

Analizando sistemas de classificação na *web* sob a perspectiva da interação social em comunidades de prática

Fernando M. Figueira Filho ¹

Paulo Lício de Geus ²

Instituto de Computação - UNICAMP
{fmarques¹, paulo²}@ic.unicamp.br

João Porto de Albuquerque

Escola de Artes, Ciências e Humanidades - USP
j.porto@usp.br

ABSTRACT

Social interaction mediated by information technology has been studied from several perspectives over the past years. This paper draws attention to the relevance of classification systems and categories as important mediation artifacts in human-computer interaction. In this sense, classification schemes provide meaning to information and mutual understanding between the parts in an interaction. However, studies in distributed cognition show us that classification schemes have a flexible nature and depend on the context in everyday human activity. Under this perspective, this paper analyzes three approaches to classification in the web: the semantic web, collaborative annotation and hybrid approaches. The analyses are based on the concept of communities of practice and on the perspective of context as a problem of social interaction in these communities. Based on these theoretical works, this paper derives two design principles for web classification systems: firstly, *collaboration* allows for participating users to be active agents in content classification. Secondly, *multidimensionality* permits the co-existence and management of multiple classification perspectives, which is of fundamental importance in complex, heterogeneous and large-scale infrastructures such as the web.

RESUMO

A interação social mediada pelas tecnologias de informação tem sido estudada a partir de diversas perspectivas ao longo dos últimos anos. Esse artigo chama a atenção para a relevância dos sistemas de classificação e suas categorias como importantes artefatos de mediação na interação humano-computador. Nesse sentido, esquemas de classificação provêm significado à informação e o entendimento mútuo entre as partes em uma interação. Entretanto, estudos em cognição distribuída nos mostram

que esquemas de classificação possuem uma natureza flexível e dependem do contexto na atividade diária humana. Sob essa perspectiva, esse artigo analisa três abordagens para classificação na *web*: a *web* semântica, anotação colaborativa e abordagens híbridas. As análises são baseadas no conceito de comunidades de prática e na perspectiva de contexto como um problema de interação social nessas comunidades. Baseado nesses trabalhos teóricos, esse artigo deriva dois princípios para o projeto de sistemas de classificação na *web*: primeiramente, a *colaboração* permite a participação dos usuários como agentes ativos na classificação de conteúdo. Como segundo princípio, a *multidimensionalidade* permite a co-existência e o gerenciamento de múltiplas perspectivas de classificação, o que é de fundamental importância em sistemas complexos, heterogêneos e de larga-escala como a *web* atual.

Author Keywords

Classification systems, web collaborative systems, communities of practice, social web.

ACM Classification Keywords

H5.3 [Information Systems Interfaces and Presentation (e.g. HCI)]: Group and Organization Interfaces—collaborative computing, web-based interaction, theory and models; H5.4: Hypertext/Hypermedia—theory, user issues; H5.m: Miscellaneous—classification systems.

INTRODUÇÃO

A *web* atual tem fomentado a criação de comunidades virtuais através de diversas ferramentas, como os *wikis*, os *blogs* e as redes sociais *online*. Essa tendência ocorre no contexto do que tem sido chamado de “*web* social”. O termo pode ser compreendido como uma mudança de paradigma que é observada, por um lado, no surgimento de aplicações que incentivam a interatividade entre as pessoas e possibilitam uma maior participação do usuário como produtor de informação, em detrimento da postura exclusiva de consumidor que era dominante na Internet de alguns anos atrás. Por outro lado, o novo paradigma tem levado a um crescente interesse em explorar, de maneira explícita ou implícita, a riqueza dos elos que se firmam com a interação social mediada pelas tecnologias de informação.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IHC 2008, October 21–24, 2008, Porto Alegre, Brazil.
Copyright 2008 ACM 978-1-60558-011-1/08/04...\$5.00

Na era da “web social”, embora observemos um crescimento no volume total de informação que é gerada de forma colaborativa, a informação possui cada vez mais um escopo local, sendo produzida com o propósito de ser compartilhada entre pessoas que têm em comum uma mesma atividade: escritores de poesia, amantes de um esporte, colaboradores em enciclopédias *online* (e.g. Wikipedia¹), desenvolvedores de software ou fotógrafos amadores. Na *web*, cada comunidade interage independentemente das barreiras geográficas que separam seus membros e o resultado dessa interação é o desenvolvimento de um conhecimento coletivo que atrai novos membros, num processo dinâmico e contínuo que é mediado pelas tecnologias de informação e sofre influência direta do *design* das ferramentas que promovem essa interatividade.

Nesse contexto, a crescente interação entre comunidades tem motivado os pesquisadores em IHC (Interação Humano-Computador) a ampliar o foco dos estudos na área, incorporando uma nova perspectiva do funcionamento da cognição humana que abandona uma visão estritamente centrada no indivíduo e propõe uma abordagem situada e coletiva, a qual abrange aspectos externos à mente humana e coloca os fatores sociais no mesmo patamar de importância para o estudo da cognição. Essa corrente, embasada nos trabalhos de pesquisadores como Hutchins [10], Star [16] e Lave [12], tem sido fundamental para um estudo mais abrangente da interação humana em sistemas de informação complexos, heterogêneos e de larga-escala, como é o caso da *web* atual.

Nesse artigo estamos interessados em analisar, sob essa perspectiva, as abordagens atualmente empregadas para classificação da informação disponível na *web*. Os impactos da classificação em sistemas de informação têm sido explorados em um largo espectro de contextos, conforme destaca o trabalho de Bowker e Star [4]. Contudo, pouco foi revelado nesse sentido com relação ao espaço informacional da *web* e sua infra-estrutura. O desafio que a *web* apresenta atualmente está evidenciado na crescente tensão entre o local – o conjunto de informações que é relevante para uma determinada comunidade – e o global – a informação que tem permeabilidade entre diferentes comunidades e serve como base para a interação entre elas. O papel dos sistemas de classificação é crucial no gerenciamento dessa tensão e o presente trabalho pretende mostrar que, embora a *web* e suas ferramentas apresentem novas abordagens para classificação da informação, os problemas encontrados nessas abordagens têm suas origens nos paradigmas e nas decisões de projeto que são utilizados para o desenvolvimento daquelas ferramentas.

