



ELSEVIER

Decision Support Systems 34 (2002) 1–17

Decision Support  
Systems

www.elsevier.com/locate/dsw

# CI Spider: a tool for competitive intelligence on the Web

Hsinchun Chen\*, Michael Chau, Daniel Zeng

*Department of Management Information Systems, Eller College of Business and Public Administration,  
University of Arizona, Tucson, AZ 85721, USA*

Accepted 1 December 2001

## Abstract

Competitive Intelligence (CI) aims to monitor a firm's external environment for information relevant to its decision-making process. As an excellent information source, the Internet provides significant opportunities for CI professionals as well as the problem of information overload. Internet search engines have been widely used to facilitate information search on the Internet. However, many problems hinder their effective use in CI research. In this paper, we introduce the Competitive Intelligence Spider, or CI Spider, designed to address some of the problems associated with using Internet search engines in the context of competitive intelligence. CI Spider performs real-time collection of Web pages from sites specified by the user and applies indexing and categorization analysis on the documents collected, thus providing the user with an up-to-date, comprehensive view of the Web sites of user interest. In this paper, we report on the design of the CI Spider system and on a user study of CI Spider, which compares CI Spider with two other alternative focused information gathering methods: Lycos search constrained by Internet domain, and manual within-site browsing and searching. Our study indicates that CI Spider has better precision and recall rate than Lycos. CI Spider also outperforms both Lycos and within-site browsing and searching with respect to ease of use. We conclude that there exists strong evidence in support of the potentially significant value of applying the CI Spider approach in CI applications. © 2002 Elsevier Science B.V. All rights reserved.

*Keywords:* Competitive intelligence; Internet searching and browsing; Internet spider; Noun phrasing; Document clustering; Experimental research

## 1. Introduction

The goal of Competitive Intelligence (CI), a sub-area of Knowledge Management, is to monitor a firm's external environment to obtain information relevant to

its decision-making process [7]. Many major companies, such as Ernst & Young and General Motors, have formal and well-organized CI units that enable managers to make informed decisions about critical business matters such as investment, marketing, and strategic planning. Traditionally, CI relied upon published company reports and other kinds of printed information. In recent years, Internet has rapidly become an extremely good source of information about the competitive environment of companies and has been reported by a Futures Group survey in 1997 to be one of the top five sources for CI professionals [6].

\* Corresponding author. Department of Management Information Systems, University of Arizona, McClelland Hall 430Z, Tucson, AZ 85721, USA. Tel.: +1-520-621-4153.

*E-mail addresses:* hchen@bpa.arizona.edu (H. Chen), mchau@bpa.arizona.edu (M. Chau), zeng@bpa.arizona.edu (D. Zeng).

Although the Internet represents significant CI opportunities, it has also brought about many technical, cognitive, and organizational challenges. Because the amount of information available on the Internet is overwhelming, CI professionals are constantly facing the problem of information overload. It is estimated that there are over 1 billion Web pages on the Internet as of February 2000 [8]. Much time and effort is required for CI professionals to search for the relevant information on the Internet and then analyze the information collected in the correct context.

Internet search engines have been useful in helping people search for information on the Internet. Nevertheless, the exponential growth of information sources on the Internet and the largely unregulated and dynamic nature of many Web sites are making it increasingly difficult to locate useful information using these search engines. It has been estimated that none of the search engines available indexes more than 16% of the total Web that could be indexed [12]. This has resulted in a low recall rate when the user is looking for obscure or unusual material. In addition, since Web page contents are extremely dynamic and may change daily or even every minute, conventional preindexed search engines suffer from the problem of providing many outdated and obsolete links.

In this paper, we present a novel approach implemented as Competitive Intelligence Spider (CI Spider) that can be used to alleviate some of the problems associated with the usual search engine approach. CI Spider accepts as input the URLs the user specifies, and follows the embedded Web links to search for user-specified keywords. After collecting on the fly a certain number (user-definable) of Web pages, CI Spider performs further text analysis to extract noun phrases from these pages. These noun phrases represent a list of key topics covered on the Web sites of interests. CI Spider also provides the functionality of visualizing the retrieved Web pages in a 2-D map where Web pages sharing similar topics are grouped together in regions. The main research hypothesis examined in this paper is that an integrated approach, such as CI Spider, can better facilitate CI professionals to analyze and summarize relevant Web sites than existing approaches using Internet search engines.

The rest of the paper is organized as follows. Section 2 briefly reviews the basic concept of Competitive Intelligence (CI) and discusses various tech-

nological supports available for CI professionals, including Internet search engine technology and related information management issues. In Section 3, we present the architectural design of the CI Spider system and give detailed technical information for the major components of CI Spider. Section 4 focuses on an evaluation methodology designed to evaluate our research hypothesis concerning the effectiveness and efficiency of an integrated approach for CI tasks. In Section 5, we report a user study performed to test our hypothesis and discuss the strength and weakness of the CI Spider system. Section 6 concludes the paper with a summary and a discussion about future research directions.

## 2. Literature review

### 2.1. Competitive intelligence

The Society of Competitive Intelligence Professionals 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” [22]. CI is different from espionage, which implies illegal means of information gathering; CI is restrained to the gathering of *public* information. Indeed, another definition of CI is “the use of public sources to develop information about the competition, competitors, and market environment” [15].

One of the main differences between CI and general business information, such as business growth rate and transaction figures, is that CI is of strategic importance on the organization. It is not only the collection of information from a variety of sources, but also the analysis and synthesis of such information, which could help the company decide the course of action to improve its position [23].

A typical CI process consists of a series of business activities that involve identifying, gathering, developing, analyzing and disseminating information [7,10,25,26]. The following list shows a typical sequence in which these activities take place.

- (1) Identify competitors, markets, customers, suppliers, or other variables in the environment to be

متن کامل مقاله

دریافت فوری ←

**ISI**Articles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات