

# Busca Semântica em Redes Sociais Inclusivas na Web

Uma abordagem baseada nos significados criados e compartilhados na Rede

Júlio César dos Reis, RA: 044415

Instituto de Computação  
Universidade Estadual de Campinas  
Campinas – SP, Brasil  
julio.reis@ic.students.unicamp.br

**Abstract—** Social networks systems on the Web is presented as a virtual space for users interact, communicate and make access and use of information. These systems could provide inclusive access to individuals by creating situations in which the difficulties of the users are minimized. One way to make this access from these networks is mainly through search engines. However, these mechanisms are currently made only through lexical-syntactic processing. So mainly due to the excess of information, context issues, and the difficulties that may be encountered by some users trying to access the digital content from this current solution, it shows that is extremely important to develop new and more powerful mechanisms of recovery of information, which takes into account in the search the meanings (semantics) created, shared and used by people in their own social network. Therefore, the objective of this paper is to investigate the basic computational requirements needed to develop methods and tools in order to computationally identify and represent the semantics that comes from the individual's interaction and communication in the social network. With this, we glimpse to create most appropriate mechanisms for sophisticated semantic search engines in inclusive social networks, which takes into account the meanings used in the network, providing more precise searches and thus helping users to access information easier.

**Keywords:** social network services; semantics; semantic search; organizacional semiotics; ontologies; copycat architecture

## I. INTRODUÇÃO

Nos últimos anos, com o avanço da tecnologia *Web*, surgiu uma categoria de *software* denominada de “software social” (*Social Software*). Estes *softwares* introduziram novas oportunidades de interação e compartilhamento de informação e conhecimento entre indivíduos. Nesta categoria estão os sistemas para redes sociais (*Social Network Services*), estes sistemas permitem principalmente indivíduos compartilharem seus interesses e atividades, além de formarem comunidades que conectam pessoas com interesses comuns.

Neste contexto, os indivíduos na rede se comunicam, trocam dados e informações e com isso significados são criados e compartilhados (emergem) entre os membros da rede ou de uma comunidade. Desta maneira, é importante a partir da interação e comunicação destes indivíduos, construir mecanismos computacionais para identificar e representar estes significados que podem ser próprios de uma rede social. Isto pode facilitar diversos fatores na interação com sistemas computacionais, além de propiciar o acesso e o compartilhamento do conhecimento de maneiras mais eficazes, podendo resultar em ganhos sociais.

As redes sociais significam uma oportunidade de interação e acesso a informação e ao conhecimento através da *Web*. Há propostas vigentes Baranauskas [1], que se propõem a lidar com barreiras que impedem o acesso do cidadão brasileiro a *internet*, e transformar sistemas para redes sociais em um motor da inclusão

digital e do desenvolvimento da cidadania. Entretanto, para que a informação e o conhecimento possam ser acessados e compartilhados de maneira adequada, não basta apenas os usuários se comunicarem e trocarem dados e informação, é necessário que estes sejam entendidos entre as partes, sejam elas indivíduos ou pelo sistema, e que haja maneiras mais eficientes e precisas para recuperar estas informações.

Portanto, é necessário o entendimento e mapeamento dos dados computacionalmente do ponto de vista semântico. Entende-se por semântica a interpretação dos signos por indivíduos e a sua ligação com elementos do mundo real. Esta interpretação está socialmente contextualizada, em outras palavras, indivíduos e comunidades podem ter diferentes interpretações para um mesmo signo, assim como um signo pode conotar significados diversos dependendo do contexto aplicado.

É proposto nesta pesquisa estudar principalmente estes aspectos semânticos em sistemas de redes sociais. Mais especificamente na identificação da emergência de semântica (significados que são criados e compartilhados pelos indivíduos com base em suas interações na rede social ao longo do tempo), e como através de um modelo estes significados podem ser melhor representados computacionalmente e conseqüentemente utilizados na busca pelos usuários da rede. Acredita-se que identificar a emergência de semântica que ocorre na rede será importante para atualizar o modelo semântico e conseqüentemente fará diferença nos resultados da busca. Para lidar com estes aspectos, objetiva-se a construção de métodos e ferramentas computacionais que analisam os dados armazenados no sistema de rede social, levando também em consideração as ações executadas pelos participantes em suas atividades de comunicação e interação na aplicação. Todavia, neste primeiro instante, visa-se principalmente investigar os requisitos teórico-metodológicos e computacionais necessários para que esta solução tecnológica seja viável e desenvolvida, para tanto, a proposta está principalmente fundamentada teoricamente na Semiótica Organizacional (SO) e na investigação do uso da arquitetura *Copycat* [2].

Vislumbra-se com esta proposta que seja possível obter mecanismos de busca mais sofisticados e adequados para redes sociais inclusivas, propiciando buscas semânticas mais precisas, que possam auxiliar os usuários iletrados digitalmente a acessarem as informações disponíveis na rede com mais facilidade e naturalidade. Para isso, este trabalho está estruturado da seguinte forma: a seção II apresenta o conceito de redes sociais inclusivas e a sua relevância para o acesso à informação; na seção III é discutida sobre buscas semânticas e a importância da representação semântica; a seção IV elucida o referencial teórico da SO na qual servirá de embasamento teórico para a proposta; a seção V apresenta os requisitos necessários para a solução em perspectiva: o desenvolvimento de um modelo semântico computacional a partir da rede social e finalmente na seção VI o trabalho é concluído.

