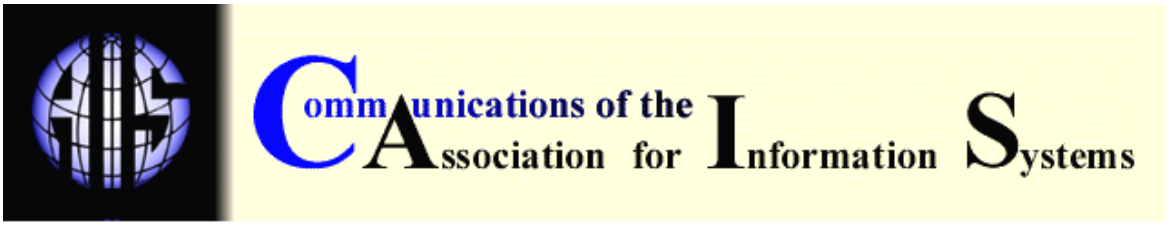

COMPETITIVE INTELLIGENCE AND THE WEB

ROBERT J. BONCELLA
Washburn University
bob.boncella@washburn.edu

ABSTRACT

Competitive intelligence (CI) is the selection, collection, interpretation and distribution of publicly held information that is strategically important to a firm. A substantial amount of this public information is accessible via the World Wide Web. This paper describes some of the difficulties in using this information resource for CI purposes, some of the solutions to these difficulties, and areas in need of research if the Web is to be used in CI.

Keywords: Competitive intelligence Internet searching and browsing Intelligence monitoring Information verification Web Mining business intelligence

I. INTRODUCTION

The intent of this paper is to provide an overview of how the Web can be used for competitive intelligence. Following a definition of Competitive Intelligence, the logical structure of the World Wide Web is reviewed to provide a foundation for understanding how information is stored on and retrieved from the web and the difficulties that arise from using this logical approach. Sections that follow detail the techniques that can be used to carry out CI projects and some of the problems associated these techniques. In particular, information gathering, information analysis, information verification, and information security are discussed as they relate to CI.

II. COMPETITIVE INTELLIGENCE

The Society of Competitive Intelligence Professionals (SCIP) defines Competitive Intelligence as

> *"the process of ethically collecting, analyzing and disseminating accurate, relevant, specific, timely, foresighted and actionable intelligence regarding the implications of the business environment, competitors and the organization itself"* [SCIP, 2003].

This process involves a number of distinct activities undertaken by a firm engaged in a CI project. An effective CI project is a continuous cycle, whose steps include: [Herring, 1998]

1. Planning and direction (working with decision makers to discover and hone their intelligence needs);

2. Collection (conducted 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).

After step 1 is completed steps 2 and 3 are the keys to a successful and efficient CI process. Many information resources are consulted to carry out steps 2 and 3. A comprehensive list of collection and analysis resources are presented by Fuld [1995].

Internet information resources are being used more frequently in the CI process. The reasons for this trend include:

1. A business Web site will contain a variety of information usually including company history, corporate overviews, business visions, product overviews, financial data, sales figures, annual reports, press releases, biographies of top executives, locations of offices, and hiring ads. An example of this information is the about page for Google. [http://www.google.com/about.html current September 1, 2003].
2. The cost of this information is, for the most part, free.
3. Access to open sources does not require proprietary software such as access to multiple commercial databases.

III. THE WEB STRUCTURE

The HTTP protocol and the use of Uniform Resource Locators (URL) determine the logical structure of the web. This logical structure provides a natural retrieval technique for the contents of the Web. The logical structure of the Web can be understood as a mathematical network of nodes and arcs. The nodes represent the web documents and the arcs are the URLs (links) located within a document. A simple retrieval technique is one that starts from a particular HTML or XML document and follows the links (arcs) from document to document (node to node). The process of "following the links" refers to document retrieval. This process is also referred to as Information Retrieval (IR). The content of the retrieved documents can be evaluated and a new set of URLs becomes available to follow.

