desta pesquisa pode ser ilustrada por diversos dados estatísticos como, por exemplo, o último Indicador Nacional de Analfabetismo Funcional (INAF) [28], divulgado pelo Instituto Paulo Montenegro (IPM) em 2009, que mostra que 73% dos brasileiros não são considerados alfabetizados plenos.

Nesse cenário, as RSOs podem ter um papel importante na adoção e uso das TICs [52], e através delas potencialmente promover o cidadão. Por meio de RSOs pessoas interagem criando conexões e constituindo comunidades; essas pessoas produzem informações, comunicam-se, colaboram e também compartilham objetivos, interesses, atividades e práticas. Logo, RSOs emergem como um espaço virtual propício para os indivíduos compartilharem experiências e conhecimento [32].

Adicionalmente, esses sistemas também podem significar uma oportunidade para a inclusão sócio-digital por intermédio de RSIs, promovendo o acesso ao conhecimento, que é uma condição básica para a vida na era digital. Esse acesso, quase necessariamente, passa pela busca de informação no sistema. Logo, mecanismos de busca são essenciais para a interação e a recuperação da informação com sucesso por pessoas em fase de letramento digital.

De acordo com o Desafio 4: "Acesso Participativo e Universal do Cidadão Brasileiro ao Conhecimento" [2], proposto pela Sociedade Brasileira de Computação (SBC)[1], promover o acesso participativo e universal ao conhecimento para todas as pessoas, de forma igualitária, por meio de sistemas computacionais, é um dos grandes desafios sócio-técnicos e científicos da sociedade contemporânea. Para isso, é necessário respeitar a diversidade cultural individual, e ao mesmo tempo, desenvolver o conceito de inclusão e de consciência de grupo.

Nesse contexto, sistemas de RSI propõem soluções universais que tendem a minimizar as barreiras de acesso, concomitantemente respeitando a diversidade dos indivíduos. Nesses sistemas, cada funcionalidade é desenhada considerando as diferenças de competência e limitações das pessoas, para que as dificuldades dos usuários possam ser vencidas durante o acesso aos conteúdos digitais. Assim, há a necessidade urgente de mecanismos de busca nesses sistemas especialmente projetados para facilitar o acesso à informação de forma inclusiva, *i.e.,* de maneira mais significativa para todos os indivíduos participantes da rede. A melhoria deste recurso pode dar-se principalmente pela facilidade e naturalidade em termos lingüísticos que o usuário terá para procurar uma informação, e também na qualidade de resposta do sistema (*e.g.* organização e ordenação dos resultados de busca).

Dessa maneira, o desafio está principalmente em como propiciar uma experiência harmoniosa entre o mecanismo e o usuário, em termos de sua linguagem cotidiana. Mecanismos de busca projetados para RSIs deveriam considerar os aspectos de inclusão e diversidade, onde não se podem pressupor usuários familiarizados com os procedimentos e algoritmos de busca da *Web*. Portanto, mostra-se essencial o desenvolvimento de mecanismos focados nos princípios de inclusão e especialmente voltados a este fim. Atualmente, mesmo os mecanismos descritos como semânticos, são baseados em representações computacionais desconectadas da realidade semântica dos usuários, uma vez que fazem uso de terminologias e vocabulários padrões. Neste trabalho o termo semântica é entendido como a interpretação de conceitos por indivíduos ou grupo de indivíduos, *i.e.,* os significados explícitos.

As principais deficiências das soluções tradicionais de busca é que elas, na maioria das vezes, já consideram que as pessoas dominam o uso da tecnologia, e são abordagens sintáticas, *i.e.,* são baseadas no processamento léxico-sintático da informação, e não em informações e representações dos significados. As soluções de busca semântica que consideram os significados muitas vezes necessitam do usuário para apontar o contexto dos significados (desambiguação), o que pode gerar dificuldades para usuários não

[1] Grandes Desafios da Computação no Brasil 2006-2016. São Paulo, 8 e 9 de maio, 2006

letrados digitalmente; essas soluções também, quase sempre, são baseadas no uso de artefatos computacionais (muitas vezes em ontologias computacionais [18]) genéricas e descontextualizadas, *i.e.,* criadas usando abordagens estritamente objetivistas [36] para a modelagem dos significados dos conceitos que compõem a ontologia. Essa abordagem é usualmente adotada na comunidade da *Web* Semântica (WS) [5]. Além disso, as soluções de busca semântica são verificadas em contextos de domínios fechados do conhecimento, *i.e.,* com conteúdos homogêneos, sem informalidades ou diversidade de assuntos, o que geralmente não ocorre no contexto de RSIs, que são constituídas por conteúdos de diversas áreas do conhecimento, sendo independentes de um domínio específico (fechado).

No contexto de RSOs, as soluções de busca quase necessariamente focam em algoritmos para diferentes abordagens de busca de pessoas na rede (*e.g.* considerando relações de amizade), mas não consideram os conteúdos criados e compartilhados pelas pessoas, nem tão pouco os significados usados e compartilhados nos conteúdos disponibilizados. Uma vez que os aspectos semânticos podem fazer a diferença para usuários compartilharem e alcançarem informação, a busca deveria considerar a diversidade de linguagem com seus aspectos semânticos, provendo resultados que façam sentido. Embora exista uma necessidade crescente de soluções que possam lidar com a semântica da informação em sistemas *Web*, e muito se tenha evoluído nesse sentido, a literatura tem apontado deficiências para as abordagens usuais da WS, o que tem demandado estudos mais aprofundados sobre a representação do conhecimento na *Web*.

A problemática descrita envolve desafios de diversas dimensões, tanto do ponto de vista social e prático, quanto do ponto de vista técnico e científico da Ciência da Computação. Dentre os problemas do ponto de vista técnico, diversos estão associados à temática de ontologias, dentre eles: métodos para representação de ontologias, técnicas para construção de ontologias, técnicas de evolução de ontologias, métodos para busca semântica baseada em ontologia, etc. Neste trabalho, além de envolver estes desafios tecnológicos do ponto de vista científico, colocamos foco no desenvolvimento de métodos para a representação e modelagem formal e computável da linguagem coloquial compartilhada em uma rede social.

1.2 Objetivo e Metodologia da Pesquisa

Frente à motivação de encontrar abordagens e soluções mais adequadas para alcançar uma "busca inclusiva", o objetivo principal e foco deste trabalho é investigar a concepção de mecanismos de busca inclusivos em RSI, que possam propiciar resultados de busca mais significativos, sendo personalizados e direcionados ao perfil dos

participantes da rede. Nesta direção, busca-se entender principalmente, como as pessoas (em fase de letramento digital, ou prospectivos usuários de uma RSI) fazem sentido de um mecanismo de busca em uma aplicação na *Web*, e como prospectar um mecanismo de busca que possa prover resultados mais adequados a elas. Assim sendo, algumas questões de pesquisa nortearam esta investigação:

- Como um mecanismo de busca inclusivo deve ser projetado considerando uma população de usuários heterogênea e iletrada digitalmente?
- Quais fundamentos teóricos e metodológicos devem informar a ação do mecanismo? Por que e como?
- Quais características o mecanismo deve conter e expressar para que os resultados de busca sejam mais apropriados e personalizados a cada indivíduo participante da rede?
- Como os resultados de busca podem ser mais bem organizados semanticamente, ordenados e apresentados aos usuários?

Como uma possível abordagem para o problema, deve-se levar em consideração, durante a modelagem semântica, o que as pessoas disponibilizam e compartilham de conteúdo digital na rede, e também as interações entre elas (*e.g.* comentários, bate-papo). Além disso, é importante que o mecanismo considere durante o processo de busca os possíveis significados dos termos, conforme quem produz e consome a informação (*i.e.* os significados para um termo devem ser interpretados conforme a pessoa que o usa).

Para que este tipo de solução seja possível, é preciso representar a semântica, mas não a semântica definida pelos vocabulários padrões, e sim a semântica advinda daquela rede social particular (*i.e.* considerar a semântica que advém das pessoas da rede, modelando o conhecimento a partir dos conteúdos digitais disponíveis na rede social em consideração). Logo, para efetuar esta representação semântica serão necessários dois passos fundamentais: primeiramente, será necessário descobrir os possíveis conceitos utilizados e suas relações semânticas; em seguida, modelar adequadamente estes significados conforme um referencial. Assim, serão essenciais tanto ferramentas que auxiliem na mineração e na descoberta dos conceitos e suas relações, assim como fundamentos que melhor representem a semântica do contexto social (modelagem do conhecimento).

Para tanto, neste trabalho foi explorado o uso do referencial teórico-metodológico base, o Método de Análise Semântica (SAM) [36] advindo da Semiótica Organizacional (SO) [67, 35]. Fundamentados neste método, foram desenvolvidos uma abordagem e um processo para modelar o conhecimento da rede social, cujo resultado final gera uma ontologia *Web* descrita em *Web Ontology Language* (OWL). O produto deste processo é usado para informar o mecanismo de busca inclusivo, e conseqüentemente prover

resultados de busca semânticos mais adequados. Assim sendo, neste trabalho objetivou-se fazer o *design* de ontologias *Web* fundamentado pelo SAM; isto visa atingir ontologias *Web* mais representativas do ponto de vista social e humano.

Como hipótese tem-se que através do uso do SAM no processo de modelagem dos significados, é possível gerar resultados de busca mais adequados para pessoas do contexto em estudo, contribuindo para melhores buscas semânticas em RSI. Adicionalmente, pretende-se verificar se pelo uso do SAM consegue-se derivar relações semânticas e ontológicas que subsidiarão resultados de busca que possam fazer mais sentido em um contexto social. Com isso, também, será possível agrupar resultados de busca relativos a significados específicos, conforme a interpretação de diferentes pessoas da rede.

A metodologia de pesquisa adotada inclui primeiramente uma revisão na literatura de propostas de busca semântica, e também abordagens de buscas para RSOs, confrontando as soluções existentes com a problemática em RSI. Em paralelo, foi desenvolvido um experimento para observar como prospectivos usuários de uma RSI interagem com um mecanismo de busca. Os estudos de caso como um todo conduzidos nesta dissertação, *e.g.* este experimento e outros que incluem a validação desta proposta, foram desenvolvidos no âmbito do projeto e-Cidadania: Sistemas e Métodos na Constituição de uma Cultura mediada por TICs [2], utilizando a RSI *VilanaRede*[2]. Este sistema de RSI foi o principal resultado do e-Cidadania. Este projeto, financiado pelo Instituto Virtual de Pesquisas FAPESP-*Microsoft Research*, visava investigar a constituição da cultura digital a partir da construção conjunta de soluções de interação e de interface, considerando as diferentes competências das pessoas na sociedade (incluindo analfabetos funcionais e pessoas com deficiência). O e-Cidadania propunha soluções para lidar com as barreiras que a maioria dos cidadãos brasileiros encontram para efetivamente utilizar a *Internet*, visando transformar RSO em um motor para a inclusão digital e a cidadania.

No experimento descrito e formalizado em Reis *et al.* [58], mostra-se que apenas busca sintática não seria suficiente para potenciais usuários de uma RSI encontrarem anúncios adequados, e explorar aspectos semânticos poderia fazer a diferença neste contexto. O estudo de caso envolveu a observação de prospectivos usuários de uma RSI perfazendo um conjunto de cenários de busca; o objetivo foi observar como os usuários se comportam e as suas dificuldades em usar o mecanismo. A partir deste resultado empírico desenvolveu-se uma proposta de representação computacional da semântica

[2] www.vilanarede.org.br

baseada no SAM, chamada de "*Semiotic Web Ontology*" (SWO). Fundamentado nesta perspectiva, um método foi proposto para criar SWOs a partir dos conteúdos de RSI.

Como parte da elaboração deste método, foram realizadas duas investigações específicas. Primeiramente, uma investigação sobre possíveis ferramentas e técnicas para auxiliar na construção semi-automática de ontologias no contexto de RSI. Paralelamente, devido à necessidade de transformar diagramas de ontologias gerados pelo SAM em ontologias *Web* descritas em OWL, foram propostas heurísticas e regras de transformação. As regras foram implementadas como uma extensão do *software* SONAR [65].

Adicionalmente, foi conduzido um estudo de caso com base no método proposto e nos resultados da análise das ferramentas de mineração de texto, e das heurísticas de transformação entre os artefatos. Para tanto, foi considerado um contexto real com usuários e infraestrutura da RSI *VilanaRede*. Neste estudo de caso, derivou-se uma ontologia SWO advinda dos conteúdos disponíveis no *VilanaRede*, que deve ser usada pelo mecanismo de busca inclusivo. Finalmente, para validar a abordagem proposta, foi implementado um protótipo do mecanismo de busca inclusivo na RSI *VilanaRede*. A partir deste protótipo, um experimento foi conduzido visando efetuar uma avaliação preliminar da proposta com prospectivos usuários alvo.

Para modelar uma representação da realidade semântica da RSI são necessários métodos que tenham o poder de capturar e representar a semântica do contexto social, construída conforme dados da rede social, e gerar artefatos computacionais processáveis. Tais fatores apontam para o uso da SO e conseqüentemente do SAM como fundamento teórico-metodológico da proposta. Sendo assim, o fato de pensar na modelagem semântica diferentemente em outra perspectiva filosófica da corrente principal de estudo da WS, usando e agregando princípios advindos da SO, junto a conceitos e resultados do SAM, pode resultar em um processo diferenciado de modelagem do conhecimento. Isto, conseqüentemente, apresentou impactos nos resultados de busca semântica, gerando uma abordagem específica para fazer desambiguação de significados.

Adicionalmente, no âmbito desta pesquisa, entende-se que novas maneiras de organizar os resultados de busca semântica podem mudar a forma como os usuários reconhecem a informação. Assim sendo, espera-se que esse trabalho de mestrado contribua nessa direção, por meio da proposição de uma nova forma de modelar o conhecimento de maneira computacionalmente tratável, de modo a permitir resultados de busca mais significativos para o público-alvo.

1.3 Contribuições e Organização da Dissertação

As principais contribuições desta dissertação envolvem:

- 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 *Web* no contexto de RSI, fundamentado em métodos da SO;
- 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 no *software* SONAR para a modelagem de ontologias *Web* baseada nos resultados e artefatos do SAM;
- Desenvolvimento de um protótipo de busca inclusiva, informado por SWO, em RSI;
- Avaliação do mecanismo desenvolvido junto ao público-alvo.

Para tratarmos as questões de pesquisa levantadas e mencionadas, apresentamos uma visão geral dos capítulos desta dissertação que, de certo modo, descrevem em cada um deles resultados parciais e etapas desta pesquisa. Esta dissertação é composta por seis Capítulos, sendo que os Capítulos 2 e 3 apresentam artigos publicados em conferências; o Capítulo 2 foi publicado em um simpósio nacional, enquanto o Capítulo 3 em um congresso internacional. O Capítulo 4 apresenta um artigo convidado para capítulo de livro internacional, enquanto que o Capítulo 5 traz um artigo submetido a um periódico internacional. O Capítulo 6 conclui este trabalho.

Cabe ressaltar, em relação aos artigos publicados, que os textos são apresentados nesta dissertação nos seus estados originais e integrais, exceto pelas alterações na sua estruturação e formatação de modo a adequá-los ao formato de publicação das teses no Instituto de Computação (IC) da Universidade Estadual de Campinas (UNICAMP). É relevante comentar, ainda, que uma vez que esta dissertação de mestrado é apresentada em formato de coletânea de publicações, os capítulos são autocontidos. Entretanto, entendemos que a leitura linear desta dissertação permite ao leitor uma compreensão do caminho de investigação percorrido, das decisões tomadas no decorrer da pesquisa, assim como dos resultados de pesquisa obtidos.

Na seqüência, são apresentadas as sínteses, ressaltando as contribuições dos artigos e como se dá o encadeamento dos trabalhos, assim como seus respectivos locais de publicação:

- **Capítulo 2**: "Busca em Sistemas Sócio-Culturais *Online*: Desafios e Recomendações" (2010). Júlio Cesar dos Reis, M. Cecília C. Baranauskas e Rodrigo Bonacin. Anais do XXX Congresso da Sociedade Brasileira de Computação. XXXVII Seminário Integrado de Software e Hardware (SEMISH). Belo Horizonte, Brasil. pp. 380-394 [56].
  Visão Geral e Contribuições: Neste capítulo efetuamos um levantamento bibliográfico a respeito de estratégias de busca para redes sociais. As duas principais abordagens investigadas foram a busca social e a busca semântica. Estes trabalhos foram discutidos, analisados e confrontados com foco na problemática em RSI, descrevendo as suas potencialidades e limites no contexto tratado. Apresentamos também os desafios técnicos e científicos relacionados à busca em RSI, e suas interseções com os Grandes Desafios da SBC. Apoiado no levantamento bibliográfico desenvolvido, nos desafios identificados, assim como nos resultados do experimento do uso de mecanismo de busca com os usuários reais da RSI *VilanaRede* [58], apontamos recomendações para o projeto (solução) de mecanismos de busca mais adequados ao conceito de RSI.

- **Capítulo 3**: "*A Semiotic-based Approach to the Design of Web Ontologies*" (2010). Júlio Cesar dos Reis, Rodrigo Bonacin e M. Cecília C. Baranauskas. *Proceedings of the 12th International Conference on Informatics and Semiotics in Organisations*. Reading - UK. pp. 60-67 [57].
  Visão Geral e Contribuições: Ainda com base nos resultados do experimento com os usuários da RSI *VilanaRede* [58], assim como a partir das recomendações descritas no capítulo 2, neste capítulo investigamos uma abordagem para o *design* de ontologias *Web* que visa alcançar um melhor poder de representação semântica e, ainda, concomitantemente ser computacionalmente tratável. Para isso, descrevemos o referencial teórico-metodológico que adotamos para desenvolver a abordagem. Com o intuito de justificar a necessidade da proposta, elucidamos, com base na literatura, problemas e limitações de ontologias tradicionais da SW. Então, uma nova abordagem para o *design* de ontologias *Web* fundamentado em aspectos teóricos da SO e técnicos da SW é prospectada. Assim, propomos o conceito de "*Semiotic Web Ontology*", ilustrando e discutindo a nova abordagem de representação semântica.

- **Capítulo 4**: "*Prospecting an Inclusive Search Mechanism for Social Network Services*". Júlio Cesar dos Reis, Rodrigo Bonacin e M. Cecília C. Baranauskas. Artigo publicado como capítulo de livro em "*ICEIS 2010 - Revised Selected Papers*", J.

Filipe *and* J. Cordeiro (Eds.), *Lecture Notes in Business Information Processing* (LNBIP) 73, pp. 555-570. Springer, Heidelberg (2011) [61].
Visão Geral e Contribuições: Com base nos resultados dos capítulos anteriores, em especial nas recomendações, este capítulo apresenta a prospecção de um mecanismo de busca inclusivo. Para tanto, neste capítulo são ilustradas as motivações e justificativas para uma busca inclusiva, descrevendo também a análise dos cenários de busca aplicados no experimento junto a RSI *VilanaRede* e seus resultados, conforme o estudo de caso descrito no relatório [58]. Com base no conceito de SWO, este capítulo mostra como desenvolver uma instancia de SWO para o contexto de RSI. Para isso, propõe um método que a partir dos dados (conteúdo) da RSI chega a uma SWO, descrevendo e exemplificando como modelar ontologias para RSI através dos passos do método; também discute como possivelmente desenvolver cada passo. Logo, para o contexto de RSI intencionamos usar SWO gerada a partir dos dados da rede social, aplicando o método proposto para criar SWO. Após a apresentação do método, é descrito um motor de busca inclusivo, ilustrando com exemplos de aplicação e possibilidades da proposta. Este trabalho foi essencial para apresentar a proposta do método, mostrar a possibilidade de uma busca inclusiva, e conclusivamente apontar a necessidade de transformação de SAM para OWL. Este capítulo descreve também brevemente as possíveis ferramentas e resultados de sua análise que podem auxiliar na construção das ontologias. Os detalhes deste estudo, contendo partes específicas e instancia em um estudo de caso do método são apresentados nos Apêndice A, B e C desta dissertação.

- **Capítulo 5:** "*Addressing Universal Access in Social Networks: An Inclusive Search Mechanism*". Júlio Cesar dos Reis, Rodrigo Bonacin e M. Cecília C. Baranauskas. Artigo submetido para um *Journal* internacional.
  Visão Geral e Contribuições: Esse capítulo apresenta de maneira instanciada as características (em relação à abordagem e método proposto) de um mecanismo de busca inclusivo e descreve como um protótipo da solução foi desenvolvido na RSI *VilanaRede*. Adicionalmente, resultados de um experimento feito junto a usuários reais da RSI *VilanaRede*, que endereça uma avaliação preliminar do mecanismo de busca inclusivo usando SWO construída a partir dos conteúdos reais desta RSI são apresentados. Os resultados, assim como uma avaliação e discussão dos mesmos, ressaltam tanto as características positivas do mecanismo junto ao método, assim como suas limitações.

Adicionalmente, o Apêndice desta dissertação é composto por três relatórios técnicos publicados no IC da UNICAMP, como descritos a seguir. Estes trabalhos não são apresentados integralmente nesta dissertação devido ao tamanho dos mesmos, mas estão disponíveis *online*. Vale salientar que esses relatórios técnicos descrevem em detalhes estudos, resultados e conclusões parciais importantes durante a pesquisa. Contudo, por não fazerem parte do núcleo substancial da pesquisa não compõem o corpo principal desta dissertação.

- **Apêndice A:** "Identificando Semântica em Redes Sociais Inclusivas *Online*: Um estudo sobre Ferramentas e Técnicas" (2010). Júlio Cesar dos Reis, Rodrigo Bonacin e M. Cecília C. Baranauskas. Relatório Técnico (IC-10-28), Instituto de Computação, Universidade Estadual de Campinas. Campinas, SP, Brasil. [53]. Derivado deste relatório técnico há o trabalho intitulado "Ferramentas e Técnicas de Mineração de Texto em Redes Sociais Inclusivas: Promovendo meios para o Acesso Universal ao Conhecimento" que foi publicado como trabalho completo na Conferência IADIS Ibero-Americana WWW/Internet 2010 (CIAWI 2010) [59].
  Visão Geral e Contribuições: Esse apêndice descreve em detalhes um estudo e seus resultados sobre a aplicação de possíveis ferramentas e técnicas de mineração de texto visando desenvolver a concepção de uma estratégia para auxiliar na construção (modelagem) das ontologias a partir dos dados de RSI. As principais dificuldades e desafios deste estudo foram lidar com o contexto particular e diversificado dos conteúdos da RSI, que contém grande informalidade e variedade de conteúdos. Este estudo com dados reais mostrou resultados promissores que podem auxiliar analistas na construção de ontologias que representem os conceitos usados na rede social.

- **Apêndice B:** "*From Ontology Charts to Web Ontologies: Heuristics and Transformation Rules*" (2011). Júlio Cesar dos Reis, Rodrigo Bonacin e M. Cecília C. Baranauskas. Relatório Técnico (IC-11-02), Instituto de Computação, Universidade Estadual de Campinas. Campinas, SP, Brasil [60]. Derivado deste relatório técnico há o poster intitulado "*Constructing Web Ontologies Informed by Semantic Analysis Method*" a ser publicado na *13th International Conference on Enterprise Information Systems* (ICEIS 2011), e o trabalho intitulado "*An Assisted Process for Building Semitoc Web Ontology*" aceito como artigo completo na *13th International Conference on Informatics and Semiotics in Organisations* (ICISO 2011).

Visão Geral e Contribuições: Neste apêndice é apresentado em detalhes como foi desenvolvido o processo de transformação entre os resultados do SAM, para ontologias *Web* descritas em *OWL* dentro do método proposto. Para isso, foram propostas heurísticas que mapeiam estruturas do digrama de ontologia da SO para códigos de ontologias em *OWL*. Fundamentado nestas heurísticas, regras de transformação foram implementadas no *software* SONAR [65], derivando código OWL a partir do diagrama de ontologia da SO (resultado do SAM). Tanto as heurísticas como as regras foram descritas e exemplificadas. Adicionalmente, resultados da aplicação do processo de transformação em um contexto de modelagem real mostram o potencial da abordagem, o valor das heurísticas e das regras implementadas para criar ontologias *Web* computacionalmente tratáveis, além de mais representativas e úteis (*i.e.* SWO).

- **Apêndice C:** "*Modeling Meanings from Inclusive Social Network Services*" (2011). Júlio Cesar dos Reis, M. Cecília C. Baranauskas e Rodrigo Bonacin. Relatório Técnico (IC-11-03), Instituto de Computação, Universidade Estadual de Campinas. Campinas, SP, Brasil [55].
  Visão Geral e Contribuições: Este apêndice mostra a aplicação e instanciação do método proposto no capítulo 4, criando uma instância da SWO no contexto real da RSI *VilanaRede*. Neste experimento foram usados tanto os resultados das ferramentas de mineração de texto aplicadas aos conteúdos da RSI *VilanaRede,* assim como foram exemplificadas as transformações e a junção de ontologias para serem usadas na busca. Neste estudo, objetivamos efetuar uma avaliação preliminar e parcial do método proposto, incluindo a qualidade dos resultados das ferramentas durante o processo de modelagem das ontologias, e o uso de ferramentas de suporte para a modelagem (*e.g.* SONAR). O experimento foi desenvolvido com 16 alunos da pós-graduação do IC da UNICAMP, que cursaram a disciplina de Semiótica da Interação Humano-Artefato Digital, oferecida no 2º semestre de 2010. Esta foi uma tentativa de observar o uso do método de modelagem e as ferramentas na prática.

Além das principais publicações mencionadas no corpo deste texto, que compõem esta dissertação, conforme resultados diretos e indiretos desta pesquisa de mestrado, e também da inserção e do envolvimento direto do autor em atividades com parceiros do projeto e-Cidadania, também foram resultados da pesquisa realizada durante este mestrado os seguintes artigos:

- REIS, J. C.; BONACIN, R.; MARTINS, M. C. 2009. Mobile Phones Interfaces for Informal Education. In: *Online Communities and Social Computing, Lecture Notes in Computer Science (LNCS)*. 13th International Conference on Human-Computer Interaction, San Diego - California, v. 5621/2009, pp. 515-524.
- REIS, J. C.; BONACIN, R.; MARTINS, M. C. 2009. Using Multimedia in the Mobile Collaborative Learning. In *V International Conference on Multimedia and Information and Communication Technologies in Education. Research, Reflections and Innovations in Integrating ICT in Education*. Lisbon - Portugal, v. 2. pp. 869-873.
- REIS, J. C.; BONACIN, R.; MARTINS, M. C. 2009. Developing Informal Education through Mobile Collaborative Learning. In the *12th IASTED International Conference on Computers and Advanced Technology in Education* (CATE), St Thomas, US Virgin Islands, pp. 1-7.
- REIS, J. C.; BONACIN, R.; BARANAUSKAS, M. C. C., 2010. Analysing Search in Inclusive Social Network Services. *Relatório Técnico (IC-10-02),* Instituto de Computação, Universidade Estadual de Campinas. Campinas, SP, Brasil.
- REIS, J. C.; BONACIN, R. 2010. Acesso Universal ao Conhecimento através de Busca em Redes Sociais Inclusivas na *Web*. In: *3º Seminário em Tecnologia da Informação PCI do CTI*, Ministério de Ciência e Tecnologia, Campinas, SP. pp. 88-93.
- REIS, J. C.; BONACIN, R.; BARANAUSKAS, M. C. C. 2010. New Perspectives for Search in Social Networks: A Challenge for Inclusion. In *Proceedings of the 12th International Conference on Enterprise Information Systems*. 12th International Conference on Enterprise Information Systems (ICEIS 2010), Funchal, Madeira Island, pp. 53-62.
- REIS, J. C.; BONACIN, R.; BARANAUSKAS, M. C. C. 2010. Search Informed by a Semiotic Approach in Social Network Services. In: *Proceedings of the 10th International Conference on New Technologies of Distributed Systems* (NOTERE 2010) - IEEE Computer Society Press. The 2010 Workshop Web2Touch - living experience through Web. Tozeur, pp. 321-326.
- REIS, J. C.; BONACIN, R.; BARANAUSKAS, M. C. C. 2010. Ferramentas e Técnicas de Mineração de Texto em Redes Sociais Inclusivas: Promovendo Meios para o Acesso Universal ao Conhecimento. In *Anais da Conferência LADIS Ibero-Americana WWW/Internet 2010* (CIAWI 2010). Carvoeiro, Algarve – Portugal, pp. 155-162.
- ALMEIDA, L. D. A; HAYASHI, E. C. S.; REIS, J. C.; MARTINS, M. C.; BARANAUSKAS, M.C.C. 2010. Conversas Online: A Synchronous

Communication Tool Integrated to Inclusive Social Networks. In *Anais do 11º Simpósio de Fatores Humanos em Sistemas Computacionais*, Belo Horizonte, Brasil, pp. 51-60.