## II. REDES SOCIAIS INCLUSIVAS E O ACESSO À INFORMAÇÃO

As redes sociais são criadas e constituídas principalmente por grupo de usuários que se relacionam, criando laços sociais, compartilhando informação e conhecimento. Esses sistemas funcionam com o primado fundamental da interação social, ou seja, busca conectar pessoas e proporcionar sua comunicação e, portanto, podem ser utilizados para forjar laços sociais [3].

Segundo Garton *et al.* [4] quando uma rede de computadores conecta pessoas ou organizações, esta é uma rede social e que pode ser uma representação explícita de amizade, relação profissional, ou troca de informação entre um grupo de pessoas. Elas representam uma forma interessante de conectar pessoas através da *internet*, constituindo grupos com interesses afins além de poder ser palco para discussões e aprendizado.

Estes sistemas são uma realidade presente nos dias atuais. As redes sociais ou as "comunidades de membros" têm grande relevância na *Web*, pois os internautas dedicam grande parte do tempo de navegação nelas, conforme revela os dados de Nielsen [5]: (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 - como o *Orkut* e o *Facebook*; (3) Os internautas brasileiros também são os que passam mais tempo neste tipo de site. Equivale a 1' - 4' minutos de navegação na *internet*; (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 [5], as redes sociais são mais populares do que o *e-mail*, com 66,8% de alcance global. No mundo figuram no quarto lugar entre os recursos mais utilizados na *internet* e 85,2% de penetração estão nos portais e comunidades de interesse geral. Adicionalmente, 85,9% dos internautas usam as ferramentas de busca que é uma das atividades mais procuradas.

Neste contexto, observa-se a importância do estudo de tais sistemas; e partindo de uma perspectiva inclusiva para a interação dos usuários com sistemas computacionais, tem-se uma grande preocupação em como possibilitar acesso irrestrito às pessoas ao conteúdo *on-line* disponível a partir destes sistemas de maneira mais natural e eficiente. As diferenças físicas, sensoriais, cognitivas e culturais dos usuários devem ser consideradas e respeitadas no desenvolvimento e uso destes sistemas. Desta forma, é extremamente importante criar métodos mais sofisticados que possibilite de maneira efetiva o acesso e uso por todas as pessoas, (por exemplo: as letras digitais ou alfabetos funcionais) da informação veiculada na mídia digital. Isto pode se dar através das redes sociais inclusivas.

Estas redes podem ser desenvolvidas como um meio de inclusão digital dos cidadãos. Segundo Baranauskas [1]:

*"No contexto de Brasil, vivenciam-se enormes diferenças sócio-econômicas, culturais, regionais e de acesso à tecnologia e ao conhecimento. Esse é um cenário onde o desafio é único: fazer com que as Tecnologias da Informação e Comunicação, via suas interfaces de usuário, beneficiem o conjunto dos cidadãos, promovendo o processo de constituição de uma sociedade mais justa e aberta às diferenças. O grande desafio da Computação para mudar esse quadro passa pela busca de métodos e design de sistemas que possibilitem o acesso e façam sentido para a comunidade de usuários, sustentando a constituição de uma cultura digital e respeitando as diversidades."*

É importante criar espaços para que as pessoas sejam incluídas de maneira natural a tecnologia. Segundo Baranauskas [1] soluções de Inclusão Digital (ID) devem ser tratadas como instrumento de transformação profunda da sociedade; dentre as

soluções possíveis estão os serviços nas redes sociais inclusivas. Este se trata de um "espaço virtual de comunicação", que seja inclusivo e que permita à comunidade compartilhar conhecimento sobre "o fazer comunitário". Esse espaço deve ser generalizado para possibilitar "trocas" (de conhecimento, bens e serviços). Enfim, não há uma definição fechada do conceito de rede social inclusiva, contudo algumas idéias podem ser melhor observadas em Hayashi *et al.* [6].

Nas redes sociais inclusivas não há usuários alvos, mas sim todos os usuários são relevantes e considerados. Assim, pode haver usuários sem habilidades para manusear os recursos do sistema e consequentemente sem conhecimento para encontrar as informações que necessitem com rapidez e sucesso. Uma maneira de efetuar o acesso e uso das informações geradas na rede pode ser feita principalmente através de mecanismos de buscas. Devido principalmente à dificuldade que podem ser encontradas por estes usuários, juntamente com o excesso de informação e problemas de contexto não resolvido por buscas sintáticas, é extremamente importante haver mecanismos de recuperação de informação mais poderosos, que levem em consideração o vocabulário e significados criados, compartilhados e utilizados por estes usuários na rede social.

A melhoria deste recurso em uma rede social inclusiva pode dar-se principalmente pela transparência que o usuário terá ao procurar uma informação e na qualidade e precisão de resposta. Desta maneira, deve-se procurar por uma solução computacional de busca situada no significado que emergiu no contexto de uso daquela rede (significados que as pessoas já têm e também que criam com o uso do sistema ao longo do tempo). Isso poderá facilitar e prover um melhor acesso ao conteúdo gerado pelos usuários da rede. Para isso estudos relativos à semântica, como ela emerge na rede e sua representação computacional devem ser desenvolvidos.

## III. BUSCA SEMÂNTICA E A IMPORTÂNCIA DAS REPRESENTAÇÕES

Para que a busca seja semântica, é necessário que a máquina tenha conhecimento sobre o domínio em trabalho, ou seja, é necessário representar o conhecimento disponível de alguma maneira que a máquina entenda. Tazi [7] defende que o conhecimento pode ser representado com os chamados Grafos Conceituais de Sowa (que é um modelo geral para representar conhecimento). Esta abordagem segue a idéia de Aristóteles de que cada conceito é representado por uma palavra ou símbolo e um conjunto de referenciais do mundo. Atuando como uma rede semântica, os nodos representando conceitos são relacionados entre si.

