Improving competitive intelligence for knowledge management systems

Kevin R. Parker*

Department of Computer Information Systems
College of Business, Idaho State University
Campus Box 8020, Pocatello, Idaho 83209, USA
Fax: +1–208–282–4367 E-mail: parkerkr@isu.edu
*Corresponding author

Philip S. Nitse

Department of Marketing, College of Business, Idaho State University
Campus Box 8020, Pocatello, Idaho 83209, USA
Fax: +1–208–282–4367 E-mail: nitsphil@isu.edu

Abstract: Knowledge management encompasses the entire intelligence cycle from planning to reporting. One aspect that is often overlooked or minimised is the inclusion of competitive intelligence. This paper proposes a new concept to fine-tune the process of electronically gathering competitive intelligence – a key activity in knowledge management systems. Most tools are nondiscriminatory in information gathering, and a structured approach is needed to assist managers at all organisational levels in the needs identification process. The proposed multiclass interest profile provides the capability of expanding the coverage of critical intelligence areas to reflect the assorted topics that make up an organisation's information needs. Each component is customisable to make the information that is gathered pertinent to the organisation, and supporting features such as profile expansion and fine-tuning are also incorporated.

Keywords: business intelligence; competitive intelligence; knowledge management; information gathering; information needs; critical intelligence needs; key intelligence topics.

Reference to this paper should be made as follows: Parker, K.R. and Nitse, P.S. (2005) 'Improving competitive intelligence for knowledge management systems', *Int. J. Internet and Enterprise Management*, Vol. 3, No. 1, pp.24–45.

Biographical notes: Kevin R. Parker is Professor of CIS at the Idaho State University. He has taught classes in both computer science and information systems over the course of his 13 years in academia. Dr. Parker's research interests include e-commerce marketing, competitive intelligence, information filtering, knowledge management, and extreme programming. He has published several papers in these areas including publications in the *Journal of Information Systems Education, Journal of Issues in Informing Science and Information Technology, European Journal of Marketing*, and *Marketing Intelligence and Planning*. Dr. Parker's teaching interests include web development technologies, programming languages, data structures, and database management systems. Dr. Parker holds an MS in Computer Science and a PhD in Management Information Systems from Texas Tech University.

Philip S. Nitse is a Professor and Chair of Marketing at the Idaho State University. His areas of research interest include competitive intelligence, knowledge management, healthcare marketing, and marketing management. He has been published in the *European Journal of Marketing*, *Competitive Intelligence Review, Marketing Intelligence and Planning, Journal of Health Care Marketing, Health Marketing Quarterly, Journal of Direct Marketing*, and *Advances in Marketing*. He has a BS in Marketing from Arizona State University, an MBA from Memphis State University, and a PhD in Marketing from The University of Memphis. In addition, he has over 18 years of sales and sales management experience with organisations such as Carrier Air Conditioning, Georgia Pacific, Mass Merchandisers, and VR Business Brokers.

An earlier version of this paper was presented to the 2001 International Symposium on Information Systems and Engineering – ISE'2001-Workshop: Knowledge Management Systems: Concepts, Technologies and Applications, Las Vegas, Nevada, 25–28 June 2001.

1 Introduction

> "The most meaningful way to differentiate your company from your competition, the best way to put distance between yourself and the crowd, is to do an outstanding job with information. How you gather, manage and use information will determine whether you win or lose." (Gates, 1999)

As organisations compete more aggressively for market share, managers are finding that they need better methods to gather and organise competitive intelligence. Businesses are increasingly recognising the need to gather and analyse information in order to compete effectively.

Both Knowledge Management (KM) and Competitive Intelligence (CI) systems are designed to enhance the information resources of an enterprise, but often target different information types and sources. Williams (2002,p.1) defines KM as:

> "The conscious strategy of putting both tacit and explicit knowledge into action by creating context, infrastructure, and learning cycles that enable people to find and use the internal collective knowledge of the enterprise."

CI, however, is concerned with gathering external, marketplace data on key competitors to enable the company to gain competitive advantage (Williams, 2002). It is important to note, however, that the two approaches complement each other, and are really two parts of the same whole (Johnson, 1999). This paper presents an approach for improving the CI-gathering aspect of KM systems.

There are many tools available for gathering information from an organisation's environment. Information gathering approaches can be found under a variety of headings, including CI, business intelligence, knowledge acquisition, knowledge discovery, knowledge harvesting, enumerative description, knowledge engineering, information retrieval, document management, and enterprise information portals, among others. However, unless those tools are provided with an adequate specification of the variables that need to be monitored, an organisation's information gathering will only be partially successful. A great deal of research has been devoted to studying *how* to locate information, whilst overlooking the equally vital issue of *what* information to locate. A recent review of software marketed toward the online intelligence community clearly illustrates that the ability of most software to determine *what* information to gather is clearly deficient (Fuld, 2001). This paper presents a structured approach for determining what information to gather by proposing a means for organisations to create a profile of critical intelligence interests that more completely expresses their information needs, thus leading to a more effective and efficient collection of external information.

2 Background

> "The modern entrepreneur does not scan all alternatives nor does he have all information about all alternatives. He invests in information only so long as the expected marginal return from the information gained exceeds the expected marginal costs." (Cyert and March, 1963,p.45)

Organisational theory suggests that managers use the information they have acquired in three strategic ways, first to make sense of an equivocal environment (Weick, 1979; 1995; March and Olsen, 1976); second to create knowledge for organisational learning (Nonaka and Takeuchi, 1995; Senge, 1990); and third to make decisions (Cyert and March, 1963; March and Simon, 1993). Strategic management's search for organisational intelligence is an attempt to make actions lead to outcomes that are consistent with their desires or their conceptions of reality. This objective is both ambiguously defined and imperfectly achieved (Leventhal and March, 1993).

Organisations learn from experience that carries over from one operating unit to another and from one activity to another (Yelle, 1979; Argote *et al.*, 1990; Udayagiri and Balakrishnan, 1993). Such learning is important since it can form the basis of an organisation's competitive advantage (Leventhal and March, 1993). However, Levinthal and March (1993) point out that experience is often a poor teacher, considering the changing nature of the world and the competitive environment. Considering all of the environmental forces that impact organisations, as well as the vast amount of information needed to make rational business decisions, any mechanism that provides timely, precise, and concise information is welcome. Organisations face constant changes in their environment, which consists of competition, technology, regulatory agencies, customers, suppliers, and employees. Organisations must communicate information and manage that information effectively and efficiently (Vibert, 2004). This has never been truer than in the electronic environment in which today's companies operate, where even a simple search can return hundreds, if not thousands, of pages of information.

