

A Inteligência Coletiva na Web: Uma Análise de Domínio para o Jornalismo Online

Straus Michalsky

Edith Zaida Sonco Mamani

Marco Aurélio Gerosa

Departamento de Ciência da Computação Instituto de Matemática e Estatística (IME) Universidade de São Paulo (USP)

strasus@ime.usp.br

esonco@ime.usp.br

gerosa@ime.usp.br

ABSTRACT

Collaboration is one of the main characteristics of the Web 2.0. From the people interaction emerges the so-called Collective Intelligence, which involves the search for knowledge from de interaction and collaboration in large scale upon the web. Online Journalism also explores the web 2.0 benefits and takes advantage of Collective Intelligence, but this exploration is not the same in each journalism site. In this paper we propose a Collective Intelligence classification supported by a Domain Analysis based in the FODA methodology and using the 3C collaboration model (communication, coordination, cooperation) 14 features were found, which the most common were: Search, Recent Activities and Sharing objects.

RESUMO

A colaboração é uma característica marcante da Web 2.0. Das interações entre as pessoas, emerge o que vem sendo chamado de inteligência coletiva. Este conceito vem ganhando notoriedade no meio acadêmico e no mercado e envolve a busca por conhecimento a partir do processamento da interação e da colaboração em larga escala que acontecem na Internet. O Jornalismo Online também explora os benefícios da Web 2.0 de forma a tirar proveito da Inteligência Coletiva. Não há um padrão e cada site de jornalismo explora essa característica de maneira própria. Neste trabalho apresentamos uma classificação da Inteligência Coletiva através da execução de uma análise de domínio utilizando a metodologia FODA baseada no modelo 3C de colaboração (comunicação, coordenação, cooperação). Foram encontradas 14 características sendo que a Busca, Atividades Recentes e Compartilhar são as mais recorrentes.

Categories and Subject Descriptors

H.5.3 [Group and Organization Interfaces]: Web, H.5.3 [Group and Organization Interfaces]: Web-based interaction

General Terms

Algorithms, Design, Human Factors.

1. INTRODUÇÃO

A Web hoje é colaborativa. Sites que possibilitavam apenas interações monousuário passaram a incorporar recursos para interação entre os usuários. Por exemplo, sites de comércio eletrônico, como Amazon (www.amazon.com), oferecem para cada produto suporte a avaliação, resenha, troca de mensagens, wiki, compartilhamento de fotos, filtragem colaborativa,

recomendação, etc. Surge uma “inteligência coletiva” a partir da análise das interações entre os usuários [18].

Esta nova forma de enxergar a web ficou conhecida como Web 2.0. Sites deixaram de ser apenas páginas estáticas de entrega de conteúdo e se tornaram aplicativos colaborativos favorecendo a participação do usuário. Alguns padrões são citados por O'Reilly [14] que caracterizam um site pertencente à Web 2.0, como por exemplo, formação de comunidade, adição de valor à aplicação pelos usuários e a aplicação acima da plataforma.

Estas alterações que ocorreram entre a web 1.0 e a web 2.0, tornando a web colaborativa, afetaram também os sites de notícias online[15].

O objetivo deste trabalho é realizar uma análise de domínio para a Inteligência Coletiva na web 2.0, considerando como ponto de partida a domínio o jornalismo online. Como objetivo secundário tem-se documentar as características recorrentes facilitando a criação de novas aplicações da Web.

2. COLABORAÇÃO NA WEB

De acordo com Ellis et al. [7] a colaboração é vista como uma combinação de comunicação, coordenação e cooperação. Esta maneira de analisar a colaboração é conhecida na literatura como modelo 3C [8], seu diagrama pode ser visto na Figura 1. A comunicação está relacionada com a troca de mensagens e informações entre pessoas; coordenação é relacionada ao gerenciamento de pessoas, suas atividades e recursos; e cooperação, é a produção que acontece em um espaço compartilhado. O modelo 3C aparece frequentemente na literatura como um meio de classificar sistemas colaborativos [5] [4].



Figura 1. O diagrama do modelo 3C de colaboração [9]

3. INTELIGÊNCIA COLETIVA

A inteligência coletiva vem recentemente ganhando notoriedade de sociólogos e teóricos, nos negócios e na ciência da computação [12][19][18][1]. Grandes empresas de tecnologia montaram recentemente grupos específicos para estudar a inteligência coletiva e a computação social, como por exemplo, o Microsoft Social Computing Group (<http://research.microsoft.com/scg>), o HP Social Computing Lab (<http://www.hpl.hp.com/research/scl>) e o IBM Social Computing Group (<http://www.research.ibm.com/SocialComputing>).

