

Searching and Analyzing Blog for Competitive Advantages

Stefano Maria De' Rossi¹

As competition increases and intensifies, organizations are discovering clever ways to systematically gather and analyze information from internal and external sources (a.k.a. “the enterprise data network”) and provide alerts of changing business conditions in real time.

Gathering and integrating real-time data from the Web is crucial for today’s enterprise. Ignoring or manually gathering this data is no longer an option for the organization that relies on current market, competitive and customer data for superior business decision

This new breed of “business activity monitoring” applications can drive greater productivity, new sources of revenue, and competitive advantage. Below are a few examples:

- **Aggregating customer and competitor information for CRM** Monitoring public opinion from cyber forums, blogs and RSS feed about products and competitors
- **Integrating competitive pricing information into pricing analytics or price alert applications** Extracting pricing information from e-business and e-government sites for trend analysis or fast response to the actions of competitors.
- **Collecting and organizing content to populate enterprise information portals** Crawling the Web, internal information sources, and subscription services to automatically populate portals with pertinent and timely content.
- **Fundamental research for risk management and Compliance.** Internet monitoring of partners, resellers, and the gray market for resale authorization, price accuracy, logo usage, logo positioning, link to and from partner sites.

While Web mining and network analysis techniques have been widely used to analyze the content and structure of the Web sites of hate groups on the Internet, these techniques have not been applied to the study of blogs.

As blogs have become one of the fastest growing types of Web-based media, bloggers can express their opinions and emotions more freely and easily than before. These blogs are microcosms of conversations happening around the world. The millions of people engaging in blog conversations are on the front lines of consumer awareness – these are the influencers shaping public opinion. Positive or negative buzz on the blogosphere can have tremendous influence and even affect stock prices.

In the blog space, many communities have emerged, which include racists and hate groups that are trying to share their ideology, express their views, or recruit new group members. It is important to analyze these cyber communities, defined based on group membership and subscription linkages, in order to monitor for activities that are potentially harmful to society.

In this research, we propose a framework to address this problem.

¹ Abstract presentato al Data Mining, Protection, Detection and other Security Technologies 2008
<http://www.wessex.ac.uk/conferences/2008/data08/index.html> .

The framework consists of four modules, namely blog spider, information extraction, network analysis, and visualization. The Blog Spider module downloads blog pages from the Web.

These pages are then processed by the Information Extraction module. Data about these blogs and their relationships are extracted and passed to the Network Analysis module for further analysis. Finally the Visualization module presents the analysis results to users in a graphical display