- REIS, J. C.; BONACIN, R.; MARTINS, M. C. 2011. Designing Mobile Phone Interfaces for Collaborative Learning in Everyday Life. In *Mobile Phones: Technology, Networks and User Issues*. Organized by the Nova Science Publishers, Inc. *Editors*: Micaela C. Barnes *and* Neil P. Meyers. ISBN: 978-61209-247-8.
- REIS, J. C.; BONACIN, R; BARANAUSKAS, M. C. C. A. 2011. Semiotic-Based Approach for Search Mechanism in Social Network Services. In *International Journal of Web Portal*s. Special Issue Web2Touch Workshop (em revisão).
- REIS, J. C.; BONACIN, R.; BARANAUSKAS, M. C. C. 2011. Beyond the Social Search: Personalizing the Semantic Search in Social Networks. In *Online Communities and Social Computing, Lecture Notes in Computer Science* (LNCS). 14th International Conference on Human-Computer Interaction, Orlando, Florida, USA (a ser publicado).
- LIMA, T. C.; REIS, J.C. 2011. Gender in the Digital Age: Women's Participation in Designing Social Software. In *Online Communities and Social Computing, Lecture Notes in Computer Science (LNCS).* 14th International Conference on Human-Computer Interaction, Orlando, Florida, USA (a ser publicado).
- REIS, J. C.; BONACIN, R.; BARANAUSKAS, M. C. C. 2011. Constructing Web Ontologies Informed by Semantic Analysis Method. In *13th International Conference on Enterprise Information Systems* (ICEIS 2011), Beijing, China. (a ser publicado).
- ALMEIDA, L. D. A; HAYASHI, E. C. S.; REIS, J. C.; COSTA, P. D. P; BARANAUSKAS, M.C.C.; DE MARTINO, J. M. 2011. Augmenting Accessibility in Social Networks: a Virtual Presenter. In *13th International Conference on Enterprise Information Systems* (ICEIS 2011), Beijing, China. (a ser publicado).
- FIGUEIREDO, A. M. C. M.; REIS, J. C.; RODRIGUES, M. A. 2011. Semantic Search for Software Architecture Knowledge: A proposal for Virtual Communities Environment. In *International Conference on Information Society* (i-Society 2011), London, UK. (a ser publicado).
- PEREIRA, R.; MIRANDA, L.C.; BARANAUSKAS, M. C. C; PICCOLO, L. S. G.; ALMEIDA, L. D. A; REIS, J. C. 2011. Interaction Design of Social Software: Eliciting Requirements through a Culturally Aware Artifact. In

International Conference on Information Society (i-Society 2011), London, UK. (a ser publicado).

- REIS, J. C.; BONACIN, R.; BARANAUSKAS, M. C. C. An Assisted Process for Building Semitoc Web Ontology. In *13th International Conference on Informatics and Semiotics in Organisations* (ICISO 2011), Leeuwarden, The Netherlands. (a ser publicado).

Capítulo 2

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

2.1 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 - Social Network Services (SNS)* [9].

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 [32]. 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 [41] 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" [2], 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 [2] 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) [45]. 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 conseqüência 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) – "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 RSIs. 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 2.2 apresenta o conceito de 'redes sociais inclusivas' visando contextualizar a pesquisa e os desafios de busca atrelados a ela; a Seção 2.3 apresenta e discute algumas das estratégias de busca possíveis para redes sociais presentes na literatura; A Seção 2.4 apresenta as recomendações e a Seção 2.5 conclui o trabalho mostrando também os trabalhos futuros.

2.2 Redes Sociais Inclusivas e Desafios Relacionados à Busca

Segundo Boyd & Ellison [9] 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 inscrevem e utilizam ativamente os sites de redes sociais como *MySpace*[3], *FaceBook*[4], *Orkut*[5], *Flickr*[6] entre outros.

Conforme revelam os dados de Nielsen [46]: (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 [46], 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)[7] assinalam que, em 2008, 65% da população brasileira não tiveram acesso à Internet. Além disso, dados do Ministério da Educação no Brasil (MEC)[8] 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 [2]. Desta forma, é extremamente relevante criar espaços e meios para que as pessoas sejam incluídas de maneira

---

[3] www.myspace.com
[4] www.facebook.com
[5] www.orkut.com
[6] www.fickr.com
[7] www.ibge.gov.br/home/estatistica/populacao/acessoainternet2008
[8] portal.mec.gov.br

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 [2] tem se desenvolvido o conceito de 'Redes Sociais Inclusivas' (RSI), constatado que redes sociais desempenham um papel importante na influencia, adoção e uso das tecnologias de informação e comunicação [52]. Segundo Baranauskas [2] 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.* [45] 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 [*e.g.* 20].

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.

2.3 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 consultas (*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 [22] 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.

2.3.1 Busca Social

O trabalho pioneiro de Yu & Singh [82] 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.* [77] 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 [23], 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 [23] 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 [22] foca em aspectos de recomendação e confiança para busca de items 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 [22] fornece as preferêcias 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 [22] 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.

2.3.2 Busca Semântica

De acordo com Kassim & Rahmany [34] 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 [39] e Wang *et al.* [79]. 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 [5] no desenvolvimento de mecanismos de busca semântica. Segundo Berkan [4] 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 [19] ontologia é uma especificação formal e explícita de uma conceitualização compartilhada, que fornece descrições sobre conhecimento. Para Studer *et al.* [72], 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 [24], Guha *et al.* [21], além de Bonino *et al.* [8] e Fang *et al.* [16] 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 [39], 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.* [8] 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 entres 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 TRUST[9] descrita em Amaral *et al.* [1]. 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.* [14] 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 *SKOS*[10], 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*[11]. Segundo Lei *et al.* [35] 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.* [35] 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 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*[12].

Mangold [39] 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

---

[9] www.trustsemantics.tip.net.pl/

[10] www.w3.org/2004/02/skos/

[11] squiggle.cefriel.it/music/

[12] www.google.com

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 Wang *et al.* [79] 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.* [15]. 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.* [15], 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.

2.3.3 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 Heflin & Hendler [24], Guha *et al.* [21], Bonino *et al.* [8] e Fang *et al.* [16] 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.* [8] atentam-se sobre a navegação na ontologia para o refinamento de consultas. Contudo, argumentamos que são necessários processos de refinamento e expansão de consulta (*query*) que vão além da navegação em ontologias que consideram apenas simples relações de generalização e especialização, conforme proposto por Bonino *et al.* [8].

Para melhorar a experiência de busca dos usuários, Bonino *et al.* [8] 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.* [21], 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.*

[35] argumentam que a solução de busca proposta 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.* [35] 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.* [15] 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 [39], 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 adéqüe às necessidades e restrições do contexto de RSI, pois nenhum destes trabalhos endereça busca semântica para redes sociais. O trabalho de Mangold [39] 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 [39] a solução de busca semanticamente mais eficaz proporciona tanto transparência para os usuários inexperientes como comportamento interativo para usuários mais experientes. Lei *et al.* [35] 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.* [1] 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.

2.4 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 [62] 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,

necessitamos 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) [80] 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.* [16] 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 2.1faz uma síntese das principais recomendações propostas para esse fim.

---

[13] www.w3.org/2001/sw/Activity

[14] protege.stanford.edu/

Tabela 2.1: Síntese das principais recomendações propostas

Elemento/Aspecto	Recomendações
Estratégia de busca	-Semântica
Conhecimento do domínio	-Extraído e baseado em atividades na rede social
Modelagem	-Feita através de ontologias, que devem ser desenvolvidas por meio de métodos e processos específicos fundamentados em aspectos humanos
Tecnologia de representação computacional	-OWL
Processo de busca	-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
Foco da Busca	-No conteúdo produzido na rede, além de pessoas
Entrada de dados	-Interface simples através de palavras-chave
Ordenação dos resultados	-Apropriada ao contexto de RSI e baseada nos aspectos semânticos modelados

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.

2.5 Conclusão

Nos dias atuais o mundo contemporâneo é caracterizado por diversos problemas sociais. 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. 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 RSIs 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 um sistema de RSI baseado nas recomendações apresentadas.

Capítulo 3

A Semiotic-Based Approach to the Design of Web Ontologies©

3.1 Introduction

Communication is a basic element for society evolution for millennia. The writing, written press, radio, television and more recently the Web are some of the greatest inventions of humanity that propitiated the information access and sharing. These inventions have transformed the society and boosted the development of the humanity as a whole. In analyzing the "emergence" and popularization of the Web, it is possible to notice various scientific and technological advances that have made it possible, among them: new physical means of communication such as optic fiber networks and wireless networks, communication protocols, computing devices such as faster processors and displays, rich and standard GUI (Graphical User Interface); and more recently a great concern in better mechanisms for managing and retrieving data and information.

Analysing the evolution of the Web, the Web 1.0 (or first-generation of the Web) provided quick access to large volumes of information. The approach in the Web 1.0 was prevalent for centuries with books and for decades with radio and television, which we had a relationship "one-to-many", *i.e.,* an information producer for many consumers. The so-called

---

Web 2.0, besides a "richer" GUI has also changed considerably the Web 1.0 approach, towards a relationship of "many-to-many", to which there are some information producers and consumers working collaboratively. Social Network Services (SNSs), Wikis, Blogs, music and video sharing sites are examples of applications where many people produce and consume information in an interactive process and usually intensively. Nowadays, literature has glimpsed the Semantic Web (SW) as an extension of the current Web, in which well-defined meaning is associated to information, enabling computers and people to work better in cooperation [5].

Web systems generate a large volume of data in various media, with complex structures highly distributed, including the immeasurable cultural diversity present in information produced by people. New opportunities for advance in the Web could be achieved through the efficient management of this information. Nevertheless, the development and use of the Web brings new problems that are dependent on scientific and technological advances in several related areas. The solution for the problem of information modeling in the Web depends on the understanding of information and knowledge "nature", and on the development of complex computational algorithms. The challenge addressed in this paper is to understand how to structure, model, organize, manage and promote means for information available in Web systems be better computationally represented, allowing more efficient ways to access and share information.

In order to deal with this challenge, it is necessary to combine fundamentals, theories and methods aiming at understanding and modeling the process of knowledge generation and sharing with new technological approaches. Conventional solutions and approaches of the SW are based on "Web ontologies". A "Web ontology" can be understood as a specification of a conceptualization which provides descriptions about knowledge [19]. Literature has shown several semantic problems and limitations related to the use of Web ontology. Therefore, the goal of this paper is to show the major deficiencies in the SW technologies by showing its failure to resolve the main issues; and to present a new approach to design ontologies in the Social Web. In this approach, we discuss how some concepts from the Semantic Analysis Method (SAM) [36] could improve the Web ontology modeling, aiming at developing an expanded and more representative Web ontology towards a 'Semiotic Web ontology'.

The paper is organized as follows: Section 3.2 presents the theoretical and methodological background of the paper; Section 3.3 presents some current problems and limitations of the SW ontologies; Section 3.4 outlines a new approach for the design of Web ontologies using SAM, and shows a brief illustration and discussion; and Section 3.5 concludes.

3.2 Theoretical and Methodological Background

In this section we present an overview of the SW concepts and its technological constraints. Besides, as a theoretical-methodological background we present an overview of the SAM from OS.

3.2.1 Semantic Web and the ontologies

The main challenge of the SW development is to represent the meaning of the content to be machine interpretable. The way this is done is at the heart of the SW study. According to Uschold [76] the most widely accepted definition for this feature is content usable by machines. This means having data on the Web defined and linked in a way that they can be used by machines, not just for displaying purposes, although for automation, integration and reuse across applications.

For that purpose, it is necessary to the machine to have a model of "knowledge" about the domain, *i.e.,* the available knowledge must be represented so that the machine can "interpret" it. Tazi [74] argues that knowledge can be represented with the Sowa's Conceptual Graphs. This approach is based on Peirce's Existential Graphs, and follows the Aristotle's idea that each concept is represented by a word or symbol, serving as a semantic network in which nodes represent concepts that are related to each other. In the SW, knowledge is represented through computing ontologies. According to Studer *et al.* [72] ontology is a shared and common understanding of some domain that can be communicated between people and computers; it is a formal specification that should be readable and understandable by machines.

The term ontology is often used to refer to the semantic understanding (a conceptual framework of knowledge) shared by individuals participating in a given knowledge domain. Semantic ontology can exist as an informal conceptual framework of types of concepts, and their relations named and defined in natural language. Alternatively, it could be constructed as a formal semantics taking into account the domain, with the types of concepts and their relationships defined systematically in a logical language. Indeed within the Web environment, ontology is not simply a conceptual framework, but a concrete syntactic structure that tries to model the semantics of a domain [31]. According to Noy & McGuinness [47], ontology along with a number of different instances of its classes constitutes a knowledge base. The classes are the focus of most ontologies. Classes describe the concepts in the domain. For instance, a class of wines represents all wines; specific wines are instances of this class. The Bordeaux wine is an instance of a class of wines. A class can have subclasses that represent concepts that are more specific than super-classes; *e.g.* we can

divide the class of all wines into red, white and rosé wines. Alternatively, we can divide the class of all wines into sparkling wines in non-sparkling wines.

At the core of the SW technology there is a language based on logic for knowledge representation and inference. Computational Languages for ontology description are designed specifically to define ontologies. According to the SW architecture proposed by Berners-Lee *et al.* [5], the ontology description languages are related to other Web languages such as *Resource Description Framework* (RDF), RDF *Schema* and the *Extensible Markup Language* (XML). According to statistics from Cardoso [12] OWL (*Web Ontology Language*) is nowadays the most common approach for modeling ontologies in software. OWL has three sub-languages with increasing expressivity: OWL Lite, OWL DL and OWL Full. OWL is currently defined by a set of recommendations of the *World Wide Web Consortium* (W3C) [80].

3.2.2 Semantic Analysis Method

As a theoretical reference of the OS for the proposed approach, we have used the Semantic Analysis Method (SAM) [36] that comes from the MEASUR (Methods for Eliciting, Analyzing and Specifying Users' Requirements) [70]. The SAM assists users or problem owners in eliciting and representing their requirements in a formal and precise model. With the analyst in the role of facilitator, the required system functions are specified in an Ontology Chart (OC). It is worth to mention that this concept of ontology is different from the SW ontology. Ontology in OS represents a business domain which can be described by the concepts, the ontological dependencies between the concepts, and the norms detailing the constraints at both universal and instance level [37]. A graphic representation of a conceptual model is called an OC. The OC describes a view of responsible agents in the focal domain and their pattern of behavior named affordances [36]. Some basic concepts of SAM adopted in this paper are based in Liu [36]:

"**The world**" is socially constructed by the actions of agents, on the basis of what is offered by the physical world itself;

"**Affordance**", a concept introduced by Gibson [18] is used to express invariant repertories of behavior of an organism made available by some combined structure of the organism and its environment. In SAM [69] the concept introduced by Gibson was extended by Stamper to include invariants of behavior in the social world;

"**Agent**" can be defined as something that has responsible behavior. An agent can be an individual person, a cultural group, a language community, a society, etc. (an employee, a department, an organization, etc.);

"**An ontological dependency**" is formed when an affordance is possible only if certain other affordances are available. The affordance "A" is ontological dependent on the affordance "B" means that "A" is only possible when "B" is also possible;

"**Determiners**" are properties which are variants of quality and quantity that differentiate one instance from another;

"**Specialization**", agents and affordances can be placed in generic-specific structures according to whether or not they possess shared or different properties;

OS adopts a subjectivist philosophical stance and an agent-in-action ontology. This philosophical position states that, for all practical purposes, nothing exists without a perceiving agent or without the agent engaging in actions. That is to say, each thing depends for its existence upon the existence of its antecedents. Words and expressions we use are names for invariant patterns in the flux of actions and events which the agents experience. The classical distinction between entity, attribute and relationship disappears to be replaced by the concepts of agents, affordances (the actions or attributes of agents) and norms (for the socially defined patterns of behaviour) related to their antecedents to indicate the ontological dependency [69]. The concepts of the Semantic Analysis are represented by means of this agent-in-action ontology.

We have investigated the design of Web ontologies to deal with their problems and limitations, as presented in the next section, inspired on this perspective.

3.3 Problems and Limitations of Semantic Web ontologies

Web ontologies (in OWL) have been widely used for many purposes, such as semantic search [*e.g.* 24, 21] and content management [*e.g.* 40]. Although literature has shown several semantic problems and limitations related to the use of these artifacts.

According to Carvalho [13], even with the advent of ontologies, there are still no tools to assist in the organization of the information in a way suitable for human mental operations in an individual or societal way. In order to facilitate the work for the computer, the organization within the ontology is formally made, creating a fixed relation of words. Carvalho [13] also argues that it is necessary to discuss the whole set of relationships and context of information contained in ontologies. This contextualization is generated from a detailed study of the topics required for understanding the subject in question. The study asks for a number of key concepts, which summarize the knowledge of the area. These concepts need to be organized as a way to produce a "knowledge tree". This tree should be able to translate that subject, representing it as accurately as possible. By establishing a hierarchy between concepts, it is difficult to accurately represent different contexts, which means that the ontology need to be attached to a well-defined domain.

Gärdenfors [17] argues that if we want to consider how humans deal with concepts and their meanings, the structures of the class relation from SW ontologies have captured only a little part of our knowledge about concepts. For example, we often categorize objects

according to the similarity between them, and similarity is not a concept that can be expressed in a natural way in a Web ontology language. Additionally, Gärdenfors [17] says that a notable characteristic of human thought is our ability to combine concepts and, in particular, understand the new combinations of these concepts. Furthermore, almost all Web applications (*e.g.* systems of question and answering) have inputs in the form of combinations of concepts. Therefore, Gärdenfors [17] states that an important criterion for the success of the computational semantic model is that it should be able to deal with combinations of concepts. This author also highlights the lack of symbolic grounding in these ontologies. The source of the problem is that each ontology (along with its terminology) works as a free floating island of reeds – it has no anchor in reality. However the "meaning" of the ontological expression does not live on these islands. Thus, Gärdenfors [17] proposes the establishment of structures called Conceptual Spaces, as a richer semantic structure underlying the representational format. Conceptual Spaces represent information through geometric structures and not through symbols.

The work of Tanasescu & Streibel [73] describes several arguments in favor of alternative models for knowledge representation in detriment of traditional ontologies, such as: (1) the inadequacy of reasoning based on categories to represent reality; (2) the need for different representations of the same identity according to the context; and also (3) the difficulty for representing psychological concepts, such as Affordances from Gibson [18] in a hierarchical structure. The authors argue that Web ontologies are not yet flexible enough to match the representational complexity of the human mind; also they are difficult to construct. Tanasescu & Streibel [73] emphasize that Web ontologies are better suited to the description of scientific fields such as medicine and biology, which are already semi-formal and organized into categories and relationships.

Tanasescu & Streibel [73] also claim that with the advent of Web 2.0 applications there has been an intensified use of non-structured notes, such as tagging and Collaborative Tagging Systems (CTS). CTS produce different results compared to using default vocabularies for tagging, and provide users with a simple way to make sense (meaning) to their own content. Consequently, the authors argue that while current investigations are still trying to alleviate the practical problems related to the use of ontologies, the WS can benefit from the techniques used by Web 2.0 applications. These techniques have spread out widely and appear to be a way to allow users to describe their own content, since the system cannot determine a priori the meaning of the content. They conclude that for a faster expansion of SW new approaches to semantic acquisition, separated from the centralized ontologies and not developed by experts, need to be explored. Thus, alternatively, they introduce the proposal of Extreme Tagging Systems (ETS), as an extension of CTS, enabling the

collaborative construction of knowledge bases over the use of formal and centralized ontologies for knowledge representation.

The work of Obitko *et al.* [48] proposes an alternative approach which remains using conventional Web ontologies for knowledge representation. They have described a strategy for designing ontologies using Formal Concept Analysis (FCA). This is a theory of data analysis that identifies conceptual structures among data sets. This method allows discovering the need for new concepts and their relationships in an ontology. FCA is based on the philosophical understanding that a concept has two parts: (1) its extension which consists of all objects belonging to the concept; and (2) its intention, which includes all attributes shared by these objects. The crucial characteristic in this method for knowledge representation is that it is not based on a priori definition of classes; nevertheless the concepts are described from their attributes. Instead to create a class and to associate attributes to it, a concept is built from their attributes.

These discussed studies propose both: (1) totally alternative methods to Web ontologies for knowledge representation in the SW; and (2) instead of using completely alternative methods some approaches just propose a differentiated design for ontologies. In the next section we propose a method to the design of Web ontologies based on SAM.

3.4 Prospecting a new approach to the design of Web ontologies

In order to produce immediate and practical results on the SW applications, our approach employs a different method which produces an agent-in-action ontology, and explores how to improve the Web ontologies using concepts from the agent-in-action perspective. In other words, we propose to develop a representational structure towards a 'Semiotic Web ontology' (SWO). It is worth to mention that it is not our goal to refute here the SW technologies of nowadays, neither to create a "perfect ontology" from a theoretical point of view; but instead we propose to expand SW techniques with methods and techniques coming from OS.

SWO is a semantic model (computationally tractable ontology) constructed from a semi-automatic method based on SAM. Some theoretical and methodological concepts of SAM are used in conjunction with other technologies from the SW to describe computationally tractable ontologies using OWL. The idea is to incorporate the concepts of particular Agents (roles) and Affordances (patterns of behavior) arising from the SAM into an expanded and more representative SW ontology.

It is also important to emphasize that we do not intend to create an OC (from SAM) in OWL or to substitute the OC at the conceptual or business level. The use of OWL is

relevant here since it is at implementation level, thus it gives us opportunities to improve the semantic models used in the existing SW applications and initiatives. We understand that this is a fast and practical way to show direct contributions from SAM to the SW. Semantic Web solutions like semantic search could take advantage of the SAM. Therefore some properties from the OC may not be fully transcribed to OWL at this time, while other aspects such as agent-affordance relationship are emphasized.

From a Semiotics perspective it is assumed that the signs are socially constructed. Thereby, a computational model that represents the semantics from a Social Web application should contain the agents that interpret the socially shared concepts. With this approach we incorporate and take to SW ontologies concerns and possible representations arising from the Ontology in a semiotic perspective. In addition to agents and affordances, we have observed that SW ontologies also do not incorporate in the model (at least explicitly) the idea of ontological dependency relations.

In order to design the Web ontology, we first create an OC using SAM. This intermediate ontology diagram is important to identify the possible agents from the context and their patterns of behavior, and thus pass these to the (computationally tractable) Web ontology using OWL. To accomplish that, a set of specific heuristics is applied to derive an initial OWL ontology. Bonacin *et al.* [7] proposed a heuristic to construct system design UML diagrams from OC; those heuristics must be adapted to our purpose. This approach does not create an equivalent ontology in OWL; instead it provides some heuristics to support the analyst during the modeling process.

In the SWO we represent the agents that have behaviour(s) (affordance) in a concept (which can have determiners), and can be important in situations of synonymous and polysemy. For instance, the concept of 'crane' can mean a bird or a type of construction equipment, and we can model it using the agents and their affordances; *e.g.* a biologist, who can be model as an agent, probably make studies about birds. To study birds is a pattern of behaviour of a biologist (in other words an affordance). As shown by Figure 1, 'crane' is a concept that can have several different meanings, although in some context, due to the agent and their affordances, the meaning of 'crane' is more closely linked to birds and not, for example, to a construction equipment, that can also be represented in the model.

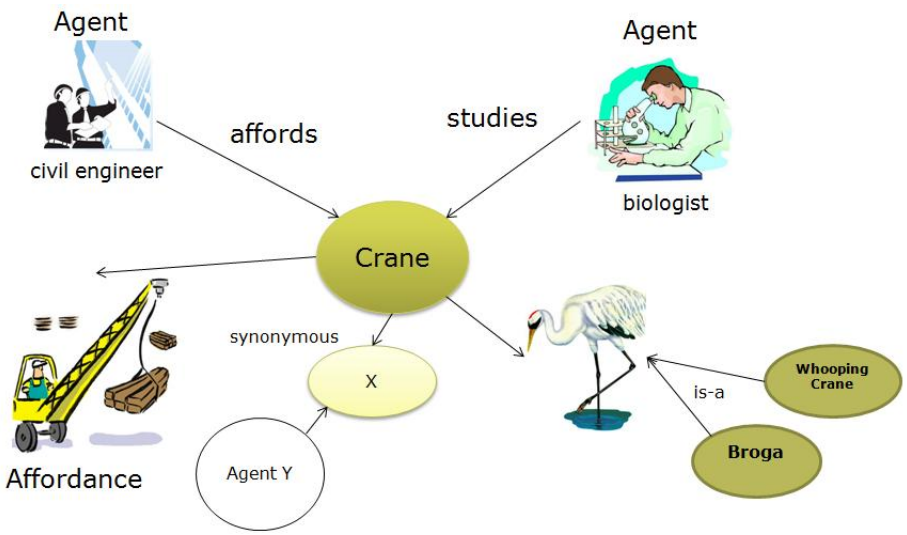


Figure 3.1: Modeling meanings in an example of polysemy using agents and affordances

Figure 3.1 illustrates an example of modeling using this approach in which the 'biologist' and the 'civil engineer' are agents that have affordances connected to specific concepts. Also this model can have relationships of specific 'is-a', *e.g.* 'Broga' and 'Whooping Crane' are specific kinds of 'crane'. This shows that concepts can be related to several agents and affordances, and with other concepts, constituting relations and representations that make more complete ontologies compared to ontologies described purely for a domain.

For instance, 'crane' can mean a construction equipment for a 'civil engineer', as well as anything else to any other agent, or have any synonym that makes sense for an agent 'Y' modeled from the data of the Web system. We can see other examples like 'Manga' (in Portuguese) can mean a fruit, a sleeve as well as a color; and we can model it using the agents and their affordances in a SWO.

In this approach, we introduce new constructions that represent agents and affordances in OWL ontology. The meanings of the concepts represented in the ontology are relative to the agents. Then, aspects such as polysemy, that is a hard problem for SW applications, could be better treated using this ontology.

