5-2010

Exploratory Competitive Intelligence through Task Complexity Analysis

Sabah Currim
Florida State University, scurrim@fsu.edu

Limin Zhang
North Dakota State University - Main Campus, limin@email.arizona.edu

Faiz Currim
University of Iowa, faiz-currim@uiowa.edu

Exploratory Competitive Intelligence through Task Complexity Analysis

Sabah Currim
Florida State University
scurrim@fsu.edu

Limin Zhang
North Dakota State University
limin.zhang@ndsu.edu

Faiz Currim
University of Iowa
faiz-currim@uiowa.edu

ABSTRACT

As one of the major resources for competitive intelligence (CI), the Internet not only provides a large amount of public data but also exposes a variety of business relations that may not otherwise be well-known. However, finding such information can be tedious and time-consuming for end-users without proper tools or expertise. In this paper, we examine the nature of CI tasks, classify and decompose them based on task complexity theories, and propose norms for a context-based approach to retrieve CI data. Our study provides a framework to further explore the relationships among CI tasks, interactive search, and context-based search systems design.

Keywords

Web search tasks, competitive intelligence, task complexity, search context.

INTRODUCTION

The goal of Competitive Intelligence (CI) is to provide a strategic advantage through information about the external agents that influence the firm's success. CI information is critical to a firm in composing a business strategy and using the strategy to succeed in today's competitive world (Bao et al. 2008). In the traditional CI process, end users (e.g., managers) communicate their requirements to trained CI professionals who interpret and perform searches for the users. The information gathering process is typically handed off to professionals because it is tedious, time-consuming and difficult (McGonagle et al. 1999). In recent years, a vast amount of public data has become available on the World Wide Web, which allows companies with limited resources (or end-users in large companies, who wish to find answers quickly) an opportunity to acquire some CI information directly from online sources. However, the lack of familiarity with appropriate search techniques and the complexity of CI tasks have prevented end-users from conducting effective CI search on the Web. In this study, we analyze certain complexity features of CI tasks and relate these features to different strategies people may use in their search process. A taxonomy of CI tasks is constructed and the role of the taxonomy in developing a context-based CI search system is discussed.

BACKGROUND

A key barrier that end-users face while gathering CI data is the complexity of the task itself. Consider the following example: "What were the sales of my competitors?" This task could involve: (1) finding out major trade associations and major licensing boards for the field; (2) searching for competitors in sources found in the previous step, as well as in other sources such as Hoover's, local yellow pages, and the Secretary of State Office's website; and (3) looking at the Security and Exchange Commission's (SEC) website if the competitor is a publicly traded firm or searching for sales information on the competitor's website and government websites if the company is private. The task complexity of this question illustrates why general search products like Google™ and Yahoo™ are limited in what they can do (Mann 2007). Once we know how to break up the question and where to look, the answers are more straightforward to assemble. However, while CI experts are capable of easily performing the decomposition, novices (business end-users) need guidance to help them through the process.

Past research tells us that *task structure* and the *strategy used for representing the task* are two important elements in web-based information seeking (Browne et al. 2007). *Task structure* refers to the degree to which the inputs, problem-solving operations, and outputs are known and recognizable to the decision maker (Byström et al. 1995; Vakkari 1999). A task can be either well-structured or poorly structured. The second element influencing complexity is the *strategy used for representing the task*, which can be either *decompositional* or *holistic* (Simon 1996). For example, "What were the sales of my competitors?" appears to be a simple and holistic task to novices, but it is a decompositional query to experts because it can be broken into multiple sub-tasks. The choice of strategy is usually influenced by the complexity of a task and the searcher's prior experience with the task. Tasks with lower complexity and where the user is experienced tend to be approached using a decomposition strategy. Tasks with high complexity where the user has limited experience tend to be approached using a holistic strategy (Browne et al. 2007).