The retrieval techniques are graph search algorithms adapted to use a document's links to implement and control the search. An example of a graph search algorithm is a breadth first search on links contained in the initial document. A modification would a best first search based algorithm. For a detail exposition of basic searching methods see Russell and Norvig [1995]

IV. INFORMATION GATHERING ON THE WEB

The most common method for gathering information from the Web is the use of "search engines"[1]. These search engines accept a user's query, generally an expression consisting of keywords, and return a set of web pages or documents that satisfy the query to some degree.

[1] Examples are:
AltaVista [http://www.altavista.com current September 1, 2003,
Infoseek [http://www.infoseek.com current September 1, 2003],
Yahoo! [http://www.yahoo.com current September 1, 2003] and

Google [http://www.google.com current September 1, 2003].

Further, this set of pages and documents are organized in some fashion. Most often this set of pages are ranked as to how well each page satisfies a query.

A Web 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. For more details on Web Crawlers, see Chen, et al. [2002].

In addition to the general search engine types, a number of domain specific search engines are available. Examples of these are

- Northern Light, a search engine for commercial publications, in the domains of business and general interest.
- EDGAR is the United States Securities and Exchange Commission clearinghouse of publicly available information on company information and filings.
- Westlaw is a search engine for legal materials.
- OVID Technologies provides a user interface that unifies searching across many subfields and databases of medical information.

A third type of search engine is the *meta-search* engine[2]. When a meta-search engine receives a query it connects to several popular search engines and integrates the results returned by those search engines. Meta-search engines do not keep their own indexes but in effect use the indices created by the search engines being searched to respond to the query.

Finally, given the success of P2P technology (e.g. Napster and Kazaa) search engines are being developed that uses the P2P technology. In this type of search, if a computer receives a request if it cannot fulfill, the request is passed on to its neighboring computer. And example of this approach is the JXTA search engine[3]. For more details on the P2P search engine technology see Waterhouse et al. [2002].

Given the size of the Web, using a graph search algorithm approach, it takes a long time to crawl and index all the relevant Web pages associated with a query, even for a domain-specific search engines. Many Web pages may be "crawled" but not indexed. As a result, information is outdated or incorrect. This "static" type of informational retrieval will not take in to account continuous updating of dynamic content Web pages. The result is information that is not current. In addition to time and currency of information, the number of pages that satisfy the user's query is a problem.

The Internet is estimated to be composed of over 552.5 billion web pages or documents, and is growing by 7.3 million pages a day [Lyman and Varian 2000]. These pages or documents can be classified into two basic types,

- the "surface Web"**,** those pages or documents that are freely available to any user. The number of these types of pages and documents is estimated to be approximately 2.5 billion; and

[2] Two examples are MetaCrawler [http://www.metacrawler.com/ current September 1, 2003] and Dogpile [www.dogpile.com current September 1, 2003].

[3] JXTA can be found at http://search.jxta.org. current September 1, 2003].

- "deep Web" pages and documents which consists of dynamic pages, intranet sites, and the content of Web-connected proprietary databases.

The number of deep Web documents is estimated to be 550 billion. Deep Web documents are generally accessible only to members of organizations that produce them or purchase them, such as businesses, professional associations, libraries, or universities. Internet search engines such as Google, AltaVista and Lycos usually do not index and retrieve deep Web pages. This distinction is important to keep in mind when doing a CI project. Some of the most valuable information, such as full-text scholarly journals, books still in copyright, business market information, and proprietary databases can only be retrieved by users with subscriptions, searching with specialized software. For a review of tools for searching the deep Web see Aaron and Naylor [2003].

A looming difficulty with gathering information using the surface Web is that a number of sites are starting to charge a fee for access to information. [Murray and Narayanaswamy, 2003].

Appendix I contains an annotated list of Web sources, some free and some not, which provide both surface and deep knowledge that would be useful when carrying out a CI project. The appendix, a summary and update of Nordstrom and Pinkerton [1999], includes the following types of information:

- Sources for general information
- Sources where you can learn about your competitors
- Sources where you can learn about industry trends
- Sources where you can learn about your firm's customers
- Chat rooms and discussions
- Sources that can help evaluate a market or an opportunity

V. INFORMATION ANALYSIS

Given the large number of pages an uncontrolled search might generate it becomes necessary to control the search. Control can be achieved by controlling the graph search techniques. Controlling the search is, in effect, a rudimentary analysis of the information being retrieved. The search should return only those Web pages that are relevant to the query. To some extent sophisticated Web search engines are able to work in this way.. This initial form of analysis is referred to as Web Mining. For a more technical discussion of web mining see Dunham [2003] and Chakrabarti [2003].