Rational choice theories indicate that managers and analysts gather and interpret data to reduce uncertainty about tactical and strategic decisions (Cyert and March, 1963; Wilensky, 1967; Larson, 1977; Guy, 1985). The data collection does not need to be comprehensive, but complete enough to reduce the decision makers' uncertainty (March and Simon, 1958). Managers periodically scan local and global news reports for pertinent events in their environments, and gather industry data on their competitors.

Search behaviour, as defined by the 'classical' model, is difficult to apply to actual search behaviour exhibited by companies (Simon, 1957). Levinthal and March (1981) distinguish between search stimulated by failure and search stimulated by slack. Search stimulated by failure is more narrow and focused, and search encouraged by slack is less controlled, broader, and has a wider variance of outcomes. Cyert and March (1963) refer to search stimulated by failure as 'problemistic' search that is initially directed towards the local environment and continues until an acceptable solution is identified. Problemistic search is very distinct because of the specific nature of the problem driving the search. Search processes have associated costs, and the broader searches have higher costs associated with them.

When companies search for solutions to specific problems, the search becomes a uni-dimensional search that covers the spectrum from exploitation to exploration (March, 1991). The search process is stimulated by the identification of a problem and ceases when a solution is recognised, although the solution may be biased toward the receiver's experience, training, and goals (Cyert and March, 1963). Cyert and March (1963) claim that search efficiency is more likely linked to personal experience, motivation and awareness than to some organisational structure.

Search behaviour has been of great interest to researchers in more recent years as well. Improving the search process to enhance information gathering is critical in several areas, most notably competitive intelligence. Many tools for gathering intelligence are profile-based, designed to sift information through a profile of intelligence needs (Berghel, 1997). These profiles are often made up of a set of topics that describe specific interests (Foltz and Dumais, 1992), and are developed early in the CI cycle and modified throughout the course of the intelligence process. Each topic can be expressed in terms of a keyword or concept. The primary weakness of this type of approach is its reliance on the completeness and accuracy of a one-dimensional or single-class profile. If the profile is insufficient in any way, the effectiveness of the filtering process is seriously diminished. For example, if the profile is too narrow in scope or omits critical intelligence topics, the competitive intelligence process will overlook much of the pertinent information that is available, leaving managers unaware of vital facts. Thus, the organisation may consistently make crucial decisions based on faulty information. If, on the other hand, the profile is too broad or general, the intelligence gathering process may be capturing irrelevant information, overwhelming the decision makers and convincing them that the CI process is ineffective. In short, the profile of information needs is the pivotal element in determining how well the CI process performs.

Needs identification requires a structured approach that takes into account multiple dimensions, or classes. The Multiclass Interest Profile (M-CLIP) described by this paper is such an approach. It helps to ensure that the process of identifying an organisation's intelligence needs considers each of the categories that make up those needs. Stadnyk and Kass (1992) proposed the development of knowledge bases of description categories over which individual models of interests can be defined. Herring (1999) proposed the concept of Key Intelligence Topics (KITs) to help identify intelligence requirements by carefully considering strategic decisions, early-warning topics, and key players. The M-CLIP is similar to these approaches, and provides a framework based on the various types of information needs in order to better specify what information should be gathered by ensuring that key items within each critical intelligence area are accounted for. By providing a structured, multidimensional framework, the M-CLIP increases the likelihood of a successful CI effort.

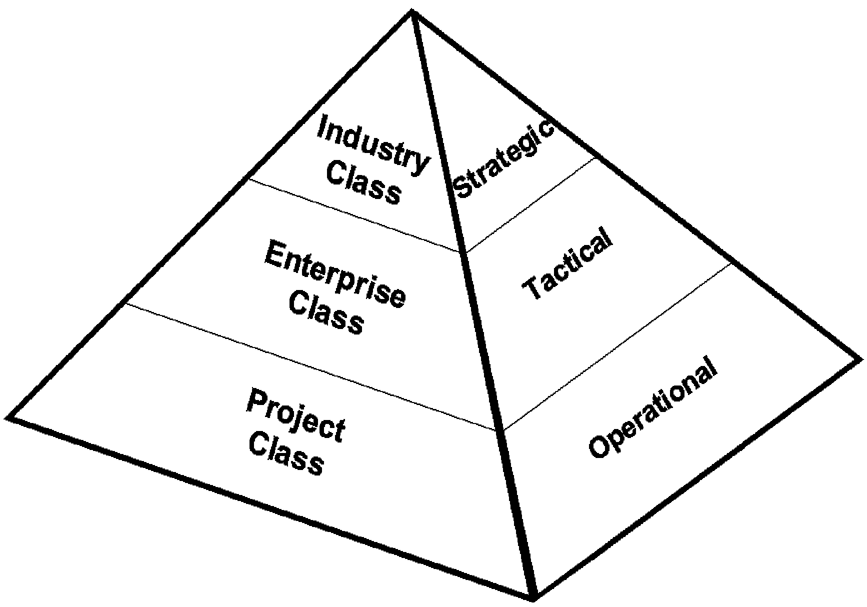
3 Design

Herring (1999) first proposed the concept of Key Intelligence Topics (KITs) on the basis of his prior work with both the government and Motorola. The KITs process helps management to identify and define critical intelligence needs. CI programmes often operate under the direction of upper management, who generally delineate the objective or need that CI must attempt to fulfil. However, CI activity should not be restricted to upper management because it can assist all organisational levels. Further, CI needs vary by company and by project. Therefore, an analysis of the information needs of an enterprise requires consideration of the types of information required by decision makers at all levels of management. Many management models, including Anthony's Managerial Pyramid (Anthony, 1965), represent organisations as having various levels of decision making: operational control, tactical control, and strategic planning, each of which has different information needs. The three main functional categories of intelligence needs proposed by Herring (1999) – Strategic Decisions and Actions, Early-Warning Topics, and Descriptions of the Key Players – can be present in each of Anthony's organisational levels. The M-CLIP, first proposed in 2001 (Parker and Nitse, 2001) and since refined, approaches the intelligence operation from a strategically aligned perspective.

The proposed design is based on the fact that the information needs of any corporation span several areas. The primary components that make up the M-CLIP were derived by taking into consideration such information-intensive activities as project management, strategic planning, competitive analysis, and environmental analysis, and then recognising a correlation between the information needs of those activities and the decision-making levels described in the Managerial Pyramid.