Outra maneira de representar conhecimento é através de ontologias. Segundo Studer *et al.* [8], uma ontologia é um entendimento comum e compartilhado de algum domínio que pode ser comunicado entre pessoas e computadores; é uma especificação formal (deve ser capaz de ser lida e entendida por máquinas) e explícita de uma conceitualização compartilhada (de um grupo e não individual; deve ser um consenso). Wiederhold [9] vê uma ontologia como um conjunto de termos e relacionamentos usados em um domínio, denotando conceitos e objetos. E de acordo com Gruber [10], as ontologias fornecem descrições sobre conhecimento.

Acredita-se que com o uso destes artefatos, tenha-se uma maneira de se ter algum tipo de semântica (conhecimento do mundo) pela máquina. Guha *et al.* [11] argumentam que a adição de semântica explícita pode melhorar as buscas. A busca semântica tenta aumentar e melhorar os resultados da busca tradicional. Segundo Bonino *et al.* [12] o ponto chave para o

processo de refinamento de uma busca semântica está na disponibilidade de uma ontologia de domínio e na capacidade de compreender as relações semânticas entre os conceitos ontológicos. Isto é importante, pois as buscas são bem dependentes de contextos devido aos vários significados que uma mesma palavra pode denotar, contendo sinônimos e polissemia, por exemplo: 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. Em uma busca que não leve em consideração o significado, resultados com ambos os conceitos poderiam ser retornados.

Segundo Botero [13] os métodos de recuperação de informação baseados na utilização de ontologias surgiram como uma resposta para as deficiências do modelo baseado somente na representação léxica de termos. As deficiências desse modelo afetam diretamente os requisitos de precisão e cobertura dos buscadores: documentos não relacionados à busca são recuperados e documentos relevantes não são. Esses problemas ocorrem porque a busca baseada na representação léxica indexa os documentos com base em termos que representam informações vagas e, algumas vezes, representam mais ruído do que informação útil. Outro problema é que a informação requisitada pelo usuário está muito mais relacionada a conceitos ou assuntos do que a palavras-chaves. Bonino *et al.* [12] argumentam que uma grande melhoria na relevância dos resultados poderia ser alcançada diminuindo a distância entre sintaxe e semântica. Sabendo “exatamente” o que o usuário quer dizer quando especifica um termo de busca e tendo a descrição do conteúdo da informação, poderia permitir sistemas de recuperação que provem resultados focados, que melhor satisfazem o usuário.

De acordo com Guha *et al.* [11] a busca semântica tem dois objetivos. Primeiro é melhorar o resultado de buscas tradicionais com dados retirados da *Web* e segundo é usar o entendimento da denotação do termo de busca utilizado para melhorar as buscas tradicionais. Além disso, descrevem como a busca semântica pode melhorar os resultados das buscas tradicionais, dizendo que há três problemas a serem endereçados: (1) Denotação: é necessário determinar o conceito denotado pelo termo de busca, se não; (2) O que apresentar: determinar qual dado relevante retornar; e (3) Apresentação: é necessário formatar apropriadamente os dados na apresentação para a inclusão nos resultados da busca. Além destes aspectos explicitados por [11] e [12], argumenta-se na importância do modelo semântico utilizado como base na busca, ou seja, para que uma busca semântica seja de qualidade e tenha melhores resultados, o modelo semântico utilizado, que pode ser uma ontologia, deve ser bem projetado, desenvolvido e evoluído.

Atualmente as ontologias vêm sendo largamente utilizadas, inclusive para buscas semânticas, contudo a literatura tem mostrado diversos problemas e limitações referentes a esta. De acordo com Carvalho [14] mesmo com o advento das ontologias ainda não há ferramentas que auxiliam na organização das informações da mesma forma em que a mesma é feita mentalmente e socialmente. A organização dentro de uma ontologia é feita de maneira formal, isto é, realizada através de uma relação fixa das palavras de forma a facilitar o trabalho para o computador. Então ao estabelecer uma hierarquia entre conceitos a mesma se mostrar incapaz de representar com precisão contextos distintos, o que obriga a ontologia estar presa a um contexto bem delimitado.

Ainda segundo Carvalho [14], é necessário discutir toda a problemática de relacionamento e contextualização das informações presentes nas ontologias. Essa contextualização é gerada a partir de um estudo aprofundado dos temas exigidos para o entendimento do assunto em questão. Esse estudo pede uma série de conceitos-chave, que resumem o conhecimento da área.

Esses conceitos precisam ser organizados de maneira a produzir um conhecimento, uma “árvore” do conhecimento. Essa árvore deve ser capaz de traduzir àquele assunto, representando o mais fiel possível. Para tanto é necessário entender o processo mental como um processo abstrato que contrapõe a subjetividade com a objetividade para conseguir apreender o fenômeno.

De acordo com Tanasescu & Streibel [15] há diversos outros argumentos que mostram os problemas da representação do conhecimento com base em ontologias, como: a inadequação de raciocínio baseado em categorias para representar a realidade, a ausência de fundamentação simbólica, a necessidade de diferentes representações da mesma identidade de acordo com o contexto e também a dificuldade de representar conceitos psicológicos, como os *Affordances* de Gibson [16] em uma estrutura hierárquica. Em Gärdenfors [17] também é apontado de maneira mais elaborada diversos problemas com base nesta representação.