De 2000 a 2009 a quantidade de usuários da internet subiu 400% [10], atingindo 1,8 bilhões de pessoas. Esta quantidade de pessoas navegando e gerando informações, direta ou indiretamente, produz uma quantidade de dados inimaginável há algumas décadas. Chris Anderson, editor chefe da revista Wired, caracteriza esta situação como Era do Petabyte[2].

De acordo com Alag [1] quando um grupo de indivíduos colabora ou compete entre si, surge uma inteligência ou comportamento que de outra forma não existiria, isto é conhecido como inteligência coletiva. Existem outras definições de inteligência coletiva e neste trabalho resolvemos nos basear nesta definição dada e definir um novo termo **inteligência coletiva emergente**. A **inteligência coletiva emergente** é a inteligência coletiva que surge da interação de usuários.

Cavalcanti & Nepomuceno [6] dividem a inteligência coletiva em 3 classes, de acordo com a forma que ela é gerada. Estas classes são: inteligência coletiva inconsciente, inteligência coletiva consciente e inteligência coletiva plena. A **inteligência coletiva emergente** é relacionada à inteligência coletiva inconsciente uma vez que os usuários normalmente não têm consciência que sua interação com o sistema está sendo utilizada para melhorar o próprio sistema.

A inteligência coletiva emergente possibilita a derivação da inteligência coletiva a partir das interações do usuário com o sistema, não sendo baseada apenas na colaboração. Com ela é possível obter várias informações como: a relevância de objetos a partir do número de visualizações, interesses de um usuário a partir de suas compras e relevância do resultado de uma busca a partir da visita dos resultados. Alguns ruídos podem aparecer neste tipo de dados e a derivação de inteligência deve sempre levar isso em conta. Um exemplo de ruído é quando um usuário acessa uma página por engano e o número de acessos é usado para recomendação.

As três dimensões da colaboração também são fontes derivadoras de inteligência coletiva emergente. Por exemplo, a adição de um comentário é uma comunicação e com a quantidade de comentários é possível descobrir qual é o objeto mais popular ou mais discutido no sistema.

Uma aplicação da inteligência coletiva muito difundida é a *folksonomia* [3]. Ela é obtida a partir da adição de *tags* / palavras-chaves a objetos possibilitando a classificação pelos usuários do sistema. Sua principal vantagem é utilizar a linguagem dos próprios usuários de forma que as buscas sejam mais precisas de acordo com a linguagem utilizada pelo grupo. Através da *folksonomia* é possível a criação de uma memória coletiva [3]. Esta memória coletiva é a classificação de um objeto em uma determinada posição no tempo. Desta forma as *tags* utilizadas em um objeto em uma data específica correspondem à memória

daquele grupo nesta data. Uma exploração direta da inteligência coletiva emergente com as *tags* seria a *Tag Cloud*. Com ela é possível observar quais *tags* são mais utilizadas dentro do sistema.

4. ANÁLISE DE DOMÍNIO

A análise de domínio é uma das etapas da engenharia de domínio. De acordo com Prieto & Rango [16] a Engenharia de Domínio possibilita que características comuns e variáveis sejam identificadas e organizadas utilizando um processo previamente definido. Ela busca identificar, coletar e organizar informações relevantes do domínio, utilizando para tal o conhecimento existente e técnicas para modelagem de informação [11].

Neste trabalho será utilizada a técnica Feature-Oriented Domain Analysis – FODA. Assim as informações obtidas na análise possibilitam determinar os limites que norteiam o domínio, as características comuns e variáveis de aplicações do domínio e gerar como resultados modelos para a representação de tais características. A modelagem FODA é orientada a características (*features*). De acordo com Kang et al. [11] essas características são atributos do sistema que afetam diretamente os usuários finais.

Para representar o domínio, jornalismo online, foi avaliado um total de vinte sites de notícias online, sendo dez nacionais e dez internacionais. Os sites internacionais foram retirados da lista de Top Sites dentro da categoria News da companhia de informações web Alexa (<http://www.alexa.com/>) ordenado por popularidade. Já a lista de sites nacionais foi retirada de uma busca no Google Brasil (<http://www.google.com.br/>) utilizando como termo de busca a palavra “notícias”.

