

## Sistemas computacionais para a identificação de especialistas em ambiente corporativo

*Um dos maiores desafios em organizações é administrar o conhecimento dos seus colaboradores, por exemplo, identificando especialistas para uma determinada necessidade de informação ou tarefa. Atualmente, uma das ferramentas utilizadas para tentar resolver este desafio é o conceito de página amarela. Páginas amarelas são implementadas criando e atualizando o perfil de cada pessoa da organização manualmente - utilizando um sistema com banco de dados ou apenas fichas em word. Este tipo de abordagem tem um custo alto e deve ser aprimorada. Este artigo apresenta algumas formas para automatizar este processo e, conseqüentemente, aumentar a sua eficiência.*

Fabrcio J. Barth

### INTRODUÇÃO

Um dos maiores desafios em organizações comerciais, educacionais ou governamentais é administrar o conhecimento dos seus colaboradores, por exemplo, identificando especialistas para uma determinada necessidade de informação ou tarefa. Encontrar uma pessoa com as habilidades e conhecimentos necessários para executar uma tarefa em uma organização é, geralmente, crucial para o sucesso de um projeto em andamento.

Reusar conhecimento está fortemente relacionado a explorar melhor o capital humano existente dentro da equipe ou organização, de modo a evitar gastos de tempo e dinheiro na aquisição do conhecimento cada vez que ele for necessário. Porém, o conhecimento nem sempre é algo facilmente acessível, visto que ele pode estar presente apenas como conhecimento tácito entre os colaboradores de uma organização.

Um sistema que perceba indivíduos que são potenciais colaboradores e os coloque em contato trará um grande benefício ao trabalho desenvolvido por ambos. Afinal, a comunicação

direta torna mais fácil e efetivo o compartilhamento dos conhecimentos tácitos e habilidades entre os indivíduos.

Um sistema para identificação de especialistas é um sistema que retorna referências para atores humanos que são identificados como especialistas no domínio requisitado.

Abordagens iniciais para o desenvolvimento de sistemas para a identificação de especialistas fazem uso de um sistema para cadastramento do perfil dos colaboradores em um banco de dados (tais abordagens também são chamadas de "Páginas Amarelas"). Este banco de dados é manualmente construído e consultado. Tais abordagens requerem um considerável esforço para a criação e manutenção dos dados. Conseqüentemente, outras iniciativas foram desenvolvidas para automaticamente construir os perfis dos colaboradores a partir de repositórios da organização. A maioria destas abordagens é focada na procura por especialistas em um determinado tópico, com base em documentos bem conhecidos deste tópico. Outras abordagens, mais recentes e que serão tratadas neste artigo, realizam a extração automática do perfil

dos colaboradores a partir de uma coleção de documentos heterogêneos, por exemplo, documentos encontrados em uma Intranet de uma organização.

O objetivo deste artigo é descrever os métodos utilizados na construção de sistemas capazes de identificar especialistas a partir de consultas feitas por qualquer usuário. As principais abordagens para a construção de sistemas utilizados na identificação de especialistas podem ser divididas em duas categorias: recuperação a partir do perfil do especialista, e; modelo de dois estágios [Barth 2009]

## RECUPERAÇÃO A PARTIR DO PERFIL DO ESPECIALISTA

Nesta abordagem é construído um perfil para cada pessoa da organização. Um perfil é composto por uma lista de palavras com as suas respectivas frequências. Esta lista de palavras é gerada a partir da concatenação do conteúdo encontrado nos documentos ou em janelas que possuem uma referência a pessoa. Geralmente, uma lista com os nomes e e-mails dos especialistas é fornecida para identificar uma citação ao candidato à especialista. No entanto, variantes dos nomes dos especialistas podem ser consideradas para melhorar o índice de recuperação dos especialistas.

Uma janela é uma sequência de palavras delimitadas: pelo número de palavras; por meta-informações, tais como as tags de páginas HTML, ou; por estruturas semânticas do texto, tais como sentenças e parágrafos.

Nesta abordagem, o Sistema de Recuperação de Informação procura pelos termos da busca nos arquivos dos perfis dos usuários, retornando os perfis mais relevantes para a consulta feita. Neste caso, qualquer função de ordenação pode ser utilizada.

Um exemplo desta abordagem é o trabalho desenvolvido por [Fu et al. 2005]. Neste trabalho é utilizada uma lista de e-mails pré-determinada para identificar os especialistas nos documentos. Para cada especialista é criado um arquivo que é formado a partir do conteúdo de todos os documentos onde a referência para o especialista foi encontrada. A recuperação de um especialista para um determinado tópico é realizada procurando-se pelos termos de uma consulta nos arquivos dos especialistas.

