

Uso de uma base de conhecimento de Senso Comum em projetos de Arquitetura da Informação de web sites

Wanderley S. Wang

Instituto de Pesquisas Tecnológicas do Estado de São Paulo (IPT)

Av. Prof. Almeida Prado, 532 São Paulo - SP
wanderleywang@yahoo.com.br
+55.11. 3767-4068

Lucia Filgueiras

Instituto de Pesquisas Tecnológicas do Estado de São Paulo (IPT)

Av. Prof. Almeida Prado, 532 São Paulo - SP
lucia.filgueiras@poli.usp.br
+55.11. 3767-4068

RESUMO

Quanto mais preciso o conhecimento prévio do que é de senso comum do público-alvo, maior é a capacidade do arquiteto da informação para projetar *sites* de forma centrada no usuário e onde as informações desejadas são facilmente encontradas. Este trabalho propõe o uso de uma base de conhecimentos de senso comum para auxiliar na tarefa de construir um sistema de organização de *sites*.

ABSTRACT

Accurate recognition of what is common sense to a target audience helps the information architect in designing sites that are focused on the user and where desired information is easily found. This paper proposes the use of a Common Sense knowledge base as a resource for building web site organization systems.

Author Keywords

Arquitetura da informação, senso comum.

ACM Classification Keywords

H.5.2. User Interfaces. H.5.3. Web-based interaction.

INTRODUÇÃO

Morville [16] considera que a qualidade de um *site* pode ser avaliada pela análise de diversas facetas: acessibilidade, valor, credibilidade, desejo, usabilidade, encontrabilidade e utilidade. Encontrabilidade (“*findability*”, em tradução livre) refere-se ao grau de facilidade com que determinada informação é localizada e que pode ser priorizada pelo projetista para melhorar a experiência que os usuários terão ao visitar um *site* [17].

Morville e Rosenfeld [18] consideram que a incapacidade de encontrar uma informação é um dos fatores que mais

desagrada os usuários, e melhorar a encontrabilidade economiza tempo, aumenta a satisfação do usuário, diminui custos e mitiga riscos. A dificuldade para se encontrar algo pode advir, por exemplo, de uma organização confusa, da falta de padrões, da ambigüidade natural existente na linguagem e em alguns esquemas de classificação, ou de perspectivas diferentes entre os criadores e usuários [18].

Um bom projeto deve ser centrado no usuário, conforme apontam diversas áreas de estudo, como Interação Humano-Computador (IHC), *Human-Information Interaction* (HII) [9], UX (*User eXperience Design*) [16] e outras, e com atenção na forma como os usuários irão interagir com a informação. Interação com a informação é o processo pelo qual pessoas interagem com o conteúdo de um sistema de informação, contemplando não apenas consultas e pesquisas, mas também ações e experiências que elas têm durante esse processo [24].

PESQUISA EXPLORATÓRIA

Segundo Pirolli [21], a pesquisa exploratória é o tipo de tarefa mais freqüente (71%) dentre as que as pessoas consideram como importantes feitas por elas na *web* comparada a pesquisar informações específicas (25%). Uma pesquisa exploratória é a que envolve a coleta de múltiplos pedaços de informação para entendimento de um assunto ou tomada de decisões.

Na *web* as pessoas não apenas procuram informação, mas mergulham elas mesmas em um corpo de informação, e a *web* transformou a busca por informação em uma “experiência” e em um processo mais rico do que aquele previamente oferecido para tarefas de informação [24]. Uma tarefa de informação é “a manifestação de um problema de alguém que procura informação e que direciona as ações de pesquisa” [14]. Modelos que estudam o comportamento das pessoas enquanto procuram informação, como o *Sense Making* [5] e o ISP (*Information Search Process*) [10] mostram como as emoções e incertezas afetam a qualidade e a natureza da pesquisa e dos seus resultados.

Entender o comportamento dos usuários diante de uma tarefa de informação é também o foco da Teoria do Forragear Adaptada para Informação [21]. Forragear

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'08 October, 21-24 - Porto Alegre, RS, Brazil.

significa “vasculhar, remexer, à procura de algo” e, também, a atividade de “um ser vivo procurar alimento, lançando mão de estratégias especializadas, desenvolvidas no âmbito da espécie” [7]. Trata-se de uma abordagem que estuda as estratégias que os usuários usam para “caçar” a informação através da análise de “pistas”, ou do “cheiro da informação”, em um ambiente informacional.