Figure 1 shows the components that comprise the M-CLIP, labelled along the front, and indicates the correspondence to the decision-making levels of the Managerial Pyramid, which are labelled along the right side.

Figure 1 Correlation between the M-CLIP components and Anthony's Managerial Pyramid



3.1 Project class

A project class consists of interest areas intended to gather intelligence that may affect current projects. This includes both long-term activities, such as tracking the daily or weekly actions of an overseas competitor, as well as shorter-term specialised projects such as the investigation of a possible acquisition or alliance prospect. A variety of internal and external factors, and the awareness of these factors, can influence the success or failure of a project. Studies indicate that only around 20% of the projects undertaken are successfully completed due to poor management, technical failure, and legislative or regulatory changes (Beidleman *et al.*, 1990).

The project component ensures that the CI process will gather information relevant to ongoing projects. Information from the external environment regarding project attributes such as international market considerations, product differentiation, economic environment, and material costs are vital to project control. Considerations such as technical requirements, manpower availability, and less-costly resources from the international marketplace must also be monitored. Increased access to pertinent information on a timely basis can help project managers anticipate problems and act accordingly, thereby modifying project objectives and planning assumptions on an ongoing basis. By being made aware of pertinent information as soon as it becomes available, project administrators can better identify and manage risks. Table 1 shows the interest areas that make up the project class. The interest areas were determined by analysing several project management studies. The M-CLIP is designed so that additional interest areas can be added to each class in the event that an organisation has specialised information needs.

Table 1 Interest categories that make up the project class

Interest area	*Description*	*Example*
Project goals or objectives	Overall focus and goals of the project includes such factors as scope, time and cost parameters, environmental, technological, and operational constraints, milestones, and control considerations	• strengthen market share • upgrade infrastructure
Project basis or background	Premise or background information on which the project is based	• availability of faster processors • improved bandwidth
Technical requirements	The technical requirements associated with the project	• broadband • high speed networking hub • GigaPop • Bluetooth
Resource requirements	The number and skill level of personnel involved in the project or the project results; the cost of materials required as well as their availability	• price of P4 • release date of Dothan • Xeon servers • Opteron

Table 1 Interest categories that make up the project class (continued)

Interest area	*Description*	*Example*
Market considerations	Progress of competition, product differentiation, and economic environment	• Internet cafes in California, Southwest and Western Canada
Project management priorities	Includes scheduling, cost control, quality control, scope management, contract management, resource management, and interface management	• access speed • faster PCs • wireless
Financial requirements	Factors that influence project financing or costs	
Return-on-investment factors	Factors that could influence the projected return-on-investment	• increased patronage
User-added interest areas	Project-specific interests that do not fit neatly into other categories	

3.2 Enterprise class

An enterprise class includes interest areas such as technological factors, investment issues, corporate news, operating expenses, *etc.*, that are necessary for tactical decision making. This includes both internal and external information at both the tactical and strategic levels that may have not been previously shared with all appropriate divisions of the organisation. For example, a regional manager in Miami may be monitoring the actions of a new competitor, unaware that the Munich office has maintained a complete file on the company since they began doing business in Germany several years earlier under a different name. Scanning internal resources, in conjunction with KM tools, helps reveal the availability of such information. This would allow the Miami office to be proactive rather than reactive to the actions of the competition. Table 2 shows the interest areas that comprise the enterprise class, including both internal and external factors. These interest areas were derived from studies in occupational research and environmental scanning literature. Again, additional interest areas can be added if further personalisation is necessary.

Table 2 Interest categories that make up the enterprise class

Interest area	*Description*	*Example*
Technological interests	Technological or scientific areas that affect the organisation	• Microsoft Exchange mail server security • Microsoft Longhorn • wireless networks • Bluetooth
Financial/investments	Stocks or other portfolio items of interest	• decrease in tech stocks • dot com failures • decrease in prime lending rate

Table 2 Interest categories that make up the enterprise class (continued)

Interest area	*Description*	*Example*
Legal issues	Legal issues that impact the organisation	• RIAA
Corporate news	News that affects the corporation	• bandwidth sharing • 802.11g • wi-fi
Unexpected operating expenses	Increases in the cost of utilities, insurance, labour, materials, and transportation.	• Public Utility Commission • utility rate hikes • NT underwriting costs
Personnel	Qualifications and training needs of personnel	• internet navigation • internet research • peer-to-peer
Infrastructure	Infrastructure maintenance and improvements	• end of free web • high-speed internet access
Industry affiliations	Professional or industry affiliations with which an organisation should be involved	• IEEE
Regulatory issues	Regulations issued by governmental or licensing bodies that directly affect an organisation	• library filters
Political issues	Political issues that affect the organisation	• online privacy legislation
User-defined categories	Organisation-specific interests that do not fit neatly into other categories	

3.3 Industry class

The industry class satisfies information needs that stem from the type of industry or organisation performing the investigation. This class will help the CI process supply intelligence related to the general external environment of the company. For instance, a corporation may want to keep a watchful eye on changes in competitive, economic, political, legal and regulatory, technological, and sociocultural forces in both the domestic and international marketplaces.

Because of its focus on external, long-range and international issues, this segment is particularly critical in the competitive intelligence process. Changes in geopolitical factors, such as trade barriers and import/export laws, can influence the availability of markets and/or expansion opportunities. Information regarding key economic, social, and technological issues that positively or negatively impact the organisation can help managers allocate attention and resources correctly (McCann and Gomez-Mejia, 1992). In any industry there are certain key factors on which all firms depend for success, such as a widespread level of manufacturing technology, the availability of low-cost raw materials from domestic and international sources, the presence of protective regulation, or certain cultural attitudes or characteristics. Changes in such key environmental variables may dramatically alter the industry as a whole. Firms that become aware of such changes early can get a head start on strategising and can consequently attain a major competitive advantage. Table 3 shows the interest areas contained in the industry class. These interest areas were derived from environmental scanning literature that focuses on the external environment and long-range planning. The class can be adapted to accommodate industry-specific variables as needed.