WEB MINING

Web mining can be categorized into three classes: Web Content Mining, Web Structure Mining, and Web Usage Mining.

Web Content Mining

Web Content Mining refines the basic search technique. Web Content Mining can be viewed as "on-line" or "off-line". In on-line Web Content Mining, the graph search algorithm is controlled by the contents of the page. Focused spiders, essentially intelligent agents, return a set of pages appropriate for the user's query. Examples of these types of intelligent agents maybe found in Chau and Chen, [2002] and a commercial product Answers On-line by AnswerChase.[4]

[4] See http://www.answerchase.com current September 1, 2003.

Off-line web content mining maybe carried out using one of two methods. An unsophisticated search engine will use keywords to control the graph search algorithm. This technique returns a set of pages that can either be searched again using a refinement of the initial search or the set of returned pages can be "text mined" using text mining techniques.

Text Mining. The goal of text mining is to perform automated analysis of natural language texts. This analysis leads to the creation of summaries of documents, determining to what degree a document is relevant to a user's query, and clusters document. Text mining applications are available commercially; for example TextAnalyst by Megaputer[5]. Another approach to text mining is taken by SITEX, software that uses an Artificial Neural Network approach to the mining operation (see Fukuda et al. [2000] for the details).

Web Structure Mining

Web Structure Mining uses the logical network model of the Web to determine the importance a Web page. One method is the *PageRank* technique [Page and Brin, 1998]. This technique determines the importance of Web information on the basis of the number of links that point to that Web page. The idea is that the more Web pages that reference a given Web page the greater the importance of the page. This technique combined with keyword search is the foundation of the Google search engine. Another technique is the Hyperlink-Induced Topic Search (HITS) [Kleinberg, 1999]. HITS finds Web pages that are hubs and authoritative pages. A hub is a page that contains links to authoritative pages. An authoritative page is a Web page that best responds to a user's query.

Web Usage Mining

Web Usage Mining performs data mining on Web logs. A Web log contains "clickstream" data. A clickstream is a sequence of page references associated with either a Web server or Web client (a web browser being used by a person). This data can be analyzed to provide information about the use of the web server or the behavior of the client depending upon what clickstream is being analyzed.

Regardless of how efficiently and/or effectively the information analysis task is performed, its usefulness is determined by the quality of the information retrieved. Because of the unsupervised development of Web sites and the ease of referencing other Web pages, the user has no easy method of determining if the information contained on a Web page is accurate. The possible inaccuracies may be accidental or intentional. Inaccuracies are a significant problem when the Web is used as an information source for a CI project. The issue is information verification.

VI INFORMATION VERIFICATION

Web search engines perform an evaluation of the information resources. The HITS and PageRank techniques evaluate and order the retrieved pages as to their relevance to the user's query. This evaluation does not address the accuracy of the information retrieved.

Confidence in the accuracy of the information retrieved depends on whether the information was retrieved from the surface web or the deep web. The deep web sources will be more reliable than the surface web sources and will require less verification than the information retrieved from surface web sources. In either case one should always question the source and if possible confirm with a non-Web source for validation. In assessing the accuracy of the information retrieved it is useful to ask the following questions:

[5] See http://www.megaputer.com/products/ta/index.php3 current September 1, 2003.

- Who is the author?
- Who maintains (publishes) the Web site?
- How current is the Web page?

Further suggestions and more detail on methods of verifying information retrieved from the Web, either deep web or surface web, can be found at the following Web sites:

http://www.uflib.ufl.edu/hss/ref/tips.html, (date of access April 18, 2003).

http://www.vuw.ac.nz/~agsmith/evaln/index.htm. (current September 1, 2003).

http://www.science.widener.edu/~withers/webeval.htm, (current September 1, 2003).

http://www.ithaca.edu/library/Training/hott.html. (current September 1, 2003).

http://servercc.oakton.edu/~wittman/find/eval.htm. (date of access April 22, 2003).