O conceito de “Cheiro da Informação” (“*Scent of Information*”), derivado da Teoria do Forragear, é uma alusão ao cheiro que os animais sentem pelo faro quando estão caçando uma presa. O cheiro da informação é produzido por sinais (pistas) associados aos *links* (como o texto do rótulo do *link*, a grafia da URL ou textos e figuras próximas) que os usuários usam de forma subjetiva para prever a possível utilidade do conteúdo distante [6, 23].

Palavras-gatilho (palavras ou frases que o usuário associa a um conteúdo) existentes nas páginas do *site* influenciam, com maior ou menor força, os usuários em suas escolhas de navegação seguindo modelos de propagação da ativação de idéias (“*spreading activation models*”) [3]. Modelos computacionais, como o WUFIS e SNIF-ACT se propõem a prever o comportamento do usuário ao navegar em busca de informação levando-se em conta a topologia de *links* do *site* e a capacidade de propagação da ativação de idéias de cada *link*, calculada em função do grau de similaridade semântica entre os sinais perceptivos próximos e o ambiente lingüístico que se espera seja representativo dos usuários com objetivos específicos [6, 21].

Segundo [3, 21], atualmente as pessoas atuam sob “racionalidade limitada” (“*bounded rationality*”), com limites de tempo e de recursos que impedem uma pesquisa exaustiva por todos os caminhos, e, por isso, se diminui sua capacidade para encontrar informações relevantes os usuários abandonam a navegação no *site*. Dessa forma, aumentar a encontrabilidade é um fator crucial para o sucesso de um *site*.

ORGANIZAÇÃO DA INFORMAÇÃO

A Arquitetura da Informação é uma disciplina que orienta a condução de projetos de *sites* possibilitando estruturar esse tipo de espaço informacional de modo a facilitar o acesso intuitivo aos conteúdos, levando-se em conta o relacionamento e a natureza de interdependência existente entre usuários, conteúdo e contexto [18].

Uma das primeiras etapas de um projeto de arquitetura da informação é coletar dados sobre o público-alvo para definir a melhor forma de organizar o *site*. O *card sorting* é uma técnica bastante sugerida por vários autores para auxiliar na tarefa de classificar e organizar o conteúdo de forma centrada no usuário [18, 22], e que auxilia a identificar a terminologia mais adequada e as diferenças de perspectivas existentes entre os diferentes grupos de usuários. Embora, em geral, não haja uma organização única que seja ideal para todos os usuários, o *card sorting* possibilita atingir uma aproximação do que seria uma

organização consensual. Porém, segundo alguns autores [15], como essa técnica não considera as tarefas que os usuários precisam realizar, a organização gerada pode não ser adequada para usuários com objetivos de uma pesquisa exploratória. Portanto, pode ocorrer desses usuários terem dificuldades para encontrar com facilidade as páginas de conteúdo relevantes.

Em contrapartida, o conjunto de abordagens da chamada *Web 2.0*, incluindo a *folcsonomia* e as redes sociais, propicia a criação de espaços colaborativos onde as pessoas podem compartilhar conteúdos e gerar novas opções de navegação e de recuperação de informações.

Folcsonomia (*folksonomy*, em inglês), neologismo cunhado por Thomas Vander Wal¹, é uma maneira de indexar informações usando uma rotulação aberta baseada em “*tags*” (etiquetas). Designa também uma taxonomia gerada pelo usuário usada para categorizar e recuperar conteúdos da *web*. O termo provém da combinação das palavras “*folk*” (povo, popular) e taxonomia. Esse tipo de taxonomia também é chamado de rotulação livre (“*free tagging*”), ou categorização colaborativa ou indexação social. O usuário rotula objetos (fotos, *sites* ou páginas de *sites*, por exemplo) com uma ou mais palavras-chave. Morville [17] considera que, como as etiquetas são públicas, servem como pivô para uma navegação social, com potencial para se tornar também um novo sistema de classificação social. Dessa forma, não há uma pessoa única ou equipe centralizada que cria a taxonomia, sendo os próprios usuários que a fazem.