Em suma, ontologias são bem adequadas para a descrição de domínios científicos como a medicina e a biologia que já são semi-formais e organizadas em categorias e relações [15]. Contudo, em outros contextos a semântica também deve ser mapeada de alguma maneira, assim ontologias devem ser criadas e mantidas não exclusivamente por humanos, mas sim geradas automaticamente e ou semi-automaticamente com base em textos e ou outras fontes. Segundo Loh & Garin [18] este ainda é um desafio existente.

Finalmente, em uma rede social, na qual a partir da interação dos usuários ocorre comunicação e conseqüentemente significados são criados e compartilhados, o modelo semântico deve melhor refletir os significados utilizados na rede social, com suas devidas relações e particularidades. Desta maneira, estratégias devem ser empregadas tanto para o projeto, aprendizado e evolução deste modelo semântico. Busca-se como resultado final um modelo semântico fiel aos significados utilizados na rede social, na qual possam sempre ser atualizado e empregado na busca e que leve em consideração não apenas os conceitos e suas relações, mas outras relações importantes do mundo real. Para isso, neste trabalho ir-se-á utilizar como base teórica o referencial da SO (veja seção IV) como uma abordagem para o *design* deste modelo. Além disso, vislumbra-se empregar a arquitetura *CopyCat* [2] como uma estratégia tanto para o aprendizado quanto para a evolução deste modelo.

#### IV. REFERENCIAL TEÓRICO DA SEMIÓTICA ORGANIZACIONAL

O principal objetivo desta seção é elucidar o referencial teórico da Semiótica Organizacional (SO), mostrando alguns conceitos desenvolvidos por esta que servirão de base para a proposta do modelo semântico computacional descrito na seção V. Semiótica como “ciência dos signos” ou teoria sobre produção de sentido e interpretação, tem sido objeto de estudo de pesquisadores em Linguística, Estudos de Mídia, Ciências Educacionais, Antropologia, Filosofia da Linguagem entre outros. A Semiótica como um todo ou em parte tem influenciado muitos destes estudos. Computação é outra área em que a Semiótica mostrou-se ser de grande relevância, especialmente nas áreas de interfaces humano-computador, semiótica computacional, entre outros trabalhos.

A SO pode ser compreendida como um dos “ramos” da Semiótica que entende uma organização (por exemplo, uma rede social) como um sistema de signos e objetiva estudá-las utilizando conceitos e técnicas baseados na Semiótica desenvolvida por Peirce [19] e Morris [20] entre outros. SO entende que todo comportamento organizado é afetado pela comunicação e interpretação dos signos pelas pessoas, de maneira individual ou em grupos.

A SO apresenta teorias e métodos que permite a análise e o *design* de sistemas de informação de acordo com três funções humanas: expressar significados, a comunicação interativa, e criação de conhecimento [21]. Estudos em SO não estão restritos a informação expressa de maneira escrita ou gráfica, eles também levam em consideração os aspectos semióticos da interação humana em uma organização. Do ponto de vista filosófico, na SO a realidade é compreendida como uma construção social baseada no comportamento dos agentes que participam dela; pessoas compartilham padrões de comportamento governados por um sistema de signos.

Métodos da SO são úteis para delinear a influência dos aspectos sociais nas organizações e na elicitação dos requisitos do sistema. Entre os métodos desenvolvidos pela comunidade da SO, está um conjunto chamado de MEASUR (*Methods for Eliciting, Analyzing and Specifying Users' Requirements*) [22], que lida com o uso de signos, suas funções na comunicação de significados e intenções, e nas consequências sociais.

Neste trabalho o principal método do MEASUR a ser considerado a princípio é o de Análise Semântica (AS) (embora não seja descartado o uso de outros métodos como complemento). Na AS o analista no papel de facilitador, especifica as funções requeridas em um diagrama de ontologia, que representa uma visão dos agentes responsáveis no domínio de negócio em foco e seus comportamentos e ações chamadas de *Affordances*. Em suma, a AS foca nos agentes e seus padrões de comportamento para descrever o comportamento organizado.

Os conceitos da AS adotados neste trabalho são baseados em Liu [23]. Neste o mundo é socialmente construído através das ações dos agentes, tendo como base o que é oferecido pelo mundo físico para ele próprio. Os conceitos são:

*Affordance*: Este foi introduzido por Gibson [16] que pode ser utilizado para expressar invariantes de repertórios de comportamento de um organismo possíveis pela estrutura do organismo combinada com a de seu ambiente. Na análise semântica, *Affordances* [23] são construções sociais válidas em um certo contexto social. Há diferentes conceitos de *Affordance*: o original de Gibson [16] e o conceito utilizado no MEASUR. O conceito de *Affordance* foi introduzido originalmente por Gibson [16] e pode ser utilizado para expressar invariantes de repertórios de comportamento de um organismo possíveis pela estrutura do organismo combinada com a de seu ambiente. “Eu quero dizer, ela (palavra *Affordance*) é alguma coisa que se refere para ambos o ambiente e o animal de uma maneira que nenhuma palavra existente o faz. Ela implica em complementaridade do animal e o ambiente.” [16, p.127]. *Affordances* não são somente propriedades físicas, eles são relativos à postura e ao comportamento do animal. Na SO, mais particularmente no MEASUR a noção que Gibson introduziu foi estendida por Stamper para incluir invariantes que são percebidas no mundo social. Estas invariantes, no mundo social, são chamadas de normas sociais por Stamper [24, p. 374]: “Gibson deu maior ênfase à percepção do mundo físico; entretanto, a noção de *Affordance* pode ser generalizada para incluir invariantes que nós percebemos no mundo social. Por exemplo, se alguém tem uma patente de direitos autorais então deveríamos ser hábeis em supor um comportamento invariante das pessoas em direção a respeitar o seu trabalho. Um “copo” tem vários invariantes sociais que são válidos, por exemplo: para permitirmos beber algo de uma maneira aceitável em uma companhia refinada e ele pode ter também a invariante de posse, assim como a patente pode . . .”. Desta maneira, na AS, *Affordances* Liu [23] são **construções sociais válidas** em um certo contexto social. Neste sentido, Stamper [25] argumenta que a realidade que conhecemos não foi construída de maneira individual, mas criada pelo desenvolvimento das culturas durante