VII. INFORMATION SECURITY

Recognizing the possibility of a firm being the focus of someone else's CI project, information security becomes a concern. These concerns include:

1. assuring the privacy and integrity of private information,
2. assuring the accuracy of its public information, and
3. avoiding unintentionally revealing information that ought to be private.

The first of the concerns can be managed through the usual computer and network security methods [Boncella 2000, Boncella 2002].

The second concern requires some use of Internet security methods. In general a firm must guard against the exploits that can be carried out against Web sites. Some of these exploits are Web Defacing, Web Page Hijacking, Cognitive Hacking, and Negative Information.

WEB DEFACING

Web Defacing involves modifying the content of a Web page. This modification can be done in a dramatic and detectable fashion. However, and perhaps more dangerous, the content can be modified in subtle ways that contribute to the inaccuracy of the information. Sidebar 1 shows an example of overt web defacing. [Cybenko, et al. 2002]

WEB PAGE HIJACKING

Web Page Hijacking occurs when a user is directed to a web page other than the one that is associated with the URL. The page to which the user is redirected may contain information that is inaccurate. Sidebar 1 is an example given by Cybenko, et al. [2002].

COGNITIVE HACKING

Cognitive Hacking or semantic attack is used to create a misperception about a firm's image. The causes of cognitive hacking maybe disgruntled customers/ employees, competition, or simply a random act of vandalism.

The two types of cognitive hacking are (1) single source and (2)multiple sources.

SIDEBAR 1. EXAMPLE OF WEB DEFACING

The following message appeared on the New York Times home page in February 2001:

Headline: Sm0ked Crew

Subhead: The-Rev|Splurge

Sm0ked crew is back and better than ever!
"Well, admin I'm sorry to say by [sic] you have just got sm0ked
by splurge. Don't be scared though, everything will be all
right. First fire your current security advisor…"
"Well, admin I'm sorry to say by [sic] you have just got sm0ked by splurge
Don't be scared though, everything will be all right, first
First fire your current security advisor, he sux

SIDEBAR2: EXAMPLE OF WEB HIJACKING

As the result of a bug in CNN's software, when people at the spoofed site clicked on the "E-mail This" link, the real CNN system distributed a real CNN e-mail to recipients with a link to the spoofed page.

With each click at the bogus site, the real site's tally of most popular stories was incremented for the bogus story.

Allegedly a researcher who sent the spoofed story to three users of AOL's Instant Messenger chat software started this hoax. Within 12 hours more than 150,000 people had viewed the spoofed page. [Cybenko, et al. 2002]

Note: CNN refers to the cable news network by that name. Their web location is http://www.cnn.com. This particular example of web hijacking can be found at http://mirrors.meepzorp.com/cnn/britney-deadhoax/ . Readers are urged to look at this URL. It is not reproduced here because the material is copyright.

Source: Cybenko et al. [2002]

Single Source Cognitive Hacking

Single source cognitive hacking occurs when a reader sees information and does not know who posted the information. Thus, the reader does not have a way of verifying the information or contacting the author of the information.

Multiple Sources Cognitive Hacking

Multiple sources cognitive hacking occurs when several sources are available for a topic, and the information is not accurate or contradictory among the sources.

These types of cognitive hacking can be further split into two categories of cognitive attacks: (1) overt and (2) covert.

Overt cognitive attack. In an overt cognitive attack no attempt is made to conceal the attack. Web page defacing would be an example of this category of attack.

Covert cognitive attack. In a covert attack false or misleading information intended to influence reader's decisions and/or activities is intentionally distributed or inserted. The misinformation appears to be reliable. See Sidebar 3 for an example.

POSSIBLE COUNTERMEASURES TO COGNITIVE HACKING

Countermeasures to cognitive hacking exploits need to be employed by a CI researcher. For example, the misleading information may be posted by a competitor as counter-CI. Proposed counter measures to single source cognitive hacking include authentication of source, information "trajectory" modeling, and Ulam games. Proposed counter measures to multiple source cognitive hacking involve determining source reliability via collaborative filtering and reliability reporting, detection of collusion by information sources, and the Byzantine Generals Model. [Cybenko, et. al. 2002]

Countermeasures: Single Source

To carry the authentication of source countermeasure the CI researcher needs to employ due diligence regarding the information source. In addition, the researcher may use implied verification of the source; for example using PKI (Digital Signature) to verify the source of the information.