Na elaboração de uma taxonomia convencional, em geral o primeiro passo é definir as categorias para depois associar os itens de conteúdo, enquanto na *folcsonomia* cada usuário analisa o item de conteúdo para depois classificá-lo com uma ou mais palavras-chave, aproximando-se, dessa forma, de abordagens de projeto focadas mais no conteúdo das páginas do que na estrutura, recomendadas por [8, 23].

Aproveitar as classificações geradas pelos usuários nessa abordagem, com o objetivo de facilitar as tarefas de pesquisa exploratória, ainda é um desafio, pois em geral, o conhecimento gerado não é estruturado e há problemas de falta de consistência [19, 20]. Iniciativas recentes objetivam prover uma abordagem semântica às iniciativas da *Web 2.0* através da criação de ontologias colaborativas [19].

Entretanto, nas abordagens da *Web 2.0* a contribuição dos usuários ao arquiteto da informação ocorre durante a administração do *site*, ou seja, *a posteriori* da implantação, quando, e quanto mais os usuários participam. Portanto, seria útil uma forma de aproveitar a colaboração dos usuários ainda na fase de projeto do *site*.

¹ <http://pt.wikipedia.org/wiki/Folksonomia>

CONHECIMENTO DE SENSO COMUM

Alguns projetos, como o OMCS² (Open Mind Common Sense - MIT) e sua versão brasileira, o OMCS-Br³ (UFSCar) coletam conhecimentos que compõem o que se denomina senso comum das pessoas. A comunidade de Inteligência Artificial usa esse termo para se referir aos fatos básicos e entendimentos que a maioria das pessoas tem [13], como saber que um quarto de hotel deve ter uma cama, pessoas possuem unhas, gelo é frio, coca-cola é um refrigerante e um filho é mais jovem que seu pai.

Sites apropriados e jogos interativos têm sido utilizados como estratégia para motivar colaboradores a participarem da coleta de senso comum via internet, possibilitando a aquisição de grande volume de conhecimento sobre diversos temas. A análise sobre o que os usuários consideram como correto e normal, e o que não parece fazer sentido, é um dos resultados que podem ser obtidos dessa coleta, definindo-se senso comum como o conhecimento compartilhado pela maioria das pessoas de uma determinada cultura [4]. Dizer que uma afirmação é senso comum em uma cultura não significa que seja cientificamente verdadeira ou que seja senso comum em outras culturas [2], mas mesmo esse tipo de afirmação faz parte do escopo de coleta desses projetos.

A representação do conhecimento em linguagem natural permite coletar modos ambíguos e redundantes de expressar uma mesma idéia e indicar diferenças contextuais, como o aspecto de formalismo em um discurso [13].

Uma das abordagens para representação do conhecimento coletado, adotada no projeto OMCS-Br, é a utilização de redes semânticas que relacionam conceitos através de relações semânticas que abrangem os vários tipos de conhecimentos (por exemplo: generalização/especialização; propriedades; causa e efeito e outras). São utilizados os 20 tipos de relações definidos com base na teoria de Minsky sobre como funciona a mente humana [4]. A rede semântica do OMCS-Br (ConceptNetBr) possui um conjunto de API's que permitem extrair diversas informações, como os conceitos mais relevantes relacionados a um determinado domínio de aplicação ou tema.

Já existem aplicações que se utilizam dessas bases de conhecimento, como o projeto *KitchenSense* [12], uma aplicação integrada à uma cozinha e a seus utensílios, conectados numa rede através de uma série de sensores embutidos. A aplicação detecta o comportamento dos usuários e prevê suas intenções para oferecer ajuda e sugestões nos procedimentos, detectar situações de distração, evitar enganos comuns do usuário e simplificar as interfaces de controle de alguns dispositivos de acordo com sua relevância para a tarefa que o usuário pretende fazer, tornando, dessa forma, o ambiente mais “inteligente”. O sistema constantemente infere o relacionamento causal e

temporal de eventos usando um algoritmo de propagação de ativação (“*spreading activation algorithm*”) [12].