3.4.1 Illustrating the approach

The use of this approach has been utilized and investigated in a scenario of Social Network Services (SNS). Experiences with users of search engines [62] point out that this kind of association, as developed in this approach, could contribute to more precise and adequate search mechanisms in SNS. We illustrate a search scenario in SNS that can be beneficiated

with the SWO. From the user profile in the SNS application, we identify the agents represented in the ontology, and make a connection between them (user and agents). Thus we can prioritize (or even limit) the search space, making a relation between the user with the ontology; *e.g.* if a biologist is logged into the system (we could find that a user is a 'biologist' based on his/her profile) and request a search with the keyword 'crane'. Whether we have a relation between the 'biologist' agent and the term 'crane' in the ontology, the results from announcements of the SNS that could be returned first (ranked first) should most likely be related to the concept of crane as a 'bird', not to other meaning(s) of this word (like a type of construction equipment).

Nevertheless to a 'civil engineer' that makes the search into the system about 'crane', probably the results that most interest him / her are about the construction equipment and not about 'crane' as a bird. We do not mean that other results are not required or may not be returned in response to the engineer search, (may be the engineer could want to know about this kind of bird). In this case the announcements from the SNS on 'crane' as construction equipment must have greater relevance in the ranking of results. However, a user that has a profile which fits a 'biologist' agent, he or she would have the announcements about 'crane' as a bird with highest priority.

The agent-affordance relation is also used to indicate the probable meaning of the terms in an announcement. For instance, we could verify whether the word 'crane' is about 'bird' or 'construction equipment' based on the user that posted such information. In this situation, whether the user who submitted the announcement fits a 'biologist' agent, 'crane' would be most likely about a 'bird'. Otherwise whether the advertiser is a 'civil engineer', in this situation 'crane' would also most likely mean 'construction equipment'. We could have relationships between agents to verify how much an agent is semantically close to another and to indicate the probable meaning based on this aspect.

3.4.2 Discussing the approach

The semantic chart (from SAM) delimits the area of operation of the context under study and identifies the basic patterns of behavior (affordances) of the agents. Understanding and modeling the invariants of behaviour of human agents, including how they communicate, interpret the signs and act in society is a key point for the construction of more accurate and flexible ontology models.

It is possible to highlight points which the OS approach deals with the shortcomings of conventional Web ontologies, such as the three deficiencies presented by Tanasescu & Streibel [73]: (1) the reasoning based just on categories to represent reality, in OS is complemented by the identification of agents and their affordances; also (2) there is no different representations of the same identity in the context, since the meaning of the identity

is relative to the agent actions, and even (3) there is no difficulty to represent psychological concepts since the concept of affordances (from the cognitive psychology) is the basis for the description of the model. Moreover, with our approach we can build more flexible ontologies, since the concepts are interpreted based on the patterns of behaviour of the represented agents, no matter whether there is a static hierarchy of concepts, because the different contexts can be identified by the agents. Similarity and combination of concepts could be done using also the agent as a way to make disambiguation. Once modeling ontologies is a hard and time consuming task, we believe that constructing geometric structures underlying it, as the Gärdenfors [17] proposes, could be not viable on a large scale. Regarding ETS approaches, they may not be feasible in some contexts in which non expert users have no ability to create and manage tags.

The understanding and modeling of ontologies using methods and techniques grounded on human cognition and behavior are also needed to build a Web with focus on human agents (and not just artificial agents). Furthermore, it is important to emphasize that we want to consider the technological work already done, looking for new modeling methods that will complement and boost the proposal of the SW. Several applications may benefit of this approach, such as new possibilities for semantic search engines in SNS that include the agents, and create new ways to more appropriate search for users.

In SNS contexts, Mika [44] has already pointed out the general advantages of incorporating the social context into the representation of ontologies. According to Mika [44] creating the link between actors and concepts into the model of ontologies brings benefits in terms of more meaningful and easily maintainable conceptual structures. Mika proposed the extension of the traditional concept of ontologies (concepts and instances) with the social dimension, extending this traditional bipartite model by incorporating actors. Mika's proposal aims at modeling networks of folksonomies using the idea of connecting the real user with the concept and their objects. By using our approach with the agents' concept and their affordances a more general and wide-ranging of applications is possible; moreover, it is based on a formal method to find out the agents, affordances and the agent-affordance relationship.

Although concepts and theories from SAM can bring benefits to the SW models, we argue that OWL models and OC do not replace each other. They present distinct views and have different proposals. While OC concerns human perception and patterns of behaviour, and can be empirically refuted, OWL concerns are the computer interpretable constructs and efficient models. In our approach, it is responsibility of the analyst to interpret and decide how to construct better computer interpretable models (such as OWL) from the OC. Tools and heuristics can be used for supporting the analyst during this process, however only the analyst is able to connect the models and examine their consistence with the real world.

3.5 Conclusion

The evolution and use of the Web over the years have brought new challenges on modeling and representing information. A better organization, management and retrieval of digital content have become a critical point to allow new opportunities for knowledge access and sharing in the Social Web. Therefore, there is a growing need for solutions that deal with semantic aspects in Web Systems trying to understand the meanings from the information and improve their use. The Semantic Web view brings practical techniques and solutions trying to create content usable by machines. Nevertheless due to the amount and complexity of data, these technologies are still insufficient to really deal with this problem, resulting on more sophisticated and adequate solutions from the human agents point of view. As presented in this paper, literature has pointed out some deficiencies of conventional Semantic Web approaches. The main goal was to raise it with a discussion for a long term work.

Hence, new approaches to better understand and model the semantic aspects of digital content in the Web are necessary. This paper presented an approach based on Organizational Semiotics to build Web ontologies. Our proposal is to design Web ontologies aided by SAM. We discussed how some concepts from SAM could improve the modeling of Web ontologies. We showed the possible contributions to improve it, indicating the practical and immediate results which the approach could be empirically demonstrated. Further work involves to develop an expanded and more human-representative Web ontology, as well as to present a practical example illustrating the use of the approach. Next steps in this research include to explore other concepts from SAM in the modeling using OWL, as well as to develop a semi-automatic software tool that materializes the ideas of the approach to create the 'Semiotic Web ontology', including the heuristics to aid creating an initial OWL ontology from the OS chart.

Capítulo 4

Prospecting an Inclusive Search Mechanism for Social Network Services©

4.1 Introduction

Web systems and portals are available to a vast number of people with socio-cultural differences. Within a Social Network people communicate and behave according to commitments, linguistic constraints, culture and other social aspects. Social Network Services (SNS) could be more appropriate to the context of people's life, considering their differences, in order to promote social and digital inclusion.

This scenario becomes even more challenging and important in contexts such as Brazil, where there are several social problems and a huge cultural diversity. In this perspective, one of the fundamental points is to provide barriers free access to information to every citizen. This could help to bring great social benefits and contribute to a profound social transformation. Thus, it is extremely important that all people have the opportunity to retrieve, access and use information provided in the digital media in a smooth way.

The SNS represents an opportunity for interaction, access to information and knowledge through the Web. These systems primarily allow individuals to share their interests and activities, constituting communities. The e-Cidadania project [3] aims at transforming a SNS into an engine for digital inclusion and citizenship. The network systems with such

characteristics can be defined as "Inclusive Social Network Services" (ISN) (see [45]). The use of search engines is one of the primary ways to find and to make access to information generated in these systems. However, search mechanisms are currently built based on comparisons of keywords and lexical-syntactical information processing (syntax search). These mechanisms are not sufficient and adequate to effectively make sense to individuals in an inclusive scenario within social networks.

Based on empirical results, which will be discussed in this paper, we have observed that people organized into virtual communities bring to this space their own vocabularies and meanings, and also develop their own local vocabularies through interaction and communication using technology. The results pointed out the need for novel search mechanisms considering the diversity of users' competencies and inclusion aspects.

A more appropriate inclusive search solution for an ISN should reflect the semantics used by participants of the social network. In few words, a search engine should take into account the local meanings created, shared and used by people organized into a community. In this paper we argue that the quality and response accuracy of a search mechanism are intrinsically associated to the proximity of the semantics shared by people. Thus, it is necessary to identify the meanings used in the network and to represent its semantic aspects. This could actually contribute to make the information accessible to everyone, including people with low educational levels that have difficulty to access online information due to their simple vocabulary or their deficiency in writing. Usually, these people use an informal (colloquial) or local vocabulary in the search. With the proposed solution they could find the correct information in an easier and more precise way, besides learning from it.

In this paper we show results of search activities within an ISN, conducted in the context of the e-Cidadania project. The goal of these activities was to observe a set of search scenarios with potential users of an ISN, and to understand how these users make sense of a search mechanism. Based on the results, we present a prospection of a more appropriate search mechanism for an ISN with foundations in Organizational Semiotics (OS) [67, 35]. In our approach the goal is to expand and to improve the search technologies and techniques of the Semantic Web field based on Reis *et al.* [57]. Besides, the representational structure (semantic model) used by the search mechanism is based on data from the interaction and communication among users in the social network system. Thus, the search engine will take into account the meanings shared and created by people (including the informal terms) in their interaction with the system aiming to provide better results.

The paper is organized as follows: Section 4.2 presents the concept of ISN and the importance of search mechanisms for the universal access to information; Section 4.3 presents the analysis of the empirical experiment with ISN users; Section 4.4 details the

proposed approach; Section 4.5 makes a discussion about the approach and related works; and Section 4.6 concludes presenting further works.

4.2 Universal Access and Inclusive Search

According to Boyd & Ellison [9] since the beginning of the Social Networking, sites such as *MySpace*, *Facebook*, *Orkut* and others have attracted millions of users and many of them have integrated these sites into their daily practices.

Online Social Networks or "communities of members" have great relevance in the Web as users spend much time navigating on them. According to Nielsen [46], social networks are more popular than e-mail, with 66.8% of global reach. Around the world, it represents the fourth most used resource in the Internet and 85.2% of penetrations are in the portals and communities of general interest. Additionally, 85.9% of Internet users use search engines, which is one of the most common activities.

Despite these great numbers and the success of Social Network sites among Internet users, in social contexts such as Brazil and other developing countries, there are yet a lot of people without access to the Internet and consequently without opportunities to access information and knowledge. Social indicators shown by the PNAD (National Survey by Household Sample) produced by the Brazilian Institute of Geography and Statistics (IBGE in its Portuguese acronym) [27] points out that in 2008, 65% of the population did not have access to the Internet.

In addition, important data from the Ministry of Education in Brazil [42] reveal that about 30 million of Brazilians are functionally illiterate, defined as the population over 15 years old and less than 4 years of schooling (21.6% of the population). Using a broader concept of functional illiteracy, according to a survey from Paulo Montenegro Institute held in 2007 [30], the majority (64%) of Brazilians between 15 and 64 years old and more than 4 years of schooling reach no more than the degree of rudimentary literacy, *i.e.*, they have only the ability to locate explicit information in short texts or make simple math; they are not able to understand longer texts. This data illustrates only part of the challenge that we face in terms of designing systems, which should include all these users. In this context it has become a major concern to allow access to online content available from SNS to all people in a more "natural" and efficient way.

Thus, it is extremely important to recreate methods to permit the effective access and use of information conveyed in digital media, for all. This could be materialized with the ISN concept. We understand ISN as a "virtual communication space" based on the concept of social networks, which is inclusive and allows the community to share knowledge about the community know-how. This space has to facilitate "exchange" (of knowledge, goods and

services) in accordance to the collaborative (project team, partners, community) system conception.

It is also worth to mention that in an ISN there are not target users, but all users are relevant and should be included without discrimination. Therefore, there may be people without skills to handle certain technological features of the system and consequently without knowledge to find information that they need in the system. Moreover, those users most often use colloquial terms to express themselves through the system. For example, they may use the term "*postinho*" (in Portuguese) instead of "Basic Health Unit" (formal). They use terms that make sense to them, but in fact, these expressions semantically mean the same. So, when someone is trying to retrieve information from the ISN, these factors must be taken into account by the search engine. On the other hand, when a user searches for something in a non-formal or not refined way and, the same concept but in its formal way (cult) is returned, this represents an opportunity for learning.

Accordingly, we should seek for a computational search solution that takes into consideration the meaning that is adopted or emerges in the context of use of that network; *i.e.* the meaning that people bring to the network, and those that are constructed by using the system over time (through interaction). This may facilitate and provide better access to the content generated by users of the network.

4.3 Analyzing Search Scenarios of an ISN

From a practical point of view, the e-Cidadania project resulted in the '*Vilanarede'* ISN system. This system has represented an opportunity to investigate the interaction behaviour of representative users in a developing country. As a direct activity of the project, we have conducted the 8th Participatory Workshop, in a telecenter located at 'Vila União', neighborhood of Campinas city, Brazil. In this workshop we developed an activity related to search in the ISN. The objective was to observe some major points including: (1) How would the users build understanding of the search engine? (2) Which keywords would they use? (3) Would they have any difficulty in completing the proposed scenarios? and (4) What would be their satisfaction with the search results?

A task sheet with 4 search scenarios was presented to each pair of participants, and a form was prepared to the observers (researchers) of the activity. Additionally, an "extra" scenario, called Scenario X, was also included in the task. We had 7 pairs of users in total. An initial instruction about the activity was given to the participants. The pairs were formed by the users themselves, and for each scenario the pair should write the words used in the search and the title of the announcements found. Resulting from this activity, we had both the sheet tasks filled by the pairs of users and the observation forms filled out by the

observers. Besides, the activity was filmed and there was audio recording of each pair during the task execution. The 4 search scenarios were:

Scenario 1: Find out announcements on how to popularize the '*VilanaRede*'.

Scenario 2: Find out announcements of mango (fruit) in '*VilanaRede*'.

Scenario 3: Find out announcements related to food in '*VilanaRede*'.

Scenario 4: Find out announcement related to religion item combined with handicraft in '*VilanaRede*'.

Each scenario intended to verify whether semantic capacity was needed for the search mechanism. The time for the completion of the scenarios was approximately 45 minutes. After the execution of the search scenarios, a general discussion was conducted in order to get the general impression from the users about the activity. During this discussion, several interesting stories were collected.

In Scenario 1, we wanted to observe whether users would use synonymous of "popularize" to find the announcements. Some pairs had difficulty to understand the scenario, as well as difficulty in choosing the terms for the search. However, some pairs associated the word "popularize" to "divulge" and quickly found related announcements. In this scenario one pair used some unusual keywords such as: "*boca-a-boca*" (a popular expression used in Brazil that means "orally passing information from person to person"), "email", "phone" and "posters". By using the term "*boca-a-boca*" in order to find announcements about how to divulge the site, unusual results also appeared as an advertisement for "*Bife de casca de banana*" (steak of banana peel). It happened because in one of the comments of this announcement we find "I'm with water in my mouth (*boca*)" in reference to the announcement of "steak of banana peel". Phrases for search like "divulgation of the *'Vilanarede*'" or verbs such as "to popularize announcement" or "advertising Vila" were also used in this scenario.

In Scenario 2, we wanted to verify whether users would find any announcement related to the mango (fruit) in the application contents. There was no announcement about the mango fruit in the system. However there was an announcement about *mangá* (cartoon) and it was written without the acute signal in the word ('*manga*' in Portuguese, which is mango fruit in English). In this scenario, users mainly used the following keywords (translated from Portuguese): "mango (fruit)", "fruit", "mango", "mango fruit", "mango / fruit". Some pairs were uncertain if they would have to put the keyword "fruit" or not. Note that in a semantic search, by putting the keyword "fruit", the application should return all the announcements with mango (fruit), in the case of announcements semantically related to fruit.

In Scenario 3, we wanted to see whether users would use the keyword "food" in the search or they would make a search for specific foods through the search engine. As a result, when users tried the keyword "food", the system returned nothing. However there are

several announcements on food in the system: the sale of "*salgadinhos*" (homemade snacks), "*pão-de-queijo*" (cheese bread) and others. Among the relevant considerations from the observers, during the execution of this scenario users said that the system should relate "*salgadinhos*" (homemade snacks), "*pão-de-queijo*" (cheese bread) and "*Bife de casca de banana*" (steak of banana peel) with the concept of food. And this makes sense since that semantically all of these are food. During the discussion phase one of the users commented: "Using food is easier because it already covers everything," *i.e.,* all types of food in the system. Another said: "To be more 'lean' and practical for those who are starting (in terms of computer literacy), like us, when we enter "food", it should return a variety of foods due to our difficulty." Yet another user says: "Maybe to use food does not help in the search for something more specific, but if it is something that we have no knowledge of the domain, or we do not know what to look for, the tool would be useful and helpful." The main keywords used in this scenario (translated from Portuguese) were "food", "*comida caseira*" (homemade cooking), "food sale", "*salgado*" (homemade snacks), "*salgadinhos*"(small homemade snacks), "*salgadinho frito*" (fried homemade snacks), pies, "*doces*" (sweets), "*pão-de-queijo*" (cheese bread), "*docinhos*" (small sweets), cake, pastel and "*brigadeiro*" (chocolate sweet). Note that users utilize several variations in words such as "homemade snack", "small homemade snack" and "fried homemade snacks".

With Scenario 4, we aimed to determine which keywords users would use when looking for a specific announcement. One of the observers indicated that the pair found the "Saint Anthony" because they already knew that this announcement was in the system. The same was reported by several other observers. The vast majority of the pairs used the keywords (translated from Portuguese): "homemade craft", "Crafts saint", "holiness", "holy" and "saints". Users found the desired information successfully. But one of the pairs put keywords like (translated from Portuguese): "Orisha", "Orisha of cloth", "religious" and "sculpture" and didn't find out any announcement. Several observers noticed that the subjects utilized terms from their own colloquial language in the search; examples can be seen as "*manga rosa*" (pink mango), "*manga coquinho*" (coconuts mango), "*tutu de feijão*" (tutu bean), "*boca-a-boca*", "small sweets", "little homemade snack" and "Orisha". Also in several occasions the pairs discussed before reaching an agreement on which word to use in the search.

Another interesting result was obtained from the interaction of a deaf user with the search mechanism. As expected this user has difficulty with the written language. We observed that he uses the same hand signal to several different words. The user had difficulty in understanding the scenario 1, since the words popularize, advertise, advertisement and disclose have the same or similar hand signals in his language. Moreover, we could see that the user has different understanding for some words that have the same meaning; his

behaviour during the search was not confident neither independent; he asked a lot of questions to the observer.

Additionally, general results indicate that users from the context under study (prospective users of an ISN) had difficulty with the search button; in other words, they do not have a clear concept about the act of "searching" in an application on the Internet. Some users had no idea about the scope of searching. They did not know whether that search referred only to the announcements in the *'Vilanarede'* system. This fact is explicit in a description from a user who said: "Search fondue because it is something chic". However another report from other user says: "Fondue is very chic, we do not have it here in our network... we will not put it in the search because the network is ours, it is "poor"... and it will not have fondue...". This statement shows that the second user has the notion of the search scope, which will be just within the announcements from that social network system; so since there were no announcements about fondue, nothing would be returned.

Even with this lack of sense about the search scope, one of the observers explains that the users were surprised with the power of the search, and they explored and tested it easily. Such surprise can be explained since most of them have never used a search mechanism before. From the forms filled by the observers, approximately 80% of the pairs felt comfortable during the task. Around 60% of the pairs did not make a lot of questions to the facilitator during the task. This point out that searching using keywords can be considered for these people. However regarding the search results the currently solutions are not enough to provide information that make sense to each user in a context.

An interesting fact reported by the observers is that sometimes users initiate the search by entering complete questions in natural language, or at least they think aloud in that way. This is confirmed with the scenario X. This scenario was described as follows: "Suppose you want to make a reservation for a medical consultation at the local hospital and go to the '*VilanaRede'* system to get information (*e.g.* phone of the hospital). How would you make the search for some announcement related to it?" In this scenario a few pairs used keywords such as (translated from Portuguese): "Hospital", "Health Center", "phone of the health clinic", "scheduling of medical appointments". However, some pairs used sentences in natural language such as: "Can anyone tell me how to make a reservation for medical consultation at PUCC[15]?" and also "What is the phone of the SUS[16] for appointments?". Observers reported that after trying natural language, users started to use terms and keywords, and sometimes they employed a combination of more than one word. During the final discussion, after the execution of the scenarios, users explained that they had learned

[15] PUCC is a hospital at Campinas city from the Pontifical Catholic University of Campinas

[16] Unified Health System in Brazil

that a complete phrase usually "does not work", and frequently only one "right word", as said by a user, is sometimes enough to return search results.

These practical results from the workshop show that users' colloquial language should be considered during the development of more appropriate search engines. Moreover, people in a social network can create their vocabulary, sharing meanings in the community. The results showed us that it is necessary to construct computationally tractable models from the semantic point of view that come out from the network itself. Semantics here is understood as the interpretation of signs [50] by individuals and their association with real world elements. This interpretation is socially contextualized; *i.e.* individuals and communities may have different interpretations for the same sign and a sign may connote different meanings depending on the context applied.

4.4 Toward an Adequate Search Mechanism for ISN

General difficulties faced by users to get information in the Web can be explained mainly by: (1) overload of information presented in the system; and (2) problems related to the contextualization of the meaning for the terms used. As an attempt to solve this problem, we have investigated an approach that can result in better and more appropriate search engines for ISN.

In a social network, the "emergence" of meaning is an ongoing process in which meanings and interpretations are constructed, used and shared through the system based on the interactions and expressions of users. These interpretations expressed by users in the system could be computationally represented. Several improvements could be achieved such as semantic models to make the social network context more faithful resulting in more adequate search engines.

In order to accomplish that, we have proposed a search engine informed by a Semiotic approach [54]. We have developed a semi-automatic method to model the semantics of the ISN using the Semantic Analysis Method (SAM); the outcome of this process is intended to be used in the search engine.

4.4.1 The Semantic Analysis Method

This section presents a brief overview of the main concepts from Semantic Analysis Method (SAM) as a theoretical-methodological background to this paper. The SAM assists users or problem owners in eliciting and representing meanings in a formal and precise model. The meanings are specified in an Ontology Chart (OC) that represents an agent-in-action

ontology. In the SAM "The world" is socially constructed by the actions of agents, on the basis of what is offered by the physical world itself [36].

It is worth mentioning that the SAM´s concepts of ontology and agent are different from the concepts in use by the Semantic Web community. An OC represents a domain in study which can be described by the concepts, the ontological dependencies between the concepts, and the norms detailing the constraints at both the universal and the instance levels. Moreover the OC describes a view of responsible agents in the focal domain and their pattern of behaviour named affordances [36]. Some basic concepts of SAM adopted in this paper are based in Liu [36], and are briefly presented as follows:

"Affordance", the concept introduced by Gibson [18], is used to express invariant repertories of behaviour of an organism made available by some combined structure of the organism and its environment. In SAM the concept introduced by Gibson was extended by Stamper [70] to include invariants of behaviour in the social world;

"Agent" can be defined as something that has responsible behaviour. An agent can be an individual person, a cultural group, a language community, a society, etc. (an employee, a department, an organization, etc.).

"Ontological dependency" is formed when an affordance is possible only if certain other affordances are available. The affordance "A" is ontologically dependent on the affordance "B" meaning that "A" is only possible when "B" is also possible. The OC represents graphically these concepts.

4.4.2 Modeling ontologies for ISN

In the '*VilanaRede'* system, users express themselves through their profiles, announcements of products, services and ideas posted; and they communicate mainly through commentary about the announcements and chats between members of the network. These data are stored in the ISN system database and from these data we represent the semantics used in the social network in a structure called 'Semiotic Web ontology' [57].

According to Reis *et al.* [57] this structure is a semantic model (computationally tractable ontology) in which the SAM is used in conjunction with other technologies from the Semantic Web field to describe computationally tractable ontologies using the Web Ontology Language (OWL) [80]. In this paper such semantic model is constructed from a semi-automatic process along with the vocabularies shared in the social network. The idea is to incorporate the concepts of particular Agents (roles) and Affordances (patterns of behaviour) arising from the SAM into an expanded and more representative Semantic Web ontology. It is worth mentioning that the goal is not to create a "perfect ontology" from a theoretical point of view, but to produce practical and immediate results for search in ISN. Therefore

some properties from the OC may not be fully transcribed to OWL, while other aspects such as agent-affordance relationship are emphasized.

This approach is justified from a Semiotic perspective, since the signs are socially constructed. Thus, a computational model that represents the semantics of a SNS should contain the agents that interpret the socially shared concepts. With this approach we incorporate and take into account the Semantic Web ontologies concerns and possible representations arising from the Ontology in a Semiotic perspective. In addition to agents and affordances, we have observed that Semantic Web ontologies do not incorporate (at least explicitly) the ontological dependency relations, an existential relation in the model. The approach is also justified by the representational limitations shown in literature [*e.g.*, 73, 5756] regarding the use of ontologies in computing and their expressivity.

Within the conceptual model of a 'Semiotic Web ontology', the agents have behaviour(s) (affordances) related to a concept. For instance, a seamstress, which is an agent, can sew a "*manga"* (it means sleeve in English). Sewing is a pattern of behaviour of a seamstress (in other words an affordance). "*Manga*" is a concept that can have several different meanings in Portuguese (It can mean sleeve, fruit, color, etc.), but in this context due to the affordance and the agent ontological dependence, the meaning of "*manga"* is possibly associated to shirt and not, to "*manga"* fruit (mango in English) that can also be represented in the model, as suggested by Reis *et al.* [57].

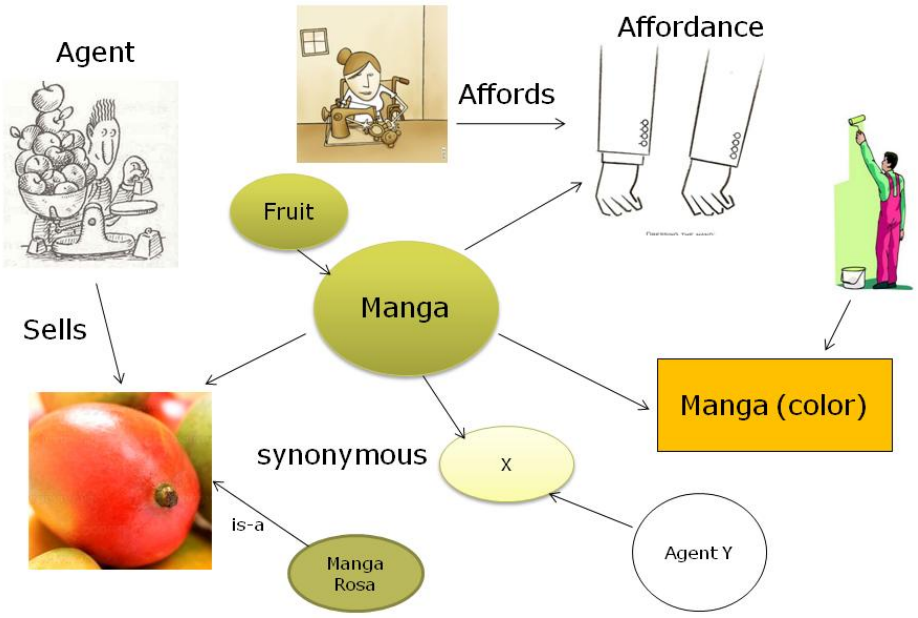


Figure 4.1: Modeling meanings according to 'Semiotic Web ontology'

Following the approach described in Reis *et al.* [57] for an SNS context imagine a scenario as illustrated by Figure 4.1. The grocer and the seamstress are agents that have affordances connected to specific concepts. This model can also have specific 'is-a' relationships; *e.g.* '*manga rosa*' is a specific kind of mango. This also shows that concepts can be related to several agents and affordances and with other concepts, constituting relations and representations that make more complete ontologies, when compared to conventional ontologies described for a domain. For example, '*manga*' can also mean a color for a painter who is searching something in the network, as well as '*manga*' can have any synonym that makes sense for an agent 'Y' modeled from the data of the social network.