SIDEBAR 3

Example of Covert Cognitive Attack

According to the US Security Exchange Commission, 15-year-old Jonathan Lebed earned between $12,000 and $74,000 daily over six months - for a total gain of $800,000. Lebed would buy a block of FTEC stock and then using only AOL accounts with fictitious names he would post a message like the one below. Repeating the post a number of times he increased the daily trading volume of FTEC from 60,000 shares to more than one million. For an entertaining account of this case see [Lewis, 2001].

FROM: LebedTG1

FTEC is starting to break out! Next week, this thing will EXPLODE . . .Currently FTEC is trading for just $21/2. I am expecting to see FTEC at $20 VERYSOON . . .

Let me explain why Revenues for the year should very conservatively be around $20 million. The average company in the industry trades with a price/sales ratio of 3.45. With 1.57 million shares outstanding, this will value FTEC at $44. It is very possible that FTEC will see $44, but since I would like to remain very conservative. . my short term price target on FTEC is still $20! The FTEC offices are extremely busy. I am hearing that a number of HUGE deals are being worked on. Once we get some news from FTEC and the word gets out about the company. It will take-off to MUCH HIGHER LEVELS!

I see little risk when purchasing FTEC at these DIRT-CHEAP PRICES. FTEC is making TREMENDOUS PROFITS and is trading UNDER BOOK VALUE!!!

This is the #1 INDUSTRY you can POSSIBLY be in RIGHT NOW.

There are thousands of schools nationwide who need FTEC to install security systems. You can’t find a better positioned company than FTEC!

These prices are GROUND-FLOOR! My prediction is that this will be the #1 performing stock on the NASDAQ in 2000. I am loading up with all of the shares of FTEC I possibly can before it makes a run to $20. Be sure to take the time to do your research on FTEC! You will probably never come across an opportunity this HUGE ever again in your entire life.

Source: Cybenko, et al., [2002]

A CI researcher may try to be aware of the information trajectory that a particular source may be following. Any significant deviation from expected information would suggest a hack in progress. In the Labed example, an experienced stock trader or broker would recognize the pattern of information flow as a variation of the classic "pump and dump" scam.

A CI researcher may employ the reasoning used in Ulam Games. This model assumes that some false information is provided by the information source. How much false information is included can be determined by using a set of questions and their answers obtained from the original source and compare them with the answers from the same set of questions asked of other related information sources. The inconsistencies among the answers to the same set of questions should reveal the false information.

COUNTERMEASURES: MULTIPLE SOURCES

The collaborative filtering and reliability reporting countermeasure is employed when a site keeps records of who and what they published on that site and reports the reliability of that information. It then uses those records to specify the reliability of future information provided by those with access to publishing on the site. A CI researcher may detect collusion by information sources by using linguistic analysis to determine if different information sources are being created by the same author. Another countermeasure is to use the Byzantine Generals model to determine the reliability of multiple sources. This model assumes that a message communicating system contains both reliable and unreliable processes. Given a number of processes from this system, the technique determines which processes are reliable and which processes are not by analyzing each process's responses to the same set of questions.

In general countermeasures to single source and multiple source cognitive hacking involve the detection of misinformation. Given the structure of the open sources in the surface web, the information source is both the provider and the editor of the information. As a result, the traditional controls used in the review and editorial process to verify information are lacking and it is up to the receiver of the information to verify that information. In the Internet age it is not so much as "caveat emptor" - buyer beware as it is "caveat lector" -reader beware.

NEGATIVE INFORMATION

A form of cognitive hacking is to build a Website that is a repository for negative information about a particular firm. A number of Websites contains the word “sucks” as part of the URL. For example, on August 8, 2003 a Google search by the author found 5360 URLs that contained the phrase "Microsoft sucks". The countermeasure to this type of attack is for the firm to monitor those sites that are trying to create a negative image of the firm and respond appropriately. Specifically a firm might employ an intelligent agent to monitor the Web for negative information and use text mining to determine the type of negative information so that an appropriate and effective response may be given.