Outras aplicações usam esse tipo de base de conhecimento para, por exemplo, facilitar o entendimento de conceitos e de vocabulário entre leigos e especialistas em determinados assuntos [11], atuar como assistentes de treinamento ou tradutores de textos que levam em consideração diferenças culturais [1]. E há as que adotam uma abordagem de “*fail-soft*”, onde o sistema apenas “sugere” idéias ou fornece dados de análise. [4] apresenta o uso da ConceptNetBr para criar ações de aprendizagem mais efetivas através do fornecimento de subsídios que permitem aos especialistas identificarem fatos equivocados, ou falta de determinados conhecimentos, no senso comum de um grupo de pessoas.

ORGANIZAÇÃO COM BASE NO SENSO COMUM

Os estudos analisados neste artigo sugerem que, para o usuário, quanto mais familiar a organização do *site*, melhor é a qualidade da encontrabilidade. Este trabalho se propõe a analisar se a organização de um *site* sugerida através da técnica de *card sorting* encontra respaldo no senso comum.

Neste trabalho, através da técnica de *card sorting* os usuários agruparam páginas de conteúdo de um site piloto seguindo uma classificação que consideram a mais apropriada. A questão a validar é se a técnica induz os usuários a escolherem uma classificação formal diferente da forma como eles relacionam os conceitos (associados às páginas de conteúdo) no senso comum, e com isso não potencializam o uso das palavras-chave e diminuem o “cheiro da informação” preconizada pela Teoria do Forragear, prejudicando a encontrabilidade.

O método proposto neste trabalho para analisar essa hipótese consiste em: (i) verificar, para cada página de conteúdo de um grupo de classificação, os termos de vocabulário mais utilizados e respectivas frequências de uso; (ii) Identificar, no senso comum através da ConceptNetBr, o contexto de palavras-gatilho que as pessoas associam aos termos encontrados na página, ponderado conforme o tipo de relação semântica e (iii) verificar se essas palavras-gatilho conduzem às outras páginas de conteúdo existentes no mesmo grupo de classificação ou se, ao contrário, conduzem à páginas classificadas em outros grupos.

Páginas identificadas como relacionadas a outras páginas de grupo diferente sugerem a necessidade de uma navegação transversal (entre seções do *site*) ou mesmo a necessidade de outra forma de organização. Resultados preliminares mostram que diversas páginas estão semanticamente mais relacionadas a páginas de outros grupos de classificação do que com as páginas do mesmo grupo, sugerido pelo experimento de *card sorting*. E que a similaridade semântica (sugerido pelo senso comum) difere da simples similaridade léxica (baseada no uso de termos iguais entre páginas, que foram utilizados pelos autores dos conteúdos).

² <http://openmind.media.mit.edu/>

³ <http://www.sensocomum.ufscar.br>

DISCUSSÃO

Análises preliminares sugerem que uma base do tipo da OMCS-Br, quando alimentada com conhecimentos sobre um determinado tema, pode fornecer subsídios para projetos de *sites* voltados para assuntos correlatos e ajudar o arquiteto da informação a projetar um sistema de organização com maior qualidade de encontrabilidade ainda na fase de desenvolvimento. Como nem todo tipo de conhecimento é considerado de senso comum, é necessário identificar qual o escopo de temas de *sites* que poderiam se beneficiar desse tipo de conhecimento.

AUTORES

Lucia Vilela Leite Filgueiras é professora assistente doutora do Departamento de Engenharia de Computação da Escola Politécnica da USP e professora do Mestrado Profissional em Engenharia de Computação do IPT.

Wanderley S. Wang é mestrando em Engenharia de Software e graduado em Licenciatura Matemática pelo IME-USP. Atua em desenvolvimento de software e é diretor da ONG Família Viva (www.portaldafamilia.org).