Table 3 Interest categories that make up the industry class

Interest area	*Description*	*Example*
Customer base and marketplace	The group or type of customer that the organisation serves	• internet savvy • coffee drinkers • Playstation
Industry	The industry type and major products that the organisation produces	• internet café • web café • cyber cafe
Socioeconomic and demographic	Socioeconomic factors that affect demand for the organisation's products	• target neighbourhoods
Competition and corporate environment	Industries that produce the same or substitute products	• Cybarea • Swish • EasyEverything
Manpower and resources	Availability of qualified personnel and organisational resources	
Technological	Technological factors that affect the organisation's performance and success	• Playstation • PC pricing trends • bandwidth
Geopolitical	Geopolitical factors that affect the organisation's markets and expansion	• Canadian expansion opportunities
Governmental	Regulations and trade barriers that affect an organisation's markets such as import/export laws, taxation issues, business law, patent and trademark law	• Tauzin-Dingell Bill • Financial Services Modernisation Act
User-defined categories	Industry-specific interests that do not fit neatly into other categories	

As indicated in Tables 1–3, each of the components that make up the M-CLIP includes a set of suggested interest areas for which multiple keywords or concepts can be specified. No such list can realistically be exhaustive, so each class is made extensible by including provisions for a set of user-defined categories that allow each organisation to personalise the M-CLIP to its specific needs.

The framework provided by the M-CLIP imposes a structure on the specification of critical information needs, and forces the developer of the profile to consider many information areas that may otherwise be overlooked. Furthermore, the framework can be adapted as needed to ensure that CI gathering targets specific information needs for a particular enterprise, industry, or project.

The M-CLIP shares many characteristics with the multiperspective KM System (KMS) described by Frank (1999). Frank notes that a KMS should offer definitions of concepts that are needed for the description and analysis of a corporation, including corporate strategy, organisational unit, and business processes, tasks and employees. These correlate quite closely with the components of the M-CLIP. Business process, tasks and employees correspond to the project class; organisational unit corresponds to the enterprise class; corporate strategy parallels the industry class.

At first glance the M-CLIP may appear to be a structured, expanded set of Key Intelligence Topics (KITs), but several features make it much more than that. Not only does the M-CLIP provide specialised templates to aid in the identification of critical intelligence needs, but it also includes an expansion mechanism to help ensure that no key concepts are overlooked, as well as an adaptive mechanism to automatically handle the removal of unproductive topics.

4 Additional features

The efficiency of information gathering can be assessed through two measures, recall and precision. Recall is the percentage of available relevant information collected by the system, whilst precision is the percentage of all the information gathered by the tool that is actually useful to the organisation (Cleverdon *et al.*, 1966). Higher values for recall and precision indicate more efficient information gathering. The multiclass design of the M-CLIP makes for a more comprehensive profile, and improvements in both recall and precision can be attained through a two-step process:

1. by augmenting the profile to make it as comprehensive as possible
2. by streamlining the profile by assessing the intelligence topics and removing those that allowed unwanted material to be returned by the CI process.

Augmentation is necessary to account for synonymy, *i.e.*, the wide variety of terms that can be used to describe the same topic. Synonymy makes it difficult to select the exact words or phrases that will result in the most successful information gathering because different individuals often describe the same concept in various ways (Furnas *et al.*, 1987). Augmentation can be accomplished by supplementing the M-CLIP with additional features such as specialised templates, an expansion mechanism, and a component-based architecture.

One of the dangers inherent in any system that relies on a larger interest profile is that it will consequently lead to the collection of extraneous information, thus decreasing precision. A profile can be fine-tuned by providing an adaptive mechanism to allow removal of unproductive, obsolete, or carelessly selected terms.

4.1 Specialised templates

Specialised templates can provide guidance to the user during the needs identification phase. Whilst the very structure of the M-CLIP forces managers to consider those areas specified by each class, additional guidance can be provided by domain-specific templates to help ensure a more comprehensive set of intelligence needs. This can be accomplished by the development of a knowledge base of sample profiles, or specialised templates, that provide a listing of suggested keywords or concepts for the user in that particular domain.

The creation of specialised templates is easier if a common vocabulary can be established for the domain. Ontologies are especially useful for this task. An ontology provides a specification of a shared conceptualisation to be used for formulating knowledge-level theories about a domain (Domingue and Motta, 1999; Guarino, 1997). If specific domains can be identified and selected for the CI process, a common ontology can be defined to map vocabularies of specified terms with generally accepted definitions (Gruber, 1991). "Ontologies inform the system user of the vocabulary that is available for interacting with the system and about the domain and the meaning that the system ascribes to terms in that vocabulary" (Farquhar *et al.*, 1997,p.707). Tools like the Ontolingua Server can assist in the construction of ontologies (Farquhar *et al.*, 1997).

4.2 Profile expansion

An expansion mechanism helps alleviate the problem of synonymy, as described previously. Useful for expanding acronyms as well as providing less common industry phrases, or even equivalent foreign phrases, the expansion mechanism supplements the selection of intelligence needs. Expansion can be accomplished through various techniques including query expansion, ontologies, thesaurus programmes, or word disambiguation techniques. Hancock-Beaulieu and Walker (1992), Robertson (1990), Ekmekcioglu *et al.* (1992), and Guarino (1997) all investigate query expansion from an information retrieval perspective. Ontologies, discussed previously as a tool for developing templates, can also be used to supplement the words or concepts that were included in the profile, or to replace the specified words with a more accurate domain-specific term. Thesaurus programmes can also assist in expansion. Jones (1993), Lee *et al.* (1994), and McMath *et al.* (1989) explore the use of thesaurus data models in retrieval systems. A relational thesaurus identifies sets of lexical relations that exist between word pairs, whilst a statistical thesaurus is constructed by analysing the co-occurence of different terms in the collection of documents or web pages being searched. Thus, words that appear frequently in the same context will appear together in the thesaurus so that the profile can be intelligently expanded. This is a form of word disambiguation. The correct meaning of a word is often dependent upon the context in which the word is found. For example, the word 'crash' when used in conjunction with an aeroplane has a vastly different meaning than when used in the context of stock markets. Over the years a variety of disambiguators have been built, and these are discussed in (Sanderson, 1994), as is the application of disambiguation to information retrieval. An excellent discussion of various approaches to thesaurus construction can be found in Crouch and Yang (1992). Any or all of these approaches can be adapted to serve as an expansion mechanism.

Profile expansion is especially useful when identifying critical intelligence needs. For example, the intelligence needs may include the term *FireWire*, but may omit the technical designation *IEEE 1394*. The expansion mechanism must be capable of supplementing provided terms (*FireWire*) with additional expansion terms (*IEEE 1394* or even *USB*). Thus, an expansion mechanism can provide synonyms or alternatives for each keyword or concept. After expansion occurs, the user must have the option of reviewing the newly added terms and removing those that are deemed unnecessary or irrelevant.