Dentro dos sites selecionados para representar o jornalismo online foram levantadas as características do domínio. Para executar este levantamento foi executada uma análise exploratória das características disponibilizadas pelos sites avaliados e uma filtragem de quais destas características possibilitam a exploração da Inteligência Coletiva. O levantamento e filtragem destas características foram executados por dois pesquisadores e foi supervisionado por um terceiro pesquisador. Estas características do domínio proposto foram descritas utilizando-se o padrão proposto por Schummer & Lukosch [17] com algumas simplificações para a descrição de padrões para interação mediada por computador.

Finalmente, cada característica foi avaliada no site em que aparece e classificada de acordo com sua função na colaboração: comunicação, coordenação, cooperação e interação pessoal. Foram também avaliadas as variabilidades das características. Sendo assim possível avaliar qual forma de colaboração utiliza a característica.

5. RESULTADOS

Os resultados obtidos na Análise do domínio bem como discussões sobre os mesmos são encontrados nesta sessão. A lista de sites obtidos utilizando os critérios definidos anteriormente no dia 29/03/2010 pode ser visualizada na Tabela 1.

A lista de características que possibilitam a exploração da Inteligência Coletiva no domínio proposto levantada a partir dos sites selecionados é descrita na Tabela 2.

5.1 Análise das Características

A Tabela 3 apresenta o resultado final onde é apresentada cada característica devidamente classificada. Todas as características que possibilitam a exploração da Inteligência Coletiva no site avaliado foram marcadas X.

Tabela 1. Lista de sites avaliados

Colocação	Sites Internacionais	Sites Nacionais
1º	news.yahoo.com	g1.globo.com
2º	cnn.com	noticias.terra.com.br
3º	news.bbc.co.uk	folha.uol.com.br
4º	nytimes.com	noticias.uol.com.br
5º	weather.com	jornalnacional.globo.com
6º	huffingtonpost.com	economia.estadao.com.br
7º	msnbc.msn.com	ultimosegundo.ig.com.br
8º	foxnews.com	oglobo.globo.com
9º	guardian.co.uk	rondoniagora.com
10º	online.wsj.com	tudoagora.com.br

Como os dados e os algoritmos que cada site utiliza não são abertos ao público, algumas características podem não utilizar a Inteligência Coletiva no contexto em que foram avaliados, porém elas possibilitam a sua utilização. Um exemplo deste tipo de característica é o comentário. Nos sites que possuem esta característica foi observada a possibilidade de adicionar comentários a notícias, mas não foi possível observar se este dado é utilizado em alguma outra situação no site.

Tabela 2. Lista de Características levantadas

Nome	Descrição
Comentário	Possibilita aos usuários comentarem as notícias.
Enquete	Possibilita aos usuários responderem perguntas escolhendo uma dentre algumas alternativas pré-definidas.
Lista de Discussão	Possibilita a um grupo de usuários a troca de mensagens via e-mail entre todos os membros do grupo.
Denúncia	Possibilita aos usuários notificarem quando encontraram algum conteúdo impróprio em objetos do sistema.
Seguir tópico	Possibilita aos usuários receberem notificações quando exista alguma atualização de um objeto ou categoria do sistema.
Compartilhamento	Possibilita aos usuários compartilharem a informação de um objeto do sistema com outros sistemas.
Avaliação	Possibilita aos usuários avaliarem objetos do sistema.
Produção	Possibilita aos usuários adicionarem novas notícias no sistema.
Indicação	Possibilita aos usuários indicarem o conteúdo de um objeto do sistema.
Upload	Possibilita aos usuários adicionarem arquivos tipo mídia (vídeo, imagens) em objetos do sistema.
Busca	Possibilita aos usuários buscarem informação no conteúdo dos objetos do sistema.
Tags	Possibilita aos usuários visualizar e/ou caracterizar com palavras-chaves objetos do sistema.
Versão de Impressão	Possibilita aos usuários imprimirem uma versão de um objeto do sistema em formato mais amigável.
Salvar	Possibilita aos usuários salvarem objetos para próximos acessos ao sistema.

Outra observação a ser feita é com relação à característica Comentário. Ela é uma forma de comunicação que pode ser utilizada para derivação de Inteligência Coletiva Emergente. Se a quantidade de comentários for utilizada para ordenar as notícias mais comentadas, por exemplo, ela é considerada Inteligência Coletiva Emergente.

Tabela 3. Avaliação das características.