In order to develop this representation for the inclusive search mechanism, we propose an assisted method (semi-automatic) with several distinct steps; the method is illustrated in Figure 4.2. It includes: (1) the extraction of terms and possible semantic relations from the database of the ISN system; (2) the creation of an OC (from SAM); and (3) the creation of the final OWL ontology.

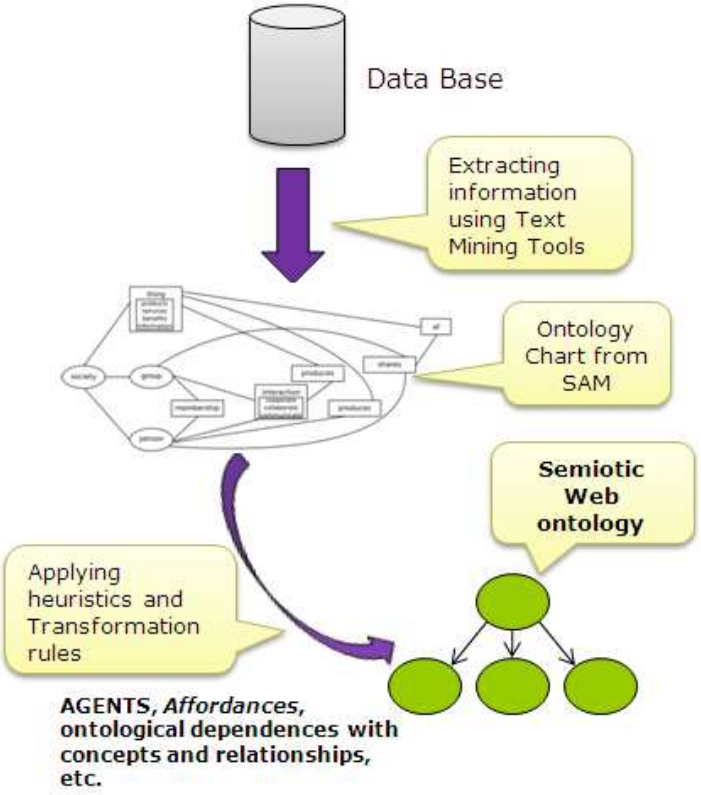


Figure 4.2: An illustration of the Semi-Automatic Method

In this assisted method, the first step deals with the data from the system database. This step takes into account the social relations in the network, and must provide the necessary well defined data (a list of concepts, agents and affordances, etc.) to build the semantic model. The next step involves the building of an OC (from SAM) by an ontology engineer. This intermediate ontology diagram is important to identify the possible agents in the ISN and their patterns of behaviour.

In the third step, from the OC, a set of specific heuristics and transformation rules must be applied to create an initial OWL ontology (computationally tractable), extending the computational development of the SONAR CASE tool [65]. Bonacin *et al.* [7] proposed a heuristic to transform OC into system design diagrams; however those heuristics must be adapted to our purpose. During the modeling of the meanings the ontology engineer can also be supported by existing tools for Ontology Learning and Engineering.

4.4.3 Extracting Information from the ISN to Build 'Semiotic Web ontology'

Regarding search in ISN, the semantic mechanism in a knowledge domain representation must consider the activity of the social network. This includes people's local and everyday language used in the network. For that, this requires tools and techniques that make extraction and text mining from the system database in order to discover and model the semantics shared by people in the social network.

Thereby, the first step of the proposed method is to extract relevant information from the database of the ISN. The objective of this step is to support the ontology modeling from data created and shared by ISN users. We have conducted a study [53] about tools and techniques for the identification of concepts and semantic relationships that come out from the ISN data. The objective was to create a designing strategy to assist the modeling of ontologies that represent the semantics shared in the social network based on the idea of Semiotic Web ontologies. In order to accomplish that, we have investigated tools and techniques that could aid in this step. We wanted to know which tools could support in discovering of relevant concepts and semantic relation between concepts. The main challenge is the heterogeneity of the content available in the ISN. The '*VilanaRede'* content (*i.e.* its announcements) was used to conduct such study. In this study we investigated several text mining tools. Among the tools described in literature, we have chosen the keyphases extractors KEA [43], tools for term extraction in the Portuguese Language such as ExATOlp [38] and algorithms for clustering like CLUTO [33].

Results showing positive and negative cases of the outcomes were elicited. Moreover, algorithm procedures were created to verify the intersection of results from the outcome of each tool. Tables illustrating the results are also described to show the terms extracted with

their relative and absolute frequency. The results indicate that the more adequate approach to analyze the network information is to verify both: the data captured by individualized announcements together with data independent of any announcement. The keyphrases extracted by KEA on each announcement inform about the subjects discussed in the network, while the approach utilized by ExATOlp provides a general vision of all considered announcements. The pos-processing done with terms organized by semantic categories shows useful as well as to analyze the terms repeated by all tools, since this can indicate the concepts that are mandatory during the ontology modeling.

Furthermore, the results allow to point out that not only the tools and techniques alone are important, but also how it is organized and used by ontology engineers in order to make decisions based on different information and perspectives. Finally, the results obtained by applying the tools and techniques on real data from '*VilanaRede'* showed promising in supporting the building of ontologies that represent the meanings used in the ISN.

4.4.4 Outlining an Inclusive Search Engine

The main objective of the proposed method is to create valuable information to inform the inclusive search mechanism. When the user is logged in the ISN and he/she enters with some search term(s) in the search engine, the system starts a process of finding out relationships of this/these keyword(s) with the available SWO. For example, suppose the user types the term "small snack". If there is nothing in the system with this expression, from the analogies and semantic relations made, the system may return some other types of food semantically close. Likewise if the user enters the word 'food', all advertisements related to food should be returned.

There are several architectural proposals for semantic search solutions, such those described in Mangold [39]; Reis *et al.* [56] describe an overview of semantic search solutions applied to SNS. The decisions and architectural strategies for resolving the semantic search in this implementation is carried out in accordance with the requirements of an ISN and follows the recommendations proposed by Reis *et al.* [56, 55]. The main difference in the search solution of this proposal is to take into account information regarding the user that is making the search (from his/her profile) and the user that produces the content; Figure 4.3 illustrates this idea.

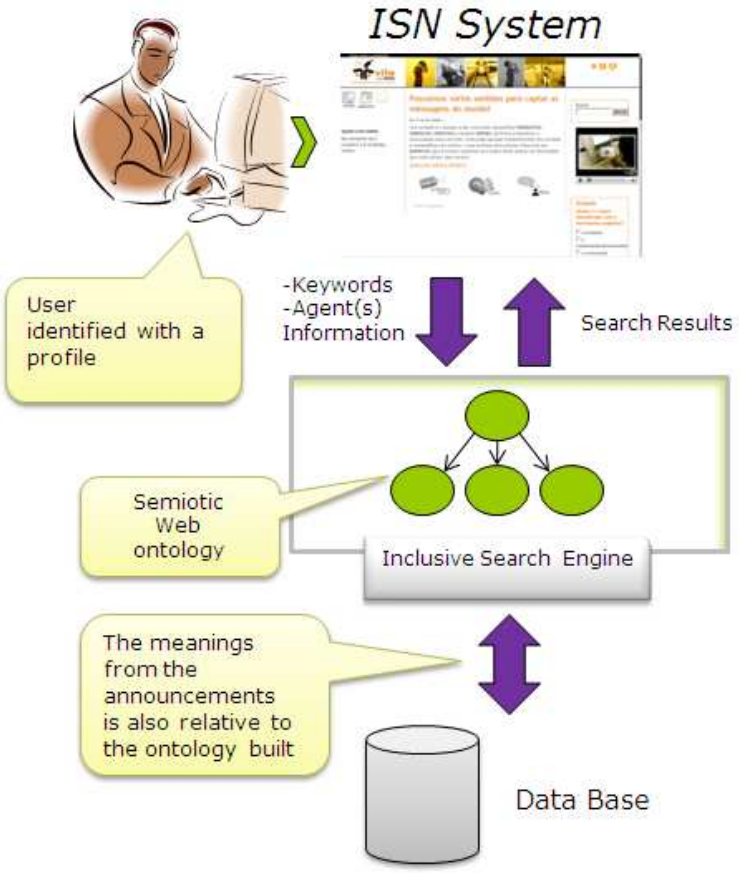


Figure 4.3: An illustration of the proposed inclusive search mechanism

In this strategy, the user profile is important due to the possibility of discovering a context for the search terms. Mainly from the user profile, we aim to identify some adequate agent(s) represented in the ontology and make a user-agent modeling as described by [54]. Thus we delineate (or even limit) the search space, making a relation between the user and the generated semantic model. For instance, imagine that a biologist is logged into the system (we could find that a user is a biologist based on his/her profile) and requests a search with the keyword 'crane'. If there is a relation between the 'biologist' agent and the term 'crane' in the ontology, most likely the results (announcements) that could be returned first (ranked first) should be related to the concept of crane as a 'bird', not the other meaning(s) of this word. However, to a civil engineer that searches the same word, maybe the results that most interest him / her refers to the construction equipment and not the bird. Semantic Web Rule Language (SWRL) [81] rules as illustrated by [54] can be used to relate agents to certain meanings enabling to deal with such situations. Besides, we do not

mean that other results are not required or may not be returned in response to the search, (the engineer may want to know about this kind of bird), but the announcements from the social network that relates 'crane' with a construction equipment must have greater relevance in the ranking of results.

The agent-affordance relation is also used to indicate the probable meaning of the words in the announcement. For example, based on the user that entered a particular announcement that mentions the word 'crane', we could verify whether the word 'crane' refers to a 'bird' or to a 'construction equipment'. If the user who submitted the announcement is a 'biologist' agent, 'crane' would be most likely a 'bird' in this announcement. In a similar way, if the advertiser is a civil engineer, 'crane' probably would mean a 'construction equipment'. According to Reis *et al.* [54] we also could have relationships between agents and could verify how much an agent is semantically close to another and indicate the probable meaning based on this aspect.

4.5 Discussion

In the proposed approach, the support for better results from the search engine demands a careful modeling procedure. Different signs with the same meaning (synonyms) coming from different virtual communities of the social network can be discovered having the opportunity to be represented in the ontology; such signs and meanings can be purely regional. Thus, they could not be present in formal dictionaries or thesauri generally used by conventional search mechanisms. Furthermore because they are cultural expressions emerging from the social network, the ontology would potentially provide smarter and richer search results when compared to ontologies based on domains or formal definitions.

The approach provides means to discover and distinguish the meanings used in the SNS, representing them through the agents in the SWO. Differently from conventional computing ontologies and other approaches to semantic representation, our proposal involves adding the agents and affordances concept in the search. This addition can cooperate for richer search results treating the polysemy problem in not restricted or controlled language contexts in ISN. Moreover, the inclusion of the agents and other concepts from SAM in the Web ontology can aid improving the search mechanism, generating more adequate results to an ISN context.

Considering users with limited literacy and with difficulty in dealing with technological artifacts (digitally illiterate), it is important to let them perform the search using their daily language since usually is what make sense to them, and to provide search results more natural and adequate to their lives. Thus, the search engine should reflect the semantic reality of the social network users. A search engine with such characteristics could create opportunities for

inclusion, since the method for building the semantic model as well as the strategies to use the ontology suggests that the returned search results will tend to make more sense for the user that searches.

Some recent studies in the literature address search solutions for SNS [*e.g.* 77, 22, 23]. These works are particularly focused on searching just the users' profile in the network; the work of Choudhari *et al.* [15] makes progress in the development of semantic search in SNS, however their work have the same limitation and does not use ontologies to perform the search. Regarding semantic search but not strictly related to SNS context, several proposals and solutions are illustrated by the survey of Mangold [39] and Wei *et al.* [79]. Ontology based semantic search solutions [*e.g.* 24, 21, 8, 16] as well as ontology based query expansion [*e.g.* 6, 26] have enhanced techniques for semantic search applications. In order to implement a solution and make improvements to a search engine of an ISN, future research includes a detailed observation of more ontology based query expansion approaches to use the 'Semiotic Web ontology' method. Other approaches [*e.g.* 75] have tried to take advantage of the 'faceted browsing paradigm', employing a solution integrating semantic search and visual navigation in graphs using the idea of social networks.

Previous work conducted by Reis *et al.* [56] have discussed the challenges related to search in ISN; the authors propose recommendations for a search engine better suited to this kind of system. Furthermore, the proposal of Reis *et al.* [54] for a search informed by a Semiotic approach in SNS is the main work that we have based on to prospect an inclusive solution. To the best of our knowledge investigations that have specifically focused on semantic search in SNS considering aspects of accessibility and inclusion were not found in the literature so far. We argue that the development of a search engine more suitable for an ISN should include these new challenges and must be informed by a Semiotic approach.

Also, the approach developed in this paper can methodologically and technologically improve and expand Semantic Web techniques, such as Web ontologies, illustrating immediate and practical results for better ISN search engines. This approach differs from others, since the search solution outlined tries to derive the meaning of the search terms and also the meaning of the terms from the ISN content produced by users based on the agents and affordances. Future experiments with real users should be conducted to verify whether our approach can bring promising benefits revealing search results more suitable to the context of social and digital inclusion, and also to SNS in general.

4.6 Conclusion and Further Work

Social network systems may provide inclusive access to digital information, creating situations where the users' diversity is respected and the access difficulties minimized. This is

the purpose of the Inclusive Social Network Services (ISN). In this context it is important to provide information retrieving in a more natural way from the user's point of view, with results that make sense to people. Therefore, more appropriated mechanisms for search should take into account the meanings created, shared and used by people in the social network.

This paper presented new perspectives for search in Social network systems which consider the inclusive social context. It showed the outcomes of an analysis regarding how to improve a search mechanism considering aspects related to the digital and social inclusion. We could verify with real users that semantic aspects can make a difference for people to reach information, and that the current syntactic search engines are not enough for an ISN context. Inspired on the practical context of ISN users, and based on the approach of 'Semiotic Web ontology' this paper outlined an inclusive search mechanism for SNS.

As further work, the goal of this research is to improve (in the implementation sense) the ideas drawn up for the search mechanism described in this paper. For that, we aim to develop the semi-automatic tool for building SWO, as an extension of the SONAR tool, including the heuristics and transformation rules to build the OWL ontology aided from OC. Furthermore, we intend to develop a pilot implementation of this search engine based on the 'Semiotic Web ontology' in the '*VilanaRede*' system, using and improving the strategies mentioned in this paper. Also the work involves new practical experiments in a case study with real users, utilizing this novel search mechanism in order to evaluate and validate the solution with empirical results.

Capítulo 5

Addressing Universal Access in Social Networks: An Inclusive Search Mechanism[∞]

5.1 Introduction

The World Wide Web (Web) may be considered as one of the more striking inventions of humanity history. Its evolution has influenced life in society, boosting radical changes in the way people behave, communicate, and interact with each other in their daily life. The contemporary world is at a unique moment in which people tend to be continuously connected to various information sources and digital artifacts. The dominance of Information and Communication Technologies (ICTs) is an essential prerequisite to be socially included in the Information Society (IS) age. Consequently, nowadays, social and digital inclusion allowing access to information for all members of society is not only necessary but critical. However, there are still many kinds of barriers that prevent people from interacting with ICTs [2].

The advent of the Social Web [9] has brought new opportunities for knowledge access and sharing. The Social Web is more than connecting pages and contents; it interconnects people, organizations and concepts. Social Network Services (SNSs), Wikis and Blogs are examples of Social Web applications that allow an intensive interaction between users. This interaction generates a huge amount of information, and consequently requires specific

[∞] Esta é uma versão de um artigo submetido para um *Journal* Internacional.

mechanisms to recover relevant information for the users. Usually, the process of finding and using the digital content in these systems depends on a search mechanism.

At the same time that SNSs impose challenges regarding information access, they may have a dominant role in influencing the adoption and use of the ICTs [52]. Moreover, SNS represents an opportunity for social and digital inclusion through Inclusive Social Network (ISN) systems [45]. We understand the ISN as a type of SNS for all; *i.e.* a system which allows the vastest diversity of people to integrate a group and interact to produce information (tangible and non tangible objects) that can be shared with other persons and groups. A SNS with resources to promote access for all, including those at the margin of the digital culture, may be defined as an ISN.

Within an ISN all people must have means for recovering information that makes sense to them. This paper proposes more adequate search mechanisms for people in the process of digital literacy and frequently with limited education. In developing countries a small part of the population has access to Internet or experience with ICTs; an even smaller part of the population has effective access to valuable and desired information through the Web. The effective access to information is much more complex than providing a computer with Internet connection; it demands, for instance, the ability to deal with computational artifacts and an adequate/minimal level of literacy. The idea should be to stimulate a beneficial cycle in a learning process, where the more that valuable information is recovered the easier it will be to recover information.

Users in the process of becoming digitally literate need a search mechanism that enables them to easily reach information, with useful and meaningful search results. The results should be associated to other words that should be understood and potentially should facilitate the next search iteration. In this sense, the search mechanism could be a tool for digital inclusion, especially in SNS where experts and novices effectively communicate and share information in a mutual development process.

The challenge addressed in this paper is how to develop a search mechanism more aligned to the ISN concept, which addresses universal access and digital inclusion issues. This search mechanism should be capable of interpreting representations/models of meanings used by people in a SNS [61]. With this in mine, we explore methods and approaches to reliably capture the social and cultural aspects of the SNS, including its colloquial language and shared meanings.

This paper proposes methods and solutions to construct search mechanisms aligned to the ISN concept, pointing out that such solutions could contribute to participatory and universal access to knowledge. Therefore, the objective is to facilitate interaction with this functionality, returning well-suited search results for all, and, at the same time, respecting the individuality of each user. We propose an ontology-based solution to design such a

mechanism. This solution is based on a novel approach for the design of Web ontology [57], which articulates aspects of semantic modeling with strategies to create the ontology based on the content that comes from the ISN system. The method was applied in a case study with real users in the context of the e-Cidadania project [3], which investigated methods and tools for the construction of ISNs. A preliminary evaluation of a prototype was made with target users in order to verify whether the solution represents more inclusive solutions. Challenges and directions for future research are also presented.

The paper is organized as follows: section 5.2 presents the motivation for this study including the current search solutions for SNS. Section 5.3 presents the theoretical methodology adopted. Section 5.4 describes an approach and method for inclusive search mechanisms (ISMs). Section 5.5 proposes a solution for an ISM. Section 5.6 presents the case study showing how the method was instantiated and the mechanism implemented. This section also presents the results of a preliminary evaluation of the mechanism with a target audience. Section 5.7 discusses the solution and the empirical findings. Finally, Section 5.8 summarizes the conclusions of the work, and points out further research and challenges to be addressed.

5.2 Search Mechanisms: A Barrier to Universal Access in SNS

In the last years the idea of universal access has become extremely important to ensure the acceptability of the emergent IS. Universal access means accessibility and usability of the technology for anyone, in any time or place. Its objective is to allow equitable access and active participation of all citizens in computer-mediated human activities [71].

Commonly referred as the 'digital divide', disparities and inequalities in access to the Internet is considerable, especially in developing countries. Universal access goes beyond providing unrestricted access to the Internet. It is necessary to eliminate the barriers in the interaction between people and computer systems, especially in the Web context. Even more important than the physical access to the Internet is universal access to the knowledge generated in digital media, which is at the core of a fairer information society. ISN still needs a novel search mechanism specially designed to facilitate access to information. This type of solution is motivated and justified by the context of significant inequalities and differences that need to be addressed in the ICT domain.

5.2.1 A Social Context of Digital Divide and Diversity

In developing countries such as Brazil, India and China, the access to information and knowledge is still restricted to a small portion of the population. These countries face a

situation of vast diversity regarding culture, geographical regions as well as differences in socio-economics and access to technology. There are technological, educational, cultural, social and also economic barriers that have prevented access and interaction with technology [2]. In this context there are still many people without access to the Internet and consequently without opportunities to access information and knowledge.

As stated by the Brazilian National Research by Household Sample produced by the Brazilian Institute of Geography and Statistics [10] (IBGE in its Portuguese acronym) in 2008, 65,2% of the population did not have access to the Internet. The Brazilian Internet Steering Committee [11] states that in 2009, 47% of the population had never used the computer and 55% had never accessed the Internet. Moreover, data from the International Telecommunication Union (ITU) [29] in 2009 reveled that in India, with a population of over 1 billion people, despite having the fastest growing rate of new Internet users, only 7% of the population use the Internet.

Regarding education, the Indicator of Functional Literacy in Brazil (INAF in its Portuguese acronym) [28] points out that in 2009, 27% of the Brazilian population between 15 and 64 years old was considered functionally illiterate, defined as the population with less than 4 years of schooling, and unable to perform simple tasks involving the understanding of reading words and phrases. Using a broader concept of functional illiteracy, according to the same Indicator, the majority (52%) of Brazilians reach the degree of rudimentary literacy, *i.e.*, they only have the ability to locate explicit information in short texts or do simple math. However, they are not able to understand longer texts, and 9% of these individuals are considered absolutely illiterates. Only 27% of the adult population is considered fully literate.

This scenario illustrates only part of the challenges that we face in terms of designing systems which should improve the social condition of these people. According to Baranauskas & Souza [2] this problem is difficult because it is unique. Its treatment requires multidisciplinary expertise, towards systems, methods and techniques to enable the establishment of a digital culture by providing barrier free access to information to every citizen, respecting their differences.

5.2.2 Search Strategies for SNS

According to Jamali & Abolhassani [32] SNSs connect a set of entities via social relationships and facilitate information exchange. There has been a rapid increase in the number of users signing up for, and actively using, social networking websites such as *Facebook*[17], *Flickr*[18],

[17] www.facebook.com

Orkut[19]*, MySpace*[20], etc. [22]. The social networks connect people with common objectives presenting a virtual space propitious for user communication and for sharing information, knowledge and social experiences [32].

Access to the information generated in the social network can be realized mainly by a search mechanism. Usually, such mechanisms are the main functionality utilized by users to find content or people in the social network, and consequently it can either facilitate or interfere with access. Thus, it is necessary to develop mechanisms specially designed and adequate for the ISN context.

The currently available search solutions for SNSs assume that the majority of people dominate the use of technology. Therefore, these solutions are typically generic, impersonal and built primarily through comparisons of keywords with lexical-syntactical information processing (syntactic search). Search engines may provide a huge number of answers in response to a user query, many of which may be completely irrelevant, whereas some of the most relevant are not found. In order to improve the relevance of search results, various alternatives and strategies have been proposed in the literature. Usually searches in SNS use approaches of social distance [*e.g.* 82, 77, 23, 51]. They do not focus on the content created by the participants nor on the semantics of the words they use.

Solutions for search mechanisms in SNS based on various strategies have been presented in the literature (*e.g.* recommendation and trust using the social distance between people in the SNS) to rank the search results. The pioneer research of Yu & Singh [82] in this area proposes a solution to find a specialist participant (reputation system) in a SNS utilizing a multi-agent system. The approach of Gürsel & Sen [22] focuses on recommendation and trust to search items in an SNS, while Vieira *et al.* [77] propose that the friendship relationship is an important evidence to find people in a SNS. They argue that the position of the search results may be modeled in function of the distance between users in a graph, more specifically, the shortest path in a friendship graph. The same idea is defended by Haynes & Perisic [23].

Motivated by the problem of information overload in Social Web systems, the work conducted by Pereira & Silva [51] presents a user-focused approach, called Folkauthority, for the improvement of the results obtained in the information recovery stage. Folkauthority is a neologism proposed to designate authority to information sources by means of folksonomy. Applying the Folkauthority concept corresponds to allowing users of the system to give cognitive authority to information sources in the system. The information is categorized by their own users. When there is a formation of a chain of authorities, at the

---

[18] www.flickr.com
[19] www.orkut.com
[20] www.myspace.com

moment of object retrieval, the prioritization of objects is given by authorities in the subject. This proposal is interesting, but requires the system users to be aware of the folksonomy concept and also to know how to categorize things using tags. In an ISN context it is not possible to assume that all people will have such 'digital proficiency'.

In an attempt to improve and apply specific approaches and techniques for information retrieval in SNSs, all the proposals mentioned above evolved with important research results. Nevertheless, they still consider the syntactical processing of information. The objective of optimizing search results has also motivated research in the semantic field by incorporating techniques from a variety of other research fields, and implementation of a number of practical systems [78]. The need to retrieve semantically enriched information suited to the user's goal gave rise to an increasing interest in research on ontologies and semantic aspects. According to Wang *et al.* [78], conventional search techniques are developed on the basis of word computation models and enhanced by link analysis. Semantic search extends the scope of traditional information retrieval (IR) paradigms from mere document retrieval to entity and knowledge retrieval. It improves the conventional IR methods by looking at a different perspective, *i.e.* the meaning of words [78].

Search mechanisms with semantic characteristics demand that the mechanism be based in the knowledge model of the domain, *i.e.,* the knowledge must be computationally represented so that the machine can "interpret" it. The knowledge can be represented by means of ontologies. It is important to distinguish between Ontology that is a branch of philosophy that studies the nature of the existence and the structure of reality [31], from ontologies in the context of Computer Science. A widely adopted definition for ontology in Computer Science is provided by Gruber [19] who explains that ontology can be understood as a specification of a conceptualization which provides descriptions about knowledge. Semantic Web (SW) ontologies usually are described by computational ontology languages such as the *Web Ontology Language* (OWL) [80].

According to Kassim & Rahmany [34] the traditional search engines are no longer able to provide precise results due to the huge volume and complexity of the information in the Web. The deficiency of these mechanisms is that they are just keyword-based, and also are not capable of dealing with polysemy and synonyms aspects. Thus, many times they return results that do not meet the users' needs.

Semantic search has become an alternative to overcome the deficiencies of such traditional mechanisms. Semantic search mechanisms try to analyze what a user desires during a search in a context through a 'logical reasoner', enabling better results. Works such as Heflin & Hendler [24], Guha *et al.* [21], besides Bonino *et al.* [8] and Fang *et al.* [16] have introduced the idea of using ontologies in semantic search mechanisms.

There are several architectural proposals for semantic search solutions. Mangold [39], Wang *et al.* [79], Hoang & Tjoa [26] and Hildebrand *et al.* [25] have made an extensive revision of the main proposals of semantic search solutions in the literature. They describe open research questions, and necessary investigations, as well as similarities, goals, applications, methodologies and technologies involved in the various proposals.

Based on the analysis of literature in SNS semantic search [56], we noticed that most of the existing work focuses on social searches (searches for users) and do not deal with the content generated and shared by the individuals. For example, Choudhari *et al.* [15] describe an architecture to develop a semantic search for SNSs using an approach based on clustering semantically related terms. While the social semantic search is certainly important to connect people and to many other activities in a SNS, it does not on its own solve the problems related to universal access to SNS content, since it does not consider the real meanings shared by people.

5.2.3 Preliminary Findings

Based on the context and the literature previous illustrated, we started our investigation with the analysis of the behavior of the prospective users of an ISN when they use search mechanisms. This study was conducted with real users using a syntactic search mechanism of an SNS system. The participants were asked to search within proposed scenarios; the objective was to observe their behavior while using the search mechanism. We observed whether semantic aspects could make a difference in their search situations. In this study a task sheet with 4 search scenarios was presented to each of 7 pairs of participants. The pairs were formed by the users themselves, and for each scenario the pair would write the words used in the search and the title of the information (announcements) retrieved from the ISN [61].

This analysis revealed the importance of considering semantic aspects in the design of a search mechanism that is more adequate for the ISN concept [61]. To improve search results for this context it is necessary to take into account the way in which the users' ordinary language influences meaning in the SNS (*i.e.* users' local/colloquial meaning should be considered during the development of more appropriate ISMs). Thus, such mechanisms should take into account the meanings created, shared and used by people through the use of the system [61]. Moreover, the results also pointed out that it is necessary to construct computationally tractable models from the semantic point of view derived from the social network itself.

We have noticed that users formulate queries that are related to a given social context. The more adequate search results (*i.e.* content from the SNS) must be meaningfully related to the user that performs the search [61]. The meanings applied to words in an inclusive

mechanism must be considered according to who produces and who consumes the information. The main challenge in this context is to provide a harmonious experience between the mechanism and the users in their colloquial language [56].

Thus, in order to develop a search mechanism more suitable for the ISN concept, which considers the users' informal language, the solution should reflect the semantics used by the participants of the system [61]. Besides taking into account the content from the ISN system to model the semantics, the meanings must be modeled according to this social context and it is necessary to clarify the pattern of behaviors shared in the SNS. In this way, methods are needed to computationally model the meanings relative to a context, and to a person or a social group that share meanings and behave according to commitments, linguistic constraints, culture and other social aspects. Once based on Semiotic theory [50], semantics is understood as the interpretation of signs by individuals in association with real world elements. This interpretation is socially contextualized*; i.e.* individuals and communities may have different interpretations.

5.3 Organizational Semiotics Theory and Methods

This section presents a brief overview of Organizational Semiotics (OS) as a theoretical background in this work. The Semantic Analysis Method (SAM) is also presented in order to clarify the proposed solution.

5.3.1 Organizational Semiotics

Semiotics, the doctrine of signs, leads to an understanding of information as properties of signs. Anything standing for another thing or used to signify something else [50] is an example of a sign: words, sentences, traffic lights, diagrams, the wave of a hand and facial expressions. OS understands an organization as a system of signs, and studies organizations using concepts and techniques rooted in Semiotics [49, 67, 36]. OS can be understood as one of the branches of Semiotics developed by Peirce [50] and others. The rationale behind OS is based on the assumption that any organized behavior is affected by the communication and interpretation of signs by people, both individually and in groups [36].

The OS discipline studies the nature, characteristics, function and effect of information and communication in organizational contexts. Organization is considered a social system in which people behave in an organized manner by conforming to a certain system of norms [36]. In this work, a SNS is also understood as an organization since there are communication rules, and some behaviors are probably expected in the communities. Social rules, often implicit, state what is allowed and what is not.

From a philosophical point of view, OS understands reality as a social construction based on the behavior of the agents who participate in it. People share patterns of behavior governed by a system of signs. This philosophical position states that, for all practical purposes, nothing exists without a perceiving agent or without the agent engaging in actions [36]. Methods from the OS are useful to delineate the influence of the social aspects in the organizations.

Among the methods employed by the OS community is a set of methods known as MEASUR (Methods for Eliciting, Analyzing and Specifying Users' Requirements) [70], which deals with the use of signs, their function in communicating meanings and intentions, and their social consequences. MEASUR involves the analysis of stakeholders in a focal problem, their needs and intentions, and the constraints and limitations related to the prospective software system. In this work the Semantic Analysis Method (SAM), from MEASUR, is adopted as part of the proposed approach to model the knowledge coming from the ISN. The following section presents a summary of SAM.

5.3.2 The Semantic Analysis Method

Using a subjectivist philosophical stance and an agent-in-action ontology the SAM determines the underlying semantics of a social context and the relationship between the human agents and their patterns of behavior [64]. SAM supports the analysis, specification and representation of a social system.

The SAM assists users or problem-owners in eliciting and representing their meanings in a formal and precise semantic model: the Ontology Chart (OC). OC is a graphic representation of a conceptual model that describes a view of responsible agents in the focal domain including their pattern of behavior, referred to as affordance(s), and the ontological dependences between them [36]. The meanings are specified in the OC, which represents an agent-in-action ontology. The OC is read from left to right; any element is dependent for its existence on the affordance(s) to its left to which it is connected. The meaning of the words used in the semantic model is treated as a relationship between the signs and the appropriate actions of the agents.

In the SAM, "The world" is socially constructed by the actions of agents, on the basis of what is offered by the physical world itself [36]. It is worth mentioning that this concept of ontology and of agent (from SAM) is not the same used by the Semantic Web (SW) community. An OC represents a domain under study, which can be described by the concepts, the ontological dependencies between the concepts, and the norms detailing the constraints at both the universal and the instance level [37]. Some basic concepts of SAM adopted in this paper are based in Liu [36], and are briefly presented as follows:

"Affordance", is a concept introduced by Gibson [18] that can be used to express the invariant repertories of behavior of an organism made available by some combined structure of the organism and its environment. In SAM [70] the concept introduced by Gibson was extended by Stamper to include invariants of behavior in the social world; and affordances are social constructs in a certain social context [36]. The social world acts as the environment that is constantly affecting the agents' behavior, at the same time that it is affected by the agents' actions.

"Agent" is a special kind of affordance, which can be defined as something that has responsible behavior. Agents are affordances that can take responsibility both for their own actions and the actions of others. An agent can be an individual person, a cultural group, a language community, a society, etc. (an employee, a department, an organization, etc.).

"Ontological dependency" is formed when an affordance is only possible if certain other affordances are available. We say that the affordance "A" is ontologically dependent on the affordance "B" to mean that "A" exists only when "B" does; *E.g.:* for a person to be able to stumble, he/she must first walk; for two people to divorce, they need to be married; thus there exist an ontological dependency between to stumble and to walk, and also between divorce and marriage.

The SAM has other important characteristics to be considered. One of the SAM principles, according to Liu [36], is that an analyst is not allowed to invent artificial terms or introduce new concepts when modeling the agents' actions in the OC. The purpose of this is to force the analyst to speak the same language as the problem-owners. Any ambiguity in the terms or concepts used in describing the problem should be resolved putting them into a context of actions which are already described and understood. When doing so, if the problem-owners are inspired with some new terms, they may be used only after a careful justification by the problem owners and the analyst. The reason for this is that the world to be modeled is constructed by the community of agents, *i.e.* the problem owners. The agents know the meaning of words in their own world, their interpretations are the only ones justified.

Figure 5.1 illustrates an example of OC [64]. In this OC, if there is a library and there is a person, then that person can have membership in the library. Note that both the library and the person are defined in respect to a specific Society and are both agents. Membership in the library (an affordance) is dependent on the existence of two antecedents: the library and the person. A person who has membership in the library is defined to have the role of member. Observe that books are defined relative to Society, although written by persons, they continue to exist after the persons no longer exist; so they are dependent on Society. Existence of the library and the book leads to the affordance 'stocks', and a combination of library membership and the library stocking books leads to

the affordance 'borrow'; members can borrow stock from the library. At this point the member has the role of borrower. These ontological dependencies are illustrated graphically in the form of a chart (Figure 5.1). Reading this OC, to borrow, the stocks and membership must exist, which depend on the existence of the library, an agent in that Society.

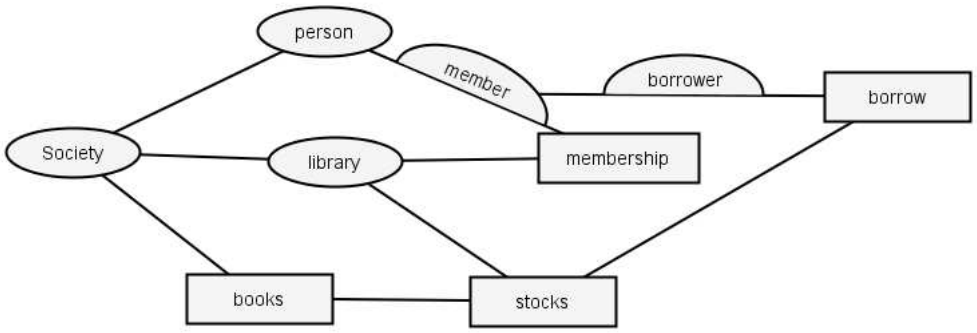


Figure 5.1: An example of OC in a library domain [64]

5.4 The Semiotic Informed Web Ontology Design

Since the meanings are socially determined, a novel approach for the design of Web ontologies grounded on semantics was proposed to inform ISMs. This proposal includes the role of the individual who interprets signs (represented as concepts in an ontology) into the Web ontology. Therefore, the concepts of 'Agents' and 'Ontological Dependences' that come from the SAM outcome are included into Web ontologies described in OWL [57]. With that, both the 'Agents' and 'Affordances' are transformed into OWL classes, and are related to each other following the ontological dependences modeled into the OC.

In this approach, instead of modeling the meanings from the network using OWL directly, they are first modeled in an OC using SAM. In order to achieve an OWL, ontology, heuristics and transformation rules presented in [60] are applied to produce a 'Semiotic Web Ontology' (SWO) [57]. SWO is a semantic model (computationally-tractable ontology) constructed by including SAM concepts as Agents, Affordances and Ontological dependences, thereby combining SAM concepts with SW technologies. The heuristics aim to transform the OC artifact into OWL code. Some relations between the models are mapped, and one model supports the construction of the other, providing benefits presented in the different viewpoints. The heuristics represent a procedure that makes explicit the relationships between the models, and makes possible the construction of one diagram from the other. They were implemented in the SONAR Case Software Tool [65].

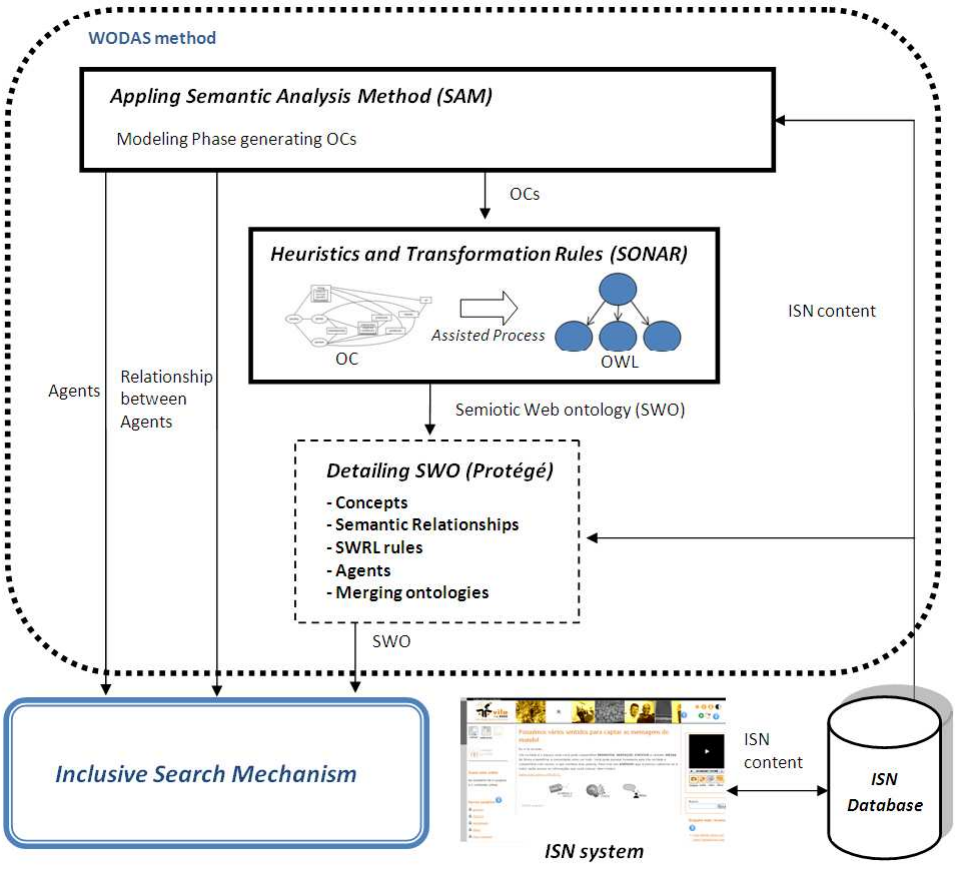


Figure 5.2: The WODAS method to create SWO

Using the SWO approach to inform a search mechanism may lead to search results that are more adequate to the ISN concept, since the shared meanings were modeled using an agent-in-action ontology grounded in the concept of affordances. In this sense, the concepts modeled in a SWO are not determined by the object itself; they are a combination of the world with who interprets it.

Figure 5.2 presents the WODAS (**W**eb **O**ntology **D**esign **A**ided by **S**emiotic) method to create SWO in a SNS context. Based on ISN content, SAM is applied generating one or more OCs. The OCs generated are used to produce the SWOs through the heuristics and transformation rules [60]. Such OWL ontologies can be merged as well as detailed (*e.g.* including more classes, instances and also SWRL - Semantic Web Rule Language - rules) using an OWL editor. Besides the detailed SWO, the ISM uses the agents modeled, and also possible relationship between the agents, during the search time.

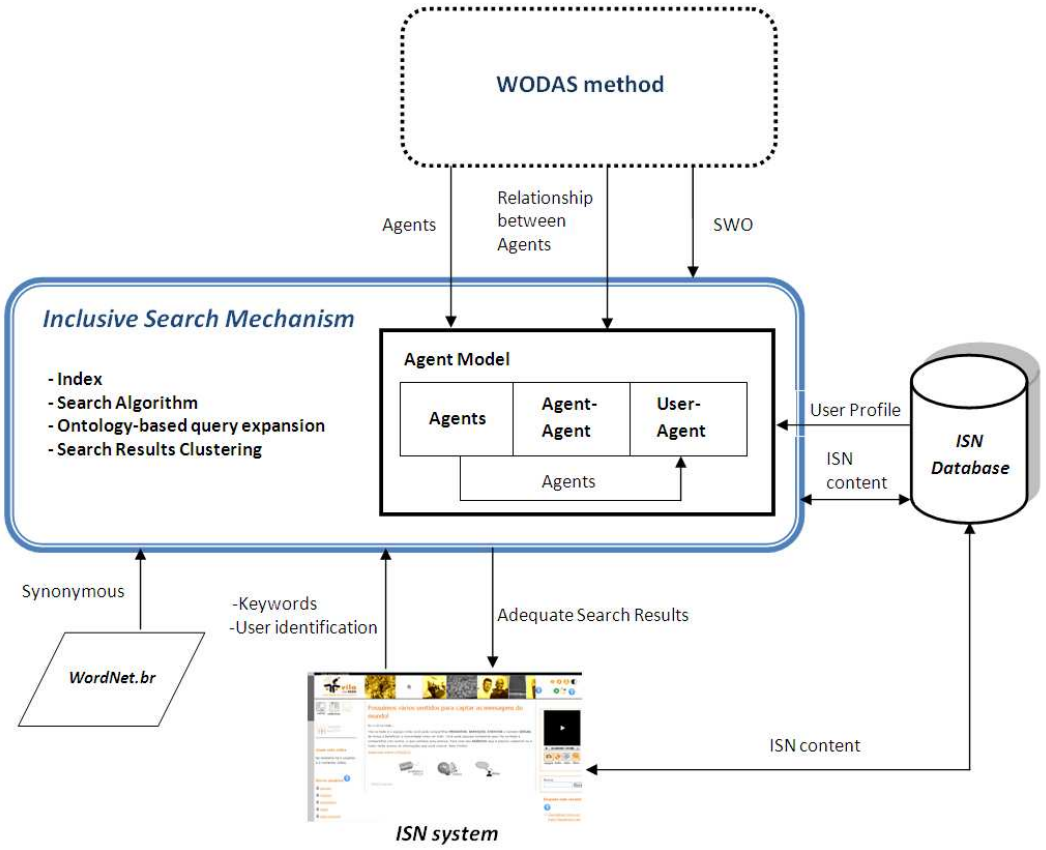


Figure 5.3: The Inclusive Search Mechanism

The search solution is based on SWO semantic information. It is used to classify words (both search keywords and words from published contents) and also to make query expansion during the search time. Besides the SWO, the ISM requires information about the relationship between each SNS user and the agents modeled, and also possible relationship between different agents. Information from the users' profile and their activity in the SNS are necessary to connect each user to a possible agent. Figure 5.3 shows a general vision of the relation between the ISM and the WODAS method.

The proposed ISM solution is built on the possible stated meanings based on the identified agents and their relationship with SNS users. This approach tries to treat polysemy, synonymous, and other aspects of human language during the search, through the identification of "local" meanings. For that, we propose to connect users to agents as explained below.

Modeling user-agent relationship: During the search time, the ISM uses the connections made between the agents represented in the ontology and the SNS users. Therefore, it is necessary to create a user-agent relationship beforehand. Users themselves specify their profiles using SNS system features. They can inform their profession, related activities in which they are involved, among other concepts that can socially characterize them. The ontology engineer makes a relationship (connection) of a particular user group in the SNS system with appropriate multiple agents represented in the OC based on this information as well as on activities of the user in the SNS (*e.g.* the content that he/she has shared). This can be modeled by using a priority rank for the different agents that suits the user, for example, from all agents related to one user, the first is the most important and the last the least. Additionally, it is also necessary to discover the relationship between the agents, *i.e.,* how much one agent is semantically close to another, *e.g.*, a 'doctor' agent is more related to 'medical laboratory scientist' agent than to a 'hairdresser' agent. The ontology engineer also has this role to model the relationship between agents.

Modeling SWRL rules to get meanings: SWRL rules are described to computationally handle the agent-term relationship. These rules are modeled into the SWO by the ontology engineer. The rules are used both during the search index processing and also during the search in an attempt to semantically classify the words.

Search Index: During the modeling phase, an inverted search index has to be created in order to associate words and their possible meanings. The possible pair (word, meaning) has a relationship with the identification of the contents that contain that word. The inverted index is automatically created from the processing of all possible contents that can be retrieved from the SNS during the search. Thus, for each content (that has an author), and this author was connected to some agent during the user-agent modeling, the words from this content are analyzed in order to remove stop words, and are confronted with the modeled SWRL rules in the ontology trying to find a possible meaning for each word based on the author' agents. An ontology class (*i.e.* a meaning) is set to a term, depending on the SWRL rules available. The same word can be set up with different meanings (OWL classes) according to the way different modeled agents conform to the SWRL rules, pointing to different SNS contents. Therefore, it is possible to recognize the appropriate meaning for a word according to the author' agents. The search index always must be automatically updated when new content is published.

5.5 An Inclusive Search Mechanism

The ISM finds out the possible meaning for the words in the SNS content (that is stored in the search index). Moreover, it matches and retrieves the content(s) that contain a word with

the meaning that is more suitable to the user who makes the search (*i.e.* matching the meaning of the input search keywords with the word/class pair in the index).

The key idea is to find out if the most likely meaning of the search keyword(s) matches the likely meaning of the word in the SNS content. Thus, the proximity or similarity scale between the search keywords and the content words is determined by the most probable agent's interpretation of the word (producer and consumer), instead of being determined by a fixed vocabulary.

A scenario to exemplify the potential and the practical aspects of the ISM proposal is presented as follows: the term 'shingle' is a polysemy that can have different meanings depending on the context. 'Shingle' means a viral disease in a medical context, as well as a roof in an architectural context. As an example, suppose that a user (a doctor) logged into the SNS system, searches for 'shingle'. If there is a relationship between such user and some agent in the OWL code, it is possible to find out the possible meaning for this keyword according to the agent, and consequently to present more suitable search results to him/her. For instance, if a user that contains a doctor agent is searching, then the mechanism should first return the results that contain 'shingle' as a disease. Thus, it is necessary to describe the Web ontology (including SWRL rules) stating that if the agent is a doctor, the meaning of 'shingle' is closer to 'disease', as well as if the agent is a 'civil engineer' the meaning of 'shingle' is relative to 'roof', and so on.

As described by the rule (1), for instance, if one instance of the Agent class has a role equal to 'doctor' and the instance of the Input class is equal to 'shingle', then this instance must be of the class 'Viral_Disease'.

(1) *Agent(?x) ^ role(?x, ?y) ^ swrlb:equal (?y, "doctor") ^ Input(?z) ^ data(?z, "shingle") → Viral_Disease(?z)*

A strategy based on semantic proximity is proposed in order to rank the search results. The search algorithm selects the results, clustering them according to their meanings. If some SNS content contains a word classified with the same meaning of the input search keyword for the user, then this content is set as an appropriate result for this user and must be presented first to the user. In the same way, if a word has the same meaning as the input search keyword, then the content that contains it is also an appropriate result.

To organize the search results, the ISM solution has two principal groups of results: (1) the group with results that are likely to be more relevant (personalized) to the user, and (2) the less relevant group. Group 1 contains search results in which the meaning of the words is the same as the input keyword. Group 2 contains all other possible meanings of the input keyword or other related word. Inside each group (1 and 2) the semantic results are

organized in: "synonymous results", "more specific results", "more generic", and "related results". Results that contain an exact matching with some input search keyword are shown first in the group; the results retrieved by the ontology-based query expansion are organized by internal groups. For example, results containing synonymous in group 1 are those in which the word has the same meaning as an input search keyword. Results that contain synonymous words for the input search term, but have no classification or a different classification (meaning), are presented in the synonymous results of the second group. Similar schemes are used for all other internal groups (semantically organized) for both group 1 and 2. The search results are organized as hide boxes and internal boxes.

5.5.1 The Search Algorithm

The search algorithm receives the input search keywords and the user's identification. The output of the algorithm is the search results organized by lists. Given the user's identification, the algorithm gets the users' agents. For each input search keyword, the first step is to find its likely meaning. This is performed through the interpretation of the representation of the agents in the Web ontology, the user-agent relationship, and SWRL rules. It is possible to have situations in which no meaning is found given the combination of the users' agents and the search keyword due to the lack of rules to treat it. Figure 5.4 shows the proposed search algorithm.

```
Input: input search keywords and the user's identification
Output: The semantic search results organized by lists
  user's agents=Get user's agent (user's identification)
  For each search keyword do
        meaning=Find meaning in the SWO (keyword, user's agents)
        index classes=Query in the index (keyword)
        expanded_words=Perform the query expansion in the SWO (keyword)
        SearchByClasses (keyword, expanded_words, index classes, meaning)
        SearchByAgents (keyword, expanded_words, user's agents)
  End
```

Figure 5.4: The proposed search algorithm

Next, the algorithm queries in the index asking for the search keyword. If the keyword exists in the index, the algorithm returns the possible classes (*i.e.* the meanings) for that keyword. However, such query made in the index may not find the input keyword, or any

defined class may be returned. This situation happens when the word is not in the contents processed by the index, or there is not a SWRL rule that treats that combination pair, *i.e.* the word with the possible agents requested. It is possible to have situations in which a word has a classification for one content (depending on the user's agents that have published it) and no classification for another content, depending on the agents and SWRL rules. Since neither (input search nor content) words may be classified using the SWRL rules, they are recorded in the index without a classification.

The algorithm performs the ('**Query expansion**') on the SWO available based on the input search keyword. The query expansion occurs for each input search keyword. The SWO is used to perform the query expansion (*i.e.* to find new possible words to be searched in the index that are semantically related to those of the input). The SWO is analyzed to find synonymous words, more generic, more specific, and also related words. The modeled ontological dependencies that come from the OC to the Web ontology are also used to find related words to be searched. For example, imagine an input keyword A; if A depends ontologically on B, then B will be used as a related word to perform the ontology-based query expansion search. All words found in the ontology for each type of operation mentioned (synonymous, specific, generic and related) are used to make the query expansion, and to return other possible semantic search results. Alternatively, to increase the synonym list, the solution uses standard vocabularies as 'WordNet'.

After the query in the index that looks for the input search keyword and the available meaning classes for this keyword, and the query expansion execution, a set of expanded words and classes may be available. The search by classes ('**SearchByClasses**') method is executed using the search keyword, the words retrieved from the ontology expansion, the meaning from the search keyword based on the user, and the classes found in the inverted index.

The algorithm queries each word in the index (including the expanded words) for results (contents) combined to each class found. For instance, if two classes were found in the index for a certain search keyword, then this keyword and also all the others found in the ontology query expansion, will be queried twice in the index combined with a class (one for each different class). If some of these retrieved classes in the index were equal to the class retrieved for the search keyword (the meaning in the algorithm), then the possible search results are set in group 1; if not, the results are set in group 2. Figure 5.5 illustrates this method step by step.

```
For each word and expanded_words do
    For each class do
        results=Query in the index (word, class)
        If (class== meaning) Then
            group_1=results
        Else
            group_2=results
        End
    End
End
```

Figure 5.5: The method search by classes

It is possible to observe that in the case that an undefined class is found in the search, but the word exists in the index, all the search results including those retrieved from the query expansion are set to group 2. Consequently, the number of search results in group 1 tends to increase in function of the number of SWRL rules. The query expansion for each keyword always happens, and the search results for the possible words found are queried in the index, combined with the possible meaning found for the input keyword search.

After the search by class execution, the search is also done by agents ('**SearchByAgents**'). If the search keyword is not found in the index, the probability of the 'SearchByClasses' to find personalized results (*i.e.* in group 1) is zero. Besides, when a meaning for the search keyword relative to the user that performs the search is not found, all the possible results found by the 'SearchByClasses' through the index are also set as less relevant results (*i.e.* group 2), regardless of the classes found in the index. This is because the classes found in the index do not match a class for the user that is performing the search, since a class was found based on the user's agents.

In order to overcome this situation, in an attempt to find results for group 1, the search results from queries in the index with the input keyword and with the expanded words are classified as group 1 or 2 according to the user's agents that have published it. In this situation the lexical-syntactical search results (regardless of the class meaning to which they belong) are selected as group 1 or group 2 according to the agent that produced the content. Therefore, if an agent that belongs to the user that has published the content found from the index matches an agent associated to the user that performed the search, then this search result is selected into group 1; otherwise the search result is set to group 2. The search mechanism will tend to return the content as an adequate result to the user that performs the search. Sometimes this agent matching is not so clear, *i.e.,* the agents (from the search and from the content) are not the same. Thus, similarities, *i.e.* relationships modeled between the

agents are used to 'infer' the meanings represented by the list of relationships between the agents. Figure 5.6 illustrates this method step by step.

```
For each word and expanded_words do
    results= Query in the index (word)
    For each results do
        agents= Get user's agent (agent result)
        If (agents == user's agents) or is_similar() Then
                group_1=result
        Else
                group_2=result
        End
    End
End
```

Figure 5.6: The method search by agents

5.6 The *Vilanarede* Case Study

The WODAS method to model SWOs and the proposed ISM were evaluated in the context of the *VilanaRede*[21] system, a SNS constructed to investigate ISN principles. The *VilanaRede* was developed as a product of *e-Cidadania's* Project, with the objective of being accessible for the widest variety of users, including those less familiar with technology and with low literacy levels. *E-Cidadania*[22] was a Brazilian research project that has taken the challenge of developing systems that allow access and that make sense to the community of users. It may contribute to the promotion of a digital culture and respect for the diversity of the population. *VilanaRede* users collaborate in the SNS by announcing and sharing goods and services, events and ideas.

The content available in *VilanaRede* as well as some users were considered in the Case Study. In this way, the WODAS method was conducted using the real content available from the *VilanaRede* system to construct the SWOs. The objective of this section is to present a preliminary evaluation of the use of the ISM implemented in *VilanaRede*. The following sections show how the SWOs were modeled using the *VilanaRede* data, the prototype developed, and its preliminary evaluation.