Diante desse problema, o estudo aqui apresentado analisa duas tendências aparentemente opostas no que tange à classificação da informação na *web*: a) a estratégia proposta

pelos adeptos de Tim Berners-Lee e da *web* semântica [3] (por vezes também chamada de *Web 3.0*), que é baseada no uso de ontologias e de vocabulários controlados para representação da informação; e b) a abordagem de anotação colaborativa, que vem sendo largamente empregada em diversos serviços, dentre eles portais para compartilhamento de conteúdo (e.g. Flickr²) e também em ferramentas de *bookmarking* social (e.g. del.icio.us³). A anotação colaborativa se baseia na associação de termos (ou *tags*) aos conteúdos digitais, representando a informação do conteúdo segundo um vocabulário não-controlado que é criado livremente pelos próprios usuários de cada serviço.

O presente trabalho analisa essas abordagens baseado no conceito de comunidades de prática proposto por Wenger [19] e na perspectiva de contexto como um problema de interação social nessas comunidades, proposta por Dourish [5]. Com base nesses trabalhos teóricos, esse artigo deriva dois princípios para o projeto de sistemas de classificação na *web*: primeiramente, a *colaboração* permite a participação dos usuários como agentes ativos na classificação de conteúdo. Como segundo princípio, a *multidimensionalidade* permite a co-existência e o gerenciamento de múltiplas perspectivas – ou sistemáticas – de classificação, o que é de fundamental importância dada a heterogeneidade do conjunto de usuários da *web*.

O restante desse artigo está organizado como se segue: primeiramente, o artigo detalha cada uma das duas abordagens citadas acima, apresentando suas características estruturais e analisando suas peculiaridades técnicas. Em seguida, introduz os referenciais teóricos relacionados à perspectiva de cognição distribuída, que servem, mais à frente, como base para uma nova análise das abordagens apresentadas, sob essa outra perspectiva. A partir dessa análise, o artigo deriva dois princípios para o desenvolvimento de sistemas de classificação na *web* – a colaboração e a multidimensionalidade – terminando com uma conclusão.

CLASSIFICAÇÃO NA WEB

Atualmente a *web* apresenta duas tendências para classificação da informação disponível. Por um lado, a *web* semântica se baseia num conjunto de padrões de representação (e.g. RDF [11]) e linguagens para descrição de ontologias (e.g. OWL [14]), que compõem uma nova plataforma para o desenvolvimento de serviços. No cerne dessa abordagem está o uso de ontologias para classificação da informação.

Por outro lado, a abordagem de anotação (ou *tagging*) cresce bastante em popularidade nas ferramentas da *web* social. Trata-se de uma estratégia de classificação distribuída, com um forte caráter colaborativo e que se

¹ <http://www.wikipedia.org>

² <http://www.flickr.com>

³ <http://del.icio.us>

sustenta através das contribuições de seus próprios usuários. Essa seção pretende analisar em maiores detalhes cada uma dessas abordagens e apresentar uma breve revisão de alguns trabalhos recentes que propõem abordagens híbridas, com o objetivo de mesclar algumas das vantagens encontradas na abordagem de ontologias com a abordagem de anotação.

Ontologias

Segundo a definição de Gruber [8], uma ontologia é “uma especificação explícita da conceituação de um domínio”. Em outras palavras, ontologias são esquemas de classificação padronizados e, nesse aspecto, se assemelham aos demais esforços de padronização que conhecemos pela história afora, como a taxonomia dos seres vivos na biologia, a classificação internacional de doenças (ICD) elaborada pela organização mundial de saúde, dentre outros. Ontologias são, portanto, classificações formais, compostas de um vocabulário controlado para representação de categorias e cujos termos devem ser avaliados pela sua precisão e acuidade. Como em outras classificações formais, o projeto de ontologias visa à estabilidade temporal, isto é, interessa aos projetistas de uma ontologia que sua estrutura, vocabulário e relações conceituais mantenham-se estáveis ao longo do tempo num dado domínio de discurso, atendendo os requisitos iniciais e evitando esforços de reestruturação que podem ser bastante custosos.

Entretanto, conforme destaca Hepp [9], ontologias apresentam algumas restrições, tanto em relação à sua criação quanto à sua atualização, as quais são resumidas a seguir: i) Em determinados domínios do conhecimento, existe um atraso significativo entre a especificação estabelecida pela engenharia de ontologias e a dinâmica conceitual, isto é, novos conceitos passam a ser relevantes para um dado domínio, enquanto que outros caem na obsolescência. Essas mudanças não são absorvidas em tempo hábil pelas ontologias correspondentes, o que dificulta o desenvolvimento de ontologias para ambientes como a *web*, em que a produção de informação segue em ritmo acelerado, dando lugar a novos nichos informacionais e a uma crescente diversidade conceitual; ii) tomando uma perspectiva econômica, os ganhos provenientes da automação que se torna possível através de uma ontologia não superam, em muitos casos, os recursos gastos para desenvolver essas mesmas ontologias. Isso porque, no estágio atual, o desenvolvimento de ontologias se dá como resultado do trabalho de um conjunto relativamente pequeno de especialistas em um domínio, sendo esse um fator limitante para a capacidade de evolução dessas ontologias; iii) por serem desenvolvidas por especialistas, a necessidade de comunicação e entendimento entre os criadores de uma ontologia e os desenvolvedores para *web* se torna crucial para o sucesso da especificação da aplicação *web*, objetivo que não é facilmente alcançado. Atualmente, a única forma de comunicação entre essas partes é através de especificações escritas em linguagens específicas (e.g. RDF e OWL), cujo entendimento limita-se

geralmente a poucos programadores e a profissionais especializados. Como consequência, mesmo entre os especialistas em um domínio de discurso, poucos estão aptos a lidar com esses formalismos, e muito embora esse problema possa ser atenuado com o uso de ferramentas gráficas para construção de ontologias, ainda assim não há garantias de que o produto final irá atender aos propósitos de quem irá utilizá-lo.