No trabalho apresentado em [Ru et al. 2006] os candidatos a especialista são ordenados pela relevância dos arquivos que descrevem os candidatos. Cada arquivo de descrição é construído com as palavras que co-ocorrem com o candidato dentro de uma mesma janela de texto em um documento, com tamanho igual a 150 palavras.

Para cada candidato a especialista [Fu et al. 2007] cria o que é chamado de documento de descrição da pessoa (PDD). Para a construção do documento de descrição da pessoa é utilizado o contexto onde uma referência à pessoa é encontrada nos documentos, ou seja, uma janela de palavras. Além disso, são utilizados documentos que são as páginas pessoais dos colaboradores da empresa.

## MODELO DE DOIS ESTÁGIOS

A maioria dos trabalhos sobre recuperação de especialistas utiliza uma abordagem composta por duas partes: um modelo de relevância, e; um modelo de co-ocorrência. O modelo de relevância representa se um documento é ou não relevante para uma consulta. O modelo de co-ocorrência representa se uma consulta (se os termos de uma consulta) é associada com uma pessoa ou não [Balog 2008].

Em [Cao et al. 2005] e [Bao et al. 2006] o modelo de relevância é implementado utilizando a função BM25 [Manning, Raghavan e Schütze 2008]. O que é variável, e testado, em ambos os trabalhos é a utilização de diferentes janelas de texto no modelo de co-ocorrência. As janelas de texto consideradas são as mesmas definidas anteriormente: tamanho de palavras, blocos semânticos e blocos gerados a partir de metadados.

Em [Cao et al. 2005] os melhores resultados foram obtidos com janelas de 50 palavras. Em [Bao et al. 2006] os melhores resultados foram obtidos utilizando as tags title, h1 e h2 ao criar as janelas. Segundo [Zhu et al. 2006], ao aumentar o tamanho da janela aumenta-se o número de associações entre pessoas e termos da consulta. No entanto, um ruído maior ao modelo é introduzido.

Em uma Intranet de uma organização alguns documentos são mais populares que outros. Segundo [Zhu et al. 2006], esta relação de popularidade pode auxiliar na identificação de especialistas em um tópico. Por isso, em [Zhu et al. 2006] o modelo de relevância considera o valor do PAGERANK [Brin e Page 1998] de cada documento.

Em [You et al. 2006] o modelo de relevância implementado é a função BM25 e o modelo de co-ocorrência utiliza é a concatenação de várias janelas baseadas em meta-informações dos documentos: janela do título do documento (TITLE), janela do cabeçalho do documento (HEAD) e janela de bloco do documento (H1 e H2).

Em [Shen et al. 2007] o modelo de relevância utiliza a função BM25 e o modelo de co-ocorrência é o número de ocorrências do especialista no documento. Apesar da simplicidade, os resultados alcançados por [Shen et al. 2007] não são muito diferentes dos resultados alcançados por [Fu et al. 2007].

## CONSIDERAÇÕES

Este artigo descreveu algumas das iniciativas para o desenvolvimento de sistemas para a identificação de especialistas em ambientes corporativos. Foi visto que as principais abordagens para a identificação de especialistas podem ser divididas em duas categorias: recuperação a partir do perfil do especialistas, e; modelo de dois estágios.

Na abordagem baseada no perfil do especialista, o Sistema de Recuperação de Informação procura pelos termos da busca nos arquivos dos perfis dos usuários. A forma para construção do perfil que retornou os melhores resultados foi utilizando o janelamento pelo número de palavras. As funções de ordenação utilizadas foram, basicamente, formadas pela função de ordenação BM25.

A maioria dos trabalhos sobre identificação de especialistas utilizam uma abordagem composta por duas partes, chamada de modelo de dois estágios. As partes deste modelo são: um modelo de relevância, e; um modelo de co-ocorrência. Geralmente, para o modelo de relevância são utilizadas ou a função BM25 ou a função BM25 com PAGERANK. Para o modelo de co-ocorrência são testadas diferentes janelas de texto. As janelas de texto consideradas são as mesmas utilizadas nas soluções baseadas no perfil do especialista: tamanho de palavras, blocos semânticos e blocos gerados a partir de metadados.

Os resultados encontrados até o momento não são conclusivos. Não sabe-se dizer se o uso da função BM25 de forma isolada é melhor que o uso da função BM25 com a função PAGERANK para a construção do modelo de relevância. Não se tem indícios suficientes para afirmar qual é a melhor equação para o modelo de co-ocorrência, nem que tipo de janela deve-se utilizar. Apesar do modelo de dois estágios ser mais utilizado que

[www.terraforum.com.br](http://www.terraforum.com.br)



a abordagem baseada no perfil do especialista, também não pode-se afirmar que o modelo de dois estágios é melhor que a abordagem baseada no perfil do especialista.

