Competitive Intelligence and Internet Sources

Alexandru BĂRBULESCU, Bucuresti, România, alexbarbulescu@ie.ase.ro

In the Knowledge Age to maintain profitability and in some cases to remain in the market, companies must focus their actions in activities such as collecting, filtering, and disseminating information about market, about competitors and their actions. Those are part of Competitive Intelligence (CI) concept. In digital age, most of the information needed for CI projects is available on the web. This paper focuses on this field and presents a mix of directions that companies need to take into consideration in their CI projects in order to achieve the goals.

Keywords: competitive intelligence, web mining, information.

Introduction

Competitive Intelligence definition from the Society of Competitive Intelligence Professionals (SCIP) is:

"the legal collection and analysis of information regarding the capabilities, vulnerabilities, and intentions of business competitors conducted by using 'open sources' and ethical inquiry."

From this definition results that CI is not espionage or an illegal activity. CI activities consist in gathering and analyzing information using available sources without breaking the law and the ethics. This line between legal and illegal and ethics and non ethics is often very hard to express.

Like the economy, the capabilities, vulnerabilities and intentions of business competitors are dynamic. The environment is changing and drives changes to the competitor capabilities, vulnerabilities and intentions. The external environment of a competitor includes the industry and the market. CI must take into consideration customers, competitors, suppliers, business environment and macroeconomic context. The objective of CI is to acquire and analyze information that enables a company to anticipate changes in its environment and make the right strategic decisions. "Right strategic decisions" are those that go the company forward to accomplish its mission. The anticipation of changes is a one of the objectives of CI. The company will "act" instead of "react" to the change of the environment. Going to the other extremity, the company makes the environment instead of "playing" in others environment by their rules.

CI Functions

CI processes involve a number of specific distinct activities undertaken by a company. According to Herring an effective CI project is a continuous cycle, whose steps include:

1. Planning and direction (working with decision makers to discover and hone their intelligence needs);
2. Collection (legally and ethically);
3. Analysis (interpreting data and compiling recommended actions)
4. Dissemination (presenting findings to decision makers)
5. Feedback (taking into account the response of decision makers and their needs for continued intelligence).

The most used functions of CI are:

- Understanding customers needs
- Providing better solutions in the market
- Anticipate changes in the market (regarding the customers and the players)
- Anticipating actions and reactions of competitors to the market anticipated changes
- Learning from others' successes and failures
- Improving acquisition activities
- Learning about new technologies, products & processes that affect business
- Learning about macroeconomic trends that affect business
- Analyzing business to make the most of your strengths

Acquisition of intelligence

This may occur from numerous sources that fit within two broad categories. The categories are Human and Published. These two categories presume two different types of skills. For human sources interviewing skills are needed. The most important is how to “ask questions to get the information”. For the published sources information retrieval skills are needed.

In this article we focus on the published sources that imply, in digital age, the internet. Internet information resources are being used more frequently in the CI process. This is because almost all of the paper publishing can be found in digital format on the internet. In some cases the information published digitally is much detailed than the information published on paper. For example the companies are publishing an article in written press about their activity, but on the web site the same information has more details, and the cost of the information is lower or even free.

The reason why I focused on the internet sources is because in many writings in this field appears the idea that much of the information needed by CI processes can be found on the internet and it is public.

The Internet

A logical structure of the internet is a graph structure consisting in nodes (web pages) and arcs (links to web pages). The most used method for gathering the information about a subject is to use search engines like google and yahoo. The search is based on some keywords. The engines retrieve links to the pages that contain the keywords. The pages are previously indexed for rapid searches. Because in many cases we talk about billions of pages that match the search criteria, a ranking model must be used in order to respond accurate. The ranking system means to order the results using a system more complex which includes more attributes, not just the fact that the keyword appears in the web page. This ranking system is based on the graph arcs, and there are used sophisticated algorithms.

According to Boncella a search engine usually consists of the following components:

1. Web Crawlers or Spiders are used to collect Web pages using graph search techniques.
2. An indexing method is used to index collected Web pages and store the indices into a database.
3. Retrieval and ranking methods are used to retrieve search results from the database and present ranked results to users.
4. A user interface allows users to query the database and customize their searches.

The methods for controlling the search in order to retrieve just the pages that are relevant for the search represent WEB Mining. This refers to Structure Mining (graph structure), Content Mining (pages content mining) and Usage Mining (usage of the web pages).