séculos ou milênios. Um copo é um artefato humano e sua utilização é possível não somente por causa de seus aspectos físicos, mas também porque ele tem *Affordances* sociais (uma criança aprende que o copo não é para jogar em outras pessoas, mas para tomar água). Mesmo o mundo físico conhecido através das ciências naturais neste sentido é predominantemente uma construção social e a ciência é a atividade social de construí-lo.

*Agente*: é um tipo especial de *Affordance*, que pode ser definido como alguma coisa que tem um comportamento responsável. Um agente pode ser um indivíduo, um grupo cultural, uma comunidade na rede social, sociedade, etc. (uma pessoa, um departamento, uma organização, etc. podem ser agentes).

*Dependência Ontológica*: é formada quando uma *Affordance* só é possível se outros *Affordances* existirem. Se o *Affordance* “A” é ontologicamente dependente do *Affordance* “B” significa que “A” somente será possível quando “B” também for possível. Por exemplo: para que uma pessoa tropece, primeiro ela deve andar; para duas pessoas se divorciarem elas devem estar casadas; assim existe dependência ontológica entre tropeçar e andar, e divórcio e casamento.

Há diversos outros conceitos como: Determinante, Especialização, Parte-Todo, Papel e Responsabilidade, contudo não foram descritos neste momento por não se mostrarem tão importantes para o entendimento do trabalho. Na próxima seção são apresentados os requisitos necessários e vislumbrados para a solução em perspectiva, mostrando os possíveis passos, metodologias e tecnologias que podem ser utilizados para se atingir o objetivo do trabalho, com base no referencial teórico apresentado nesta seção.

## V. REQUISITOS NECESSÁRIOS PARA A SOLUÇÃO: VISLUMBRANDO UM MODELO SEMÂNTICO COMPUTACIONAL A PARTIR DA REDE SOCIAL

Ao menos que os conceitos utilizados em uma rede social sejam identificados e extraídos, a semântica que existe e que emerge nesta rede não pode ser mapeada a priori em uma ontologia ou em qualquer outra representação do conhecimento. Em uma rede social, a emergência de semântica é um processo contínuo, na qual significados, conceitos e interpretações são construídos, transferidos e compartilhados através do sistema a partir das interações e expressões dos usuários por meio de comunicação. Desta maneira, é necessário criar mecanismos e ferramentas computacionais, para que estas interpretações expressas através das interações dos usuários com o sistema e também da interação entre usuários possam ser identificadas e representadas. Isto pode ocorrer através da análise desta interação e comunicação ao longo do tempo com o desenvolvimento da rede. A investigação neste aspecto mostra-se de suma importância, podendo gerar diversos avanços: modelos semânticos mais fiéis à realidade e como consequência buscas semânticas mais precisas, que levam em consideração os significados utilizados e emergidos na rede, que fazem sentido para os indivíduos daquele contexto.

Os usuários na rede podem utilizar um vocabulário próprio para a busca, assim a partir de uma representação computacional da semântica e o tratamento de sua evolução e atualização, objetiva-se facilitar e melhorar a qualidade e precisão na recuperação da informação. Um mecanismo desta natureza em uma rede social inclusiva pode facilitar o acesso do usuário à informação e a conteúdos veiculados na rede. Novas abordagens para esta representação computacional e sua evolução devem ser investigadas e propostas. Para isto, neste trabalho vislumbra-se 3 requisitos necessários para a solução.

O primeiro requisito (subseção A) trata do modelo semântico de maneira conceitual, objetiva-se neste desenvolver um modelo

que tenha o poder de melhor retratar a semântica utilizada na rede social, para isso ir-se-á utilizar conceitos da SO. Tendo um modelo formal, visa-se a partir deste no segundo requisito (subseção B), identificar as informações necessárias da rede social como: usuários e suas relações, anúncios, comentários de anúncios, bate-papo entre os usuários, sendo os dados extraídos com algoritmos de mineração de dados da base de dados, para gerar um modelo semântico e desenvolver uma representação computacional da semântica utilizada naquela rede social. O terceiro requisito (subseção C), dado que a rede social está em constante atividade e modificação, é necessário ter um maneira de evoluir este modelo semântico, para isto neste trabalho objetiva-se utilizar a arquitetura *CopyCat* [2].

### **A. Requisito 1: Modelo Conceitual Semântico dos significados utilizados na Rede Social**

Visa-se desenvolver um modelo conceitual para a representação semântica na rede social com base na SO. A idéia é que os conceitos da SO (descritos na seção IV) possam ser utilizados em conjunto com outras tecnologias computacionais, por exemplo, da *Web Semântica* [26] para descrição computacionalmente tratável das ontologias com o uso da linguagem *OWL* [27] para representar a semântica utilizada na rede. Contudo, neste trabalho ir-se utilizar destes conceitos junto com a arquitetura da *CopyCat* [2].