Embora apresentem dificuldades para se adequar ao dinamismo da *web*, ontologias têm o potencial de solucionar uma série de problemas hoje enfrentados principalmente pelos mecanismos de busca de informação. Dentre esses problemas, podemos destacar a ambigüidade de termos causada pela polissemia – quando um só termo está associado a múltiplos sentidos (e.g. *mouse* pode referir-se ao dispositivo de computador, a uma espécie de roedor ou, ainda, ao sobrenome do roedor mais famoso do mundo) – e pela sinonímia – quando múltiplos termos representam o mesmo sentido (e.g. carro e automóvel). A argumentação dos adeptos da *web* semântica é que a linguagem hipertextual *per se*, composta de termos que estão dissociados de qualquer especificação formal de conceitos, leva naturalmente à imprecisão nos resultados de busca e a outros problemas relacionados à recuperação de informação.

Anotação colaborativa

Em outro extremo, a abordagem de anotação colaborativa cresce em popularidade e, assim como as ontologias, também apresenta vantagens e desvantagens. Na *web*, o produto da atividade colaborativa de anotação leva o nome de folksonomia⁴ e podemos observar o uso da abordagem em diversos serviços hoje disponíveis, dentre eles portais de compartilhamento de conteúdo (e.g. Flickr, You Tube⁵) e ferramentas de *bookmarking* social (e.g. del.icio.us).

Embora cada vez mais populares e abundantes, os sistemas de anotação são muito semelhantes entre si quanto às suas funcionalidades. Basicamente, uma vez autenticado no sistema, cada usuário tem a opção de associar termos (ou *tags*) aos conteúdos que compartilha ou mesmo aos conteúdos compartilhados por outros usuários. Não há controle sobre os termos que podem ser associados (vocabulário não-controlado), de modo que qualquer seqüência de caracteres, a princípio, pode ser utilizada.

Do ponto de vista do usuário consumidor de conteúdo, existem hoje duas maneiras de se utilizar uma folksonomia. A primeira é equivalente às tradicionais buscas por palavras-chave. Nesse caso, as palavras-chave são usadas como um filtro para selecionar os conteúdos previamente classificados com os termos usados na busca. A segunda é

⁴ Termo cunhado por Vander Wal [18], a partir da junção das palavras inglesas folk e taxonomy.

⁵ <http://www.youtube.com>

através das chamadas nuvens de termos (ou *tag clouds*), que representam um subconjunto de termos de uma folksonomia com pesos que variam de acordo com a frequência de associação de cada termo. Na nuvem, os termos mais frequentes geralmente possuem maior destaque (i.e. aumento da fonte).

As folksonomias possuem algumas vantagens principais: i) como produto de uma atividade coletiva de classificação, o resultado de uma busca pode trazer não somente conteúdos classificados pelo próprio usuário que realiza a busca, mas também conteúdos classificados por outros usuários. Essa característica confere maior abrangência às folksonomias, no sentido de que os esforços de classificação podem ser distribuídos entre uma gama abundante e variada de usuários; ii) as folksonomias estão relacionadas a um paradigma de organização que apresenta algumas vantagens com relação a outras abordagens mais tradicionais como, por exemplo, a de diretórios. Do ponto de vista estrutural, diretórios são hierárquicos e exclusivos, enquanto que folksonomias são planas e não-exclusivas [6], de maneira que, numa folksonomia, um conteúdo pode estar associado a diversas categorias ao mesmo tempo e não pertencer a nenhuma categoria exclusivamente. iii) os serviços de anotação oferecem uma baixa barreira de entrada aos seus usuários e isso funciona como um incentivo à colaboração na criação de folksonomias [1]. Em particular, sistemas de anotação são fáceis de usar, podem ser integrados aos navegadores na forma de *plug-ins* gratuitos (e.g. *del.icio.us*) e seus usuários não são obrigados a usar nenhum vocabulário controlado para classificar um conteúdo, ao contrário da abordagem de ontologias.

Entretanto, folksonomias têm em comum com a *web* tradicional as mesmas características que são alvo de críticas por parte dos defensores do emprego de ontologias e da *web* semântica. Particularmente, os termos de uma folksonomia não possuem nenhuma relação explícita entre si, ao contrário de um esquema de diretórios, no qual os nomes apresentam uma relação hierárquica. Além disso, os termos também não possuem nenhuma relação com conceitos formalmente especificados e, dessa forma, não é possível estabelecer nenhum tipo de relação semântica entre termos (e.g. não existe nenhuma regra que relacione os termos *carro* e *automóvel*). Essas características levam a ambigüidade que, do ponto de vista da recuperação de informação, é responsável pela geração tanto de problemas de precisão (i.e. recuperação de informação irrelevante), quanto de problemas de revocação (ou *recall*, i.e. a omissão de informações relevantes). A seção seguinte revisa alguns trabalhos que procuram solucionar esses problemas através de abordagens híbridas.

Abordagens híbridas

Folksonomias são estruturalmente planas, isto é, não há relações de qualquer natureza entre os termos que as compõem. Extrair relações semânticas entre os termos de uma folksonomia possibilitaria, por exemplo, a extensão de

consultas para termos relacionados ao contexto da pesquisa, bem como a desambiguação de consultas, quando necessário. Nas nuvens de termos, destaques poderiam ser dados não somente aos termos mais frequentes, mas também a termos semanticamente relacionados.

Com esse objetivo, algumas abordagens híbridas procuram integrar a estratégia de anotação com a abordagem da *web* semântica, na tentativa de, com isso, aumentar a capacidade de inferência sobre o espaço folksonômico. Nessas abordagens, a folksonomia é complementada por estruturas de dados auxiliares que estabelecem relacionamentos entre os termos de uma folksonomia.

Alguns trabalhos têm aplicado métodos estatísticos com esse objetivo. [13, 20, 2] exploram a co-ocorrência de termos na classificação de diferentes conteúdos. Begelman et al. [2] extrai do espaço folksonômico um grafo não-direcionado, com termos em vértices e com arestas ponderadas unindo termos relacionados, cujo peso indica o número de vezes que os termos co-ocorrem na folksonomia. Wu et al. [20] apresenta um modelo probabilístico, também com o objetivo de gerar agrupamentos (ou *clusters*) de termos relacionados semanticamente. Schmitz [13] extrai relações do tipo pai-filho entre pares de termos baseado no cálculo de probabilidades de um termo x ocorrer dado um termo y em contraste com a probabilidade de y ocorrer dado x .

