



PROTOTIPAÇÃO DE UM SISTEMA DE RECOMENDAÇÃO DE TAGS PARA O TAGMANAGER

Lucas Pupulin Nanni (PIBIC/UEM), Valéria D. Feltrim (Co-Orientadora)
Sérgio Roberto P. da Silva (Orientador), e-mail: srsilva@uem.br.

Universidade Estadual de Maringá / Centro de Tecnologia - Departamento
de Informática / Maringá, PR.

Ciências Exatas e da Terra / Ciência da Computação

Palavras-chave: Social Tagging, Personomia, Folksonomia.

Resumo:

Como forma de gerenciar a enorme quantidade de informação disposta na *Web*, foi proposta a utilização de *tags*, que possibilitam aos usuários organizarem seus próprios conteúdos publicados, originando os sistemas baseados em *tagging*. No entanto, isto tem gerado problemas relativos à dificuldade de encontrar o que se deseja em meio ao emaranhado de *tags* que acaba se formando quando estes sistemas crescem. Este crescimento leva a formação de um vocabulário de *tags* problemáticas. Para minimizar isso, propõe-se a utilização de sistemas de recomendação de *tags*. Estes sistemas orientam o usuário na escolha de uma *tag* para rotular um determinado conteúdo, possibilitando a criação de um vocabulário mais coerente e coeso, o que auxilia na etapa de recuperação de informação. Desta forma, este trabalho visa a prototipação de um sistema de recomendação de *tags* para um sistema denominado *TagManager*, um gerenciador de vocabulário do usuário em sistemas baseados em *tagging*.

Introdução

A facilidade com que usuários comuns de computadores publicam informações e documentos diariamente na *Web* produz um crescimento exponencial de conteúdo digital, gerando uma grande dificuldade de encontrar o que se deseja em meio a toda esta informação. Em resposta a essa dificuldade surgiram técnicas alternativas para se organizar os conteúdos na *Web*, como o uso de *tagging* [1] (etiquetagem), que tem se mostrado muito promissor. A etiquetagem possibilita que os usuários comuns anotem com *tags* (rótulos) os conteúdos que lhes interessam, facilitando, assim, sua posterior recuperação nos sistemas que os organizam. O conjunto de rótulos usado pelo usuário cria nos sistemas baseados em *tagging* um vocabulário próprio, que é comumente denominado de **personomia** [2]. O compartilhamento de personomias – possibilitado por alguns sistemas de organização de conteúdos como o *Delicious* (<http://www.delicious.com>) e o *Flickr* (<http://www.flickr.com>) – é denominado de **folksonomia** [3]. As folksonomias são categorizações livres,

nas quais o usuário pode usar qualquer *tag* desejada para categorizar um conteúdo. Devido a essa liberdade, e a facilidade de se usar os sistemas baseados em *tagging*, surgem inconvenientes que dificultam a busca e a própria formação do vocabulário, visto que o usuário nem sempre utiliza anotações adequadas ao categorizar recursos na *Web*, utilizando para isso “*tags* mascaradas” [4] ou completamente inconsistentes.

Desta forma, inserido em um projeto maior, este trabalho tem por fim estudar técnicas de recomendação de *tags* para sistemas baseados em *tagging* e sugerir o protótipo de um sistema recomendador de *tags*, em língua inglesa, para o sistema *TagManager* [5], o qual está sendo desenvolvido pelo Grupo de Sistemas Interativos Inteligentes (GSII - <http://www.din.uem.br/gsii>) do DIN/UEM. Tal esforço é justificado devido ao grande uso atual de sistemas baseados em *tagging* e às dificuldades encontradas pelos usuários para categorizar um recurso na *Web*, cabendo ao protótipo auxiliá-los no gerenciamento adequado de seus vocabulários nestes sistemas.

Materiais e métodos

Para a prototipação do sistema foi utilizada a linguagem de programação *Python* (<http://www.python.org>) em sua versão 2.6.5, a ferramenta de processamento de linguagem natural *NLTK* (<http://www.nltk.org>) e a ferramenta de *parsing HTML/XML BeautifulSoup* (<http://www.crummy.com/software/BeautifulSoup>). A técnica de recomendação abordada por esse trabalho foi a **recomendação semântica**, utilizando-se de termos de três fontes: da personomia, da folksonomia e do texto da página *Web* a ser categorizada.

Em linhas gerais, para cada uma das três fontes, foi extraído um conjunto de termos representativos, com os quais se emergiu uma ontologia referente a cada recurso, seguindo a abordagem proposta por Basso [6]. Finalmente, os termos foram identificados e recomendados de acordo com as características do usuário. A Figura 1 ilustra os recursos utilizados e sua combinação para a recomendação das *tags*.