---

[21] www.vilanarede.org.br
[22] www.nied.unicamp.br/ecidadania

5.6.1 Building SWOs from V*ilanaRede* Content

The content domain of the announcements available in *VilanaRede* is wide-ranging; *i.e.* there are announcements about various domains. Actually, one of the main challenges in modeling it is to deal with the open and informal domain of the announcements. Such contents created by users of the *VilanaRede* system form a "reference *corpus*", which is constituted by content from various areas and subjects independent of a specific domain. Thus the content presented in this system tends to be heterogeneous with respect to: content's subject, ways of expression, users, identity, among other factors. Such announcements are diversified regarding the contents such as: sale of various handmade products, meals, electronic products, advocacy services, events including debate about education and "June parties"[23]; as well as ideas that cover various subject such as: recipes, environmental awareness, health tips, and so on.

This investigation considered 230 announcements in *VilanaRede* distributed among products, services and ideas. The SAM was applied using these announcements created by users of *VilanaRede* in an attempt to model the shared meanings. Commentaries of users about the announcements were also considered during the modeling. Altogether 10 groups of announcements were selected according to subjects' sets. Based on these, SAM was applied to each group resulting in 1 or 2 OC(s) for each group. The subjects of the groups are: cooking and meal ordering, sale of products and services, cultural events, announcing *VilanaRede*, physical exercises and health promotion, social projects including inclusion and citizenship, offer of courses and seminars, health-oriented food, handicraft and environment.

Using the WODAS method to create SWO, an activity was carried out by sixteen graduate students in Computer Science, in the role of analysts, who were studying OS. They were divided into groups of two or three people. Each group received one subject (one of those already mentioned) collected from the *VilanaRede's* announcements. Each set of data includes approximately twenty to thirty announcements.

After learning how to apply SAM to construct OCs, each group would model an OC that might represent the semantics of the *VilanaRede* announcements. The students could also use the SONAR tool [65] to model the OCs. They were free to define their own strategies to build up the OCs. The groups had about one week to explore the announcements and to construct their OC. At the end of the activity, each group presented their results they achieved during the modeling process. Six OCs in an initial version were created by the students. Figure 5.7 shows one of the OCs modeled from the *VilanaRede* content. It models courses and seminars offered in *VilanaRede*.

---

[23] typical parties that occur in some regions of Brazil in the winter season

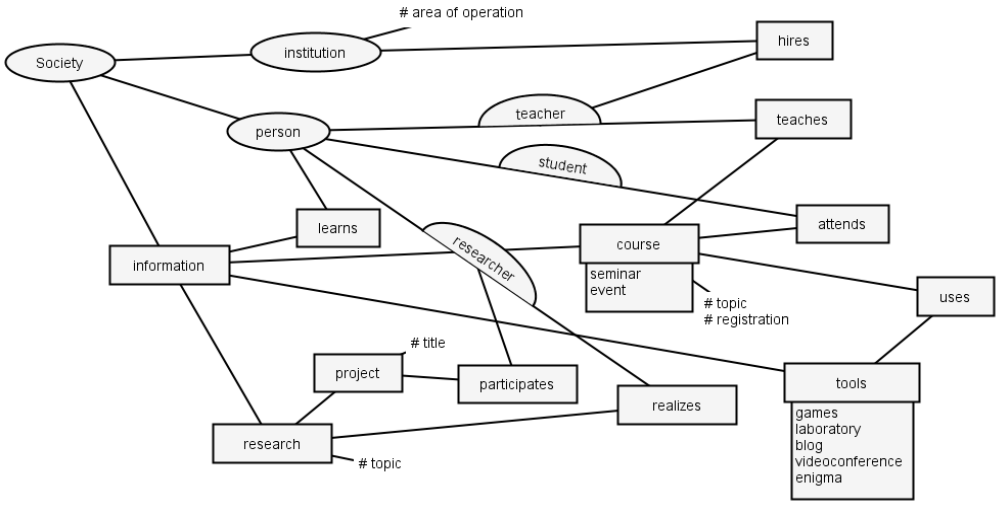


Figure 5.7: An example of an OC modeled from *VilanaRede* content [55]

The next step involved modeling an SWO from the OC. For that, the semi-automatic process that includes heuristics and transformation rules for deriving an initial Web ontology described in OWL from OC was applied. In this process the analyst specified the affordances that should be mapped to OWL classes and the affordances that should be mapped to object properties using the SONAR [65] tool. From that, the implemented transformation rules were able to construct an OWL file from the OC. This OWL file includes classes, object properties and data properties derived from the heuristics. The tool constructs one OWL for each OC.

At this point the diagrams represent pieces of knowledge in a fragmented way, since it is difficult to model the variety of contents at the same time. However, these diagrams need to be centralized and detailed for a wider coverage, since this knowledge representation must be used in practice by the ISM. For this purpose, the ontologies were merged into a single ontology. After that, the details were developed because the process at this point delivered an initial version of OWL ontologies, and additional information was necessary about the modeled classes in the final SWO. Therefore, new classes were modeled to build new relationships and instances of classes were created, range of values added, as well as SWRL rules defined, among other information to complement the ontology. The details of how the transformation occurs, and also an example of OWL codes generated using the tool, can be found in [60]. Details of applying the method to the *VilanaRede* context can be found in [55].

5.6.2 The Prototype Construction

The *VilanaRede* system uses a Content Management System[24] Drupal 5 to manage users' accounts, content types and other general functionalities; a JBoss Application Server[25] 4.2 hosts specific functionalities of the ISN such as password by images, multimedia comments, collaborations with others contents as well as the ISM services. The ISM was developed as Web Services connected to the Drupal modules and the ISN modules. Information such as the index, the agents and user-agent modeling are stored in a PostgreSQL[26] 8.3 database.

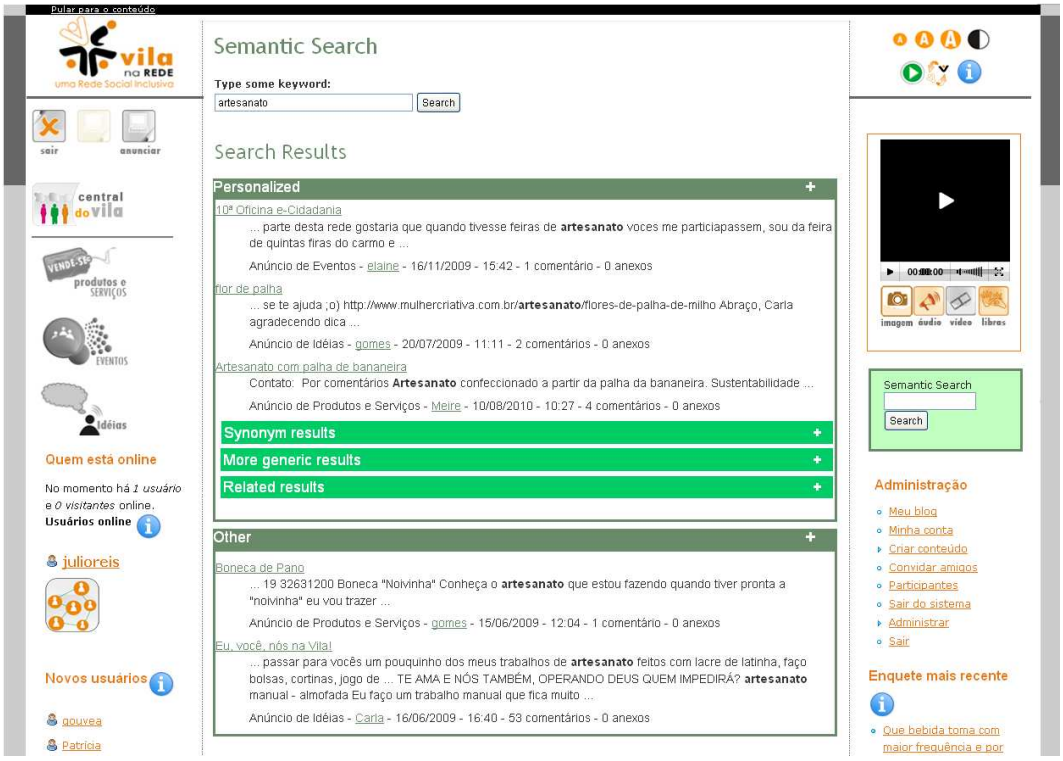


Figure 5.8: UI of semantic search results in VilanaRede

The ISM Web Services manage all the necessary information and provides all the services to handle the search request, the index processing, the adding of agents and user-agent relationship, as well as the ontology management for getting meanings and query expansion. The search services include the search algorithm that provides the search results organized by

[24] www. drupal.org
[25] www.jboss.org
[26] www.postgresql.org

lists that are interpreted in the Drupal side after the search service request. The index and the search algorithm were developed following the design of an ISM as presented in the sections 5.4 and 5.5. Moreover, in the Profile User Interface (UI) at *VilanaRede* users can choose a possible agent as its social profile that is stored using these services.

The UI of search results in *VilanaRede* was developed following the organization of search results as proposed in the section 5.5. Figure 5.8 shows an example of UI after the search request processing. The 'Personalized' box contains the results for group 1 while the 'Other' box contains the results for group 2. The intern boxes (*e.g.* synonym results) into both 'Personalized' and 'Other' boxes organize the semantic search results.

5.6.3 The Preliminary Evaluation of the Search Mechanism

Besides the semantic aspects regarding the search results, it is also important to analyze how users make sense of the UI solution. For example, would people recognize the search boxes and the organization proposed for the semantic search results? Would they easily interact with the boxes that contain the results?

Subjects: An evaluation was carried out with 25 real users. The evaluation included both people that had already interacted with *VilanaRede* before (16 people) as well as people that had never used the *VilanaRede* (9 people). The group included technology informed people, and also people with few experience with computers. The participants live in two cities of São Paulo state, Brazil: Pedreira (4 people), and Campinas (21 people). All people did the proposed activity individually. People from Pedreira carried out the activity in a telecenter, while people from Campinas did the activity in their house or in their work environment.

Regarding their age, 48,0% were between 51 and 60 years old, and 40,0% between 21 and 30 years old; 68,0% of the participants were women. Considering the schooling level, 36,0% of the participants had university degrees, 24,0% had high school degrees while 8% had not completed high school; 20,0% had just elementary school education and 12,0% had not completed even elementary school. The experiment included people of different social profiles such as: housewife, cook, handicraftsman, hairdresser, seamstress, retired people, teacher, student, researcher and others.

The majority of these people (about 72,0%) had a computer at home, and 60,0% had connection to the Internet; but just 40,0% of the participants use the computer frequently. When they use the computer they are helped by a close younger relative such as their son or daughter.

Material: Five distinct groups of search scenarios were proposed to conduct the experiment. For each group three specific search scenarios were elaborated. People of the same group performed the same three search scenarios. The search scenarios were designed specifically for each group aiming at observing and evaluating different aspects of the search

(*e.g.* polissemy in words). The search scenarios were also designed to encompass the different subjects from the *VilanaRede* content. Each group was constituted according to their social identity (*i.e.* regarding the agent concept).

Some scenarios asked the participant to search for things within a wide range of subjects (*e.g.* to find announcements related to 'project'), since there are different kinds of projects announced in *VilanaRede* such as social projects and sustainable projects. Other scenarios asked the participant to find very specific information (*e.g.* to find announcements that contain diseases that physical exercises may avoid).

Since the results tend to be personalized, the profile set during the user-agent modeling for each person that was in a group aimed at observing whether the results would make sense for each person. For example, a scenario asked to find announcements related to 'material' (*i.e.* any kind of material available in the announcements). Since there are various announcements related to different kinds of material (*e.g.* didactic material for courses, handicraft materials, recyclable material, *etc*) and there were various profiles of people in the group (*e.g.* student, environmentalist, handicraftsman) the results regarding material should suit their profile. For example, for a student, the personalized results should be related to didactic material and so on for the other profiles.

Moreover, the aim of the scenarios was to observe whether people would recognize the synonymous search results provided, as well as the more generic, specific and related results that could be returned; *e.g.* if someone searched for 'handicraft', then the 'specific results' should return the different examples of handicraft available at *VilanaRede*.

Procedure: During the activity each participant of the same group performed the three search scenarios twice. Each participant did not necessarily perform the activity at same time. Each participant used the ISM developed in the *VilanaRede* system, named here "mechanism *I*", and also used the default syntactic search mechanism available by Drupal adopted in the *VilanaRede* system, named here "mechanism *S*". Half of the people were chosen to start the activity using mechanism *I*, while the other half started the activity using mechanism *S*. This division was carried out within each group, *i.e.* while half of the group started with mechanism *I*, the other half started with mechanism *S*. In groups that were composed of an odd number of people, the difference was compensated between the groups, *i.e.* in a group of five people, two people started with mechanism *I* and three with mechanism *S*, while in another group of five people, three of them would start with mechanism *I* and two with mechanism S. The choice of which mechanism people would start with was made randomly.

Each participant had three attempts to find all possible announcements related to the question made by each proposed scenario (*i.e.* a maximum of three attempts for mechanism *I* and three attempts for mechanism *S* or vice-versa). Participants received support from the researcher throughout the activity. For each scenario executed they would fill out a form

answering a few questions about the search results retrieved for each mechanism in each search attempt. They would answer whether the results provided by the search mechanism were sufficient and relevant according to the scenario goal, and would point out the more relevant announcement of the search. The participants might use any keywords for each attempt.

The participants would also answer a few questions after the finalization of each scenario (*i.e.* after completing the attempts with both search mechanisms *I* and *S*). The question was, for example: In what search mechanism did he/she find the better results (*i.e.* provided more important and relevant search results)?, and in which attempt?; In what mechanism did the participant find the required announcements more rapidly according to the scenario goal (*i.e.* without many attempts)?; Did the search results ranking attend his/her expectations in both kinds of mechanisms?; Did the participant have any difficulty in interacting with the search boxes in the UI during the mechanism *I* search (*i.e.* Did he/she recognize the search boxes or did he/she have any difficulty opening or closing it?).

In addition, at the end of the activity, after performing all three search scenarios, each participant was asked which of the two search mechanisms, I or S, he/she preferred and if he/she would use the search mechanism while interacting in *VilanaRede*. Participants were also asked which kind of organization for the search results did they prefer to use during interaction with *VilanaRede* search, the *I* or *S* mechanism.

Once we conducted the experiment with 25 people and each of them executed 3 search scenarios answering the questionnaires, we had 75 scenario executions for both mechanisms *I* and *S*. It is worth mentioning that not all 3 attempts to search were used by the participants. Results are presented in the following section.

Results

According to the participant answers, 87% of the scenario executions indicate that better search mechanism results were found using mechanism *I*. From these indicated scenarios, 60% found the more important and relevant results in the first attempt to search, while 29% found the more important results in the second attempt.

During the execution of the search attempts with mechanism *S*, 25,3% of the scenario executions returned sufficient results in the first attempt to search, while 52% of the scenario executions returned relevant search results in this first attempt. On the first attempts, but using mechanism *I*, 54,6% of the scenario executions were considered to return sufficient search results, and 80% were considered to return relevant search results. For the first search attempt mechanism *I* was superior to mechanism *S* and showed good behavior. Observe Figure 5.9 and Figure 5.10.

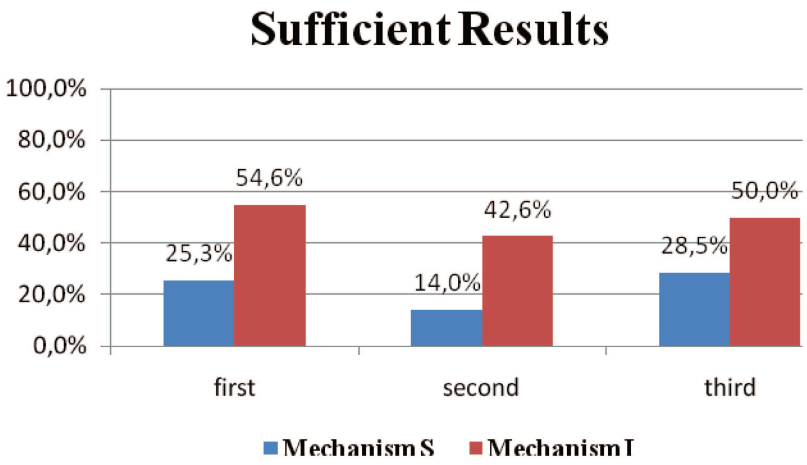


Figure 5.9: Comparing the mechanisms regarding sufficient results

In the scenario executions of the second attempt using mechanism *S*, 14% were reported to return sufficient results, and 34,3% returned relevant results. In the second attempt using mechanism *I*, 42,6% of the executed scenarios returned sufficient search results, and 66,0% were considered to return relevant results. Figure 5.9 compares the mechanism regarding sufficient results returned, while Figure 5.10 shows a comparison of the mechanisms regarding the relevant results.

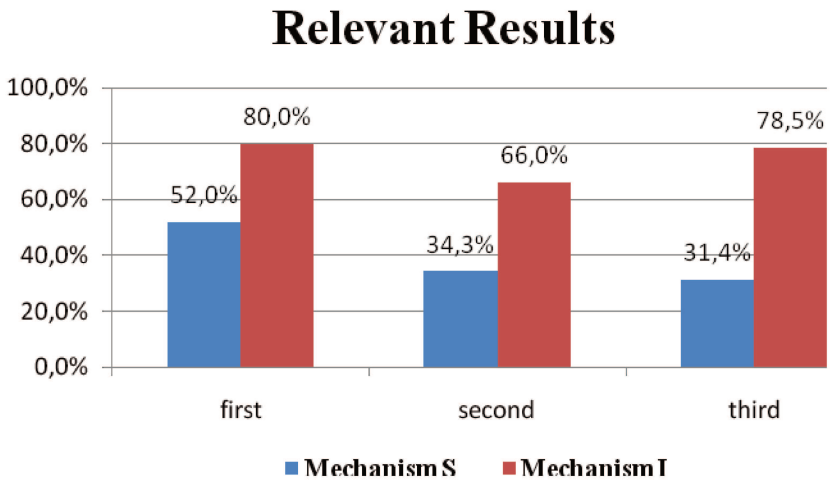


Figure 5.10: Comparing the mechanisms regarding relavant results

In the third attempt, 28,5% of the scenario executions using mechanism *S* returned sufficient results, while using mechanism *I* 50% of the executions presented sufficient results. In the third attempt 31,4% of the scenarios executed with mechanism *S* returned relevant search results while with mechanism *I* about in 78,5% of the executions the results were considered relevant. It is important to note that with mechanism *S*, during the three attempts, the number of executions that returned relevant results decreased for each attempt, while with mechanism *I* despite a small drop in the second attempt, the third attempt maintained about the same level as the first (see Figure 5.10).

Figure 5.11 compares the proportion of executed scenarios in each search attempt. In all the scenarios the first attempt was executed using both mechanisms. As shown in Figure 5.11, using mechanism S, in 85,4% of the scenarios it was necessary to execute a second attempt (according to the participants), while using mechanism I, the second attempt was necessary in just 72% of the scenarios. This shows that a second search attempt was necessary for more scenarios with mechanism *S*, and that the first attempt using mechanism *I* was more effective. During the third attempt, while in 46,7% of the scenarios it was necessary to perform attempt executions using mechanism S, attempt executions were necessary in just 37,4% of the scenarios using mechanism *I*.

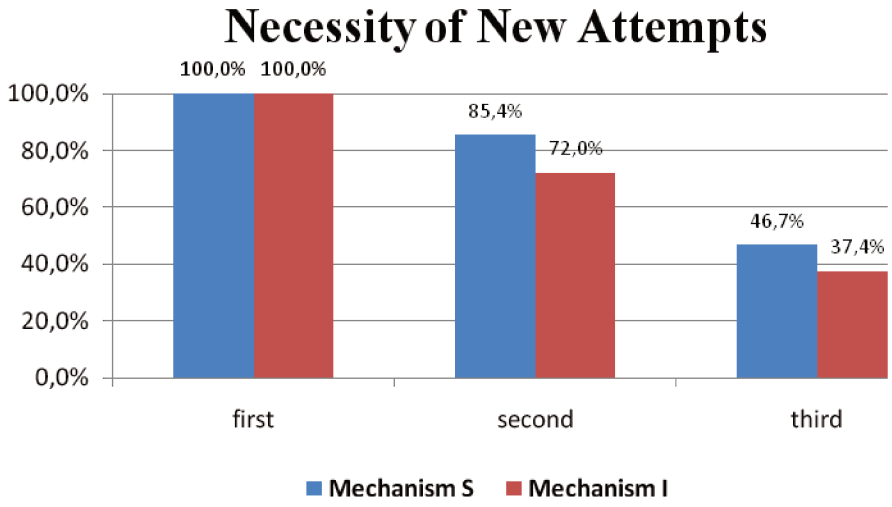


Figure 5.11: Comparing the mechanisms regarding necessity to execute new attempts to search

Moreover, based on the participant's answers, the results indicate that with mechanism *I*, in 88% of the scenario executions the results were found quickly (*i.e.* the participant considered that they did not have to make a lot of attempts to reach the desired information).

For mechanism *S* just 38,6% of the executions were considered to reach the desired announcements quickly.

Regarding the search results ranking, in 86,6% of the scenario executions, the search results ranking was considered to match the users' expectation with mechanism *I*, while with mechanism *S* this number is 49,3%. Moreover, the users wondered whether there were other relevant announcements, beyond those presented by the search results in 49,3% of the executions with mechanism *I*, while with mechanism *S* the users wondered about other relevant announcements in 82,6% of the executions. This means that when using mechanism *I* users believed they had retrieved more relevant search result than when using mechanism *S*.

Regarding general answers after the experiment, 96% of the participants answered that they would prefer mechanism *I* to use daily at *VilanaRede*. About 76% of the users answered that they preferred the search results divided by boxes than a traditional search list; but in 24% of the scenario executions, they noted having some difficulty interacting with the search boxes. The difficulties included understanding the purpose of the search boxes, as well as their labels and how to open or close them. However, we noticed that such difficulty was concentrated in the first search attempts.

From the qualitative point of view, participants that preferred the search results divided by boxes said that they organize the search results better, facilitating the query, and the results visualization. The boxes give a better view of the results improving their understanding as well as the semantic organization in more specific or more comprehensive results. One participant said: "*It is easier to locate what you are looking for or something similar (a related subject)".* Those participants who preferred the traditional search list claimed that this way is easier for them since that is how they learned, and they are more familiar with it. One of the participants claimed that it would be interesting to join both approaches providing the options with mechanism *I* on an interface side. Thus, it would be possible to join the more restricted and controlled results provided by mechanism *S* according to the input search keywords with the broader results provided by mechanism *I,* that could favor the discovery of useful and important related search results.

The main arguments given by the participants about their preference for mechanism *I* include that it returns more complete and wider search results with more alternatives. One participant said: "*The information is returned in a more detailed way, more complete, but in an organized form presenting all possible search results related to the input search keywords*". They also said that using mechanism *I* they could find more satisfactory results, and that the 'Personalized box' and the 'Other box' can lead to a better division of the results and more practical to find information. Additionally, one participant explained that mechanism *I* is mainly useful to suggest related semantic search results to the input search keywords.

5.7 Discussion

Ontologies have a fundamental role in representing semantics. To create useful ontologies with a suitable representation of reality is still an ongoing goal, and how to use them in services such as semantic searches involves many challenges still to be addressed. In this work we have investigated a method for the construction of more representative Web ontologies from the universal access point of view. For that, we have based on SAM, which tries to model the agents and their respective affordances from the social network. The approach is different since the meanings of the words are synthesized in existential relationships.

The proposed method explored the combination of OS concepts with SW technologies aiming to reach more representative Web ontologies to be used by an ISM following the ISN concept. The instantiation of the proposed method using real ISN contents as well as its use for an ISM developed in the *VilanaRede* ISN has shown the potential of the proposed approach. The results show the impact of the proposed method in the prototype developed at *VilanaRede*.

Many particular cases that represent the social network culture may not be present in dictionaries or formal thesauruses generally used by conventional search mechanisms. One ontology modeled based on data from the ISN system potentially may lead to more appropriate semantic relations that will provide possibilities for semantically useful and meaningful search results. The proposal provides means to discover and to distinguish the meanings from the ISN, representing them through the agents into the ontology.

It was possible to construct ontologies from the contents of an ISN system through the WODAS method as well as to find agents that are used by the ISM. The developed approach, grounded on SAM, represent new possibilities, influencing positively on a more inclusive search solution. Such an approach enables identification of the possible agents and their relationships, bringing a new approach to represent the semantic information that supports the search mechanism with a different paradigm to represent meaning. The identified and modeled agents are part of the search solution, being a special cornerstone and differential of the proposed solution. By associating it to the search, the OC may enable new semantic search possibilities besides the approaches based on a strict hierarchy of classes. The OC enables finding existential relationships, discovering more representative search keywords to be used by the ontology query expansion. Besides the existential relationships, the agents and their possible pattern of behavior (affordance) enabled by the method are used by the search algorithm to find more adequate and personalized search results.

The proposed method (WODAS) that generates OWL ontology (SWO) was used by the prototype developed in the *VilanaRede* ISN. The case study conducted evaluated the

prototype approach informed by the SWO. The proposed organization for the semantic search results as well as the labels, and UI mechanism to interact with the search results were evaluated during the experiment. First, it is important to mention the understanding of the users on the proposed search activity with the scenarios. We noticed that most of users in an initial phase of digital literacy do not have a clear concept about a search mechanism. Sometimes it was difficult for them to answer whether the search results were relevant and sufficient, since the participants would like some results that do not exist among the *VilanaRede* announcements. They even do not have the idea that if the system does not contain a specific announcement, no results could be returned regarding it. Although they did not have the exact meaning of the search concept, *e.g.* the possible universe of results; they liked to see their own announcements among the search results and also they would search for their own announcements.

Due to these limitations, sometimes they had difficulty to express their opinion and also to suggest a search keyword during the activity. In this same context, regarding keywords, some participants expressed how they make sense of keywords; for example, some users expected that a keyword in singular would return just one result (*e.g.* 'plate' will return just one result for plate and 'plates' will return various results).

Regarding the UI proposal to organize the semantic search results, users presented problems recognizing the label's meaning on the search boxes. This was observed mainly with those not used to digital concepts, but also with a few more technically informed users. Most of the novice users did not recognize the search results in the boxes at the first attempt to search. It was necessary for the researcher to call their attention to the boxes, and then they could understand and recognize them. They usually knew that more search results were there, and also recognized the icon to open the boxes, but they did not know the kind of results that could exist inside the boxes by just reading their labels. When observing the results inside the boxes some people recognized the concepts described in the labels as: "more generic" and "more specific results".

Some users also did not recognize the purpose of the 'Personalized box', but most of the times they liked the results available in this box, and considered the results from mechanism *I* richer than those generated by mechanism *S*. The approach to include the personalized results based on the agents connected to users was positive. For example, a handicraftsman recognized the results in the 'Personalized box' from other handicraftsman that are also *VilanaRede* users. But the Personalized box cannot be considered as the unique aspect in the personalization and should be combined with relations and other strategies.

Moreover, the possibility to find specific concepts from generic ones, and the visa versa, was relevant and interesting in mechanism *I* since users could find other related results. However, essential for this mechanism are the details and quantity of classes and instances

available in the ontology, since the terms must be found in the ontology to provide a possibility of results.