The structure mining refers to the pages (nodes) and links (arcs). When a page has more incident arcs (links from other pages) will mean that this page is important and relevant and the rank will be increased. The model is based on the topology of the hyperlink. This model can be used to categorize the Web pages and is useful to generate information such as similarity and relationships between Web sites. And the link structure of the Web contains important implied information, and can help in filtering or ranking web pages.

Web content mining is more accurate that the standard search and refine the initial search results, using text mining algorithms and techniques. It consists in but not limited to structured data extraction, information integration, knowledge synthesis, template recognition and page segmentation. In the last decade there were developed algorithms using web agents and mining techniques to achieve the goal of structuring web data. Those are ordered in agent-based approach (ABA) and database approach (DA). The agent-based approach involves sophisticated systems that can act autonomously or semi-autonomously using customized rules, to discover and organize web data. The agent-based systems can be placed into the following categories: Intelligent Search Agents (Intelligent Web Agents are developed to search and organize relevant information using customized information regarding a particular domain), Information Filtering/Categorization (Web based agents use retrieval techniques and characteristics of open hypertext Web documents to automatically retrieve, filter, and categorize them), Personalized Web Agents (Web agents that learn user preferences over multiple categorizations and discover web data that correspond to those preferences).

The database approach is focused on techniques for integrating and organizing the heterogeneous and semi-structured data on the Web into more structured and high-level collections such as relational databases, and using standard database querying mechanisms and data mining techniques to access and analyze this information. In the field there are two approaches: multilevel databases and web query systems.

Web Usage Mining

Usage mining is represented by activities involving techniques for extracting relevant information from the web logs containing page references associated with either a Web server or Web client. Web usage mining is the activity that involves the automatic discovery of user access patterns from one or more Web servers. It mines the secondary data (Web server access logs, browser logs, user profiles, registration data, user sessions or transactions, cookies, user queries, mouse clicks and any other data as the result of interaction with the Web) derived from the interactions of the users during certain period of Web sessions.

In any way the information is retrieved, one most important CI principle says that the information must be verified using multiple sources or other way. Using web sources one important problem is the inaccuracies of information, either accidental or intentional.

Conclusions

This article presents an overview of the concepts associated with CI projects using the Web. The methods and techniques associated with information gathering and information analysis are to a great degree automated by using personalized or focused intelligent web agents. Web searches return a large set of pages that require an automated approach, like text mining, to information analysis. The main idea of this paper is that the search is more accurate when there are combined usage mining techniques with content mining and structure mining. The results are even better when the algorithms are parametric and are customizable with user profile.

References

[1] Aaron, R. D. and E. Naylor "Tools for Searching the 'Deep Web' ", Competitive Intelligence Magazine, (4:4), 2003.

[2] Boncella, R. J., 2003, Competitive Intelligence and the web

[3] Bouthillier, F., Jin, T. CI Professionals and their interactions with CI technology. Journal of Competitive Intelligence and Management, Vol. 3

[4] Chen, H., Chau, M., and Zebg, D. (2002) "CI Spider: A Tool for Competitive Intelligence on the Web", Decision Support Systems.

[5] Chen, H., M. Chau and D. Zebg, (2002) "CI Spider: A Tool for Competitive Intelligence on the Web".

[6] Fleisher, C. S. and B. E. Bensoussan, (2000) Strategic and Competitive Analysis, Upper Saddle River, Prentice Hall, 2003.

[7] Fleisher, C.S. & Blenkhorn, D.L., 2001, Managing Frontiers in Competitive Intelligence, Quorum Books, Connecticut, USA

[8] Goujon, B., 1999, Competitive Intelligence: extraction of relevant information with the contextual exploration method.

[9] Herring, J. P. (1998) "What Is Intelligence Analysis?" Competitive Intelligence Magazine

[10] Hohhof, B. The Information Technology Marketplace in Miller, J. (Ed). Millennium Intelligence: Understanding and Conducting Competitive Intelligence in the Digital Age, Cyber Age books, 2000.

[11] Miller, Jerry. (2000). Millenium Intelligence: Understanding and Conducting Competitive intelligence in the Digital Age. Medford, NJ: Cyberage Books.

[12] SCIP (Society of Competitive Intelligence Professionals) http://www.scip.org/.