Semântica como um problema de representação

O que há em comum entre a abordagem de ontologias e as abordagens híbridas recém-apresentadas é que ambas enxergam a semântica como um problema de representação. Nessa visão, categorias (ou termos) são relacionadas semanticamente a partir de estruturas de dados que, no caso das ontologias, são construídas *a priori* através do trabalho e da revisão de um grupo de especialistas, e no caso das abordagens híbridas, são automaticamente extraídas do espaço folksonômico a partir de métodos estatísticos. Ao relacionar duas categorias dessa maneira num esquema de classificação, estabelecemos uma ligação estática que é independente do contexto da atividade em que aquelas categorias foram usadas de fato.

Segundo Dourish [5], ver o contexto como um problema de representação é apenas uma das maneiras de enxergar a contextualidade. Nesse ponto de vista, o contexto ou as relações semânticas são um tipo de informação que pode ser representada através de estruturas de dados, como grafos, vetores etc. As relações descritas nessas estruturas são estáveis e, dessa forma, são aplicáveis em qualquer situação. Por essa razão, contexto e atividade, sob essa perspectiva, são completamente separáveis.

Seguindo a mesma linha argumentativa de Dourish, a seção seguinte analisa o conceito de comunidades e sua relação com uma prática regular. Essa concepção será fundamental para a crítica dessa perspectiva de semântica – como um problema de representação – e servirá como base para

introduzirmos uma nova concepção, na qual a interação tem um papel central no conceito de semântica e de contextualidade para a atividade de classificação em sistemas de informação complexos, como é o caso da *web* atual.

COMUNIDADES E CLASSIFICAÇÃO

Comunidades virtuais são estudadas em diversas disciplinas, e embora haja uma noção intuitiva do que elas representam, não há um consenso quanto à sua definição. Nesse artigo estamos particularmente interessados em comunidades virtuais cujos membros estão de alguma forma envolvidos em uma prática regular. Essa seção visa apresentar o conceito de comunidade de prática e sua relação com os sistemas de classificação. Essa noção fundamenta a análise das abordagens de classificação para *web* sob a perspectiva da interação entre membros de uma mesma comunidade e entre membros de diferentes comunidades.

Comunidades de prática e semântica

Segundo Wenger [19], “comunidades de prática são grupos de pessoas que compartilham um interesse ou uma paixão por algo que fazem e aprendem como fazê-lo melhor através da interação regular entre os membros do grupo”. Esses grupos não estão necessariamente associados a uma organização, podem não apresentar uma hierarquia definida entre seus componentes e em geral não são regidos por regras ou normas formais como nas organizações tradicionais. Nesse aspecto, os conceitos desenvolvidos por Wenger a partir da observação dos processos informais de aprendizagem podem ser aplicados ao estudo da interação entre comunidades que se dá mediada pelo aparato técnico das redes de informação e seus *softwares*.

Cabe aqui entender o conceito de prática, na concepção de Wenger. Para ele, “prática é primeiramente e acima de tudo um processo através do qual nós podemos experimentar o mundo e nosso envolvimento com ele de modo compreensível” [19]. Em outras palavras, o mundo se torna compreensível para nós à medida que adquirimos experiência através do nosso envolvimento em atividades que realizamos regularmente. Como um exemplo, suponha que uma pessoa que queira aprender a tocar um instrumento e participe de ensaios regulares com uma banda, mas não tenha nenhum conhecimento preliminar em música. No primeiro contato com o grupo essa pessoa talvez tenha dificuldades de distinguir os instrumentos, salvo pelo reconhecimento de características como cor, formato ou tamanho. Nesse estágio da interação com o grupo, a pessoa pouco sabe sobre a função desempenhada por cada instrumento na banda. Entretanto, à medida que a pessoa vai se familiarizando com o seu próprio instrumento e praticando com os demais, outros aspectos passam a fazer sentido e ela se torna capaz de diferenciar, por exemplo, um som agudo de um grave ou um timbre metálico de um aveludado. Mais do que isso, a partir da interação gradual com os demais membros da banda, a pessoa reconhece a

função do seu instrumento dentro da música. Não só os seus conhecimentos técnicos evoluem, mas também o seu vocabulário (i.e. o membro passa a usar os jargões usados pelo grupo) e suas preferências musicais e artísticas (i.e. músicos trocam álbuns entre si, recomendam artistas). Ao se tornar um músico experiente, a pessoa não domina apenas a técnica, mas, acima de tudo, passa a olhar o mundo sob a perspectiva de um músico.

Nesse sentido, a participação em uma comunidade de prática, seja em sua periferia, ainda como aprendiz, ou em seu núcleo, como praticante envolvido e experimentado, revela novos aspectos do ambiente e possibilidades de ação que passam a ser relevantes para as atividades que o membro realiza no seu dia-a-dia. Isso inclui os sistemas de classificação e suas categorias. Segundo Bowker e Star [4], à medida que uma pessoa ganha envolvimento numa comunidade de prática, categorias até então consideradas estranhas ou diferentes passam a ser familiares ao membro, num processo chamado de naturalização. Por exemplo, o músico que antes estranhava os termos aveludado ou metálico para classificação de um timbre sonoro, passa a entender o significado dessas categorias através do seu envolvimento e participação em comunidades de prática formadas por músicos.

A prática, portanto, é o processo pelo qual as coisas e as ações ganham significado. Membros de uma mesma comunidade de prática interagem em um mesmo processo e, por isso, desenvolvem esquemas de classificação próprios que refletem suas crenças, interpretações e visões de mundo. Dessa forma, uma categoria não possui um significado em si própria. Antes, a associação entre informação e uma categoria só possui sentido – ou evoca significado – no contexto de uma prática.

O local versus o global

Na introdução apontamos a tensão entre o local – a informação que tem significado e é relevante para uma comunidade específica – e o global – a informação que pode ser utilizada na interação entre diferentes comunidades. Na *web*, o gerenciamento dessa tensão para veiculação da informação apresenta diversos desafios em várias áreas da computação. Nesse artigo, estamos interessados em contextualizar o problema com o conceito de comunidade de prática. Até então discutimos a interação entre os membros de uma mesma comunidade, processo que é fundamental para o desenvolvimento de um conhecimento coletivo e para elaboração de esquemas de classificação locais à comunidade. Nessa seção analisaremos, sob essa perspectiva, como se dá a interação entre diferentes comunidades de prática e qual o papel dos sistemas de classificação nesse processo.