REFERENCES

1. Anacleto, J. C.; Lieberman, H.; Tsutsumi, M.; Neris, V.; Carvalho, A. F. P.; Espinosa, J.; Godoi, M.; Zem-Mascarenhas, S. H. Can Common Sense uncover cultural differences in computer applications? IFIP, V. 217, 2006, <http://www.media.mit.edu/~lieber/Publications/Cultural-Differences-IFIP06.pdf>
2. Anacleto, J. C.; Carvalho, A. F. P. de; Pereira, E. N.; Ferreira, A. M.; Carlos, A. J. F. Machines with good sense: How can computers become capable of sensible reasonig? (prelo). In: BRAMER, M. (Org.). Artificial Intelligence in Theory and Practice II - WCC 2008 (b). Berlin: Springer-Verlag, 2008, v. 1, p. 1-10.
3. Budiu, R.; Royer, C.; Pirolli, P. Modeling Information Scent: A Comparison of LSA, PMI, and GLSA Similarity Measures on Common Tests and Corpora. Palo Alto Research Center, CA, 2007. <http://www.cs.cmu.edu/~ralucav/papers/riao.pdf>
4. Carvalho, A. F. P. Utilização de Conhecimento de Senso Comum no Planejamento de Ações de Aprendizagem Apoiado por Computador. 2007. 257f. Dissertação (Mestrado em Ciência da Computação) - UFSCar, 2007.
5. Case, D.O. Looking for information: A Survey of Research on Information Seeking, Needs, and Behavior. San Diego, Academic Press, 2002.
6. Chi, E. H.; Pirolli, P. L.; Chen, K.; Pitkow, J. E. Using information scent to model user information needs and actions and the Web. Proceedings of ACM (CHI 2001). <http://www.parc.com/research/publications/files/821.pdf>
7. Dicionário Houaiss, 2001
8. Halland, A. G. U., Halland, M. S. Core+Paths: A Design Framework for Findability, Prioritization and Value. ASIS&T 2007, <http://www.iasummit.org/2007/>
9. Jones, W. et al. "It's about the information stupid!": Why We need a separate field of Human-Information Interaction. CHI 2006 Panel.
10. Kuhlthau, C.C. Seeking Meaning: A process approach to library and information services. Westport, CT: Libraries Unlimited, 2004.
11. Kumar, A. Sundararajan, S. Lieberman, H. Common Sense Investing: Bridging the Gap between Expert and Novice. CHI 2004. <http://doi.acm.org/10.1145/985921.986015>
12. Lee, C. J.; Bonanni, L.; Espinosa, J. H.; Lieberman, H.; Selker, T. Augmenting kitchen appliances with a shared context using knowledge about daily events. Proceedings of IUI '06, 2006.
13. Liu, H.; Singh P. ConceptNet: a practical common sense reasoning toolkit. BT Technology Journal, v. 22, n. 4, 2004.
14. Marchionini, G. Information seeking in electronic environments. Cambridge University Press, 1997.
15. Maurer, D.; Warfel, T. Card sorting: a definitive guide. Boxes and Arrows, 2004. http://www.boxesandarrows.com/archives/card_sorting_a_definitive_guide.php
16. Morville, P. User Experience Design. Semantic Studios, 2004. <http://semanticstudios.com/publications>
17. Morville, P. Ambient Findability. 1.ed. Cambridge: O'Reilly, 2005. 188p.
18. Morville, P.; Rosenfeld, L. Information Architecture for the World Wide Web: Designing Large-Scale Web Sites. 3.ed. Cambridge: O'Reilly, 2006. 504p.
19. Noy, N. F.; Chugh, A.; Alani, H. The CKC Challenge: Exploring tools for collaborative knowledge construction. IEEE Intelligent Systems, vol. 23, no. 1, pp. 64-68, Jan/Feb., 2008
20. O'Reilly, T. What is web 2.0 – Design Patterns and Business Models for the Next Generation of Software. O'Reilly, 2005. <http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html>
21. Pirolli, P. Information Foraging Theory: Adaptive Interaction with Information. New York, NY: Oxford University Press, 2007. 204p.
22. Reis, G. A. dos. Centrando a Arquitetura de Informação no usuário. 2007. 250f. Dissertação (Mestrado em Ciência da Informação) – Escola de Comunicações e Artes / USP, 2007.
23. Spool, J.; Perfetti, C.; Brittan, D. Designing for the Scent of Information. User Interface Engineering, 2004. 31p. <http://www.uie.com/reports/>
24. Toms, E. G. Information Interaction: Providing a Framework for Information Architecture. Journal of the American Society for Information Science and Technology, Vol. 53, August 2002, p. 855 – 862.