

# Ferramentas para detecção de grupos em WIKIS

Fabício J. Barth<sup>1</sup>

<sup>1</sup>Apontador

fabricao.barth@lbslocal.com

Rua Funchal, 129, 1º andar, Vila Olímpia  
São Paulo - SP

08 de outubro de 2010

# Vantagens do uso de WIKI

[O'Leary 2008]

- WIKIS auxiliam na criação de documentos consensuais;
- através de WIKIS é fácil gerar informações a partir de diversas fontes, e;
- WIKIS geram uma rede de conhecimento ligando pessoas ao conteúdo criado.

# Vantagens do uso de WIKI

[O'Leary 2008]

- WIKIS auxiliam na criação de documentos consensuais;
  - através de WIKIS é fácil gerar informações a partir de diversas fontes, e;
  - **WIKIS geram uma rede de conhecimento ligando pessoas ao conteúdo criado.**
- 
- Quais são as características desta rede?
  - Como fazer uso da informação existente nesta rede?

# Objetivo do trabalho

## Objetivo

O objetivo deste trabalho é explorar duas técnicas, agrupamento hierárquico de documentos e análise de grafos, para a detecção de grupos em WIKIS.

# Objetivo do trabalho

## Objetivo

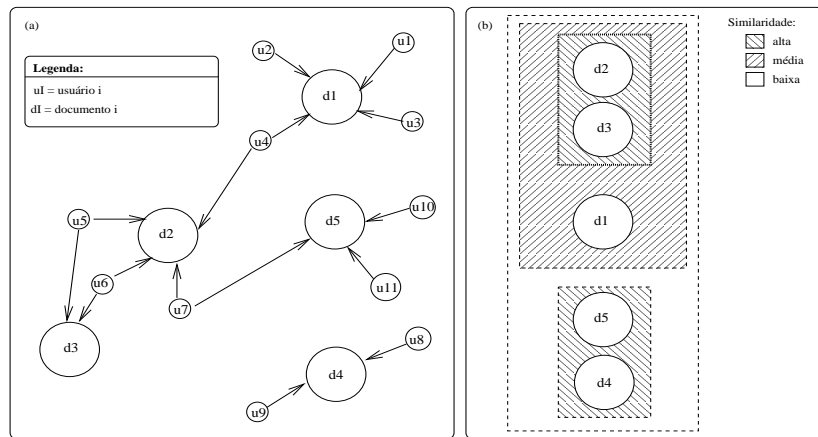
O objetivo deste trabalho é explorar duas técnicas, agrupamento hierárquico de documentos e análise de grafos, para a detecção de grupos em WIKIS.

Exemplo de histórico de criação e alteração de páginas em um WIKI.

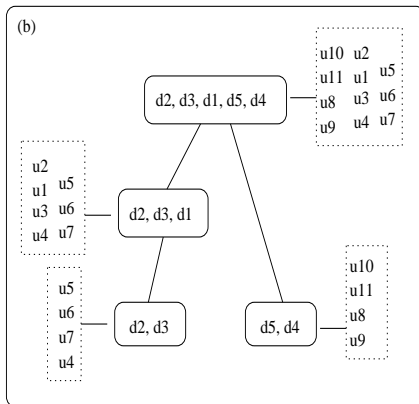
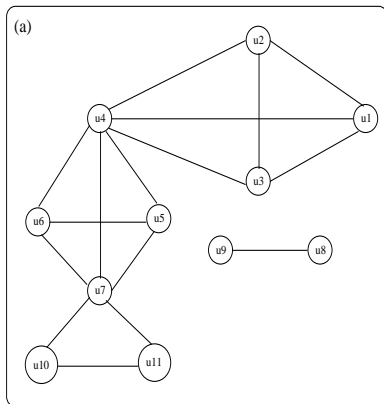
Documento	Versão	Editor	Data	Documento	Versão	Editor	Data
$d_1$	1	$u_1$	...	$d_2$	4	$u_7$	...
$d_1$	2	$u_2$	...	$d_3$	1	$u_5$	...
$d_1$	3	$u_2$	...	$d_3$	2	$u_6$	...
$d_1$	4	$u_3$	...	$d_3$	3	$u_6$	...
$d_1$	5	$u_4$	...	$d_4$	1	$u_8$	...
$d_2$	1	$u_4$	...	$d_4$	2	$u_9$	...
$d_2$	2	$u_5$	...	$d_5$	1	$u_{10}$	...
$d_2$	3	$u_6$	...	$d_5$	2	$u_{11}$	...

# Relações

Exemplos da ligação entre usuários e documentos e da relação de similaridade entre documentos.



# Grafos e Agrupamentos



## Implementação (1/2)

- A ferramenta acessa o conteúdo produzido em WIKIS através de uma **interface XML-RPC**.
- Foram acessados ambientes colaborativos de escrita utilizados por aproximadamente **300 pessoas** em aproximadamente **15 projetos** distintos.
- Cada nodo  $N$  em  $G(N, E)$  representa uma pessoa. Para cada pessoa em  $N$  são identificadas as arestas  $E$  que ligam esta pessoa a outras pessoas que manipularam os mesmos documentos.
- Para a visualização dos grafos foi utilizada a **API Jung**<sup>1</sup>.
- O algoritmo utilizado para a identificação dos sub-grupos no grafo foi o algoritmo **VoltageClusterer** [Wu e Huberman 2004].

---

<sup>1</sup><http://jung.sourceforge.net/>









## Implementação (2/2)

- Para a implementação do agrupamento hierárquico foi utilizado um algoritmo da classe **UPGMA** [Jain, Murty e Flynn 1999, Manning e Schütze 2003].
- O **pré-processamento** de cada **documento** inclui algoritmos de *stemming* [Porter 1980], lista de *stop-words* [Manning e Schütze 2003] e a transformação de cada documento em um vetor utilizando a equação *TF-IDF* [Salton e Buckley 1988].

# Resultados

- Os grupos encontrados utilizando a **análise de grafos** foram grupos muito similares aos formalmente definidos no ambiente de trabalho.
- Foi possível identificar alguns **hubs** (pessoas que conectam grupos) durante a análise.
- A partir do **agrupamento hierárquico** de documentos foi possível identificar grupos de pessoas com interesses e atividades em comum que nunca trabalharam nos mesmos projetos.
- Implementação disponível em <http://trac.fbarth.net.br/wikiAnalysis> com licença GPL.

# Referências

-  JAIN, A. K.; MURTY, M. N.; FLYNN, P. J. Data clustering: a review. *ACM Computing Surveys*, v. 31, n. 3, p. 264–323, 1999.
-  MANNING, C. D.; SCHÜTZTE, H. *Foundations of Statistical Natural Language Processing*. [S.l.]: MIT Press, 2003.
-  O'LEARY, D. E. Wikis: from each according to his knowledge. *IEEE Computer Society*, p. 34–41, 2008.
-  PORTER, M. An algorithm for suffix stripping program. *Program*, v. 14, n. 3, p. 130–137, 1980.
-  SALTON, G.; BUCKLEY, C. Term weighting approaches in automatic text retrieval. *Information Processing and Management*, v. 24, p. 513–523, 1988.
-  WU, F.; HUBERMAN, B. Discovering communities in linear time: A physics approach. *European Physics Journal*, p. 331–338, 2004.