UNINTENTIONAL DISCLOSURE OF SENSITIVE INFORMATION

An important concern of information security in a CI environment is unintentionally revealing sensitive information. In the course of doing business in public, a firm may reveal facts about itself that individually don’t compromise that firm but, when taken collectively, reveal information that is confidential. For an example see Hulme [2003] that describes the information collected by a hacker from open sources about a computer system prior to an attack on that computer system. Another example of unintentional disclosure of information can be in the listing of position openings on a public Website. This information may reveal details about that firm’s plans to enter a new market that ought to be held private. For an example see Krasnow [2000]. A countermeasure to these types of security breaches is for the firm to carry out a CI project against itself.

VIII. CONCLUSION

This article presents an overview of the issues associated with implementing a CI project 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 Web Spiders. Nonetheless such searches may return a large set of pages that require an automated approach, like text mining, to information analysis.

The assurance of the validity of results based on these actives is not well automated. In particular, information verification, in the form of due diligence, at this stage, requires human intervention.

To maintain information security against CI, assure the accuracy of a firm's public information, and provide countermeasures to cognitive hacking, a firm may need to monitor its "information presence" on the Web

With respect to CI, the boundaries between the phases of information gathering (information analysis; information verification and information security) are not well defined. Table 1 is a summary of how these phases and their associated problems and solutions relate to CI steps of collection and analysis.

Table 1. Summary

CI Step	Techniques	Problems	Solutions
Collection	Web Search Engines on Open Sources General Search Engine Meta-Search Engines Personalized Web Crawlers P2P Search Engines	Too Many Irrelevant Responses	Use Advanced Search Methods within Search Engines
Analysis **Summarization**	Web Mining Content On-line Focused Spiders Intelligent Agents Off-line Text Mining	Relevance of Response Validity of Technique	Research on focus of search techniques Research on analysis of unstructured data
Verification	HITS & Page Rank Techniques Due Diligence Who is author? Who maintains? How current? How reliable is the source?	Validity of Hub Cognitive Hacking Overt Covert Single Source Multiple Source Negative Information Revealing Private Information	Due Diligence None required By definition detectable Due Diligence Information Trajectory Modeling Ulam Games Assurance of Source Reliability Byzantine Generals Model Detection of Collusion by Linguistic Analysis Monitor relevant websites Run CI project against your firm

This study shows that using the Web for CI involves limitations that need to be resolved through research. Among the needed streams of research are:

1. Development of methods that improve the efficiency and accuracy of text mining for information analysis.
2. Automating the process of information verification of Web sources in general and surface Web sources in particular.
3. Develop methods for improving security including the automatic detection of false information, inaccurate information, and negative information.

Editor's Note: This article is based on a tutorial of the same title presented by the author at AMCIS 2003 in August 2003. The article was received on September 1, 2003 and was published on September 29, 2003.

REFERENCES

Aaron, R. D. and E. Naylor "Tools for Searching the 'Deep Web' ", *Competitive Intelligence Magazine*, (4:4), Online at http://www.scip.org/news/cimagazine_article.asp?id=156. (current April 18, 2003).

Boncella, R. J. (2000) "Web Security For E-Commerce" *Communications of the Association for Information Systems,* Volume (4)11, November 2.

Boncella, R. J. (2002) " Wireless Security: An Overview" *Communications of the Association for Information Systems*, Volume (9)14, October

Calishain, T. and R. Dornfest (2003) *Google Hacks: 100 Industrial-Strength Tips & Tools*, Sebastopool, CA: O'Reilly & Associates.

Chakrabarti, S. (2003) *Mining the Web: Discovering Knowledge from Hypertext Data,* San Francisco, CA: Morgan Kaufmann.

Chen, H., M.l Chau and D. Zebg, (2002) "CI Spider: A Tool for Competitive Intelligence on the Web", *Decision Support Systems*, (34)1, pp. 1-17.

Cybenko,G., A., Giani and P. Thompson. (2002) "Cognitive Hacking: A Battle for the Mind", *IEEE Computer* (35)8, August, pp. 50–56.