Características Sites de Notícias	Comunicação			Coordenação		Cooperação						Interação Individual		
	Comentário	Enquete	Lista de Discussão	Denúncia	Seguir Tópico	Compartilhamento	Avaliação	Produção	Indicação	Upload	Busca	Tags	Versão Impressão	Salvar
G1	X					X	X				X		X	
Terra Notícias	X	X		X		X	X	X		X	X		X	
Folha Online	X	X		X		X	X			X			X	X
UOL Notícias	X		X	X		X	X			X	X		X	
Jornal Nacional						X	X			X			X	
Economia & Negócios	X	X				X				X			X	X
Último Segundo	X			X		X		X	X	X	X		X	X
O Globo	X	X		X		X	X	X		X	X		X	
Rondoniagora	X					X				X			X	
Tudo Agora	X					X				X	X		X	
Yahoo news	X			X		X	X		X	X	X		X	
CNN	X	X		X	X	X	X	X	X	X	X		X	X
BBC news	X		X	X		X	X		X	X			X	
The New York Times	X				X	X	X		X	X	X		X	
The Weather Channel						X				X	X			
The Huffington Post	X		X	X		X	X	X		X			X	
MSNBC	X		X			X	X			X			X	
Fox news	X			X		X			X	X			X	
Guardian	X			X		X			X	X			X	
The Wall Street Journal	X		X			X	X		X	X			X	X

Na característica Tag existe uma peculiaridade uma vez que apenas a CNN e o Último Segundo implementam esta característica de forma a realmente aproveitar a Inteligência Coletiva Emergente. Estes sites são os únicos a possibilitarem a adição de tags pelos usuários e apenas em matérias escritas pelos próprios usuários. Em todos os outros casos a Inteligência Coletiva Emergente não é explorada nesta característica, pois as palavras-chaves não são adicionadas pelos usuários.

5.2 Discussão

A utilização da Inteligência Coletiva não pode ser verificada plenamente nos sites analisados. Porém como Cavalcanti & Nepomuceno [6] explica é necessário comunidades (redes) bem estruturadas para que ela possa ser explorada. A aparição destas comunidades já é citada por O'Reilly [14] como uma característica da Web 2.0 e pode ser observado o início da sua exploração em

sites como o da CNN onde o usuário pode se cadastrar e criar um perfil. A partir do momento em que estas comunidades forem estruturadas dentro destes sites de notícias surgirão mais possibilidades de utilização da Inteligência Coletiva.

Outro ponto que está ligado a estas comunidades dentro do jornalismo é a terceira fase do jornalismo online descrita por Pires & Baldessar [15]. Uma vez que para ela ser alcançada é necessário que esta comunidade alimente o site com notícias.

Assim percebemos que a terceira fase do jornalismo online, a web 2.0 e a Inteligência Coletiva estão intimamente ligadas. Sua base é a comunidade que é criada por estes sites de notícias. Para que a Inteligência Coletiva possa ser melhor aproveitada é necessária uma melhor estruturação desta comunidade, de forma a incentivar os usuários a participarem.

6. CONSIDERAÇÕES FINAIS

Este artigo apresentou a análise de domínio da Inteligência Coletiva no jornalismo online através do FODA, realizando uma classificação das características dentro das formas de colaboração apresentadas e das formas de Inteligência Coletiva que podem ser exploradas. É importante mostrar que, apesar deste estudo ser focado no jornalismo online, ele reflete de uma forma geral o uso da Inteligência Coletiva na web.

Há várias possibilidades de exploração da Inteligência Coletiva nas características levantadas, mas nem todas são utilizadas. Algumas destas características precisam de uma melhor estruturação da comunidade para serem exploradas, porém algumas outras já podem ser aplicadas na estrutura atual dos sites. A fim de que os usuários tenham melhores experiências dentro destes sites, estas características devem ser exploradas de acordo com as preferências e necessidades dos usuários.

Um dos possíveis motivos da subutilização da Inteligência Coletiva é a complexidade dos algoritmos utilizados na sua derivação. Este trabalho é um primeiro passo para a redução deste problema uma vez que ele executou a Análise de Domínio e a sequência dentro da Engenharia de Domínio é a implementação destas características de forma reusável.

Um trabalho futuro é o mapeamento dos algoritmos utilizados na derivação da Inteligência Coletiva com as características levantadas neste trabalho. Como por exemplo, na Categorização pode ser utilizados algoritmos como: K-nearest neighbor, Clusterização hierárquica, K-Means ou mesmo Redes Neurais [1][18]. O mapeamento dos Algoritmos com as Características é de grande ajuda para quem pretende utilizar estas características em seus sites.