A idéia é incorporar principalmente os Agentes (papeis) e *Affordances* (padrões de comportamento) juntamente com os conceitos e as relações semânticas de ontologias tradicionais em um novo modelo ontológico. Isto se justifica a principio, pois segundo Liu [23] os conceitos são socialmente construídos, ou seja, são definidos a partir dos Agentes e seus *Affordances*. Assim um modelo semântico computacional deveria refletir estas premissas. Além dos Agentes e *Affordances*, observa-se que nas ontologias tradicionais não incorporam a idéia de relações de dependência ontológica (relativo ao existencial) no modelo. Desta maneira, seria interessante que tais dependências fossem representadas, deixando o modelo mais rico quanto a sua representatividade.

Logo, neste modelo, além das principais relações semânticas existentes em ontologias tradicionais, que não consideram os papeis dos indivíduos em sociedade e seus padrões de comportamento, fica descrito nesta nova proposta que, os Agentes tem um comportamento (*Affordance*) sobre um conceito (que possui atributos). Por exemplo: Uma costureira, que é um Agente, pode costurar, que é um padrão de comportamento da costureira, (ou seja, um *Affordance*) uma manga de camisa. Neste caso, manga é um conceito que pode ter diversos significados distintos, mas neste contexto, devido ao *Affordance* e Agente, o significado de manga é sobre camisa e não, por exemplo, sobre manga (fruta). Adicionalmente, estes conceitos podem estar relacionados a outros Agentes e *Affordances* além de outros conceitos. Na próxima subseção é apresentado mais detalhes dos requisitos necessários quanto a implementação deste modelo, sendo um segundo requisito para a solução.

### **B. Requisito 2: Mineração dos dados e a Criação do Modelo Computacional**

Para desenvolver uma representação semântica a partir da rede social e com base na SO é necessário a principio identificar algumas informações da rede para a construção do modelo semântico. Na rede social há expressão e comunicação dos usuários através principalmente dos anúncios, comentários e bate-papos. Estes dados estão armazenados em uma base de dados, na qual deve-se aplicar algoritmos de mineração de texto e dados.

Estes algoritmos devem de alguma maneira efetuar identificação de padrões para a formação dos conceitos, com algumas restrições e variáveis. Devem levar em consideração os Agentes, *Affordances*, dependências ontológicas e outras necessidades impostas pelo modelo conceitual. Eles não devem apenas fazer processamento de texto retirando os dados não necessários, mas também efetuar correlações e ter aprendido, por exemplo, o algoritmo tem que saber qual conceito já foi identificado e qual não foi e se são os mesmos. Estes algoritmos devem levar em consideração as relações sociais na rede, além de disponibilizar as entradas necessárias com formato específico (bem definido) para a construção do modelo semântico inicial e sua evolução.

A partir de pesquisas iniciais, foram encontradas algumas possibilidades de utilização de alguns *softwares* e algoritmos. O *software* Kea [28] é uma possibilidade. Este é um algoritmo para a extração de palavras chaves de texto de documentos. Ele pode ser usado para indexação livre ou indexação com vocabulário controlado. Outra possibilidade é a técnica *Latent Semantic Analysis* (LSA) [29]. Contudo, pesquisas e testes mais apurados devem ser feitos com estes para realmente verificar sua viabilidade ou se extensões poderiam e ou deveriam ser feitas nestes para se adequar às necessidades deste trabalho.

Observa-se que a partir destes algoritmos será necessária a extração de diversos requisitos. Estes algoritmos deverão ter o poder de retornar uma lista de possíveis Agentes, lista de possíveis *Affordances* (e suas possíveis relações com Agentes), uma lista de conceitos (e suas possíveis relações com *Affordances* e Agentes). Outras possibilidades seria a sugestão de sinônimos e polissemia criados a partir da rede.

Após a mineração inicial dos dados da rede social, têm-se as entradas necessárias para a criação de um modelo semântico. O desenvolvimento deste seria feito através de um processo semi-automático (com intervenção e ajuda humana). Contudo, seria interessante a ferramenta poder dar uma sugestão inicial da organização da representação com as relações semânticas, e também levar em consideração a última versão da representação desenvolvida agregando novos conceitos e relações encontrados. Este deve utiliza-se do Requisito 1 como base (modelo conceitual) e dos resultados gerados pelos algoritmos de mineração para o seu desenvolvimento. A principio objetiva-se utilizar técnicas de *Ontology Learning* no processo automático e o humano produziria como resultado final um modelo semântico consistente e estável. Este modelo seria a entrada inicial para o Requisito 3 (Evolução do Modelo), ou seja, a *Slipnet* inicial do *Copycat*. A próxima subseção ilustra os detalhes.

### **C. Requisito 3: Evolução do modelo com a Arquitetura Copycat e seu uso na busca semântica**

Segundo Hofstadter & Mitchell [2] a arquitetura *Copycat* é uma arquitetura cognitiva que objetiva ser capaz de descobrir analogias de uma maneira psicologicamente realística. É um processo para a categorização de algo que entra pelo aparelho sensorial através de um mecanismo de analogia. A arquitetura é baseada em 3 elementos: *Slipnet*, *Workspace* e o *Coderack*.

Ainda de acordo com Hofstadter & Mitchell [2] a *Slipnet* funciona como uma rede semântica de conceitos relacionados que são representados como nós, na qual se pode saltar de um conceito para outro da rede, tentando encontrar qual o conceito mais apropriado dado uma entrada. Cada relação conceitual é feita por um *link* que tem um valor numérico associado, que representa a distancia conceitual entre os dois nós envolvidos. A *Slipnet* não é estática, ela é dinâmica e responde a situações em questão. Os nós adquirem níveis variados de ativação – em função do que está no *Workspace* (que podem ser pensados como uma medida de

relevância para uma situação em questão). Assim irá espalhar um número de ativações para os conceitos vizinhos, e com o tempo a ativação pode diminuir. Esta ativação pode variar continuamente. No entanto, quando a ativação de um nós ultrapassa um limiar, este nó tem a probabilidade de pular para um estado de total ativação e também pode decair. A ativação, ou seja, a relevância percebida de cada conceito é uma função sensível, que muda com o tempo e com a maneira que o programa correntemente entende a situação que esta sendo enfrentado.