This study is based on the assumption that novice CI searchers often use a holistic rather than a decompositional search strategy due to the fact that most CI tasks are relatively complex and novice searchers have limited experiences with the tasks and the search domain. This assumption implies that a search system targeted towards inexperienced CI end-users needs to (1) support the analysis and decomposition of complex CI tasks; (2) provide critical domain knowledge when such knowledge is needed; and (3) formulate queries that best represent users' search tasks.

THE ANALYSIS OF CI TASKS

Complexity of CI Tasks

The nature of tasks has been extensively studied in social science, psychology, and information science. A number of classification schemes have been proposed to categorize and characterize tasks. Some of these classification schemes are applicable to general tasks (Algon 1997; Byström et al. 1995; Campbell 1988; Xie 1998), while some are specific to information search (Kellar et al. 2007; Kim 2006; Marchionini 1989). Among the well-known works on task classification is Campbell's typology of complex tasks (Campbell 1988), which allows complexity to be defined objectively and independently of the person performing the task. This typology captures the complexity of a task in terms of its information load and rate of information change. We chose Campbell's typology of task complexity to analyze and categorize CI tasks because it is objective, detailed, and useful in decomposing tasks and reducing task complexity. Moreover, this typology has been used in the MIS field for analyzing online tasks (Browne et al. 2007), and its objective nature makes it feasible to be operationalized and implemented into a computer system to aid novices accomplish the task more effectively.

Campbell's typology presents four factors that influence task complexity, namely *uncertainty between potential paths and potential end states*, *multiple desired end states*, *multiple paths to a desired end state*, and *conflicting interdependence among paths*. We now briefly describe these four factors. (1) *Multiple paths to a desired end state*: As the number of possible ways to reach a desired outcome increases, information overload, and hence complexity increases. For example, if Google returns 3 million results for a query, and pieces of the answer are in 50 different pages (but few of those 50 are on the first page of results returned), the user experiences information overload. (2) *Multiple desired end states:* As the number of desired end states increase, complexity increases. A simple CI user query like 'Who are my competitors?' has a different set of answers depending on the user's company and which product / group of products the user is looking at. For example, the list of Microsoft's competitors varies based on what product or group of products under consideration; Word, Office Suite Enterprise, Windows, Zune have different set of competitors. It is therefore necessary for the system to understand the exact context of the user query before the system processes the query. (3) *Uncertainty between potential paths and potential end states:* As the uncertainty of the connection between the path and end state increases, the complexity increases. In the scope of CI queries that our study currently focuses on, this factor of complexity is mainly caused by the fact that the user is a novice CI searcher and does not know how to break the original task into individual pieces and how to structure the query for CI search on the Internet. (4) *Conflicting interdependence among paths*: When there are negative relationships among outcomes, i.e., achieving one goal conflicts with reaching another, complexity increases. In Web search, increasing recall in search engines often affects precision of the results. Another example of this factor is the domain-dependent nature of some CI queries (different approaches need to be taken to find relevant results in each domain).

Based on Campbell's complexity theory, we conducted a case study on a real client, MeGa Home & Wedding, a Denver-based startup company specializing in floral decoration and home accessories. MeGa Home & Wedding, like most startups, faces competition from both local business and national chains. The owner of MeGa wanted to improve her knowledge of the existing market to gain a competitive edge in the challenging economic times. The company agreed to let us help them do some research about their competitors. More specifically, they wanted to gather basic CI information such as the competitors' names, their products, clients, and partners. The company would then use such information to make critical decisions such as setting product prices, expanding customer base, and adjusting marketing strategies. Next, we analyze common CI search tasks based on the aforementioned four factors of Campbell's task complexity model, using MeGa as an example.

Uncertainty between potential paths and potential end states: The owner of MeGa wants to keep track of her competitors' new products and prices of current products. The problem associated with this task is the presence of *uncertainty between* *potential paths and end states*: this task cannot be answered by a single search because to answer this query, we need to first find out who the competitors are. Then based on the names of the competitors, we can search for products of the competitors and current prices of their products. To find MeGa's competitors, we submitted the query "competitors of Mega Home & Wedding" to Google, Yahoo, and Turbo10[1]. We chose these search engines for our test queries because Google and Yahoo are the most popular search engines by the number of searches during August 2009[2], and Turbo10 is a metasearch engine that searches the invisible Web. However, none of the results returned in the top-10 list of these search engines provided any clue on who are the company's competitors. The problem is caused by the fact that MeGa is a new start-up and it does not have much Web presence. To solve this problem, the company name could be replaced by the industry name.