As características levantadas serão implementadas na forma de componentes para o Groupware Workbench <http://www.groupwareworkbench.org.br/>. Com estes componentes prontos à criação de aplicativos para web 2.0 é facilitada uma vez que a complexidade encontrada na derivação da Inteligência Coletiva fica encapsulada dentro dos componentes criados e quem desenvolve a aplicação pode manter o foco nas funcionalidades do sistema.

7. AGRADECIMENTO

Agradeço a FAPESP pelo apoio (Processo 2009/05013-8).

8. REFERÊNCIAS

- [1] Alag, S. (2008) *Collective Intelligence in Action*, Manning Publications, ISBN 1933988312.
- [2] Anderson C. (2008) *The End of Theory: The Data Deluge Makes the Scientific Method Obsolete* http://www.wired.com/science/discoveries/magazine/16-07/pb_theory
- [3] Aquino, M. C. (2007) *Hipertexto 2.0, folksonomia e memória coletiva: Um estudo das tags na organização da web*. E-Compós Vol. 9.
- [4] Borges R. M., Pinto S. C. C. S., Barbosa J. L. V., Barbosa D. N. F. (2007) *Usando o modelo 3C de colaboração e Vygotsky no ensino de programação distribuída em pares*. Workshop em Informática na Educação (sbie) 2007 XVIII Simpósio Brasileiro de Informática na Educação – SBIE.
- [5] Borghoff U. M., Schlichter J. H. (1996) *Computer-supported cooperative work: introduction to distributed applications*
- [6] Cavalcanti M., Nepomuceno C. (2006) *O Conhecimento em Rede*. Editora Campus, Rio de Janeiro, 2007.
- [7] Ellis, C.A., Gibbs, S.J. & Rein, G.L. (1991) *Groupware - Some Issues and Experiences*. *Communications of the ACM*, Vol. 34, No. 1, pp. 38-58.
- [8] Fuks, H., Raposo, A., Gerosa, M.A., Pimentel, M. & Lucena, C.J.P. (2007) "The 3C Collaboration Model" in: *The Encyclopedia of E-Collaboration*, Ned Kock (org).
- [9] Gerosa, M.A., Raposo, A.B., Fuks, H. & Lucena, C.J.P. (2006) "Component-Based Groupware Development Based on the 3C Collaboration Model", *Anais do XX Simpósio Brasileiro de Engenharia de Software – SBES 2006*, Florianópolis-SC, ISBN 85-7669-079-9.
- [10] Internet World Stats (2009) *Internet Usage Statistics - The Internet Big Picture World Internet - Users and Population Stats em: <http://www.internetworldstats.com/stats.htm>*
- [11] Kang, K. C. et al. (1990) *Feature-Oriented Domain Analysis (FODA) Feasibility Study*. : CMU/SEI.
- [12] Levy, P. (2003) *A Inteligência Coletiva: por uma Antropologia do Ciberespaço*, Loyola, ISBN 8515016133.
- [13] Oliveira C. A. F.; Reis J. F. T. (2004) *Filtragem Colaborativa uma forma de personalizar informações*. *CienteFico*. Ano IV, v. I, Salvador, janeiro-junho 2004
- [14] O'Reilly, T., *What is Web 2.0*, (2005) <http://www.oreilly.com/go/web2>
- [15] Pires, E. W. & Baldessar (2006), M. J. *A internet como fator de mudança no jornalismo*, *Revista Científica em Ciências da Comunicação da Faculdade Estácio de Sá Santa Catarina* http://www.sc.estacio.br/revistas/comunica_artigos.html
- [16] Prieto-Diaz, R.; Arango, G. (1991) *Domain Analysis Concepts and Research Directions*. In: Prieto-Diaz, R., Arango, G. (eds), *Domain Analysis and Software Systems Modeling*. : IEEE Computer Society Press.
- [17] Schummer, T.; Lukosch, S. (2007) *Patterns for Computer-Mediated Interaction*. West Sussex: John Wiley & Sons Ltd.
- [18] Seragan, T. (2007) *Programming Collective Intelligence: Building Smart Web 2.0 Applications*, O'Reilly, ISBN 0596529325.
- [19] Shirky, C. (2008) *Here Comes Everybody: The Power of Organizing Without Organizations*, Penguin Press HC.