O *Coderack* armazena um pequeno conjunto de certos agentes autônomos chamados de *codelets* (pequenos trechos de código executável que executam uma função) e eles podem executar em um determinado tempo. Ao executar, os *codelets* executam uma ação no *Workspace*. Esta ação no *Workspace* irá causar uma mudança na *Slipnet*, que enviará novos *codelets* para o *Coderack*, fechando o ciclo da arquitetura *Copycat*. Em suma, o problema em questão excita a *Slipnet*, cada um dos nós da *Slipnet* recebe uma ativação em função do que está no *Workspace*, que passa para a ativação do *codelets*. A seleção da execução dos *codelets* é probabilística até convergir para uma solução ideal [2].

A idéia neste trabalho para a evolução do modelo semântico é utilizar esta arquitetura do *Copycat* e possivelmente efetuar uma melhoria desta. Utilizando o resultado da mineração como entrada para a arquitetura, estes seriam possíveis conceitos a serem incorporados na *Slipnet*. Busca-se a partir do processo das analogias feita pela arquitetura, sugerir relações semânticas entre as novas entradas e a *Slipnet* já formada. Ou seja, o objetivo é tentar evoluir a *Slipnet* a partir das próprias relações semânticas encontradas na *Slipnet* e do *Workspace*.

Um problema observado é que no *CopyCat* original, tem-se regras iniciais para ser fazer as analogias, no contexto deste trabalho, não há estas regras, apenas a *Slipnet* e as novas entradas. Assim a arquitetura de alguma maneira deve tentar criar novas relações semânticas dos conceitos que já estão na *Slipnet*. A versão inicial desta é dado pelo modelo consistente do Requisito 2, e sua evolução deve seguir o modelo conceitual do Requisito 1. Vislumbra-se que esta seria uma nova técnica para *Ontology Evolution*.

Além da evolução do modelo semântico com base em analogias, objetiva utilizar a arquitetura *CopyCat* também para o processo de busca. Ou seja, após o usuário entrar com um termo para a busca, a ferramenta tentaria fazer um processo de analogia e análise de relações semânticas deste com o modelo da *Slipnet*, assim, o objetivo é inferir quais outros conceitos e ou ações, devido aos *Affordances*, são possíveis. Com isto poderia-se efetuar uma busca utilizando-se deste resultado, retornando resultados semanticamente correlatos. Por exemplo, um usuário pode digitar na busca o termo “salgadinho”. Caso não haja nada no sistema com “salgadinho”, a partir das analogias e relações semânticas, o sistema pode retornar “pão-de-queijo”, ou algum outro tipo de comida semanticamente próxima.

Nesta solução de busca com o *CopyCat*, deve-se levar em consideração diversos fatores: a entrada de informação requerida (palavra-chave); informações do usuário que está fazendo a busca (perfil); A representação semântica construída, instanciada pela *Slipnet* e em constante evolução, junto com a arquitetura que fará as analogias e análise semântica, além dos dados na base de dados (conteúdo em si). Nestas informações, o perfil do usuário é importante devido a uma possível descoberta de contexto. Ou seja, a partir do usuário, busca-se identificar o Agente representado no modelo semântico, desta maneira pode-se limitar o espaço de busca, fazendo uma relação do usuário com o modelo semântico gerado; por exemplo, caso uma costureira esteja logada no sistema (descobre que é costureira com base no seu perfil) e solicite uma busca com a palavra chave “manga” e no modelo semântico há

uma relação entre o Agente “costureira” e o conceito “manga”, muito possivelmente os resultados que devem ser retornados devem estar relacionados a manga de camisa, e não a outros significados da palavra manga. Não que outros resultados não devam e ou possam ser retornados, mas este em específico sobre manga de camisa deve ter maior relevância no *ranking* dos resultados.

## VI. CONCLUSÃO

As redes sociais na *Web* atualmente são uma realidade e deveriam ser inclusivas, ou seja, proporcionar acesso inclusivo a todos os usuários. Neste contexto, as buscas semânticas podem ter grande relevância ao permitir o acesso à informação a partir de sua recuperação mais precisa e natural. Para isso, neste trabalho argumentou-se que desenvolver e evoluir um modelo semântico computacional adequadamente a partir da rede social mostra-se de extrema relevância para a qualidade e sofisticação destas buscas. Logo, buscou-se investigar os requisitos teórico-metodológicos e computacionais necessários para que esta solução tecnológica seja viável utilizando o referencial teórico da Semiótica Organizacional e a arquitetura *Copycat*. Esta abordagem revela-se como uma alternativa interessante tanto para o *design* diferenciado quanto para a evolução de um modelo semântico, que poderão gerar impactos diretos na busca. Contudo aprofundamentos em como viabilizar esta abordagem com implementações computacionais são necessários, podendo verificar de fato a viabilidade da proposta. Estes constituem os próximos passos do trabalho.