Em uma comunidade de prática, membros não são legitimados como tal de maneira abrupta, mas através de um processo contínuo de participação. O envolvimento de uma pessoa em uma comunidade de prática parte de uma participação periférica e, nesse estágio, o membro ainda não

está familiarizado com os objetos e com os esquemas de classificação daquela comunidade (i.e. o membro ainda não é totalmente um membro, ele não fala como um membro, não age como um membro e não interpreta o mundo como um membro). Do ponto de vista da comunidade, a pessoa ainda é um “estranho no ninho”, alguém “de fora”. Do ponto de vista do participante, tudo o que vem daquela comunidade parece estranho, ambíguo ou incompreensível. Entretanto, com a participação ativa na comunidade e com o crescimento dos laços de relacionamento com seus membros, os objetos e as categorias peculiares àquela comunidade passam a ser familiares. Ninguém determina que uma pessoa seja membro de uma comunidade. Isso ocorre através de um processo gradual e informal que tem como base o conceito de prática.

Até agora tratamos do conceito de local sob a perspectiva de uma comunidade de prática singular. Mas o que então corresponderia ao global? O global é o que permite a interação entre comunidades, é o que dá permeabilidade à informação entre diferentes contextos. No processo dinâmico de formação de novas comunidades, o global confere um grau de abertura à imigração de novos membros [4]. O global alivia a tensão do estranho *versus* o familiar, do naturalizado *versus* o incompreensível. Em outras palavras, o global faz a ponte entre o que é local a diferentes comunidades de prática, aproximando essas comunidades. Eventualmente, membros exclusivos de uma comunidade, através desse processo de mediação, passam gradualmente a fazer parte de outras comunidades. Nós, como membros de múltiplas comunidades de prática, somos agentes na construção de um contexto global entre elas. Fazemos isso, por exemplo, ao convidar os amigos do laboratório de computação a participar de um coral e ao ajudar os colegas do coral a resolver problemas com o computador de casa. A importância desse trabalho de articulação já havia sido reconhecida por sociólogos como Granovetter em [7], trabalho no qual ele destaca a importância dos elos fracos na interação entre grupos sociais distintos.

Nesse sentido, o conceito de objetos de fronteira (ou *boundary objects*) é central nessa discussão e vem enriquecer o papel do contexto global na interação entre comunidades de prática. Durante estudo de caso realizado em um museu de história natural com cientistas membros de diferentes comunidades de prática, Star e Griesemer [17] perceberam que os mesmos espécimes de pássaros mortos tinham significados bem diferentes entre biólogos especialistas e amadores apreciadores de pássaros. A cooperação entre diversas partes, no entanto, faz-se através da negociação de um significado comum que permita a comunicação e o trabalho em conjunto. Por outro lado, essa negociação não implica na imposição da perspectiva de uma comunidade sobre a outra, nem no uso de uma fonte externa de padronização.

O papel dos sistemas de classificação nessa negociação é o de estabelecer categorias que sejam de comum

entendimento entre diferentes comunidades de prática. Na interação entre especialistas e amadores no caso do museu, a categoria “pássaro” é suficientemente adequada para promover a comunicação, mas provavelmente não o seria na interação entre biólogos com diferentes especialidades. Nesse sentido, categorias e suas fronteiras precisam ser elásticas o suficiente para se adaptar às necessidades locais e robustas o suficiente para manter uma identidade global entre comunidades de prática diferentes, isto é, as categorias se tornam *objetos de fronteira*. Esse aspecto dá origem a esquemas de classificação que devem ser fortemente estruturados no escopo local e fracamente estruturados no escopo global.

Semântica como um problema de interação

Na *web*, comunidades e informação habitam múltiplos contextos. Nesse sentido, prover a semântica adequada para a informação através de diferentes contextos tem se tornado uma preocupação crescente na área de IHC. Conforme vimos anteriormente, a semântica por vezes é colocada como um problema de representação, concepção que é fruto da idéia de que o contexto pode ser “capturado”, “representado” e “modelado”. Essa concepção motivou a crítica de pesquisadores como Dourish em [5], no qual o conceito de contexto é dividido entre uma concepção mais positivista – contexto como representação – e outra mais fenomenológica – contexto como interação. Se o contexto tem um caráter representacional, atividade e contexto são separáveis. Nessa visão, a semântica da informação é definida *a priori* e sua representação permanece relativamente estável com o tempo. Em oposição a essa perspectiva, Dourish propõe uma visão alternativa da contextualidade que considera a interação como insumo primário para determinação da semântica da informação.

Nessa nova perspectiva, a contextualidade é uma propriedade relacional que se firma entre a informação e a ação. Uma informação não possui um significado pré-determinado ou está em uma categoria pré-definida. Ao contrário, a informação pode ter um significado ou pode estar associada a uma categoria, dependendo da atividade que está em consideração. Por essa razão, contexto e significado são características determinadas dinamicamente, e não *a priori*. O contexto é uma consequência, não uma premissa.

Dourish encontra a relação entre ação e significado no conceito de prática explorado anteriormente nessa seção. Essa unificação, segundo ele, é central para questão do contexto, uma vez que “o contexto está essencialmente relacionado às maneiras segundo as quais as ações podem ser compreendidas” [5]. Nessa concepção, questões como – o que constitui o contexto, como a relevância da informação pode ser gerenciada, como a semântica da informação pode ser levada através de diferentes contextos etc. – são questões relacionadas ao conceito de prática. A prática, nessa visão, é o processo através do qual ação e significado podem ser relacionados. Nesse aspecto, a semântica da

informação não é algo a ser representado, mas sim negociado a partir da interação entre membros de uma mesma comunidade de prática – no escopo local da interação – ou entre membros de diferentes comunidades de prática – no escopo global da interação.