The complexity of this task can be further reduced by capturing how prices of the competitors' products change over time. Capturing changes over time allows sophisticated monitoring and quick updating of competitor information. The owner may notice that a particular competitor adjusts her prices every Tuesday and that she has recently doubled the price for red rose stems. Such information changes the uncertain path to a more straightforward one. As a consequence, the owner now has an edge about pricing and will be able to adjust her offerings dynamically.

The complexity of uncertainty between potential paths and potential end states is caused by the entangled structure and dynamic characteristics of CI tasks and can be reduced by breaking the tasks into multiple sub-tasks, as most CI tasks are well-structured, and the structure is obvious to experts, but not to novices. Therefore, a task taxonomy such as the one described in Section 3.2 can be built into the search system to help search novices perform task decomposition

Multiple desired end states: Replacing the company name with industry name in the query "competitors of Mega Home & Wedding" will help reduce the uncertainty between potential paths and potential end states. However, the company is within multiple industries: silk flowers, home decoration and wedding accessories, which results in different sets of competitors (*multiple desired end states*), depending on the context. Therefore, the user needs to clarify which industry she is interested in finding competitors in. Moreover, there is an added dimension of complexity if the answer is dependent on context (Campbell 1988). For example, the user also needs to specify the spatial granularity for her query: the list of competitors will vary depending on whether she is interested in local, national or international competitors. After the context information is gathered, the original query is transformed to "competitors Denver silk flowers" (given the user chose "local" as the geographic preference and "silk flowers" as the industry of interest.) The new query has reduced complexity in (1) uncertainty between potential paths and end states dimension and (2) multiple desired end states dimension. It has yielded improved search results from Google and Yahoo which revealed six and two competitors in the silk flowers business in Denver, respectively.

The complexity of multiple desired end states is closely related to the context-sensitive nature of search queries. CI queries are especially context-dependent compared to general web queries because the user is always interested in CI information relative to her own organizational context. To reduce this dimension of complexity, a search system may use a combination of an ontology, a lexicon, CI domain knowledge, and user profiles to identify ambiguous words as well as context-sensitive words. The user is then presented with a number of options (such as spatial granularities and industries) to help disambiguate context and reduce the number of end states.

Multiple paths to a desired end state: Although the search results have been greatly improved, the query "competitors Denver silk flowers" still yielded a number of irrelevant results. For example, a blog on the 2009 Denver Home and Garden Show was included because the source said: "Artificial flowers are NOT permitted." This problem indicates the presence of *multiple paths to a desired end state*, as the search engine gives users a false sense of coverage when it misses relevant sources and buries good sources in a pile of irrelevant results (Mann 2007). An increase in the number of possible ways to arrive at an outcome increases information load, and hence increases complexity. Novice searchers get frustrated when the results do not match their question. To reduce the number of paths, particularly the number of non-quality paths to a desired end state, a search system specialized in retrieving CI information should have the capability of identifying the complexity dimension in queries and narrowing down the search space by redirecting the queries to appropriate sources. For example, the query "competitors Denver silk flowers" may be redirected to Denver's Yellow Pages in order to achieve higher recall.

Conflicting interdependence among paths: If there are negative relationships between outcomes, i.e., achieving one goal conflicts with reaching another, complexity increases. In Web search, increasing recall in search engines often affects precision of the results. For example, to locate all the clients of 1800flowers[3], we would like to collect as much relevant

[1] The searches reported in the paper were performed on February 23, 2010 and April 22, 2010.

[2] According to a report by SearchEngineWatch.com: http://searchenginewatch.com/3634991.