4.3 Component-based architecture

The M-CLIP has, by design, a component-based architecture. Such architectures allow components to be removed and interchanged with other components. This allows "user configuration of systems, reusability of components, and a more versatile and potentially robust 'building block' approach to system architectures" (Grundy *et al.*, 1998,p.427).

The multiple classes that comprise the M-CLIP provide a set of customisable components. This personalisation allows each manager to specify the type of information that best suits the nature of decisions that must be made. This customisation can take place on a variety of levels. If managers are individually responsible for identifying their own information needs, then the organisation can develop component templates and provide them to users for guidance during information needs identification. These templates, or components, can be designed to target a particular specialisation, situation, or project. Once created, these components can be reused throughout the organisation with minimal modifications. This makes it possible to create a complete profile of intelligence topics by integrating a set of pre-constructed components. This helps to ensure that individuals in key positions throughout the organisation are performing the type of information acquisition necessary for optimal decision making.

Furthermore, components can be removed and replaced with components that better fit the specific needs of a particular organisation. It is possible to incorporate additional components beyond those provided, or remove some of the provided components in order to work with a subset. This feature is provided to make the M-CLIP customisable to any organisation. The separation of content and design allows the concept of multiple interest classes to be utilised in a variety of ways.

4.4 Profile fine-tuning

Whilst profile expansion helps to increase recall, it can also lead to a decrease in precision. Whenever more intelligence topics are included in the profile, there almost inevitably will be an increase in the number of documents returned, with some of those documents being irrelevant. Some of the topics in the profile may have been poorly selected, whilst others may become obsolete as interests evolve. Frisse and Cousins (1989) note that information that once seemed essential may suddenly become irrelevant. An adaptive mechanism can alleviate that problem by ensuring that only relevant information needs are included in the M-CLIP. When information needs change, the system must be able to detect those changes and adapt in response (Sheth, 1994).

One technique for adapting a profile in response to evolving information needs is to analyse user behaviour patterns to establish a basis for modification of the profile to maintain its accuracy (Fischer and Stevens, 1991). Based on Anderson's discourse on the Rational Analysis of Human Memory (Anderson, 1990), this approach recognises that past usage patterns can be used to predict future usage. One approach that can be used to determine which portions of the profile are responsible for the return of irrelevant material from the CI process is to solicit user feedback. Relevance feedback allows users to assess the relevance of gathered material to their information needs. One form of relevance feedback, incremental feedback, is especially applicable to

CI because judgements are made on a case-by-case basis rather than all at once (Allen, 1996). The user evaluates and rates the documents returned by the system, and those ratings are then associated with the specific information needs that led to the selection of that material. A cumulative average rating is maintained for every term or concept in the profile so that the M-CLIP can track and assess performance. The system can periodically review user ratings and automatically reduce the relative weight of unsatisfactory terms. If those ratings drop below a pre-determined threshold, the system can either arbitrarily remove those terms or can request user confirmation prior to their elimination.

An alternative to relevance feedback is implicit feedback, which can be inferred from user behaviour without any additional work on the part of the user (Oard and Kim, 1998; Kim *et al.*, 2000; Balabanovic, 1998). Regardless of the approach selected, an adaptive mechanism is essential to optimising precision.

5 Pilot study

A pilot study was used to evaluate the effect of the M-CLIP on information gathering. Note that the study was not designed to evaluate a specific CI system, but rather to assess whether use of the M-CLIP can improve information gathering in general. A comparison study was designed to measure the performance of a typical information gathering system and the performance of one that uses an M-CLIP. Comparison studies are performed by selecting one or more measures of a system's performance, providing the same input to both the system and a standard, and then comparing the results on the measures (Cohen and Howe, 1988; 1989). This comparison study employed a conventional profile of information needs as the standard against which the M-CLIP was compared. The conventional profile was designed to be representative of those profiles used by many commercial information gathering systems, and as such was both semi-structured and single-dimensional. The measures used were recall and precision, as defined in the preceding section. These measures are widely used to assess the efficiency of many search strategies, including CI systems (Laplanche *et al.*, 2003), and are applicable to the evaluation of information gathering. A performance improvement exists if one profile attains a higher recall value and an equivalent or greater precision value. The study tested the assertion that no difference exists between the performance of an information gathering system driven by a conventional user profile and one based on an M-CLIP. In this comparison study, if the assertion of no difference is rejected, then it can be concluded that one type of profile leads to improved information gathering.

To eliminate extraneous variables from the comparison study, the information gathering system was required to accommodate either a conventional profile or an M-CLIP without restructuring the profiles or the system. Because commercially available information gathering systems are not able to seamlessly accommodate multiple profiles, a prototype information gathering system was developed. The information gathering prototype was used to simultaneously filter an information stream through both a conventional profile and an M-CLIP.

Five professionals volunteered to participate in the study: two CPAs, two researchers from a semiconductor manufacturer, and one construction firm manager. The subjects were not compensated, but volunteered because of an interest in how CI could benefit their companies. Each participant developed both a conventional profile of information needs and an M-CLIP. Like many commercial systems, the information needs identification for the conventional profile provided tips for profile creation, but little structured guidance. The M-CLIP prototype system was used to guide the construction of an M-CLIP for each participant. Table 4 illustrates a sample conventional profile used in the study, whilst Table 5 provides a sample M-CLIP used in the study. In both profiles the asterisk wildcard at the end of a keyword or phrase is used to assure that the system will match any and all variations of the word in the information set. These profiles were created by the subject simply to test the M-CLIP performance, so neither is particularly refined nor well thought out from a competitive intelligence or information gathering tools perspective. Note however that the sample M-CLIP is a great deal more structured than the sample conventional profile, and also contains considerably more intelligence topics.