The case study conducted with users using the developed ISM in the *VilanaRede* ISN pointed out satisfactory results showing the potential of the proposed approach for ISMs. It was important to preliminarily evaluate the solution showing possible improvements and limitations of the approach. One limitation of this evaluation is the fact of comparing it with a syntactic search mechanism, as no similar approach (concerning universal access and social network) in the SW field was found. In fact, the syntactic search still has been the usual approach for SNS content search, and the social and digital inclusion aspects are not considered in the design of such search mechanisms. In this sense, we expected to contribute with one step further by providing ISM based on methods that model the user colloquial language. However, we also recognize that much work is still necessary to create a solution truly adequate for the context under study, which brings many big challenges and further research.

5.8 Conclusion

Search mechanisms may be decisive for recovering adequate information. These mechanisms are even more crucial in ISN systems, which presuppose to promote the universal access to knowledge. The semantic model that underlies and informs the search mechanism is a key point to enable a suitable search mechanism for ISN. In this context, Web ontologies have a great potential to be very useful for the creation of Web applications which are more adequate for diverse populations and also for their inclusion. However, it is necessary to investigate methods that are able to reflect complex social concepts and meanings. In general, the application of Web ontologies to domains that are not clearly delineated is still a challenge, since such artifacts are very hard to be created and sustained (*i.e.* they evolve over time).

In this work we proposed a method to construct representative Web ontologies to be used by semantic search mechanisms following the ISN concept. The SWO approach generated and constructed from real ISN content were used to inform an ISM. A prototype in the *VilanaRede* ISN was developed following the proposal for an ISM. A case study using the ISM prototype created in the *VilanaRede* ISN was conducted with real users in order to evaluate the proposal. This evaluation included users of different social profiles as well as experience with computers as they used the solution constructed. In this way, it was possible to show that the method generates artifacts with characteristics capable to reach positive results for inclusive searches. Looking for a solution that can reach search results that make sense for each user, *i.e.* being personalized, we obtained positive feedback from users

involved in the case study. In order to achieve this, the knowledge representation must also progress toward a human and social vision. In the long term, we expect to produce an adequate process for modeling knowledge, as well as techniques to explore it, to provide search results that make sense for all.

Further work involves trying to produce larger empirical results to deeply evaluate the proposed approach, as well as to face new challenges to the ontology evolution maintaining consistency over time between the ISN content and the ontologies. Our future work also includes exploring other techniques to develop the ontology-based query expansion and to design more adequate visualization techniques for the search results.

Capítulo 6

Conclusão

O contexto único de multiplicidade de riquezas culturais e problemas sociais presente na sociedade contemporânea, principalmente em países em desenvolvimento como o Brasil e a China, exige que a criação de aplicações computacionais seja abordada de maneira socio-técnica e inclusiva. Considerando este cenário, a pesquisa relatada nesta dissertação abordou o desenvolvimento de mecanismos de busca mais adequados ao conceito de RSI. Neste trabalho buscaram-se respostas para a criação de métodos e mecanismos inclusivos que pudessem gerar resultados de busca com mais sentido para os usuários, uma vez que consideram a heterogeneidade cultural e significados compartilhados socialmente. Logo, na proposta apresentada, os significados compartilhados pelos indivíduos na rede social são considerados na modelagem da semântica dos dados. O acesso irrestrito à informação mesmo quando a busca é realizada através de termos locais ou coloquiais (regionais ou informais) usados na rede, representa uma oportunidade de aprendizagem, além de inclusão digital e social.

Embora 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* (*i.e.* a *Web* Semântica), as abordagens para os mecanismos de busca ainda são principalmente fundamentadas em dados sintáticos e vocabulários padrão (*e.g.* tesauros). Estes vocabulários muitas vezes não fazem parte da linguagem cotidiana das pessoas que usam esses serviços. Ainda há grandes limitações nas soluções de busca para tratar problemas do tipo polissemia, na qual uma palavra-chave pode ter diversos significados distintos. Isto se torna ainda mais crítico quando se consideram usuários em fase de letramento digital e/ou analfabetos funcionais. Soluções que exigem muito destes usuários como, por exemplo, navegar em uma ontologia, podem não ter sucesso. Assim, há

necessidade urgente de soluções de interação que não considerem apenas os vocabulários formais e padrões, mas também a linguagem coloquial, naturalmente compartilhada pelos usuários e mediada pela interação com os sistemas computacionais no dia-a-dia.

Nos últimos anos, as pesquisas para melhorar os mecanismos de busca na *Web* (Semântica) têm envolvido diferentes fatores e abordagens (*e.g.* uso de ontologias para busca semântica). No entanto, no contexto de busca semântica, os métodos e técnicas tradicionais que se fundamentam principalmente em tecnologias da *Web* Semântica (*i.e.* ontologias *Web*), ainda são estritamente fundamentados em abordagens que não consideram adequadamente os aspectos cognitivos, sociais e humanos na modelagem do conhecimento.

Nesta dissertação, como um possível caminho para o problema de busca visando a inclusão, foram articulados métodos e artefatos que possibilitam o entendimento, a discussão e a modelagem de conceitos compartilhados em RSOs. Estes métodos são fundamentados em uma visão subjetivista, que está mais alinhada à compreensão dos aspectos humanos presentes no contexto sócio-cultural onde o *software* está inserido. A abordagem proposta, que informa e subsidia o mecanismo de busca proposto, foi construída pela integração, articulação e expansão de conceitos e técnicas de duas áreas de pesquisa principais: a SO e a WS, desenvolvendo-se a concepção de mecanismos de busca inclusivos norteados pelo conceito de RSIs. A solução utiliza estratégias de busca semântica fundamentada em um modelo diferenciado que representa o conhecimento advindo da rede social. Este modelo incluiu uma estrutura ontológica de suporte à busca, baseada em um método proposto para o *design* de ontologias *Web* que utiliza artefatos da SO.

A abordagem desenvolvida nesta dissertação (*i.e.* o uso de artefatos da SO junto a tecnologias da WS), assim como o método proposto para efetivar esta abordagem, foram subsidiados por resultados práticos de estudo de caso com usuários reais da RSI *VilanaRede*. Ficou evidente que os resultados de busca, assim como a sua organização, resultante de mecanismos estritamente sintáticos, não eram suficientes para auxiliar potenciais usuários de uma RSI a alcançar informação com facilidade, uma vez que pessoas em uma rede social podem expressar e criar o seu próprio vocabulário, compartilhando significados em uma comunidade. Conseqüentemente, estes resultados evidenciaram ser necessário construir modelos computacionalmente tratáveis do ponto de vista semântico para lidar com estes aspectos lingüísticos.

Ainda fundamentado nestes resultados empíricos, nesta dissertação apontou-se os desafios relacionados à busca em RSI e foram descritas recomendações para uma solução de mecanismo mais adequado e especialmente voltado para RSI. O principal desafio endereçado pela abordagem foi desenvolver uma solução computacional de busca situada nos significados que emergem 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). Para tanto, foram estudadas possíveis ferramentas e técnicas de mineração de texto para auxiliar na construção (modelagem) das ontologias a partir dos dados (conteúdos) da RSI, e desenvolvido um experimento com dados reais da RSI *VilanaRede*. Neste experimento objetivou-se analisar a aplicação de possíveis ferramentas e técnicas para a identificação de termos mais relevantes e freqüentes em RSI, visando a concepção de uma estratégia para auxiliar na construção de ontologias para este contexto. Um dos desafios foi lidar com informação em linguagem natural, não estruturada, além de conter características de informalidade e abrangência de assuntos. O processo de extração apontou resultados que demonstram a importância da aplicação de métodos apropriados ao contexto considerado e a limitação do estado da arte. Além de tentar identificar os conceitos e suas relações a partir da linguagem natural usada na rede, foi necessário efetuar formalmente a representação desta semântica.

Considerando o contexto em estudo, uma nova proposta de abordagem para a representação da semântica da linguagem utilizada em redes sociais foi necessária, uma vez que desenvolver representações mais adequadas dos significados que as pessoas organizadas em rede compartilham pode ser um fator determinante para o desenvolvimento de mecanismos de busca inclusivos. Assim sendo, foi investigada e proposta uma nova abordagem e método para o *design* diferenciado de ontologias *Web,* visando uma maneira mais representativa e adequada para a modelagem dos significados compartilhados pelas pessoas em redes sociais. Esta abordagem foi fundamentada principalmente no SAM da SO, desenvolvendo o conceito de SWO. Objetivando levar em consideração os aspectos sócio-culturais presentes na rede social visou-se potencialmente fornecer resultados mais adequados de busca semântica quando comparado com outras abordagens tradicionais.

A abordagem incluiu principalmente o conceito de 'Agentes' (conceito advindo do SAM) que é traduzido como classes nas ontologias *Web*. Fundamentado nesta abordagem proposta, foi desenvolvido um método semi-automático que inclui a extração de conceitos advindos dos conteúdos do sistema, e heurísticas para transformar a Ontologia conceitual resultante do SAM em uma ontologia *Web* computacionalmente tratável descrita em OWL.

Um estudo de caso foi conduzido para instanciar o método proposto como um todo em um contexto real da RSI *VilanaRede*. Assim, uma SWO foi desenvolvida com base nos conteúdos reais desta RSI. Neste estudo de caso, tanto o resultado da aplicação das ferramentas de mineração de texto quanto a transformação entre artefatos SO para WS foram utilizados. Uma avaliação preliminar do resultado destas ferramentas durante a modelagem das ontologias do *VilanaRede* também foi conduzida neste estudo de caso.

A abordagem fornece meios para descobrir e distinguir os possíveis significados utilizados em redes sociais, representando-os através dos 'Agentes' na SWO por meio de regras descritas usando SWRL. A proposta melhora a representação semântica por

acrescentar conceitos advindos do SAM que conectam o significado aos agentes, e conseqüentemente resulta em uma busca mais inclusiva. Isso gerou oportunidades de resultados de busca semanticamente organizados e personalizados aos usuários, conforme seu perfil social, tratando problemas de polissemia e sinônimos, uma vez que este serviço de busca considera desde aspectos semânticos lexicais (*e.g.* sinônimos, homônimos) até conceituais (*e.g.* generalização, especialização e conceitos relacionados).

Visando avaliar a solução proposta, um protótipo de mecanismo de busca inclusivo na RSI *VilanaRede* foi implementado; este mecanismo usa a SWO construída a partir dos conteúdos reais deste sistema. Adicionalmente, um estudo de caso com usuários reais usando *VilanaRede* foi conduzido. Pela análise deste estudo de caso foi possível avaliar a abordagem e identificar tanto aspectos positivos quanto negativos da proposta.

O restante deste capítulo está assim organizado: a seção 6.1 traz uma síntese das principais contribuições desta dissertação e a seção 6.2 aponta os desafios e os trabalhos futuros.

6.1 Contribuições da Pesquisa

Para responder as principais questões de pesquisa levantadas e apresentadas no início deste trabalho, esta dissertação lidou com vários desafios presentes no desenvolvimento de mecanismos de busca. As contribuições desta dissertação serão organizadas conforme estas questões de pesquisa:

Como um mecanismo de busca inclusivo deve ser projetado considerando uma população de usuários heterogênea e iletrada digitalmente?

Esta questão foi principalmente abordada e explorada nos Capítulos 2 e 4 desta dissertação. O Capítulo 2 mostrou os desafios envolvidos e gerou as recomendações que nortearam a criação de um caminho para uma solução mais adequada. O Capítulo 4 avançou na proposta de solução. A principal inovação explorada foi utilizar os conteúdos da RSI para a construção de ontologias *Web*. Foi realizado um estudo e análise de uma estratégia usando ferramentas e técnicas de mineração de texto para a construção de ontologias no contexto de RSI (Apêndice A). Esta análise foi uma contribuição relevante no sentido de investigar a viabilidade de semi-automatizar o processo e facilitar um meio para a solução proposta.

Quais fundamentos teóricos e metodológicos devem informar a ação do mecanismo? Por que e como?

Este trabalho adotou uma visão subjetivista para o *design* de ontologias *Web*. Logo, buscou-se o uso de teorias e métodos que pudessem embasar a solução. A abordagem desenvolvida usando o SAM como fundamento é o principal diferencial deste trabalho, assim como o desenvolvimento e avaliação desta abordagem é a principal contribuição desta dissertação. Isto porque, historicamente, as modelagens no contexto da computação

consideram relações causais e objetivistas, enquanto a modelagem proposta pelo SAM exige que se pense em relações existenciais durante a modelagem semântica, *i.e.* na existência de cada *affordance* e em suas relações ontológicas, além de relacionar *affordances* socialmente responsáveis (Agentes) com os *affordances* socialmente interpretados. As relações existenciais amplificam a visão do domínio no qual o sistema computacional está inserido, favorecendo a especificação da semântica compartilhada naquele contexto. O Capítulo 3 mostrou a proposta desta abordagem articulada com ontologias *Web*, enquanto o Capítulo 4 apresentou uma proposta de método para a construção semi-automática de ontologias *Web* no contexto de RSI. Para viabilizar o método e alcançar artefatos computacionalmente tratáveis foram desenvolvidas heurísticas e regras de transformação no *software* SONAR. Estas heurísticas resultaram na modelagem de ontologias *Web* baseada nos resultados e artefatos do SAM (Apêndice B).

Quais características o mecanismo deve conter e expressar para que os resultados de busca sejam mais apropriados e personalizados a cada indivíduo participante da rede?

Nesta dissertação desenvolveu-se o conceito de "mecanismo de busca inclusivo" como referência a um mecanismo de busca informado por SWOs, que conecta os possíveis agentes identificados durante o SAM a usuários da rede social. Adicionalmente, a idéia de representar as possíveis interpretações para um termo conforme os agentes por meio de regras SWRL nas ontologias possibilitou gerar uma abordagem para agrupar semanticamente os resultados de busca. Ainda, considerar os resultados de busca segundo conteúdo(s) de usuários relacionados aos mesmos agentes do usuário que faz a busca contribuiu para o agrupamento e seleção dos resultados mais apropriados. Conseqüentemente, a busca tende a apresentar os resultados que fazem mais sentido para o usuário que faz a busca. Estas características são expressas no protótipo de busca inclusiva, desenvolvido na RSI *VilanaRede* (descrito no Capítulo 5 desta dissertação).

Como os resultados de busca podem ser mais bem organizados semanticamente, ordenados e apresentados aos usuários?

De acordo com a descrição do mecanismo inclusivo apresentado no Capítulo 5, os resultados de busca semântica foram organizados de forma a diferenciar os resultados mais adequados e os menos adequados de acordo com o usuário que faz a busca. Uma maneira de inferir um possível contexto automaticamente para a busca é determinar uma possível classe na ontologia para um termo de entrada conforme os agentes relacionados ao usuário. Assim, o(s) possível(is) agente(s) ou papel(is) relacionados ao usuário é(são) o(s) responsável(is) por determinar um contexto do termo (*i.e.* seu significado).

Adicionalmente, teve-se a preocupação de agrupar e organizar os resultados de busca em caixas expansíveis que representam resultados semanticamente relacionados. Através da avaliação do mecanismo junto ao método proposto (Capítulo 5) pode-se verificar se tal

organização desenvolvida, assim como os resultados retornados com base na ontologia SWO efetivamente auxiliaram os usuários a encontrarem e a reconhecerem a informação mais rapidamente e facilmente. Os resultados preliminares da avaliação com usuários finais do *VilanaRede* sugeriram e apontaram para a eficácia da abordagem proposta, contudo ainda revelam que melhorias devem ser tratadas.

6.2 Desafios e Trabalhos Futuros

Há diversas frentes de estudo sob diferentes perspectivas que podem representar uma evolução desta pesquisa. Considerando os resultados apresentados, esta dissertação representa os primeiros passos para alcançar os desafios científicos e tecnológicos do problema de pesquisa abordado. Podemos de imediato identificar dois desafios e três frentes principais de estudos futuros que podem representar aprofundamentos no assunto. A investigação destes permitirá um refinamento da proposta e a sua adequação a um escopo mais amplo de problemas.

Os desafios estão principalmente relacionados às propostas de soluções para manter as ontologias e a modelagem usuário-agente (perfil para a busca) atualizadas (semi) automaticamente no decorrer do tempo. Isto é importante para manter a proposta apresentada nesta dissertação mais dinâmica e automatizada. O primeiro desafio é pesquisar métodos e técnicas aplicáveis ao contexto em estudo que possam gerar maneiras mais ágeis para construir e gerenciar as ontologias SO e OWL, assim como propor técnicas que permitam a evolução e mapeamento de ontologias integrando diferentes versões, enquanto mantém a consistência entre elas. Isto é necessário uma vez que a rede social apresenta uma atividade dinâmica, nas quais novos usuários se cadastram, novos conteúdos são disponibilizados e novas buscas são realizadas. Este desafio também envolve o estudo de como tratar a coexistência e evolução entres as ontologias SO e OWL, de forma a representar o conhecimento da rede atualizada.

Outro desafio envolve manter a relação dos usuários da rede com os agentes descobertos e modelados por meio do SAM. No presente trabalho, a solução para este problema foi feita manualmente, uma vez que o engenheiro de ontologia relaciona manualmente cada usuário com possíveis agentes e papéis modelados. "Como descobrir o(s) agente(s) relacionado(s) aos usuários de maneira mais automática visando conseguir gerar melhores níveis de personalização?" Isto envolve tratar sistematicamente as atividades do usuário desde o seu cadastramento no sistema e ao longo do tempo, uma vez que com o tempo o usuário pode desenvolver novos papéis sociais ou reprimir alguns já assinalados. Para tanto, observa-se ser necessário estratégias para determinar o perfil do usuário dinamicamente. Uma possível solução a ser investigada é tentar identificar e filtrar através de serviços de *log* as ações dos

usuários na rede, *e.g.* os anúncios postados, os termos de busca usados, as relações sociais desenvolvidas, entre outras. No entanto, encontrar automaticamente as relações entre agentes e usuários e as relações válidas entre agentes ainda é um ponto em aberto. O desafio está em como fazer isso sistematicamente e corretamente.

Algumas soluções mais específicas e pontuais também podem ser desenvolvidas no futuro: *e.g.* uma interface para o cadastramento dos agentes que advém do SAM e sua conexão com o sistema SONAR. Isto permitiria o cadastro automático no sistema de rede social a partir do SONAR usando Serviços *Web,* conforme a arquitetura do protótipo implementado. As relações entre agentes também podem ser automaticamente derivadas e sugeridas a partir do diagrama de ontologia.

Além destes desafios, as frentes de estudo identificadas como trabalhos futuros estão relacionadas a: 1) mineração e extração de informação; 2) modelagem e representação do conhecimento; 3) mecanismo de busca.

1. *Mineração e extração de informação:*

Esta frente está diretamente relacionada com os desafios supramencionados. Além da importância da atualização das ontologias e também dos perfis, é importante conseguir informações que possam enriquecer a ontologia, e ao mesmo tempo diminuir a carga de trabalho em modelagem do engenheiro de ontologia. Estudos relativos ao processamento de linguagem natural e extração de informação poderiam endereçar a identificação e sugestão de possíveis relações de significados (*i.e.* um agente com um termo remete a uma possível classe da ontologia que representa um significado para o termo). Isto poderia auxiliar o engenheiro a criar regras SWRL, que interpretam os possíveis significados para os termos. Poderia haver uma ferramenta para o engenheiro mais facilmente poder modelar as regras e ser auxiliado por algoritmos de extração que indicassem possíveis regras que seriam descritas em SWRL, propiciando uma maneira mais automática e rápida de descobrir e alimentar a base de conhecimento (ontologias OWL) com regras SWRL e também novos conceitos (affordances da rede social). Tais regras também poderiam ser vislumbradas a partir da análise do diagrama de ontologia.

2. *Modelagem e representação do conhecimento:*

Neste trabalho foram principalmente explorados conceitos do SAM na abordagem desenvolvida. Contudo, o Método de Análise de Normas – *Norm Analysis Method* (NAM) da SO também poderia ser explorado e agregado. Poder-se-ia também desenvolver uma metodologia geral, incluindo NAM para criar SWOs destinadas a contextos além das redes sociais; ou também levar em consideração outros tipos de relações sociais desenvolvidas na rede (*e.g.* amizade) durante a modelagem. Também pode ser explorado o mapeamento e referencias com ontologias globais, tornando o conhecimento modelado disponível a outros domínios. Outra possibilidade que pode diretamente impactar na busca é usar o conceito de

proximidade na modelagem e não apenas relações estritamente booleanas. Desta maneira conseguiríamos, por exemplo, saber o grau de relacionamento entre conceitos, assim como quanto um Agente está semanticamente próximo a um conceito ou a outro.

3. *Mecanismo de busca:*

Além destas ações que poderão diretamente impactar para um melhor mecanismo de busca semântico e inclusivo, podemos também desenvolver pesquisas diretamente relacionadas ao algoritmo de busca, e estratégias de busca. Primeiramente, seria interessante e relevante efetuar um estudo em paralelo com a solução desenvolvida neste trabalho para uma arquitetura distribuída da WS, onde os Serviços *Web* desenvolvidos fossem totalmente independentes de um sistema fechado (e.g. sistema de gerenciamento de conteúdo). É igualmente relevante desenvolver trabalhos que documentem de forma mais detalhada o algoritmo de busca desenvolvido, assim como a maneira que a SWO foi usada no processo de busca, fazendo expansão da consulta (*query*) baseada na SWO. Pode-se também investigar, para o contexto em estudo, novas maneiras dos usuários expressarem os dados de entrada na solução de busca, *i.e.* soluções de interfaces que auxiliam a introduzir os dados de entrada. Estas informações poderiam subsidiar resultados melhores, assim como novas abordagens para a organização e ordenação, e a interação com os resultados (*e.g.* de maneira visual).

Do ponto de vista de avaliação empírica deste trabalho, semanticamente alguns outros experimentos podem ser realizados. É importante ressaltar que nesta dissertação buscamos efetuar uma avaliação do mecanismo proposto com usuários reais. Contudo, outros experimentos sobre o mecanismo podem ser propostos para analisar, por exemplo: precisão e cobertura dos resultados e análise dos *logs* dos resultados semânticos durante o experimento desenvolvido. Finalmente, podemos conduzir novas avaliações com usuários reais usando o mecanismo em outros contextos, observando outros detalhes na prática, discutindo e articulando resultados qualitativos e quantitativos do uso do mecanismo.

Apêndice A

Identificando Semântica em Redes Sociais Inclusivas *Online*: Um Estudo sobre Ferramentas e Técnicas[Σ]

Abstract

Access to knowledge is a basic condition for living in the digital age and Social Network Services are a reality nowadays. Search mechanisms are increasingly essential for interaction and information retrieval in such systems. Appropriate representation of the meaning that people use in SNS can be a determining factor for the development of more adequate search engines. The identification of concepts and semantic relationship that come out from the network data are even more relevant for Inclusive Social Network Services (ISN), which presuppose respect for the diversity of users, including those in the process of digital literacy. This work studies tools and techniques for the identification of concepts and semantic relationships in ISN, aiming at designing a strategy to assist the building of ontologies that model the shared semantics in the social network, toward more adequate search mechanism for ISN. The extraction process points out results which demonstrate the importance of applying appropriated methods to the considered context.

Σ Este é um relatório técnico que, originalmente, foi publicado pelo IC/UNICAMP: J.C. dos Reis, R. Bonacin e M.C.C. Baranauskas, "Identificando Semântica em Redes Sociais Inclusivas *Online*: Um estudo sobre Ferramentas e Técnicas". Instituto de Computação, Universidade Estadual de Campinas, Brasil, Relatório Técnico, IC-10-28, 2010 [52].
Disponível em: <www.ic.unicamp.br/~reltech/2010/10-28.pdf>

Apêndice B

From Ontology Charts to Web Ontologies: Heuristics and Transformation Rules[Δ]

Abstract

The evolution of the Semantic Web depends on novel techniques and methodologies that can handle and better represent the meanings of the huge amount of information available nowadays. Recent proposals in literature have explored new approaches based on Semiotics. The 'Semiotic Web Ontology' (SWO) is an attempt to model the information in a computer-tractable and more adequate way, and, at the same time to be compatible with the Semantic Web (SW) standards. This work presents a computer assisted process for building SWOs. The process includes heuristics and transformation rules for deriving an initial Web ontology described in Web Ontology Language from Ontology Charts produced by the Semantic Analysis Method. Moreover, the entire process is discussed; results of the application of the process to a real context show the potential of the approach and the value of the proposed heuristics and implemented rules to create more representative Web ontologies.

[Δ] Este é um relatório técnico que, originalmente, foi publicado pelo IC/UNICAMP: J.C. dos Reis, R. Bonacin e M.C.C. Baranauskas, "*From Ontology Charts to Web Ontologies: Heuristics and Transformation Rules*". Instituto de Computação, Universidade Estadual de Campinas, Brasil, Relatório Técnico, IC-11-02, 2011 [60]. Disponível em: <www.ic.unicamp.br/~reltech/2011/11-02.pdf>

Apêndice C

Modeling Meanings from Inclusive Social Network Services[α]

Abstract

Inclusive Social Network Service (ISN) can be defined as a Social Network Service (SNS) with resources that promote access for all, including those in the margin of the digital culture. An ISN must include adequate means to recover information that make sense for all. A search mechanism capable of understanding the shared meanings used by the ISN users is still needed. In this sense methods and approaches should support capturing the social and cultural aspects from the ISN including its colloquial language and the shared meanings. In order to achieve a better understanding and representation of the semantics utilized by ISN members, this technical report presents the application and the analysis of a semantic modeling method proposed to represent meanings of terms adopted by ISN users. The outcome of the method is intended to be used by an inclusive search mechanism. This approach can enable novel ontology-based search strategies that potentially provide more adequate semantic search results.

[α] Este é um relatório técnico que, originalmente, foi publicado pelo IC/UNICAMP: J.C. dos Reis, M.C.C. Baranauskas e R. Bonacin, "*Modeling Meanings from Inclusive Social Network Services*". Instituto de Computação, Universidade Estadual de Campinas, Brasil, Relatório Técnico, IC-11-03, 2011 [54]. Disponível em: <www.ic.unicamp.br/~reltech/2011/11-03.pdf>

Referências Bibliográficas

1. AMARAL, C.; LAURENT, D.; MARTINS, A.; MENDES, A.; PINTO, C. **Design and implementation of a semantic search engine for Portuguese**. In Proceedings of the 4th International Conference on Language Resources and Evaluation, LREC 2004, Lisbon, Portugal, Vol. I, 2004, p. 247–250.
2. BARANAUSKAS, M.C.C. & SOUZA, C.S. **Desafio 4 Acesso Participativo e Universal do Cidadão Brasileiro ao Conhecimento**. Computação Brasil, ano VII, núm. 23, 2006, p.7.
3. BARANAUSKAS, M.C.C. **e-Cidadania: Systems and Methods for the Constitution of a Culture mediated by Information and Communication Technology**. Proposal for the Microsoft Research-FAPESP Institute. 2007.
4. BERKAN, C. R. **Semantic Search: An Antidote for Poor Relevancy**. 2007 Disponível em: <http://www.readwriteweb.com/archives/semantic_search_antidote_for_poor_relevancy.php> Acessado em Fevereiro de 2009.
5. BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. **The Semantic Web**, Scientific American, May 2001 issue.
6. BHOGAL, J., MACFARLANE, A., SMITH, P. **A review of ontology based query expansion**. Information Processing & Management, Vol. 43, No. 4, 2007, p. 866-886.
7. BONACIN, R., BARANAUSKAS, M. C. C., LIU, K. **From Ontology Charts to Class Diagrams: semantic analysis aiding systems design**. In Proceedings of the 6th International Conference on Enterprise Information Systems. 6th International Conference on Enterprise Information Systems (ICEIS), Porto, Portugal, 2004.
8. BONINO, D.; CORNO, F.; FARINETTI, L.; BOSCA, A. **Ontology Driven Semantic Search**. WSEAS Transaction on Information Science and Application, Issue 6, Vol. 1, 2004, p. 1597-1605.
9. BOYD, D. M & Ellison, N. B. **Social Network Sites: Definition, History, and Scholarship**. Journal of Computer-Mediated Communication, 13. International Communication Association, 2008, p. 210–230.