[3] 1800flowers is a large online vendor that also has a silk flower business.

results as possible using queries such as "clients of 1800flowers," "1800flowers clients" and "1800flowers report customers." However, these queries were not very helpful. Searching Google, Yahoo and Turbo10 using these queries yielded results that consists most of the query terms, but none of the results provided information relevant to the search task. This is a common situation CI researchers encounter during their tasks as certain information is not widely publicized by enterprises and may only be revealed through webpage hyperlinks, specialized web services, new portal, business forums, etc. (Kassler 1999). However, by removing the dependence on a specific competitor and generalizing the search to the industry, we may be able to provide useful information to the user. For example, the query "silk flowers industry report customers" improves the search by including links to industry reports such as Hoovers and IBISWorld.

In addition, this particular query "clients of [a company]" is domain-dependent. We found that the answers were easy to find for some companies (such as IBM and Oracle) than for some other companies. We hypothesize that the availability of clients for specific competitors depends on the barriers to entry. When the barriers are high, companies feel comfortable publishing their client list for marketing purposes, but when the barriers to entry are low, then such information is hard to find. A CI search system may use a profitability indicator such as *benefit-cost ratio* (also called *profitability index*) as an estimate to entry barriers and search results. The higher the benefit-cost ratio is, the higher is the entry barriers, and there is higher possibility that the query will return good results. When the profitability and entry barriers are low, there is a greater chance that the query will not produce good results. In this case, the query may be reformulated to yield better results. An area of future research would be to identify domain-dependent CI tasks and use more sophisticated mechanism for the prediction of search results.

A Task Taxonomy

To operationalize the task decomposition approach that is proposed in the previous section to reduce the complexity caused by uncertainty between potential paths and potential end states, we analyzed approximately 50 search tasks described in the CI literature (including work by Prescott (Prescott et al. 2001) and Rugge and Glossbrenner (Rugge et al. 1995)) and decomposed them into sub-tasks. A task taxonomy is constructed to capture unique characteristics of each task and the relations between different tasks. The task taxonomy is formally represented as a graph in which tasks are mapped into graph vertices and relations between the tasks are mapped into graph edges. Currently we have defined three types of relations between a pair of tasks: hierarchical relation, sequential relation, and associative relation. If task t_j is a subtask of task t_i, then t_i and t_j is connected through a hierarchical relation, expressed as $hr(t_i, t_j)$. If the execution of t_j requires the answer to t_i, that is, t_j needs to be executed after t_i, then t_i and t_j is connected through a sequential relation, expressed as $sq(t_i, t_j)$. For relations other than hierarchical and sequential relations, we link the two tasks through an associative relation, expressed as $ac(t_i, t_j)$. In the task taxonomy, each relation is labeled as one of the three types of relations. Figure 1 shows part of the graph with four tasks. Task t_1 "What are the prices of competing products?" can be decomposed into two sub-tasks: t_2 "Who are our competitors?" and t_3 "What products do our competitors have?" In addition, task t_4 "What are recent products introduced by our competitors?" is related to t_3. Both t_3 and t_4 require the answers to t_2 in order to be executed.