Table 4 Sample conventional profile used in study

Interest area	*Example*
People's names	• Dick Cheney
Company names	• Ernst & Young
Places	• West Texas
	• Texas
Product names	• Primavera
Industry phrases	• CPA
	• audit report*
	• generally accepted*
	• federal income tax returns
	• industry practice
	• CFMA
	• GAAP
	• AICPA
Action verbs	• auditing
	• funding
	• compliance
	• taxing
	• manage*
	• consult*
	• audit*

Table 5 Sample M-CLIP profile used in study

Project class	
Project goals or objectives	• timely completion • quality • service
Project basis or background	• loss of profits • review
Technical requirements	• tax abilities • account* software
Resource requirements	• certified public accountant • communication* • training*
Financial requirements	• cost benefit* • allowability
Return-on-investment factors	• decrease in funding • net income*
Enterprise class	
Industry affiliations	• TSCPA • AICPA
Regulatory issues	• grant requirements • independence • peer review • regulations • GAAP
Political issues	• loss of government funding
Industry class	
Customer base and marketplace	• Texas business
Industry	• Public accounting
Socioeconomic and demographic	• Texas • West Texas
Competition and corporate environment	• Ernst & Young • Rival firms
Manpower and resources	• CPA*
Geopolitical	• reciprocity

During the two-week study, approximately 250 items from the Associated Press newswire information feed were filtered through both the conventional profile and the M-CLIP. The process resulted in a retained-information set, which contained documents that matched the interests specified by the particular profile, and a bypassed-information set, containing documents rejected by the system. Participants were required to read every document in both the retained and bypassed document sets and evaluate each with regard to how well it pertained to their actual information needs. These evaluations were recorded for each document, concept, and profile so that the system could determine how many relevant documents *should* have been retained, as well as how many documents actually *were* retained. The following variables were recorded.

- N_{HIT} – total number of documents actually retained by the system, including both relevant and irrelevant documents
- N_{BAD} – number of documents retained by the system but rejected by the user
- N_{MISS} – number of relevant documents incorrectly rejected by the filtering system.

These three variables were used to calculate precision and recall for both the conventional profile and the M-CLIP using the commonly accepted formulas below. The results are shown in Table 6.

$$Precision = \frac{N_{HIT} - N_{BAD}}{N_{HIT}} \tag{1}$$

$$Recall = \frac{N_{HIT} - N_{BAD}}{N_{HIT} - N_{BAD} + N_{MISS}} \tag{2}$$

Table 6 Results of the comparison test

	Recall		*Precision*	
Subject	*Conventional (%)*	*M-CLIP (%)*	*Conventional (%)*	*M-CLIP (%)*
1	22.88	82.20	71.05	85.84
2	42.86	94.24	41.87	97.06
3	54.72	67.92	44.62	48.00
4	70.31	71.88	41.67	52.87
5	2.13	14.89	100.00	66.67
Overall	45.18	66.23	59.84	70.09

In designing the comparison study the authors were aware that the M-CLIP might lead to higher recall simply because the M-CLIP would be, in most cases, larger than a conventional profile of information needs. However, the resulting set of information needs associated with the M-CLIP should also be more complete due to the specialised templates and profile expansion feature. In addition, even an improved profile will degrade precision if the adaptive fine-tuning mechanism fails to work properly. Therefore the study results should not be skewed.

5.1 Results

For each participant, the M-CLIP resulted in higher values for recall than did the conventional profile. All but one participant experienced better precision. The overall rating for each profile in the final row of Table 4 shows that the M-CLIP achieved a recall value of 66.23% as compared to 45.18% for the conventional profile. The M-CLIP had a precision value of 70.09%, whilst the conventional profile scored only 59.84%.

Recall that in order to validate the conceptual model, the M-CLIP was required to attain a higher recall value than the conventional profile, and an equivalent or greater precision value. Those conditions were satisfied, so the assertion that there is no difference between the performance of an information gathering system using the two profiles can be rejected, which indicates that one type of profile, in this case the M-CLIP, leads to improved information gathering.

These results suggest that the M-CLIP will collect both more, and more useful, information. A comparison study applying the process to an actual CI operation will be required to test whether those results hold for the CI process as well.

5.2 Analysis

Table 6 indicates that in some cases the M-CLIP performed markedly better than the conventional profile, whilst in other cases it resulted in relatively small improvements. An analysis of the behaviour of the participants in the study helps to explain these discrepancies.

The scores attained by the M-CLIP were most impressive for Subject 1 and Subject 2. Each of these subjects determined their information needs immediately upon receiving the prototype system and initiated the information gathering process the day that the test began. An examination of their profiles indicates that there is approximately a 4:1 ratio between the number of information needs in the M-CLIP and the number of information needs in the conventional profile. These subjects performed the comparison test as directed and devoted careful preparation to their profiles. As a consequence, the M-CLIP performed far better than its conventional counterpart for these subjects.

The performance measures were much closer for Subject 3 and Subject 4. Like all of the test subjects, Subject 3 and Subject 4 were business professionals. However, unexpected demands on their time prevented them from creating their profiles of information needs until the test period reached the halfway point. Instead of devoting adequate time to needs identification, they accepted the minimal set of information needs provided by the templates. Consequently, their M-CLIPs are only slightly larger than their conventional profiles (a 1.5:1 ratio), reflecting the subjects' rush to complete the profiles. Because these subjects failed to devote adequate time to creating a thorough M-CLIP, their M-CLIPs performed only marginally better than the conventional profile.

Subject 5 had the most unexpected results. The recall achieved by both profiles was remarkably low, whilst the precision achieved was extremely high. Examination of the two profiles revealed that this individual's profiles were overly specific. Because the profiles were so specific, very few articles were retained; but those that were retained were usually of interest to the subject. Thus, the recall was low whilst the precision was high.

The variation in the performance measures underscores how critical the needs identification process is. Herring (1999,p.14) states that: "…well-defined intelligence needs are the prescription for planning and carrying out the right intelligence operations and producing the appropriate intelligence products." Needs identification is a difficult process, even with the assistance provided by the M-CLIP, and it requires careful consideration and reflection. If the time is not taken to develop an adequate set of information needs, then the performance gains that are made possible by the M-CLIP will not be realised.

6 Discussion

Although the margin of difference varied from subject to subject, the M-CLIP consistently performed better than the conventional profile. The study determined that the more complete set of information needs identified and refined with the aid of the M-CLIP did improve the gathering of information, as indicated by the improvement in recall and precision measures. However, to say that the M-CLIP is an improvement over conventional needs identification approaches is not enough. The factors that contribute to that improvement must be examined. First and foremost among those factors is the comprehensiveness of the M-CLIP. By comparing the M-CLIP and conventional profile of each of the participants in the comparison study, it was seen that whilst the size differential between the profiles may not always be appreciable, the M-CLIP is consistently broader in scope and more complete. This can be attributed to the fact that the M-CLIP provides more guidance to the user during needs identification. Potential keywords or concepts are suggested via templates to make the user reflect upon essential information needs. Requiring the user to carefully consider his or her information needs in each area should lead to a more comprehensive set of intelligence topics. Keyword expansion, in the form of a synonym feature, also helped to make the M-CLIP more complete. Finally, the adaptation mechanism helped to improve precision. In some cases the participants relied heavily on the templates, using the profile suggestions as a basis for their initial profile, and modified them through relevance feedback as the information gathering process progressed. These features combine to make for a more complete and comprehensive specification of information needs, expanding the coverage and accuracy of the information gathering process.