Dunham. M. H. (2003), *Data Mining: Introductory and Advanced Topics,* Upper Saddle River, NJ: Prentice Hall.

Fleisher, C. S. and B. E. Bensoussan, (2000) *Strategic and Competitive Analysis,* Upper Saddle River, NJ: Prentice Hall, 2003.

Fukuda, F.H., etal., (2000) "Web Text Mining using a Hybrid System", *Proceedings of the Sixth Brazilian Symposium on Neural Networks (SBRN'00)*

Fuld, L. (1995) *The New Competitor Intelligence*, New York: Wiley.

Herring, J. P. (1998) "What Is Intelligence Analysis?" *Competitive Intelligence Magazine*, (1:2), pp., 13-16. http://www.scip.org/news/cimagazine_article.asp?id=196 (current September 1, 2003)

Hulme, G. W. (2003) "Hack in Progress" , Information Week,

http://www.informationweek.com/story/showArticle.jhtml?articleID=14400070 (current September 15,2003).

Kleinberg, J. M. (1999), "Authoritative Sources in a Hyperlinked Environment", *Journal of the ACM* (46)5, pp. 604-632, September.

Krasnow, J. D. (2000), "The Competitive Intelligence and National Security Threat from Website Job Listings" http://csrc.nist.gov/nissc/2000/proceedings/papers/600.pdf. (current September 1, 2003).

Lewis, Michael (2001), *Next: The Future Just Happened*, New York: W.W. Norton & Company,

Lyman, P. and Varian, H.R. (2000) "Internet Summary" Berkeley, CA: How Much Information Project, University of California, Berkeley, http://www.sims.berkeley.edu/research/projects/how-much-info/internet.html. (current September 1, 2003).

Murray, M. and R. Narayanaswamy, (2003) "The Development of a Taxonomy of Pricing Structures to Support the Emerging E-business Model of 'Some Free, Some Fee'", *Proceedings of SAIS 2003*, pp. 51-54.

Nordstrom R. D. and R. L. Pinkerton "Taking Advantage of Internet Sources to Build a Competitive Intelligence System" *Competitive Intelligence Review*, Vol. 10(1) 54–61

Page, L., and S. Brin, "The Anatomy of a Large-Scale Hypertextual Web Search Engine", http://www-db.stanford.edu/~backrub/google.html , 1998.(current September 1, 2003).

Russell, S. and P. Norvig, (1995), *Artificial Intelligence: A Modern Approach,* Upper Saddle River, NJ: Prentice Hall.

Schneier, Bruce (2000) “Semantic Attacks: The Third Wave of Network Attacks”, *Crypto-gram Newsletter,* October 15, 2000, http://www.counterpane.com/crypto-gram-0010.html. (current September 1, 2003).

SCIP (Society of Competitive Intelligence Professionals) http://www.scip.org/. (current September 1, 2003).

Waterhouse, S. et al., (2002): "Distributed Search in P2P Networks". *IEEE Internet Computing*, 6(1), 68-72.

APPENDIX I. TYPES OF COMPETITIVE INTELLIGENCE INFORMATION AVAILABLE

Note: Some of these sources are free; others charge a fee.

http://www.scip.org - Society of Competitive Intelligence Professionals - offers assistance, articles, and advice [current September 1, 2003].

SOURCES FOR GENERAL INFORMATION

http://www.usnews.com- Weekly changes make this a very good news source [current September 1, 2003]..

http://www.wsj.com - Daily access to the leading stock market newspaper [current September 1, 2003].

http://www.tollfree.att.net - This AT&T internet directory provides a listing of 800 and 888 telephone numbers. It is possible to search using key words such as a product type [current September 1, 2003].

http://www.epa.gov - Hyperlinks to environmental financing plus speeches, reports, regulations, laws, and more [current September 1, 2003].