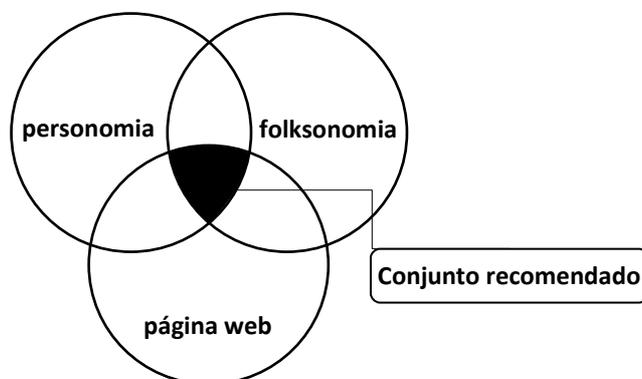


Figura 01. Combinação dos recursos para a recomendação

Para a extração de termos da personomia e da folksonomia foi utilizada a técnica de *page scraping* nas páginas geradas pelo sistema *Delicious*. Já para o recurso textual, foi procedida a extração de termos por meio da técnica de *HTML parsing* da página *Web* categorizada. O grupo de termos extraídos foi classificado – segundo métricas de termo-frequência, similaridade termo-título e posicionamento textual –, lematizado e reduzido a 20% de seu tamanho, utilizando o Princípio de Pareto [7].

A recomendação das *tags* foi realizada a partir da interseção das três ontologias emergidas, sendo que os termos pertencentes à personomia receberam maior apreço. A interseção foi construída relacionando-se os termos da personomia com os termos da folksonomia e da página *Web*. A escolha das *tags* seguiu a frequência que cada termo possuía, sendo que os recomendados eram os mais relacionados.

Resultados e Discussão

Para a extração de termos da página *Web*, foi verificado que ao utilizar métricas de posicionamento textual, junto à classificação por termo-frequência, termos mais significativos obtiveram melhor classificação. Além disso, quando aplicada a técnica de similaridade texto-título, termos com baixa classificação foram reconsiderados e reclassificados. A lematização restringiu a diversidade semântica dos termos ao serem reduzidas as flexões gramaticais dos mesmos, permitindo ao conjunto extraído se amortizar a 20% sem que termos representantes do contexto fossem perdidos, favorecendo desta forma a emersão da ontologia.

Um ponto avaliado na emersão das ontologias foi a desambiguação de sentido proposta por Basso [6]. Este processo, como foi abordado, é discutível, já que não releva o contexto dos sentidos e não aborda um critério de desempate entre os mesmos.

Por fim, na Tabela 1 foram listadas algumas categorizações de páginas *Web* experimentadas pelo sistema prototipado e parte de seus respectivos conjuntos de *tags* recomendadas. Para tanto, foi utilizada a personomia de *lucas.nanni* (<http://www.delicious.com/lucas.nanni>) no sistema *Delicious*.

Tabela 1 – Categorizações e suas *tags* recomendadas.

<i>URL</i> da categorização	<i>Tags</i> recomendadas
http://www.youtube.com/	web (object); sharing (distribution); entertainment (diversion); video (visual_communication); youtube ()
http://www.apple.com/	store (hardware); iphone (); technology (discipline); computer (machine); apple ()
http://www.wikipedia.org/	encyclopedia (reference_book); community (group); education (content); opensource (); wikipedia ()
http://vanderwal.net/	blog (diary); information (collection); tagging (attach); folksonomy (); vanderwal ()

Pode-se verificar que, para cada *tag* recomendada, junto a ela segue o seu composto semântico entre parêntesis.



O resultado obtido foi considerado satisfatório, pois atendeu as expectativas de uma recomendação semântica relacionada com a personomia do usuário, mesmo fornecendo alguns compostos semânticos descontextualizados com o recurso categorizado, devido a ineficácia da desambiguação de sentidos utilizada, ou ausentes, devido a sua inexistência na *WordNet* (<http://wordnet.princeton.edu/>).

Conclusões

Com este trabalho ficou evidenciado que a recomendação semântica pode orientar um usuário na escolha de *tags* que rotulam com maior eficácia um recurso *Web*. Além disso, verificou-se que o pré-processamento dos termos e a maior ponderação da personomia contribuíram com um conjunto de *tags* mais relacionadas às características do usuário, permitindo fornecer a este um vocabulário mais consistente.

Agradecimentos

Agradeço de forma única a todos que mostraram interesse e se comprometeram com o desenvolvimento desse trabalho.

Referências

- [1] Smith, G. *Tagging: People-powered Metadata for the Social Web*. Voices That Matter, New Riders Press, 2008.
- [2] Hotho, A.; Jäschke, R.; Schmitz, C. e Stumme, G. Information Retrieval in Folksonomies: Search and Ranking. In *Anais do ESWC 2006*, Budva, 2006, 411--426.
- [3] Wal, T.V. *Folksonomy Coinage and Definition*. Disponível em <http://vanderwal.net/folksonomy.html>. Acessado em 18 de março de 2010.
- [4] Amaral, A. e Aquino, M.C. Práticas de Folksonomia e Social Tagging no Last.fm. In *Anais do IHC 08 - VIII*, Porto Alegre, 2008, disponível em http://www.inf.pucrs.br/ihc2008/pt-br/assets/files/Praticas_Folksonomia_Social_Tagging_Lastfm.pdf
- [5] Da Silva, J.V. Gerenciamento do vocabulário do usuário em sistemas baseados em Tagging. *Dissertação de Mestrado - Universidade Estadual de Maringá*, 2009.
- [6] Basso, C.A. Uma proposta para a evolução de ontologias a partir de personomias em sistemas baseados em tagging. *Dissertação de Mestrado - Universidade Estadual de Maringá*, 2009.
- [7] Newman, M.E.J. Power laws, Pareto distributions and Zipf's law. In *Contemporary Physics* 46, 323--351, 2005.