10. BRAZILIAN INSTITUTE OF GEOGRAPHY AND STATISTICS. Disponível em: <www.ibge.gov.br> Último acesso em Outubro de 2010.
11. BRAZILIAN INTERNET STEERING COMMITTEE. **Survey on the use of Information And Communication Technologies in Brazil.** 2009. Disponível em: <www.cgi.br/english/index.htm> Acessado em Janeiro de 2011.
12. CARDOSO, JORGE. **The Semantic Web Vision: Where Are We?** Intelligent Systems, IEEE. Volume 22, Issue 5, 2007, p: 84 - 88
13. CARVALHO, M. L. B. **Web semântica e semiótica: ontologias e aplicação a permacultura**. Master Dissertation - Centro Universitário Eurípides da Marília, 2005.
14. CELINO, I; VALLE, E, D; CERZZA, D; TURATI, A. **Squiggle: A semantic search engine for indexing and retrieval of multimedia content.** In Proceedings of the SAMT 2006, pp. 20–34.
15. CHOUDHARI, AJAY; JAIN, MOHIT; SINHAROY, AVIK; ZHANG, MIN. **SmartSeek: A Semantic Search in Social Networks**. 2008. Disponível em: < www.cc.gatech.edu/projects/disl/courses/8803/2008/project/project_deliverables/group22/final_report.pdf>. Acessado em Janeiro de 2011.
16. FANG, WEI-DONG; ZHANG, LING; WANG, YAN-XUAN; DONG, SHOU-BIN. **Toward a Semantic Search Engine based on Ontologies**. In Proceedings of the 4th International Conference on Machine Learning and Cybernetics. 2005, p. 1913-1918.
17. GÄRDENFORS, P. **How to Make the Semantic Web More Semantic**. In Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS), 2004, p. 17-34.
18. GIBSON, J.J. **The Theory of Affordances**. In Perceiving, Acting, and Knowing. Eds. Robert Shaw and John Bransford, 1977.
19. GRUBER, T. R. **A translation approach to portable ontologies**. Knowledge Acquisition, Vol. 5, Num. 2, 1993, p. 199-220.
20. GUELFI, NICOLAS; PRUSKI, CÉDRIC; REYNAUD, CHANTAL. **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, 2010, p. 297-302.
21. GUHA, R.; MCCOOL, R.; MILLER, E. **Semantic Search**. Proceedings of the 12th international conference on World Wide Web. Budapest, Hungary, 2003, p. 700-709.
22. GÜRSEL, ANIL & SEN, SANDIP. **Improving search in social networks by agent based mining**. In Proceedings of the 21th international joint conference on Artifical intelligence table of contents. Pasadena, California, USA. 2009, p. 2034-2039
23. HAYNES, J. & PERISIC, I. **Mapping Search Relevance to Social Networks. International Conference on Knowledge Discovery and Data Mining**. In

Proceedings of the 3rd Workshop on Social Network Mining and Analysis. Paris, France. 2009, Article Nº 2.

24. HEFLIN, J. & HENDLER, J. **Searching the web with SHOE**. Artificial Intelligence for Web Search. In AAAI Workshop, WS-00-01, AAAI Press, Menlo Park, CA, 2000, p. 35–40.
25. HILDEBRAND, M.; OSSENBRUGGEN J., AND VAN HARDMAN, L. An **analysis of search-based user interaction on the semantic web**. Report, CWI, Amsterdam, Holland, 2007.
26. HOANG, H. H. & TJOA, M. **The State of the Art of Ontology-based Query Systems**: A Comparison of Existing Approaches. In Proceedings of the IEEE International Conference on Computing and Informatics (ICOCI). Kuala Lumpur, Malaysia, 2006.
27. IBGE. **Acesso à Internet e posse de telefone móvel celular para uso pessoal 2008**. 2008. Disponível em: <http://www.ibge.gov.br/home/estatistica/populacao/acessoainternet2008> Acessado em Janeiro de 2010.
28. INAF – Brasil. **Indicador de Alfabetismo Funcional**. Acessado em Dezembro de 2010. Disponível em: <www.ipm.org.br/download/inaf_brasil_2009_relatorio_divulgacao_revisto_dez-10.pdf>, 2009.
29. INTERNATIONAL TELECOMMUNICATION UNION. Disponível em: <www.itu.int/en/pages/default.aspx> Acessado em Outubro de 2010.
30. IPM. **Indicador Nacional de Analfabetismo Funcional**. 2007. Disponível em: <http://www.smec.salvador.ba.gov.br/site/documentos/espaco-virtual/espaco-dados-estatisticos/indicador%20de%20analfabetismo%20funcional%202007.pdf> Acessado em Janeiro de 2010.
31. JACOB, ELIN K. **Ontologies and the Semantic Web**. Bulletin of the American Society for Information Science and Technology. Vol. 29, Issue 4, 2005, p. 19-22.
32. JAMALI, M. & ABOLHASSANI, H. **Different aspects of social network analysis**. In Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence, Washington, DC, USA, IEEE Computer Society. 2006, p. 66–72.
33. KARYPIS, G. **CLUTO: a clustering toolkit. Technical Report.** 02-017, Dept. of Computer Science, University of Minnesota, 2002. Disponível em: <http://glaros.dtc.umn.edu/gkhome/views/cluto/ > Último acesso em Setembro de 2010.
34. KASSIM, J. M. & RAHMANY, M. **Introduction to Semantic Search Engine**. In International Conference on Electrical Engineering and Informatics. Selangor, Malaysia. Vol. 02, 2009, p. 380-386.

35. LEI, Y.; UREN, V.S.; MOTTA, E. **SemSearch: a search engine for the Semantic Web.** In Proceedings of the EKAW 2006, Managing Knowledge in a World of Networks. POdebrady, Czech Republic. 2006, p. 238-245
36. LIU, K. **Semiotics in information systems engineering**. Cambridge University Press, 2000.
37. LIU, K., SUN, L., FU, Y. **Ontological Modeling of Content Management and Provision**. In Information and Software Technology, V.50 N.11, 2008, p.1155-1164.
38. LOPES, L., FERNANDES, P., VIEIRA, R., FEDRIZZI, G. **ExATOlp: An Automatic Tool for Term Extraction from Portuguese Language Corpora.** In Proceedings of the 4th Language and Technology Conference, 2009, p. 427-431.
39. MANGOLD, C. **A survey and classification of semantic search approaches**. Int. J. Metadata, Semantics and Ontology, Vol. 2, No. 1, 2007, p.23–34
40. MAO, M., HE, D., PENG, Y. **Ontology-Based Content Management and Access Framework for Supporting E-Learning Systems**. In Proceedings of International Conference on Web Intelligence and Intelligent Agent Technology. Washington, DC, USA. 2006, p. 220-226.
41. MAULDIN, M. L. **Conceptual Information Retrieval: A Case Study in Adaptive Partial Parsing.** Kluwer Academic Publishers Norwell, MA, USA, 1991.
42. MEC, 2007. Disponível em: http://portal.mec.gov.br, Acessado em novembro de 2009.
43. MEDELYAN, O. AND I. WITTEN, H. **Domain-independent automatic keyphrase indexing with small training sets**. Journal of the American Society for Information Science and Technology. V. 59, I. 7, 2008, p. 1026- 1040.
44. MIKA, PETER. **Ontologies are us: A unified model of social networks and semantics**. In Proceedings of the 4th International Semantic Web Conference (ISWC 2005), LNCS 3729, Springer-Verlag, 2005.
45. NERIS, V. P. A.; ALMEIDA, L. D. ; MIRANDA, L. C. ; HAYASHI, E. ; BARANAUSKAS, M. C. C. **Towards a Socially-constructed Meaning for Inclusive Social Network Systems**. In: International Conference on Informatics and Semiotics in Organisations. Beijing. 2009, p. 247-254.
46. NIELSEN. **Global Faces and Networked Places**. The Nielsen Company USA. 2009. Disponível em: <http://blog.nielsen.com/nielsenwire/wp-content/uploads/2009/03/nielsen_globalfaces_mar09.pdf> Acessado em Julho de 2009.
47. NOY, N. F. & MCGUINNESS, D. L. **Ontology Development 101: A Guide to Creating Your First Ontology**. Knowledge Systems Laboratory, 2001.
48. OBITKO, M., SNASEL, V., SMID, J. **Ontology design with Formal Concept Analysis**. In Proceedings of the International Workshop on Concept Lattices and their Applications, 2004, p. 111–119.

49. OSW. **The circulation document in the Organizational Semiotic Workshop**, The 1st International Workshop on Organizational Semiotics, Enschede, The Netherlands, 1995.
50. PEIRCE, C.S. **Collected Papers**. Cambridge, Mass: Harvard University Press, 1931-1958.
51. PEREIRA, R & SILVA, S.R.P. **Folksonomias: uma análise crítica focada na interação e na natureza da técnica.** In Proceedings of the VIII Brazilian Symposium on Human Factors in Computing Systems. ACM International Conference Proceeding Series; Vol. 378. Porto Alegre, RS, Brazil, 2008, pp. 126-135
52. RAMACHANDRAN, DIVYA; KAM, MATTHEW; CHIU, JANE; CANNY, JOHN; FRANKE, JAMES L. **Social Dynamics of Early Stage Co-Design in Developing Regions.** In Proc. of the SIGCHI Conference on Human Factors in Computing Systems. California, USA. 2007, p. 1087-1096.
53. REIS, J. C., BONACIN, R., BARANAUSKAS, M. C. C. **Identificando Semântica em Redes Sociais Inclusivas Online: Um estudo sobre ferramentas e técnicas.** Relatório Técnico (IC-10-28), Instituto de Computação, Universidade Estadual de Campinas. Campinas, SP, Brasil, 2010.
54. REIS, J. C., BONACIN, R., BARANAUSKAS, M. C. C. **Search Informed by a Semiotic Approach in Social Network Services.** In: Proceedings of the IEEE Computer Society Press. Workshop Web2Touch - living experience Through Web, Tozeur, 2010. p. 321-326.
55. REIS, J. C.; BARANAUSKAS, M. C. C; BONACIN, R. **Modeling Meanings from Inclusive Social Network Services.** Relatório Técnico (IC-11-03), Instituto de Computação, Universidade Estadual de Campinas. Campinas, SP, Brasil, 2011.
56. REIS, J. C.; BARANAUSKAS, M. C.; BONACIN, R. **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.
57. REIS, J. C.; BONACIN, R.; BARANAUSKAS, M. C. C. **A Semiotic-based Approach to the design of Web Ontologies**. In: Proceedings of the 12th International Conference on Informatics and Semiotics in Organisations. 12th International Conference on Informatics and Semiotics in Organisations (ICISO), Reading - UK. 2010, p. 60-67.
58. REIS, J. C.; BONACIN, R.; BARANAUSKAS, M. C. C. **Analyzing Search in Inclusive Social Network Services.** Relatório Técnico (IC-10-02), Instituto de Computação, Universidade Estadual de Campinas. Campinas, SP, Brasil, 2010.
59. REIS, J. C.; BONACIN, R.; BARANAUSKAS, M. C. C. **Ferramentas e Técnicas de Mineração de Texto em Redes Sociais Inclusivas: Promovendo Meios para o Acesso Universal ao Conhecimento**. In Anais da Conferência IADIS Ibero-

Americana WWW/Internet 2010 (CIAWI 2010). Carvoeiro, Algarve – Portugal, 2010, p. 155-162.

60. REIS, J. C.; BONACIN, R.; BARANAUSKAS, M. C. C. **From Ontology Charts to Web Ontologies: Heuristics and Transformation Rules**. Relatório Técnico (IC-11-02), Instituto de Computação, Universidade Estadual de Campinas. Campinas, SP, Brasil, 2011.
61. REIS, J. C.; BONACIN, R; BARANAUSKAS, M. C. C. **Prospecting an Inclusive Search Mechanism for Social Network Services**. In "*ICEIS 2010 - Revised Selected Papers*". J. Filipe *and* J. Cordeiro (Eds.), Lecture Notes in Business Information Processing (LNBIP), Vol. 73. Springer, Heidelberg, 2011, p. 555-570.
62. REIS, J.C.; BONACIN, R.; BARANAUSKAS, M.C.C. **Analysing Search in Inclusive Social Network Services**. Relatório Técnico (IC-10-02), Instituto de Computação, Universidade Estadual de Campinas. Campinas, SP, Brasil, 2010.
63. SALTER, A. **Semantic Modelling and a Semantic Normal Form**. SOCTR/01/01. School of Computing. Staffordshire University.
64. SALTER, A; LIU, K. **Using Semantic Analysis and Norm Analysis to Model Organisations.** In 4th International Conference on Enterprise Information Systems (ICEIS), 2002, pp. 847-850.
65. SANTOS, T. M.; BONACIN, R.; BARANAUSKAS, M. C. C.; RODRIGUES, M. A. **A Model Driven Architecture Tool Based on Semantic Analysis Method**. In Proc. of the 10th International Conference on Enterprise Information Systems – ICEIS 2008. Barcelona, Spain, v. 2., 2008, p. 305-310.
66. SILVA S.R.P. AND R. PEREIRA. **Aspectos da Interação Humano-Computador na Web Social**. In VIII Simpósio Brasileiro de Fatores Humanos Sistemas Computacionais. ACM Vol. 378, 2008, p. 350-351.
67. STAMPER, R. K. **Organisational Semiotics: Informatics without the Computer?** In In-formation, Organisation and Technology: Studies in Organisational Semiotics, eds. K. Liu, R. Clarke, P. B. Andersen and R. K. Stamper. Kluwer Academic Publishers, 2001.
68. STAMPER, R. K.; ALTHANS, K.; BACKHOUSE, J. **Measur: Method For Eliciting, Analysing and Specifying User Requirements.** In: Computerized Assistance During the Information Systems Life Cycle. North-Holland, 1988, p. 67-115.
69. STAMPER, R., LIU, K; M. HAFKAMP, Y., ADES. **Understanding the Role of Signs and Norms in Organisations: A semiotic approach to information systems design**. Journal of Behaviour and Information Technology, 19(1), 2000, p. 15-27.
70. STAMPER, R.K. **Social Norms in requirements analysis - an outline of MEASUR**. In Jirotka M, Goguen J, Bickerton M. (eds) Requirements Engineering, Technical and Social Aspects. New York, 1993.

71. STEPHANIDIS, C. & EMILIANI, P. L. **Universal Access to Information Society Technologies: Opportunities for People with Disabilities**. K. Miesenberger, J. Klaus, W. Zagler (Eds.): ICCHP 2002, LNCS 2398, 2002, p. 8–10.
72. STUDER, R.; BENJAMINS, V. R; FENSEL, D. **Knowledge engineering: principles and methods**. Data & Knowledge Engineering, v.25, n.1/2, 1998, p. 161-197.
73. TANASESCU, V. & STREIBEL, O. Extreme **Tagging: Emergent Semantics through the Tagging of Tags**. In International Workshop on Emergent Semantics and Ontology Evolution. Busan, Korea, 2007.
74. TAZI, S. **Using Sowa's conceptual graphs for enhancing hypertext readers' performances**. Intelligent Hypertext Workshop, Washington. citeseer.ist.psu.edu/12096.html, 1994
75. TVAROŽEK, M., BARLA, M., FRIVOLT, G., TOMŠA, M., MÁRIA, B. **Improving Semantic Search via Integrated Personalized Faceted and Visual Graph Navigation**. In SOFSEM 2008: Theory and Practice of Computer Science. Lecture Notes in Computer Science - Springer Berlin / Heidelberg. Vol. 4910, 2008. p. 778-789.
76. USCHOLD, M. Where are the semantics in the semantic web? American Association for Artificial Intelligence. AI Magazine. Vol 24, Issue 3. Menlo Park, CA, USA, 2003, p. 25-36.
77. VIEIRA, M. V.; FONSECA, B. M.; DAMAZIO, R.; GOLGHER, P. B.; REIS, D. C.; NETO, B.R**. Efficient Search Ranking in Social Networks**. In Proceedings of the 16th ACM Conference on information and knowledge management. Lisbon, Portugal. 2007, p. 563-572.
78. WANG, W.; BARNAGHI, P. M.; BARGIELA, A. **Search with Meanings: An Overview of Semantic Search Systems**. International Journal of Communications of SIWN, Vol. 3, 2008, p. 76-82.
79. WANG, W.; BARNAGHI, P. M.; BARGIELA, A. **The Anatomy and Design of A Semantic Search Engine**, Tech. rep., School of Computer Science, University of Nottingham Malaysia Campus, 2007.
80. WORLD WIDE WEB CONSORTIUM (W3C). **OWL-Web Ontology Language**, Recommendation 10 February 2004, Disponível em: <http://www.w3.org/TR/owl-features>, Último acesso em fevereiro de 2011.
81. WORLD WIDE WEB CONSORTIUM (W3C**). SWRL: A Semantic Web Rule Language Combining OWL and RuleML**, W3C Member Submission 21 May 2004, Disponível em: <http://www.w3.org/Submission/SWRL/>, Último acesso em fevereiro de 2011.
82. YU, BIN & SINGH, MUNINDAR P. **Searching social networks**. In Proceedings of the 2th International joint Conference on Autonomous agents and multiagent systems. New York, NY, USA. ACM Press. 2003, p. 65–72

Anexo A

Autorizações para Publicação

A.1 Sociedade Brasileira de Computação

4/4/2011 Gmail - informe para usar artigo em ...

Gmail

Julio Cesar Reis <juliocesardosreis@gmail.com>

informe para usar artigo em dissertação de mestrado

2 messages

Julio Cesar Reis <juliocesardosreis@gmail.com> **31 mars 2011 13:39**
À: sbc@sbc.org.br

Sociedade Brasileira de Computação (SBC),

Eu envio este e-mail com o intuito de informar a SBC que irei incorporar um artigo que foi publicado no SEMISH 2010 no corpo de minha Dissertação de Mestrado. O artigo será publicado na íntegra, sem alterações (com exceção da formatação do texto, numeração da figuras e tabelas, e formato de citações).

o artigo é:

Title: Busca em Sistemas Sócio-Culturais Online: Desafios e Recomendações
Authors: Júlio Cesar dos Reis, Maria Cecília Calani Baranauskas and Rodrigo Bonacin
Conference: XXX Congresso da Sociedade Brasileira de Computação
Publication: *Anais do* XXX Congresso da Sociedade Brasileira de Computação
Publisher: Sociedade Brasileira de Computação
Place/Date: Belo Horizonte, Brazil; 20-23 July, 2010.
ISBN: 2175-2761
Pages: 380-394

Atenciosamente

–
Júlio Cesar dos Reis

Sociedade Brasileira de Computação <sbc@sbc.org.br> **31 mars 2011 14:13**
À: Julio Cesar Reis <juliocesardosreis@gmail.com>

Boa tarde, Júlio César,

Tudo bem, a SBC tem a Revista Eletrônica de Iniciação Científica (REIC), que é um periódico destinado a publicação de artigos técnico-científicos produzidos por alunos de graduação das áreas de computação e informática. Podes encontrar maiores informações em http://www.sbc.org.br >> Publicações.

Qualquer dúvida, estou à disposição

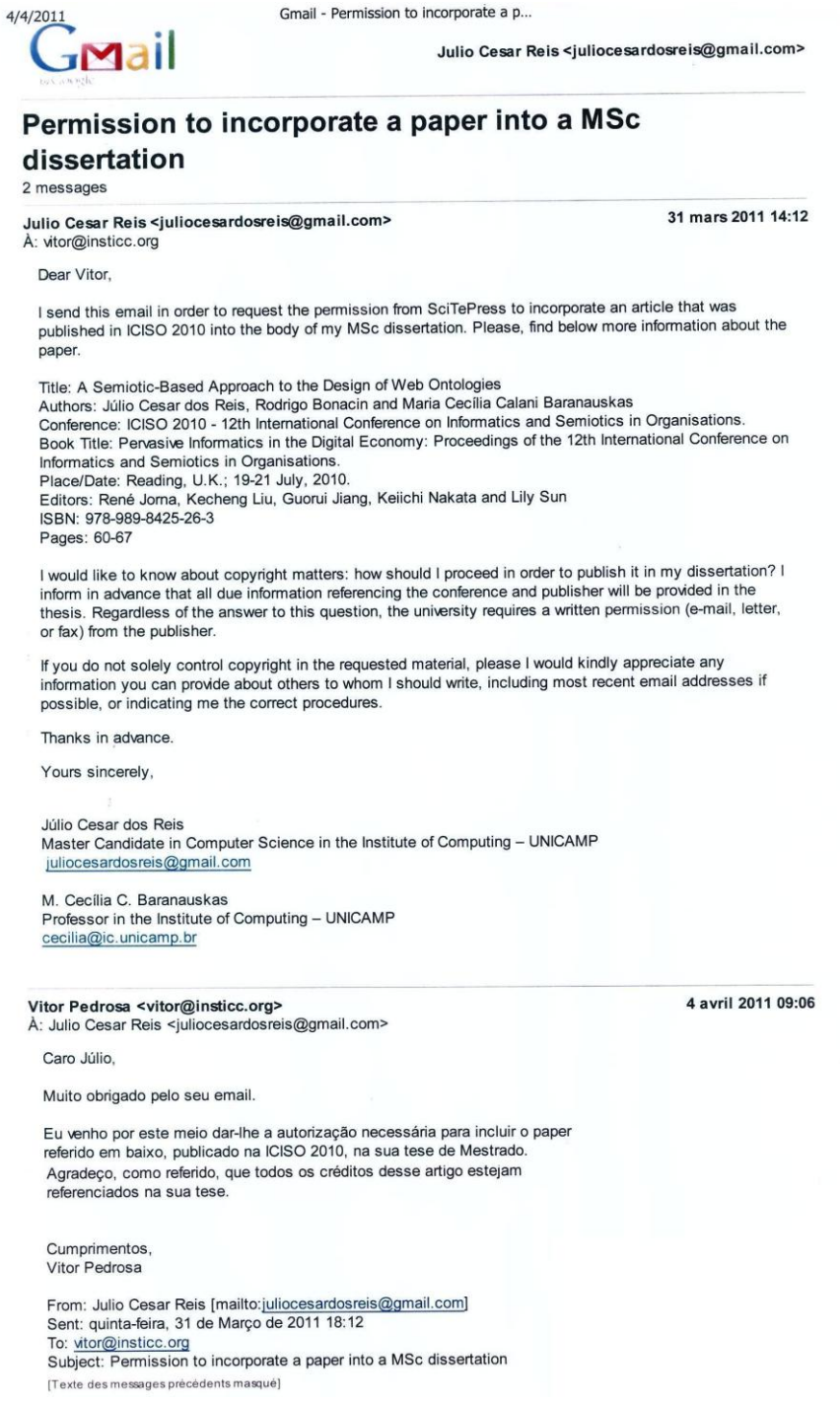
SBC
www.sbc.org.br

Neiva da Silva
Secretaria
sbc@sbc.org.br
Sociedade Brasileira de Computação
Av. Bento Gonçalves, 9500 - Prédio 43412 - Sala 219
CEP: 91509-900 - Agronomia - Porto Alegre/RS
Tel.: +55 51 3308-6835 Fax: +55 51 3308-7142

Em 31/3/2011 13:39, Julio Cesar Reis escreveu:
[Texte des messages précédents masqué]

https://mail.google.com/mail/?ui=2&... 1/1

A.2 SciTePress



4/4/2011 Gmail - Permission to incorporate a p...

Gmail

Julio Cesar Reis <juliocesardosreis@gmail.com>

Permission to incorporate a paper into a MSc dissertation

2 messages

Julio Cesar Reis <juliocesardosreis@gmail.com> **31 mars 2011 14:12**
À: vitor@insticc.org

Dear Vitor,

I send this email in order to request the permission from SciTePress to incorporate an article that was published in ICISO 2010 into the body of my MSc dissertation. Please, find below more information about the paper.

Title: A Semiotic-Based Approach to the Design of Web Ontologies
Authors: Júlio Cesar dos Reis, Rodrigo Bonacin and Maria Cecília Calani Baranauskas
Conference: ICISO 2010 - 12th International Conference on Informatics and Semiotics in Organisations.
Book Title: Pervasive Informatics in the Digital Economy: Proceedings of the 12th International Conference on Informatics and Semiotics in Organisations.
Place/Date: Reading, U.K.; 19-21 July, 2010.
Editors: René Jorna, Kecheng Liu, Guorui Jiang, Keiichi Nakata and Lily Sun
ISBN: 978-989-8425-26-3
Pages: 60-67

I would like to know about copyright matters: how should I proceed in order to publish it in my dissertation? I inform in advance that all due information referencing the conference and publisher will be provided in the thesis. Regardless of the answer to this question, the university requires a written permission (e-mail, letter, or fax) from the publisher.

If you do not solely control copyright in the requested material, please I would kindly appreciate any information you can provide about others to whom I should write, including most recent email addresses if possible, or indicating me the correct procedures.

Thanks in advance.

Yours sincerely,

Júlio Cesar dos Reis
Master Candidate in Computer Science in the Institute of Computing – UNICAMP
juliocesardosreis@gmail.com

M. Cecília C. Baranauskas
Professor in the Institute of Computing – UNICAMP
cecilia@ic.unicamp.br

Vitor Pedrosa <vitor@insticc.org> **4 avril 2011 09:06**
À: Julio Cesar Reis <juliocesardosreis@gmail.com>

Caro Júlio,

Muito obrigado pelo seu email.

Eu venho por este meio dar-lhe a autorização necessária para incluir o paper referido em baixo, publicado na ICISO 2010, na sua tese de Mestrado.
Agradeço, como referido, que todos os créditos desse artigo estejam referenciados na sua tese.

Cumprimentos,
Vitor Pedrosa

From: Julio Cesar Reis [mailto:juliocesardosreis@gmail.com]
Sent: quinta-feira, 31 de Março de 2011 18:12
To: vitor@insticc.org
Subject: Permission to incorporate a paper into a MSc dissertation
[Texte des messages précédents masqué]

A.3 Springer

Springer
the language of science

Springer-Verlag GmbH
Tiergartenstrasse 17
69121 Heidelberg

Postfach 10 52 80
69042 Heidelberg

tel 06221 / 487 – 0
fax 06221 / 487 – 83 66
www.springer.com

Springer-Verlag | Postfach 10 52 80 | 69042 Heidelberg

Júlio Cesar dos Reis
Institute of Computing
at University of Campinas (UNICAMP)
1251 ave Albert Einstein
Cidade Universitária
13083 970, Campinas, São Paulo
Brazil

Ralf Gerstner
tel +49 (0)6221/487-8144
fax +49 (0)6221/487-8144
ralf.gerstner@springer.com

Heidelberg, April 4th, 2011

Dear Júlio Cesar dos Reis,

Springer-Verlag herewith grants permission to include a re-print of your paper:

Prospecting an Inclusive Search Mechanism for Social Network Services

which was published in LNBIP 73, ISBN 978-3-642-19802-1, pp. 555- 570, 2011,

in your MSc dissertation. Please include a reference to the LNBIP version.

Best Regards,

Ralf Gerstner
Computer Science Editor