Outro fator importante nessa concepção, destaca Dourish, é considerar a prática como um processo dinâmico, que evolui e se adapta. Dessa forma, apenas representar formas particulares de prática seria voltar à concepção representacional de contexto. Nesse aspecto, uma abordagem de classificação para *web* deve suportar a evolução dos esquemas de classificação e a adaptação desses esquemas no nível de interação, permitindo a emergência de novas relações entre ação e significado.

A seção seguinte se baseia no referencial teórico apresentado nessa seção e na perspectiva de contextualidade proposta por Dourish para analisar, sob outra lente, as abordagens de classificação na *web*.

CLASSIFICAÇÃO NA WEB: UMA NOVA PERSPECTIVA

Na análise inicial das abordagens de classificação na *web* foram considerados os seus aspectos técnicos e detalhes operacionais. Aqui pretendemos aprofundar aquela análise e rever cada abordagem sob a ótica do conceito de prática, da tensão entre o local e o global nos sistemas de classificação e da perspectiva de contexto como problema de interação.

Web semântica e a padronização

Ontologias

Ontologias são produzidas através de esforços de padronização, estabelecendo uma especificação da conceitualização sobre um domínio de conhecimento particular. Nesse sentido, ontologias procuram solucionar o problema da tensão entre o local e o global, impondo um padrão único de representação. Entretanto, estudos anteriores sobre o uso da padronização em sistemas de informação de larga-escala revelam que os padrões não se sustentam como tal por muito tempo e que o padrão para alguns, pode se tornar a confusão de outros [15].

Em uma infra-estrutura distribuída e heterogênea como a *web*, o uso da padronização dificulta ou mesmo impossibilita a interação entre diferentes comunidades de prática. Um requisito primário para que haja a interação nesse nível é o compartilhamento de estruturas de categorização de escopo global. Categorias, nesse sentido, são objetos de fronteira com um importante trabalho de articulação, permitindo a compreensão mútua entre as partes de uma interação e veiculando a semântica da informação através das fronteiras de cada comunidade. Entretanto, tais estruturas não emergem de um esforço unilateral de uma comunidade particular, mas da negociação entre diferentes comunidades.

O problema das ontologias está na estratégia de padronização que está por trás da abordagem. Um esquema

de classificação padronizado reflete a crença em um estilo particular de classificação ou, ainda, uma visão particular de um domínio de discurso de acordo com a prática de seus criadores. Em outras palavras, ontologias têm o objetivo de funcionarem globalmente, mas não escondem a parcialidade, que é o reflexo da perspectiva de uma minoria de especialistas que as produzem. A imposição de um padrão por uma minoria de especialistas em um domínio automaticamente exclui a perspectiva de uma maioria que também interage em um mesmo sistema de informação.

Por essa razão, a *web* semântica e o uso de ontologias para classificação da informação não oferecem apenas desafios técnicos e econômicos como sugere Hepp [9], mas também importantes desafios de caráter social a serem transpostos.

Abordagens híbridas

No contexto da *web* semântica, as abordagens híbridas procuram unir as vantagens dos sistemas colaborativos com a capacidade de inferência das ontologias. Os trabalhos nessa direção têm o objetivo de extrair relações semânticas entre os termos de uma folksonomia, construindo estruturas de dados que representam essas relações. Diferentemente da estratégia de ontologias, a representação aqui não é elaborada e gerenciada por um conjunto de especialistas, mas extraída automaticamente através de métodos estatísticos baseados na co-ocorrência de termos em diferentes anotações.

A estratégia não deixa de ter uma relação com a padronização. Em vez de padrões serem elaborados como no caso das ontologias, nesse caso padrões são extraídos, de modo que as estruturas de dados são construídas automaticamente a partir do reconhecimento de padrões (i.e. co-ocorrência de termos). Nessa estratégia, a relação semântica entre termos é avaliada quantitativamente, a partir de distribuições das frequências dos termos numa folksonomia e sem levar em consideração o contexto de classificação. Dessa forma, nada é revelado com relação aos diversos esquemas locais de classificação, uma característica em comum com a estratégia de ontologias. Em resumo, a abordagem de ontologias ao menos considera uma perspectiva real de classificação – aquela formada pelos especialistas que constroem a ontologia. No caso das abordagens híbridas, não existe uma perspectiva definida e a confiabilidade dos esquemas produzidos estará sempre sujeita aos erros inerentes ao método estatístico empregado.

Anotação colaborativa e a multiplicidade

O resultado da atividade de anotação é a produção de folksonomias, que são esquemas de classificação produzidos colaborativamente e, por isso, representam um lugar comum para o acúmulo de múltiplas perspectivas de classificação. Nesse sentido, folksonomias apresentam uma importante e decisiva vantagem sobre a estratégia de padronização empregada pela *web* semântica.

Entretanto, folksonomias também apresentam problemas. Os mais salientes estão relacionados à ausência de relacionamento entre os termos e a falta de estrutura, o que leva a uma baixa capacidade de inferência e aos problemas relacionados à ambigüidade. Conforme foi analisado anteriormente, esse conjunto de características traz problemas de recuperação de informação, diminuindo a eficiência da abordagem.

Porém, ao mudarmos de perspectiva, percebemos problemas de outra natureza. Em primeiro lugar, uma folksonomia não representa um único esquema de classificação, mas diversos deles. Dessa forma, o problema da ambigüidade não está apenas no nível léxico, mas num nível mais abstrato, relacionado aos esquemas de classificação. Diversas perspectivas de classificação encontram-se miscigenadas no espaço folksonômico, não sendo possível escolher qual delas é a mais adequada para uma determinada ocasião. Nesse ponto, voltamos ao problema do contexto.

Quando realizamos uma busca em uma folksonomia utilizando, por exemplo, o termo A como palavra-chave, poderíamos traduzir a consulta para a seguinte pergunta: “Quais conteúdos foram classificados com o termo A?”. Entretanto, sabemos que, pelo problema da ambigüidade, o sistema poderá trazer resultados irrelevantes. Agora suponhamos uma situação hipotética, em que a semântica do termo A pudesse ser especificada através do contexto no qual o termo foi usado para classificação. Isso eliminaria os resultados ambíguos, aumentando a eficiência da busca. Então uma forma de melhorar os resultados da pesquisa seria refinar a pergunta: “Quais conteúdos foram classificados com o termo A no contexto X?”. Dessa forma, estaríamos especificando a semântica exata do termo, através da indicação explícita do contexto em que o termo se aplica.