SOURCES WHERE YOU CAN LEARN ABOUT COMPETITORS

http://www.marketguide.com - Market Guide is a good source for financial information on 10,000 publicly traded companies [current September 1, 2003].

http://www.moodys.com - Useful to check out the credit rating of competition [current September 1, 2003].

http://www.databaseamerica.com - Provides current information on competitor’s products and strategies [current September 1, 2003].

http://www.lifequote.com - LifeQuote Check out your competition pricing. Over 250 insurance companies scanned for quotes for life insurance. Similar sites can be found for other industries [current September 1, 2003].

http://www.hispanicbusiness.com - Hispanic Small Business Magazine [current September 1, 2003].

http://www.lexisnexis.com/ - [current September 1, 2003].

http://www.dnb.com/ - Dun & Bradstreet [current September 1, 2003].

http://www.hoovers.com/ - Hoovers provides profiles on 12,000 corporate firms listed in one directory. A great deal of financial data is available [current September 1, 2003]

SOURCES WHERE YOU CAN LEARN ABOUT INDUSTRY TRENDS

http://http://www.dol.gov - Department of Labor offers a wide range of material from many sources on many industries [current September 1, 2003].

http://www.fedworld.gov - Fed World is an easy access to nearly all government information sources [current September 1, 2003].

http://www.nist.gov - National Institute of Standards and Technology (NIST) provides information on research in a wide variety of industries [current September 1, 2003].

http://www.internetnews.com - What's happening on the Net and with Net businesses [current September 1, 2003].

SOURCES WHERE YOU CAN LEARN ABOUT YOUR OWN CUSTOMERS

http://www.perseusdevelopment.com - Perseus Developers of on-line survey software [current September 1, 2003].

http://www.surveysite.com - Survey Site is another on-line business that specializes in on-line survey preparation [current September 1, 2003].

http://www.sotech.com - Socrates Software for developing your own on-line survey [current September 1, 2003].

http://www.demographics.com - American Demographics is a good business book store [current September 1, 2003].

http://www.gallup.com- Polls, polls, and more polls. Poll results and an opportunity to take a poll on-line [current September 1, 2003].

http://www.acnielsen.com - A.C. Nielsen worldwide web site [current September 1, 2003].

CHAT ROOMS AND DISCUSSION GROUPS

http://www.ListServe.com - If you already have a group of people with common interests, link them together with a listserve site. Let them know about it with some publicity and see what comes up [current September 1, 2003].

SOURCES TO HELP MAKE AN EVALUATION OF A MARKET OR OPPORTUNITY

http://www.bizweb.com - BizWeb is a comprehensive resource. It includes product, company, and industry information [current September 1, 2003].

http://www.iriinc.org - Industrial Research Institute is a good source for high-tech data research [current September 1, 2003].

http://www.uspto.gov - U.S. Patent and Trademark Office lets you search patents, obtain statistics, look at publications, and/or join a forum [current September 1, 2003].

http://www.morebusiness.com - A very good place to help you make a checklist of things to consider before entering or dropping an export market [current September 1, 2003].

http://www.yahoo.com/government/countries- A collection of governmental resources from 70 countries. Change the countries to agencies and you get a complete list of U.S. government agencies and hyperlinks to their home pages. Includes hyperlink to executive branch offices [current September 1, 2003].

http://www.ustr.gov/index.html- A collection of reports, speeches, testimony, etc., on foreign trade. A very good review of tariff policies for more than 40 nations [current September 1, 2003].

LIST OF ACRONYMS

CI	Competitive Intelligence
HITS	Hyperlink-Induced Topic Search
HTML	Hypertext Markup Language
IR	Information Retrieval
P2P	Peer to Peer
SCIP	Society of Competitive Intelligence Professional
URL	Uniform Resource Locator
XML	Extensible Markup Language

ABOUT THE AUTHOR

Robert J. Boncella (http://www.washburn.edu/cas/cis/boncella) is Professor of Computer Information Science at Washburn University, Topeka, KS. Dr. Boncella has a joint appointment in the Computer Information Sciences Department, where he conducts classes in Data Communications and Computer Networks, and in the School of Business, where he offers instruction on Computer Based Information Systems in the school's MBA program. He holds a Ph.D. and Masters degrees in Computer Science from the University of Kansas and a Master of Arts in Philosophy from The Cleveland State University. He is a member of ACM, AIS, AAAI, and IEEE. His current areas of interest are web based information systems, intelligent agents, and decision making under uncertainty, and computer security and privacy.

Communications of the Association for Information Systems

ISSN: 1529-3181