## REFERÊNCIAS

- [1] Baranauskas, M.C.C. (2007): e-Cidadania: Systems and Methods for the Constitution of a Culture mediated by Information and Communication Technology. Proposal for the Microsoft Research-FAPESP Institute.
- [2] Hofstadter, Douglas & Mitchell, Melanie. (1995): The Copycat project: a model of mental fluidity and analogy-making in Hofstadter, D., "Fluid concepts and creative analogies: computer models of the fundamental mechanisms of thought", Basic Books, Inc. New York, NY, USA, pp. 205 – 267.
- [3] Recuero, R. C. (2004): Redes sociais na Internet: Considerações iniciais. Disponível em: <pontomidia.com.br/raquel/intercom2004final.pdf> Último acesso em: Junho de 2009
- [4] Garton, L.; Haythornthwaite, C.; B. Wellman. (1997): "Studying Online Social Networks," JCMC, vol. 3.
- [5] Global Faces and Networked Places. (2009): The Nielsen Company, USA.
- [6] Hayashi, E.C.S.; Neris, V.P.A.; Almeida, L.D.A.; Rodriguez, C.L.; Martins, M.C.; Baranauskas, M.C.C. (2008): Inclusive Social Networks: Clarifying Concepts and Prospecting Solutions for e-Cidadania. Technical Report IC-08-029, Institute of Computing, University of Campinas - UNICAMP
- [7] Tazi, Saïd. (1994): Using Sowa's conceptual graphs for enhancing hypertext readers performances. Intelligent Hypertext Workshop, Washington. citeseer.ist.psu.edu/12096.html
- [8] Studer, R. et al. (1998): Knowledge engineering: principles and methods. Data & Knowledge Engineering, v.25, n.1/2.
- [9] Wiederhold, G. (1996): Foreword: intelligent integration of information. Journal of Intelligent Information Systems, v.6, n.2/3.
- [10] Gruber, T. R. (1993): A translation approach to portable ontologies. Knowledge Acquisition, v.5, n.2.
- [11] Guha, R.; McCool, R.; Miller, E. (2003): Semantic Search. Proceedings of the 12th international conference on World Wide Web. pp. 700-709. Budapest, Hungary
- [12] Bonino, D.; Corno, F.; Farinetti, L.; Bosca, A. (2004): Ontology Driven Semantic Search. WSEAS Transaction on Information Science and Application, Issue 6, Volume 1, pp. 1597-1605.
- [13] Botero, S. W. (2008): Extração de relações semânticas via análise de correlação de termos em documentos. Dissertação de Mestrado – Faculdade de Engenharia Elétrica e de Computação. Universidade Estadual de Campinas - UNICAMP.

- [14] Carvalho, M. L. B. (2005): Web semântica e semiótica: ontologias e aplicação a permacultura. Dissertação (Mestrado em Programa de Pós-Graduação em Ciência da Computação) - Centro Universitário Eurípides Da Marília
- [15] Tanasescu, V. & Streibel, O. (2007): Extreme Tagging: Emergent Semantics through the Tagging of Tags. In International Workshop on Emergent Semantics and Ontology Evolution (ESOE2007). Busan, Korea.
- [16] Gibson, J. J. (1979): The Ecological Approach to Visual Perception. Houghton Mifflin Company, Boston, Massachusetts, 127-143.
- [17] Gärdenfors, P. (2004): How to Make the Semantic Web More Semantic. In Formal Ontology in Information Systems, Proceedings of the Third International Conference (FOIS), pp. 17-34.
- [18] Loh, S. & Garin, R. S. (2001): Web Intelligence: Inteligência Artificial para Descoberta de Conhecimento na Web. In: V Oficina de Inteligência Artificial, Pelotas, RS. EDUCAT - UCPEL, pp. 11-34.
- [19] Peirce, C.S. (1931-1958): Collected Papers, Cambridge, Mass: Harvard University Press.
- [20] Morris, C. W. (1938): Foundations of the theory of signs, International Encyclopedia of Unified Science, University of Chicago Press, 1 (2).
- [21] Stamper, R. K. (2001): 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.
- [22] Stamper R.K. (1993): Social Norms in requirements analysis - an outline of MEASUR. In: Jirotko M, Goguen J, Bickerton M. (eds) Requirements Engineering, Technical and Social Aspects. Academic Press, New York.
- [23] Liu, K. (2000): Semiotics in information systems engineering, Cambridge University Press.
- [24] Stamper, R. K. (1996): Signs, Information and Systems, in B. Holmqvist, et. Al. (Eds) Signs of Work Semiotics Information Processing in Organisations, Walter de Gruyter, N. Y.
- [25] Stamper, R. K. (2000): Information Systems as a Social Science: An Alternative to the FRISCO Formalism. In E. D. Falkenberg, K. Lyytinen, A. A. Verrijn-Stuart (eds) Information System Concepts: an Integrated Discipline Emerging, Kluwer Academic Publishers, USA, pp.1-51.
- [26] Lee, T. B.; Hendler, J.; Lassila, O. (2001): The Semantic Web, Scientific American, May 2001 issue.
- [27] W3C (2004): "OWL-Web Ontology Language", Recommendation 10 February 2004, <<http://www.w3.org/TR/owl-features/>>, Accessed March, 2009.
- [28] Witten, I. H.; Paynter, G. W.; Frank, E.; Gutwin, C.; Nevill-Manning, C. G.. (1999): KEA: practical automatic keyphrase extraction. In: Proceedings of the fourth ACM conference on Digital libraries table of contents. Berkeley, California, United States. pp. 254-255.
- [29] Deerwester, S.; Dumais, S.; Furnas, G. W.; Landauer, T. K.; Harshman, R. (1990): "Indexing by Latent Semantic Analysis". *Journal of the American Society for Information Science* 41 (6): pp. 391-407.