Porém, essa abordagem hipotética nos revela dois problemas. Em primeiro lugar, a composição das folksonomias atuais só é capaz de armazenar o relacionamento entre três entidades: conteúdo, termo e usuário. Consultas devem levar em conta apenas os relacionamentos dessas entidades e, dessa forma, a especificação de um contexto exigiria a inclusão de uma quarta entidade no espaço folksonômico, que hoje não existe. Em segundo lugar, ao indicar o contexto na pergunta, estamos tratando o contexto como uma forma de informação. Dessa maneira, independentemente do aparato tecnológico e dos sistemas de *software* envolvidos, o contexto requereria uma forma de representação, o que permitiria comunicar ao sistema de busca qual o contexto que nos é relevante para aquela dada consulta. Isso significa tratar a semântica do termo A como um problema de representação, o que tem sido alvo de crítica em toda argumentação apresentada nesse artigo. Ainda que as folksonomias comportassem o contexto como uma quarta entidade, o contexto não é algo que possa ser representado, conforme argumentamos ao longo desse artigo.

COLABORAÇÃO E MULTIDIMENSIONALIDADE

Entre as três abordagens de classificação analisadas na seção anterior, podemos confrontar a estratégia de padronização empregada pelas ontologias e abordagens híbridas, e a estratégia de colaboração dos sistemas de anotação. Em outras palavras, traçamos um paralelo, no que tange às abordagens para classificação da informação, entre duas tendências encontradas na *web* atual: a *web* semântica e a *web* social. Nessa análise, mostramos que a estratégia de colaboração empregada nas ferramentas da *web* social permite o acúmulo de múltiplas perspectivas de classificação na forma de folksonomias e que essa característica representa um aspecto positivo com relação à abordagem de padronização defendida na *web* semântica.

Entretanto, a estratégia de colaboração que dá origem às folksonomias não é apenas caracterizada por um conjunto de funcionalidades que incentivam a participação dos usuários como classificadores de informação. Mais do que isso, a estratégia de colaboração tem suas origens em uma mudança de paradigma no desenvolvimento de aplicações para *web*. Ao interpretar a estratégia de colaboração como um novo paradigma, podemos expandir a nossa compreensão do problema e, assim, desvinculá-lo de uma particular ferramenta ou aplicação.

Nesse sentido, o objetivo dessa seção é concluir a linha argumentativa defendida nesse artigo, elevando-a ao nível dos princípios utilizados no projeto de aplicações. A vantagem disso é poder enxergar o problema da classificação da informação na *web* segundo princípios básicos que podem ser aplicados no desenvolvimento de qualquer serviço ou aplicação. Dessa maneira, podemos tratar as folksonomias não apenas como o resultado da atividade de classificação em sistemas de anotação, mas como fruto de um novo paradigma, que privilegia o papel do usuário como agente classificador da informação.

Dando continuidade a análise feita na seção anterior, as folksonomias, por outro lado, sofrem de problemas relacionados à ambigüidade, uma vez que o contexto da classificação é perdido no ato da colaboração – ou anotação. Foi argumentado que tais problemas não podem ser resolvidos através da representação do contexto, mas pela contínua negociação da semântica através da interação no nível das comunidades de prática.

Nessa nova abordagem, uma folksonomia passaria de uma estrutura unidimensional, na qual múltiplas perspectivas estão miscigenadas em um único espaço, para uma estrutura multidimensional, na qual cada conteúdo é classificado segundo múltiplas perspectivas – ou sistemáticas – de classificação. Em vez de se misturarem em uma única dimensão, essas múltiplas perspectivas *co-existem* no mesmo espaço folksonômico, de modo que a classificação de um conteúdo também se torna multidimensional: há uma dimensão comum – ou global – que é compartilhada por diversas comunidades de prática; e diversas outras dimensões particulares – ou locais – a cada comunidade de

prática. A diferença para abordagem convencional de folksonomias é que, com o emprego da multidimensionalidade, sistemas de classificação podem separar uma perspectiva de classificação da outra, permitindo o gerenciamento da multiplicidade e o uso da multidimensionalidade estrutural para solução do problema da perda do contexto no ato da classificação.

Dessa forma, o problema do contexto é tomado sob a perspectiva da interação e, assim, podemos definir um novo conjunto de funcionalidades para o desenvolvimento de abordagens de classificação de conteúdo na *web*, de modo que novas aplicações passem a suportar, além da atividade colaborativa, a multidimensionalidade na atividade de classificação. Em primeiro lugar, essas aplicações devem oferecer um espaço compartilhado para interação entre membros de uma mesma comunidade de prática, disponibilizando recursos como fóruns de discussão, visualização de perfis dos membros, troca de mensagens etc. Como exemplo, muitos serviços para formação de redes sociais (e.g. Orkut⁶) oferecem suporte a formação e ao gerenciamento de comunidades. Nesses serviços, usuários podem filiar-se a diversas comunidades, sendo algumas delas moderadas e outras abertas ao público.

Nesse contexto, cada comunidade teria um espaço informacional próprio, que poderia ser privado (para acesso exclusivo dos membros de uma mesma comunidade) ou público (disponível também para usuários externos). Nesse espaço próprio, cada membro poderia compartilhar um conteúdo e classificá-lo com as categorias (ou termos) de sua escolha, no escopo local à comunidade. Outros membros da mesma comunidade poderiam complementar ou até mesmo corrigir determinadas classificações (i.e. num funcionamento similar ao dos *wikis*).

Além dos espaços locais a cada comunidade, haveria um espaço compartilhado que permitiria a interação entre comunidades no escopo global. Nesse espaço, cada conteúdo seria representado por categorias de articulação, que teriam a função dos objetos de fronteira explorados anteriormente nesse trabalho. Essas categorias poderiam ser estabelecidas automaticamente pelo sistema, a partir da escolha das categorias mais comuns utilizadas para classificação de um dado conteúdo nos diversos escopos locais (i.e. o sistema poderia escolher, por exemplo, as categorias com maior frequência de associação para um mesmo conteúdo). Alternativamente, essas categorias de articulação poderiam ser estabelecidas colaborativamente, através de uma classificação compartilhada por membros de diferentes comunidades de prática.

