

Aplicação de técnicas de descoberta de conhecimento em textos na inteligência competitiva

Leandro Krug Wives*

José Palazzo Moreira de Oliveira**

Resumo

O nível de globalização proporcionado pela Internet disponibiliza um mercado de alcance mundial a qualquer empresário. Independente de seu tamanho, a micro, pequena e média empresa coloca-se no mesmo patamar das grandes empresas, obtendo as mesmas oportunidades e alcances. Esse fato caracteriza tanto uma vantagem quanto uma desvantagem, já que novos concorrentes podem surgir a cada momento e causar impacto mercadológico mesmo estando localizados fisicamente em um extremo oposto do mundo.

Nesse cenário, a qualidade e a oferta de produtos cresce e os clientes passam a ser mais exigentes. Com isso, as empresas devem inovar, oferecendo produtos e serviços de qualidade a fim de satisfazer às exigências dos clientes e manter-se competitiva no mercado.

Para manter-se competitiva uma empresa necessita obter informações (tanto internas quanto externas) sobre o seu ambiente. Isso significa que o empresário deve estar sempre a par dos seus produtos, clientes, fornecedores e concorrentes. Ele deve, além disso, estar sempre monitorando novas tecnologias e avaliando se e como elas poderiam ser utilizadas para obter alguma espécie de vantagem sobre seus concorrentes. Isso deve ser feito antes do que seus concorrentes pois o controle da concorrência é uma tarefa recíproca.

Existem atualmente numerosas fontes de informação ao dispor de um empresário, o que faz com que a quantidade de informações que uma empresa tenha a sua disposição ultrapasse a capacidade de processamento e compreensão. Logo, os empresários devem selecionar de alguma forma as informações mais relevantes ao seu negócio, descartando as demais.

Inteligência Competitiva [BAL 98] é uma área da administração de empresas que busca justamente solucionar os problemas relacionados à coleta e análise de informações relevantes a uma empresa para que esta possa adquirir vantagem competitiva sobre seus concorrentes. Essa área utiliza muitas técnicas de coleta e análise de informações que são estudadas pelas áreas de Recuperação de Informação [WIV 97] e de Descoberta de Conhecimento em Textos [FER 97; MOS 98].

Um dos maiores problemas enfrentados por essas áreas consiste em localizar e analisar informações relevantes para o usuário de forma rápida. O problema é agravado

* wives@inf.ufrgs.br

** palazzo@inf.ufrgs.br

pelo fato de grande parte da informação disponível estar em um formato textual que possui ambigüidades inerentes à linguagem natural utilizada neles [CHE 94].

O agrupamento (*clustering*) de informações textuais [WIV 98] é uma técnica capaz de identificar correlações entre documentos e gerar grupos de documentos que tratem de um mesmo conteúdo. Ela pode ser aplicada em informações coletadas por uma empresa a fim de facilitar o processo de identificação de informações relevantes. A identificação de correlações entre documentos também pode auxiliar o empresário a compreender melhor as informações de que possui, já que a distribuição de assuntos de um conjunto de documentos é identificada facilmente pelo método de agrupamento.

Uma vez selecionado o grupo de documentos mais relevantes para a empresa, é possível selecioná-los e reaplicar o método, refinando os resultados a fim de identificar subgrupos e subassuntos de interesse. É possível também identificar as características (termos) mais importantes (representativos) de um grupo e, com elas, elaborar regras que permitam realizar classificação e filtragem de novas informações, facilitando o processo de monitoração. A identificação de características ajuda também o empresário a compreender melhor determinado assunto (ou ramo) e indica quais termos devem ser utilizados a fim de que informações pertinentes a esse ramo sejam recuperadas.

Dependendo do conjunto de documentos utilizados como entrada para o método de agrupamento é possível identificar tendências de mercado (realizando comparações com resultados anteriores), mercados ainda não explorados, ligações entre empresas e fornecedores, além de exceções (por exemplo, identificar o porquê de determinado produto surgir em um grupo aparentemente sem ligação).

A técnica de *clustering* não é uma solução definitiva para os problemas envolvidos no processo de Inteligência Competitiva, que envolve as fases de coleta, organização, análise, difusão e identificação de soluções. O *clustering* auxilia as fases de coleta, organização e análise de informações sob alguns aspectos e de forma um pouco intuitiva, ou seja, muito dependente de quem utiliza essa técnica.

Como em todo processo de descoberta, para que novas oportunidades sejam identificadas, é necessário preparar corretamente os dados e informações e conhecer um pouco sobre eles. As técnicas de análise e descoberta devem ser aplicadas com algum objetivo [FEL 97] e de forma sistemática, ou do contrário nada será descoberto.

Mesmo que determinada informação ou conhecimento tenha sido adquirido durante um processo de descoberta, o fato de possuí-los não é o bastante para que o empresário mantenha-se estrategicamente posicionado no mercado. Para tanto é necessário não só possuir a informação, mas sim, também, utilizá-la no momento e local apropriados.

Referências

- [BAL 98] BALESTRIN, Alsones; VARGAS, Lilia Maria. **Monitoramento do ambiente concorrencial na indústria metal-mecânica da região de Caxias do Sul (RS)**. 1998. Disponível por WWW em <http://osse.inf.ufrgs.br/intcomp/artigos/art-01.html>.
- [CHE 94] CHEN, Hsinchun. The vocabulary problem in collaboration. **IEEE Computer: Special Issue on CSCW**, Los Alamitos: IEEE Computer Society. v.27, n.5, p.2-10, 1994. Disponível por WWW em <http://ai.bpa.arizona.edu/papers/cscw94/cscw94.html> (em 22/05/2000).
- [FEL 97] FELDENS, Miguel Artur. Descoberta de Conhecimento em Bases de Dados e Mineração de Dados. In: **OFICINA DE INTELIGÊNCIA ARTIFICIAL, I**, 1997, Pelotas, RS. Proceedings... Pelotas: EDUCAT, 1997. p.51-59.
- [FER 97] FELDMAN, Ronen; HIRSH, Haum. Exploiting background information in knowledge discovery from text. **Journal of Intelligent Information Systems**, Netherlands: Kluwer Academic Publishers. v.9, n.1, p.83-97. 1997.
- [MOS 98] MOSCAROLA, Jean; BOLDEN, Richard. **From the data mine to the knowledge mill: applying the principles of lexical analysis to the data mining and knowledge discovery process**. France: Université de Savoie, 1998a. (Note de Recherche, n. 98-15).
- [WIV 97] WIVES, Leandro Krug. **Um Estudo sobre Técnicas de Recuperação de Informações com ênfase em Informações Textuais**. Porto Alegre, 1997. 55p. Trabalho Individual (TI-672) -- CPGCC, UFRGS.
- [WIV 98] WIVES, Leandro Krug. **Um Estudo sobre Agrupamento de Documentos Textuais em Processamento de Informações não Estruturadas Usando Técnicas de "Clustering"**. Porto Alegre, 1998. 102p. Dissertação (Mestrado em Ciência da Computação) -- PPGC, UFRGS.