Acredita-se que toda organização deve utilizar sistemas que facilitam a identificação de especialistas. No entanto, a implantação destes sistemas deve ser apoiada por um processo que identifica todas as características da organização e cria mecanismos de busca personalizados para a mesma.

## REFERÊNCIAS BIBLIOGRÁFICAS

[Balog 2008] BALOG, K. People Search in the Enterprise. Tese (Doutorado) -- Dutch Research School for Information and Knowledge Systems, 2008.

[Bao et al. 2006] BAO, S. et al. Research on expert search at enterprise track of trec 2006. In: The Fifteenth Text REtrieval Conference Proceedings (TREC 2006). 2006.

[Barth 2009] BARTH, F. J. Recuperação de documentos e pessoas em ambientes empresariais através de árvores de decisão. Tese (Doutorado) -- Escola Politécnica da Universidade de São Paulo, 2009.

[Brin e Page 1998] BRIN, S.; PAGE, L. The anatomy of a large-scale hypertextual web search engine. Computer Networks and ISDN Systems, v. 30, n. 1-7, p. 107-117, 1998.

[Cao et al. 2005] CAO, Y. et al. Research on expert search at enterprise track of trec 2005. In: The Fourteenth Text REtrieval Conference Proceedings (TREC 2005).

[Fu et al. 2007] FU, Y. et al. Thuir at trec 2007: Enterprise track. In: The Sixteenth Text REtrieval Conference (TREC 2007).

[Fu et al. 2005] FU, Y. et al. Thuir at trec 2005: Enterprise track. In: The Fourteenth Text REtrieval Conference Proceedings (TREC 2005).

[Manning, Raghavan e Schütze 2008] MANNING, C. D.; RAGHAVAN, P.; SCHÜTZE, H. An Introduction to Information Retrieval. Cambridge University Press, 2008.

[Ru et al. 2006] RU, Z. et al. Bupt at trec 2006: Enterprise track. In: The Fifteenth Text REtrieval Conference (TREC 2006).

[Shen et al. 2007] SHEN, H. et al. Research on enterprise track of trec 2007. In: The Sixteenth Text REtrieval Conference Proceedings (TREC 2007).

[You et al. 2006] YOU, G. et al. Ricoh research at trec 2006: Enterprise track. In: The Fifteenth Text REtrieval Conference (TREC 2006).

[Zhu et al. 2006] ZHU, J. et al. The open university at trec 2006 enterprise track expert search task. In: The Fifteenth Text REtrieval Conference Proceedings (TREC 2006).

\*\*\*

**Fabício Barth** é consultor da TerraForum.  
Seu e-mail é [fabricao.barth@terraforum.com.br](mailto:fabricao.barth@terraforum.com.br)

[www.terraforum.com.br](http://www.terraforum.com.br)

## ARTIGOS RELACIONADOS

A Intranet voltada para as necessidades dos funcionários

Tecnologias para a Colaboração e o Sistema de Organização do Conhecimento

O blog como agente de comunicação corporativa

Recursos Humanos, Negócios e Aprendizado

Portais corporativos: instrumento de gestão de informação e conhecimento

Software para Gestão de Conteúdo: Sete Características Críticas

Portais Corporativos: múltiplos conceitos, perspectivas e desafios de estruturação

Determinantes da Necessidade de Portais Corporativos

Tendências emergentes em mecanismos de busca

Portais Corporativos: novo instrumento de gestão da informação

## A EMPRESA

A TerraForum Consultores é uma empresa que atua em consultoria, tecnologia e design, fundada em agosto de 2002. Os clientes da empresa são, em sua maioria, grandes e médias organizações dos setores público, privado e terceiro setor. A empresa possui escritórios em São Paulo, Curitiba, Rio de Janeiro, Porto Alegre e Toronto, no Canadá. O crescimento da empresa tem se pautado por contínuo re-investimento no desenvolvimento de sua equipe, metodologias e infra-estrutura. É dirigida pelo Dr. José Cláudio Terra, pioneiro e maior referência em Gestão do Conhecimento no país. Além disso, conta com uma equipe especializada e internacional de consultores.

## PUBLICAÇÕES TERRAFORUM

Winning at Collaboration Commerce

Gestão do Conhecimento e E-learning na Prática

Portais Corporativos, a Revolução na Gestão do Conhecimento

Gestão do Conhecimento - O Grande Desafio Empresarial

Gestão do Conhecimento em Pequenas e Médias Empresas

Realizing the Promise of Corporate Portals: Leveraging Knowledge for Business Success

Gestão de Empresas na Era do Conhecimento

Gestão de Conteúdo 360° - Integrando Design, Negócios e Tecnologia

Gestão 2.0 - Como integrar a colaboração e a participação em massa para o sucesso nos negócios