O sistema poderia oferecer uma interface de navegação na qual fosse possível visualizar múltiplas perspectivas de classificação em simultâneo. Essa funcionalidade orientaria

o usuário novato a encontrar comunidades de prática afins e o auxiliaria na atividade de recuperação de conteúdo.

Do ponto de vista da atividade de recuperação de conteúdo, os algoritmos fariam uso da estrutura multidimensional, dando maior relevância às categorias e aos conteúdos compartilhados no escopo local do usuário, diminuindo, com isso, a ambigüidade no nível dos esquemas de classificação.

CONCLUSÃO

Nesse artigo apresentamos uma análise das abordagens de classificação atuais para *web*, primeiramente sob uma perspectiva mais conservadora, e depois utilizando um referencial teórico fundamentado em estudos na área das ciências sociais e na área da psicologia cognitiva. Através dessa nova perspectiva multidisciplinar, pudemos revelar problemas nas abordagens de classificação analisadas que, até então, eram pouco explorados na literatura em computação.

A linha argumentativa utilizada defende que o espaço informacional da *web* não deve ser categorizado segundo estratégias de padronização. Sistemas de informação de larga-escala como a *web* precisam estar preparados para gerenciar múltiplas perspectivas de classificação, dando suporte a um grupo cada vez mais heterogêneo de usuários. O problema que surge é: quando escolher uma perspectiva em detrimento das outras? Essa pergunta nos levou para uma discussão aprofundada sobre a contextualidade nos sistemas de classificação e como tratar o problema nos sistemas de informação.

Nesse sentido, o conceito de comunidades de prática promove o elo entre o aprendizado e a prática. Novos esquemas de classificação surgem para promover o significado da informação entre membros de uma mesma comunidade – no escopo local da informação. Ao mesmo tempo, a interação entre diferentes comunidades se dá através da negociação de estruturas de articulação que promovem o entendimento mútuo entre as partes – no escopo global da informação. A compreensão do caráter distribuído da cognição humana [10] nos faz repensar as decisões de projeto e os modelos para o desenvolvimento de *software* atuais.

Com esse objetivo, o presente artigo relaciona o problema da classificação com os atuais paradigmas usados para o desenvolvimento de aplicações na *web*. Após análise baseada no conceito de comunidades de prática proposto por Wenger [19] e na perspectiva de contexto como um problema de interação proposta por Dourish [5], o artigo propõe dois princípios que deverão fundamentar uma abordagem de classificação na *web*: o primeiro – da colaboração – é fundamental na formação das folksonomias e no acúmulo de múltiplas perspectivas de classificação. Entretanto, a identificação de cada perspectiva em uma folksonomia é prejudicada pela mistura dessas diversas perspectivas em uma estrutura unidimensional. Nesse

⁶ <http://www.orkut.com>

sentido, o artigo propõe o princípio da multidimensionalidade, através do qual múltiplas perspectivas de classificação podem co-existir e ser gerenciadas em sistemas de classificação. Nesse artigo, propusemos a relação dessa multidimensionalidade com a interação social em comunidades de prática e, por fim, sugerimos a incorporação de novas funcionalidades nos serviços atuais que permitam a formação de comunidades de prática e a classificação multidimensional de conteúdo.

REFERÊNCIAS

1. AMES, M., AND NAAMAN, M. Why we tag: motivations for annotation in mobile and online media. *Proceedings of the SIGCHI conference on Human factors in computing systems* (2007), 971–980.
2. BEGELMAN, G., KELLER, P., AND SMADJA, F. Automated Tag Clustering: Improving search and exploration in the tag space. *Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland* (2006).
3. BERNERS-LEE, T., HENDLER, J., AND LASSILA, O. The semantic Web. *Scientific American* 284, 5 (2001), 28–37.
4. BOWKER, G., AND STAR, S. *Sorting Things Out: Classification and Its Consequences*. MIT Press, 1999.
5. DOURISH, P. What we talk about when we talk about context. *Personal and Ubiquitous Computing* 8, 1 (2004), 19–30.
6. GOLDER, S. A., AND HUBERMAN, B. A. Usage patterns of collaborative tagging systems. *J. Inf. Sci.* 32, 2 (2006), 198–208.
7. GRANOVETTER, M. The Strength of Weak Ties. *American Journal of Sociology* 78, 6 (1973), 1360.
8. GRUBER, T. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies* 43, 5/6 (1995), 907–928.
9. HEPP, M. Possible Ontologies-How Reality Constrains the Development of Relevant Ontologies. *IEEE INTERNET COMPUTING* (2007), 90–96.
10. HUTCHINS, E. *Cognition in the Wild*. Bradford Books, 1995.
11. LASSILA, O., SWICK, R., ET AL. Resource Description Framework (RDF) Model and Syntax Specification. *W3C Recommendation* 22 (1999), 2004–03.
12. LAVE, J. *Cognition in Practice: Mind, Mathematics and Culture in Everyday Life*. Cambridge University Press, 1988.
13. SCHMITZ, P. Inducing ontology from flickr tags. *Collaborative Web Tagging Workshop at WWW2006, Edinburgh, Scotland, May* (2006).
14. SMITH, M., WELTY, C., AND MCGUINNESS, D. OWL Web Ontology Language Guide. *W3C Recommendation* 10 (2004).
15. STAR, S. Power, Technologies and the Phenomenology of Standards: On Being Allergic to Onions. *J. Law, ed. A Sociology of Monsters* (1991), 27–57.
16. STAR, S. The trojan door: Organizations, work, and the “open black Box”. *Systemic Practice and Action Research* 5, 4 (1992), 395–410.
17. STAR, S., AND GRIESEMER, J. Institutional Ecology, Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science* 19, 3 (1989), 387.
18. WAL, T. V. Folksonomy :: vanderwal.net. Disponível online: <http://www.vanderwal.net/folksonomy.html>. Último acesso: 26/06/2008.
19. WENGER, E. *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press, 1999.
20. WU, X., ZHANG, L., AND YU, Y. Exploring social annotations for the semantic web. *Proceedings of the 15th international conference on World Wide Web* (2006), 417–426.