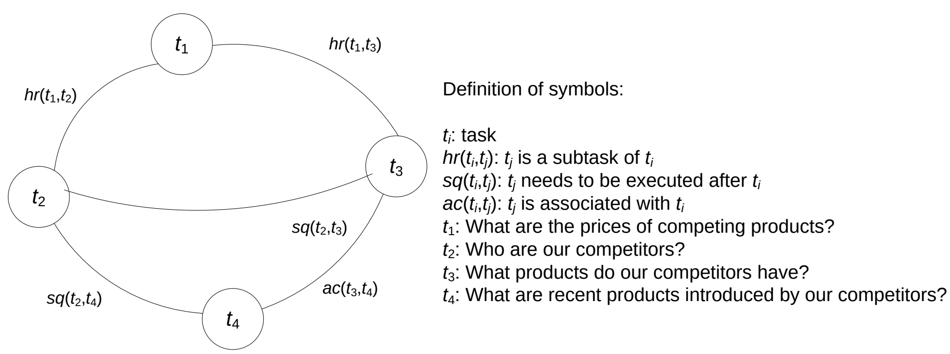

Figure 1: A Partial Graph of the Task Taxonomy

METHODOLOGY

As our work falls within the realm of design science research (DSR), we briefly summarize what we have done to be consistent with the guidelines for such research (Hevner et al. 2004). The *design artifacts* in our work comprise the taxonomy of CI task decomposition (based on an application of the Campbell's typology) and the CI search system model which incorporated the framework into its architecture. *Problem relevance* is important in DSR. Our view is that CI can provide small and medium businesses with information to make strategic decisions. However, the task complexity and lack of familiarity with searching for such information serves as a barrier to non-professionals accessing needed CI. Most existing tools cater to CI professionals rather than novices. Our *design evaluation* was presented using a case study done with MeGa Home & Wedding. The key *research contributions* of our work are an understanding why searching for relevant CI is complex, and what needs to be done to simplify the procedure based upon Campbell's typology of task complexity. This led to our development of a taxonomy of CI tasks with an associated task decomposition (which can be incorporated into a software system, such as the prototype we developed). *Research rigor* in our work involved choosing relevant and well-accepted theory to aid in understanding and modeling CI tasks. We also took care in designing the case study and evaluating the proposed CI task taxonomy. As is well-recognized, *design is a search process*. In our case, the design process began with gathering requirements from novice users to understand which search tasks proved challenging. Experts were consulted to both establish a baseline of good results as well as understand how skilled users approached the search process. Recognition of task complexity for novice users led to our development of the system architecture to incorporate the CI task and sub-task decomposition[1]. Further, we integrated user profiles for search personalization and inclusion of important terms, and employed a CI-relevant ontology based on WordNet and ResearchCYC (Fellbaum 1998; Matuszek et al. 2006). Finally, in terms of *communication of research*, we include in this paper a description of both the business relevance and theoretical underpinnings of our approach. In addition to communication of results to the research community, we provided feedback to business users as well.

CONCLUSIONS AND ON-GOING WORK

Having the ability to obtain answers to common CI questions is valuable to end-users. However, to do so effectively, we need a tool that will assist novice searchers with context-dependent search. This area has not been sufficiently explored in the CI literature. In this paper, we describe how task complexity literature can provide us with valuable insights on how to assist CI end-users with their Web search tasks. More specifically, we adopt Campbell's task complexity theory to analyze and decompose CI search tasks based on their complexity levels. We also present a task taxonomy that consists of common CI search tasks, their subtasks, and relations between different tasks and subtasks. To demonstrate the feasibility and practicality of the proposed approach, we have developed a prototype system that performs task analysis and decomposition for CI end-users. The system is integrated with the task taxonomy, a lexicon, a knowledge base, and user profiles.

Our ongoing and future work is focused on (1) expanding and validating the task taxonomy by consulting CI professionals through interviews and Delphi survey; (2) considering the dimension of cognitive complexity (in addition to task complexity) to improve the classification of CI search tasks and the performance of the associated tools, and (3) designing the system so that it will adapt to users' evolving search skills and experience.