Several lessons were learned from the pilot study. The multiple classes of the M-CLIP, in conjunction with the provided templates, expansion mechanism, and adaptive mechanism, performed well and improved information gathering, as demonstrated by improved measures of recall and precision. Whilst the approach is promising, additional tests are called for. Future iterations of this study will require that the needs identification process be completed before the information gathering process gets under way. In addition, the comparison study must be applied in an actual competitive intelligence operation. The strengths and limitations of the M-CLIP approach are listed in Table 7.

Table 7 Strengths and limitations of the M-CLIP

Strengths
• provides structured approach (framework) for needs identification
• provides templates for guidance during needs identification
• provides better gathering of CI due to better structure
• higher recall/higher precision
• provides expansion feature
• component-based architecture provides personalisation, interchangeability
• adaptive mechanism provides evolution and adaptation
• provides coverage of multiple dimensions of critical intelligence needs
• spans more management levels of the organisational pyramid than does KITs

Table 7 Strengths and limitations of the M-CLIP (continued)

Limitations
• not a CI process, but a structured approach to needs identification and evolution
• more time consuming to develop fuller set of information needs
• requires involvement by more levels of management than KITs
• larger set of information is needed due to broader search requirements
• larger set of information needs has the potential of improving recall but decreasing precision (if the adaptation mechanism fails to function properly)
• requires analysis of data collected before using information collected

7 Conclusion

CI tools can be improved if they are augmented with the M-CLIP. The pilot study indicates that the enhanced profile of information needs improves the performance of profile-based information gathering tools by allowing more relevant information to be gathered, whilst at the same time being more discriminating. The enhanced profile can help CI efforts become even more valuable by providing decision makers with a more complete set of information, enabling them to assess domestic and international issues in an efficient, accurate, and timely manner. By encompassing a broader spectrum of corporate interests, the M-CLIP provides the means to access a greater percentage of relevant online information, enhancing CI efforts.

References

Allen, J. (1996) ‘Incremental relevance feedback for information filtering’, *Proceedings of the Nineteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Zurich, pp.270–278.

Anderson, J.R. (1990) *The Adaptive Character of Thought*, Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Anthony, R. (1965) *Planning and Control Systems: A Framework for Analysis*, Boston: Harvard University Press.

Argote, L.S., Beckman, L. and Epple, D. (1990) ‘The persistence and transfer of learning in industrial settings’, *Management Science*, Vol. 36, pp.140–154.

Balabanovic, M. (1998) ‘An interface for learning multi-topic user profiles from implicit feedback’, *Proceedings of the AAAI Workshop on Recommender Systems*, Madison, WI, pp.7–11.

Beidleman, C., Fletcher, D. and Veshosky, D. (1990) ‘On allocating risk: the essence of project finance’, *Sloan Management Review*, Vol. 31, No. 3, pp.47–55.

Berghel, H. (1997) ‘Cyberspace 2000: dealing with information overload’, *Communications of the ACM*, Vol. 40, No. 2, pp.19–24.

Cleverdon, C., Mills, J. and Keen, M. (1966) *Factors Determining the Performance of Indexing Systems*, Cranfield, UK: Cranfield Institute of Technology College of Aeronautics, Vols. 1–2.

Cohen, P. and Howe, A. (1988) ‘How evaluation guides AI research’, *AI Magazine*, pp.35–43.

Cohen, P. and Howe, A. (1989) 'Toward AI research methodology: three case studies in evaluation', *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 19, No. 3, pp.634–645.

Crouch, C. and Yang, B. (1992) 'Experiments in automatic statistical thesaurus construction', *Proceedings of the Fifteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Copenhagen, Denmark, New York: ACM Press, pp.77–88.

Cyert, R.M. and March, J.G. (1963) *A Behavioral Theory of the Firm*, Englewood Cliffs, New Jersey: Prentice-Hall.

Domingue, J. and Motta, E. (1999) 'A knowledge-based news server supporting ontology-driven story enrichment and knowledge retrieval', *Proceedings of the 11th European Workshop on Knowledge Acquisition, Modeling, and Management (EKAW 1999)*, Dagstuhl Castle, Germany, Berlin: Springer-Verlag, pp.103–120.

Ekmekcioglu, F.C., Robertson, A.M. and Willett, P. (1992) 'Effectiveness of query expansion in ranked-output document retrieval systems', *Journal of Information Science*, Vol. 18, No. 2, pp.139–147.

Farquhar, A., Fikes, R. and Rice, J. (1997) 'The ontolingua server: a tool for collaborative ontology construction', *International Journal of Human-Computer Studies*, Vol. 46, No. 6, pp.707–728.

Fischer, G. and Stevens, C. (1991) 'Information access in complex, poorly structured information spaces', *Proceedings of CHI'91 Conference on Human Factors in Computing Systems*, New Orleans, LA, New York: ACM Press, pp.63–70.

Foltz, P. and Dumais, S. (1992) 'Personalized information delivery: an analysis of information-filtering methods', *Communications of the ACM*, Vol. 35, No. 12, pp.51–60.

Frank, U. (1999) 'An object-oriented architecture for knowledge management systems', *Arbeitsberichte des Instituts für Wirtschaftsinformatik*, Koblenz, No. 16.

Frisse, M.E. and Cousins, S.B. (1989) 'Information retrieval from hypertext: update on the dynamic medical handbook project', *Proceedings of Hypertext '89*, Pittsburgh, PA, New York: ACM Press, pp.199–212.

Fuld, L. (2001) 'Intelligence software: reality or still virtual reality?', *Competitive Intelligence Magazine*, pp.22–27.

Furnas, G., Landauer, T., Gomez, L.M. and Dumais, S. (1987) 'The vocabulary problem in human-system communication', *Communications of the ACM*, Vol. 30, No. 11, pp.964–971.

Gates, B. (1999) *Business @ the Speed of Thought*, New York: Warner Books.