REFERENCES

1. Algon, J. (1997) Classifications of tasks, steps, and information-related behaviors of individuals on project teams, P. Vakkari, R. Savolainen and B. Dervin (eds.) *International Conference on Research in Information Needs, Seeking and Use in Different Contexts*, Taylor Graham, Los Angeles, CA.
2. Bao, S., Li, R., Yu, Y., and Cao, Y. (2008) Competitor Mining with the Web, *IEEE Transactions on Knowledge and Data Engineering*, 20, 10, pp 1297-1310.
3. Browne, G.J., Pitts, M.G., and Wetherbe, J.C. (2007) Cognitive stopping rules for terminating information search in online tasks, *MIS Quarterly*, 31, 1, pp 89-104.
4. Byström, K., and Järvelin, K. (1995) Task Complexity Affects Information Seeking and Use *Information Processing & Management*, 31, 2, pp 191-213.
5. Campbell, D.J. (1988) Task Complexity: A Review and Analysis, *Academy of Management*, 13, 1, pp 40-52.
6. Fellbaum, C. (1998) WordNet: An Electronic Lexical Database (Language, Speech, and Communication) The MIT Press, p. 423.
7. Hevner, A.R., March, S.T., Park, J., and Ram, S. (2004) Design science in information systems research, *MIS Quarterly*, 28, 1, pp 75-105.
8. Kassler, H. (1999) Competitive intelligence on the Web, in: *EContent*, pp. 16-24.
9. Kellar, M., Watters, C., and Shepherd, M. (2007) A field study characterizing Web-based information-seeking tasks, *Journal of the American Society for Information Science and Technology*, 58, pp 999-1018.
10. Kim, J. (2006) Task as a predictable indicator of information seeking behavior on the Web, Rutgers University, Unpublished dissertation.

[1] Due to space limit, the implementation of the prototype system is reported in a separate paper.

11. Mann, T. (2007) The Peloponnesian War and the Future of Reference, Cataloging, and Scholarship in Research Libraries.
12. Marchionini, G. (1989) Information seeking strategies of novices using a full-text electronic encyclopedia, *Journal of the American Society for Information Science*, 40, 1, pp 54-66.
13. Matuszek, C., Cabral, J., Witbrock, M., and DeOliveira, J. (2006) An Introduction to the Syntax and Content of Cyc *AAAI Spring Symposium*.
14. McGonagle, J.J., and Vella, C.M. (1999) The Internet age of competitive intelligence Quorum Books, Westport, CT.
15. Prescott, J.E., Miller, S.H., and Society of Competitive Intelligence Professionals (2001) Proven strategies in competitive intelligence : lessons from the trenches Wiley, New York.
16. Rugge, S., and Glossbrenner, A. (1995) The information broker's handbook McGraw-Hill, New York.
17. Simon, H.A. (1996) The Sciences of the Artificial, (Third Edition ed.) MIT Press, Cambridge, MA.
18. Vakkari, P. (1999) Task complexity, problem structure and information actions - Integrating studies on information seeking and retrieval, *Information Processing & Management*, 35, 6, pp 819-837.
19. Xie, H. (1998) Planned and situated aspects in interacting IR: Patterns of user interactions and information seeking strategies, Unpublished dissertation.

APPENDIX 1: MEETING DESIGN SCIENCE GUIDELINES

This appendix summarizes what we have done to be consistent with the guidelines for design science research.

Guideline	Research Component
Design as an Artifact	A taxonomy of CI task decomposition (based on an application of the Campbell's typology) which was incorporated into the CI search system architecture
Problem Relevance	Competitive intelligence (CI) can provide small and medium businesses with information to make strategic decisions. The task complexity and lack of familiarity with searching for such information serves as a barrier to non-professionals accessing needed CI. Most existing tools cater to CI professionals rather than novices.
Design Evaluation	Case study with MeGa Home & Wedding
Research Contributions	An understanding why searching for relevant CI is complex, and what needs to be done to simplify the search process using Campbell's typology of task complexity. Development of a taxonomy of CI tasks with an associated task decomposition that can be incorporated into a software system.
Research Rigor	Choosing relevant and well-accepted theory to aid in understanding and modeling the CI task process. Careful design of case study and evaluation of the proposed CI task taxonomy.
Design as a Search Process	The design process began with gathering requirements from novice users to understand which search tasks proved challenging. Experts were consulted to both establish a baseline of good results as well as understand how skilled users approached the search process. Recognition of task complexity for novice users led to a development of the system architecture to incorporate the CI task and sub-task decomposition. Further integration of user profiles for search personalization and inclusion of important terms, and usage of a CI-relevant ontology based on WordNet and ResearchCYC.
Communication of Research	Description of both business relevance and theoretical underpinnings of our approach. Communication of results to both novice business users and the research community.