Gruber, T. (1991) 'The role of common ontology in achieving sharable, reusable knowledge bases', in J.A. Allen, R. Fikes and E. Sandewall (Eds.) *Principles of Knowledge Representation and Reasoning*, Morgan Kaufmann, San Mateo, CA, pp.601–602.

Grundy, J.C., Mugridge, W.B. and Hosking, J.G. (1998) 'Static and dynamic visualisation of software architectures for component-based systems', *Proceedings of the Tenth International Conference on Software Engineering and Knowledge Engineering (SEKE'98)*, San Francisco, CA: KSI Press, pp.426–433.

Guarino, N. (1997) 'Semantic matching: formal ontological distinctions for information organization, extraction, and integration', in M.T. Pazienza (Ed.) *Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology*, Berlin: Springer Verlag, pp.139–170.

Guy, M.E. (1985) *Professionals in Organizations: Debunking a Myth*, New York: Praeger Publishers.

Hancock-Beaulieu, M. and Walker, S. (1992) 'An evaluation of automatic query expansion in an online library catalogue', *Journal of Documentation*, Vol. 48, No. 4, pp.406–421.

Herring, J.P. (1999) 'Key intelligence topics: a process to identify and define intelligence needs', *Competitive Intelligence Review*, Vol. 10, No. 2, pp.4–14.

Johnson, A.R. (1999) 'Two parts of the same whole" in "your say: competitive intelligence and knowledge management', *Knowledge Management Magazine*, Vol. 3, No. 3.

Jones, S. (1993) 'A thesaurus data model for an intelligent retrieval system', *Journal of Information Science*, Vol. 19, No. 3, pp.167–178.

Kim, J., Oard, D.W. and Romanik, K. (2000) *Using Implicit Feedback for User Modeling in Internet and Intranet Searching: Technical Report*, College Park, Maryland: University of Maryland at College Park College of Library and Information Services.

Laplanche, R., Delgado, J. and Turck, M. (2003) 'Beyond keyword search: how a "concept" search can improve information gathering in a CI context', *SCIP.Online*, Vol. 1, No. 30.

Larson, M.S. (1977) *The Rise of Professionalism: A Sociological Analysis*, Berkeley and Los Angeles, CA: University of California Press.

Lee, J., Kim, M.H. and Lee, Y.J. (1994) 'Ranking documents in thesaurus-based boolean retrieval systems', *Information Processing and Management*, Vol. 30, No. 1, pp.79–91.

Leventhal, D. and March, J.G. (1993) 'The myopia of learning', *Strategic Management Journal*, Vol. 14, pp.95–113.

Levinthal, D. and March, J.G. (1981) 'A model of adaptive organizational search', *Journal of Economic Behavior and Organizations*, Vol. 2, pp.307–333.

March, J.G. (1991) 'Exploration and exploitation in organizational learning', *Organizational Science*, Vol. 2, pp.71–87.

March, J.G. and Olsen, J.P. (1976) *Ambiguity and Choice in Organizations*, Bergen, Norway: Harald Lyche Publishers.

March, J.G. and Simon, H.A. (1958) *Organizations*, New York: John Wiley and Sons.

March, J.G. and Simon, H.A. (1993) *Organizations*, 2nd edition, Oxford, UK: Blackwell.

McCann, J. and Gomez-Mejia, L. (1992) 'Going "online" in the environmental scanning process', *IEEE Transactions on Engineering Management*, Vol. 39, No. 4, pp.394–399.

McMath, C., Tamaru, R. and Rada, R. (1989) 'A graphical thesaurus-based information retrieval system', *International Journal of Man-Machine Studies*, Vol. 31, No. 2, pp.121–147.

Nonaka, I. and Takeuchi, H. (1995) *The Knowledge-Creating Company*, Oxford, England: Oxford University Press.

Oard, D. and Kim, J. (1998) 'Implicit feedback for recommender systems', *Proceedings of the AAAI Workshop on Recommender Systems*, Madison, WI, Menlo Park, California: AAAI Press, pp.81–83.

Parker, K.R. and Nitse, P.S. (2001) 'Improving competitive intelligence gathering for knowledge management systems', *Proceedings of the 2001 International Symposium on Information Systems and Engineering – ISE'2001-Workshop: Knowledge Management Systems: Concepts, Technologies and Applications*, Las Vegas, Nevada, pp.122–128.

Robertson, S.E. (1990) 'On term selection for query expansion', *Journal of Documentation*, Vol. 46, No. 4, pp.359–364.

Sanderson, M. (1994) 'Word sense disambiguation and information retrieval', *Proceedings of the 17th ACM SIGIR Conference*, pp.142–151.

Senge, P. (1990) *The Fifth Discipline: The Art and Practice of the Learning Organization*, Doubleday, Chatham, Kent, UK.

Sheth, B.D. (1994) 'A learning approach to personalized information filtering', *Unpublished Master's Thesis*, Massachusetts Institute of Technology, retrieved November 2003 from the World Wide Web, http://agents.media.mit.edu/publications/sheth-thesis.pdf

Simon, H.A. (1957) *The Sciences of the Artificial*, 3rd edition, Cambridge, Massachusetts: The MIT Press.

Stadnyk, I. and Kass, R. (1992) 'Modeling users' interests in information filters', *Communications of the ACM*, Vol. 35, No. 12, pp.49–50.

Udayagiri, N.D. and Balakrishnan, S. (1993) 'Learning curves and knowledge spillovers: the case of semiconductor memories', *Jones Center Working Paper*, Wharton School, University of Pennsylvannia.

Vibert, C. (2004) *Competitive Intelligence: A Framework for Web-Based Analysis and Decision Making*, South-Western, Mason, Ohio.

Weick, K.E. (1979) *The Social Psychology of Organizing*, 2nd edition, New York: Random House.

Weick, K.E. (1995) *Sensemaking in Organizations*, Thousand Oaks, California: Sage Publishers.

Wilensky, H.L. (1967) *Organizational Intelligence*, New York: Basic Books.

Williams, R. (2002) 'Applying KM lessons to competitive intelligence: creating a user-driven competitive intelligence culture', *American Productivity and Quality Center*, retrieved March 2003 from the World Wide Web http://old.apqc.org/free/articles/Applying%20KM%20 Lessons%20to%20Competitive%20Intelligence.doc

Yelle, L.E. (1979) 'The learning curve: historical review and comprehensive survey', *Decision Sciences*, Vol. 10, pp.302–328.