

Theory and Support for Process Frameworks of Knowledge Discovery and Data Mining from ERP Systems

Elliot Bendoly

Decision and Information Analysis, Emory University

Abstract. Existing theory has framed the process of information extraction and agglomeration, also referred to as the knowledge discovery (KD) process, as a series of strategic search decisions, subject to constraints, with the objective of attaining a sufficient level of domain-specific knowledge for use in strategic planning. Supported by the experiences of firms representative of client, developer, and third-party segments of the data mining community, this work provides an extension to this basic framework. The implications provided suggest a wealth of untapped opportunities in the area of KD research.

Keywords: Data mining, Knowledge discovery in databases, Process, ERP

Section: Research

Original submission date at Information & Management : September 8, 2001

Published in Information & Management 40(7), August 2002

Room 411, Goizueta Business School

1300 Clifton Road, Emory University

Atlanta, GA 30322-2701

Phone 404-727-7138; Fax: 404-727-2053

elliott_bendoly@bus.emory.edu

1. Introduction

The adoption of enterprise resource planning (ERP) systems over the last fifteen years has been accompanied by an explosion of readily available transactional data. Sales figures, human resource activity, stock-out, and defect occurrences are only a part of the issues accessible to and empowering the modern corporation. Often however, the relevance of this data comes from the information that can be *derived* from examining multiple issues simultaneously and the ability to draw inferences critical to strategic planning [2]. Benefits contribute to the firm's business intelligence and subsequently to the overall competitive advantage of the firm [9, 20]. The challenge faced by firms, and the analysts charged with manipulating this data, rests in their ability to provide acceptable levels of strategically applicable information as the result of allowable effort, in time and money.

One of the fundamental problems of information extraction is that the formats of available data sources are often incompatible, requiring extensive conversion efforts. In an attempt to reduce this difficulty, several ERP systems, such as SAP and Baan, have embedded means by which to organize and archive the transactional data in their application databases. The incorporation of such *data warehousing* schemes has interested both researchers and practitioners. *Knowledge discovery* (KD) describes both the overall process by which information is then extracted / agglomerated and the domain dedicated to research on it [8, 27].

There has also recently been a concentrated effort to provide *data mining* (DM) tools able to assist analysts faced with unstructured KD tasks. Distinction between KD and DM concepts remain ambiguous, however, in spite of efforts by various researchers in giving good definitions and examples of their differences [3]. Of the distinguishing features discussed, the most common are the iterative and process-oriented nature of KD, as well as its emphasis on the development of strategic knowledge and domain understanding. Data mining on the other hand is discussed in specific applications and tools for finding rules and relationships among the data.

From a knowledge management standpoint, DM tools allow for the creation of well-defined transferable information [18]. In contrast, KD processes are also characterized by data retrieval, data cleansing, criteria specification, and performance analysis. KD processes agglomerate interim information found by such techniques as data mining in generating understanding and domain knowledge. In an adaptation of the scheme proposed by Haekel and Nolan in 1993 [10] and inspired by the linkages to decision theory as proposed by Kuhlthau, the relationships can be depicted hierarchically as in Figure 1.



Figure 1. Hierarchical Depiction of Knowledge Discovery and Related Concepts

There may be many different goals of a particular KD task, including such objectives as the derivation of dependent relationships, development of forecasts, and classification. The product is an agglomeration of such information, organized in a format that can be applied as knowledge ultimately relevant to a downstream planning activity. At any one iteration in the process, alternate levels of task information become applicable and thus alternate extraction techniques (DM tools) may prove superior. However, while the majority of recent KD literature has focused on the efficiency of the algorithmic tools, a review of the literature reveals that little attention has been given to the strategic nature and internal dynamics of the discovery process. Since the overall success of a KD process is dependent upon both process efficiency and quality of output, and since certain constraints, such as time, may apply to an overall knowledge discovery

task, an analysis of the dynamic nature of these strategies is needed to effect improvements.

2. KD Process Theory

In 1989, the term knowledge discovery in databases (KDD) was coined to represent the process by which resources are applied towards the transformation of available data into strategic information [5, 12]. The task of this transformation can be characterized by a number of issues related to the nature of the data and the features of the outputs desired. The discipline encompasses research including the study and development of data mining (DM) tools.

Although the large number of DM papers and books might not seem a problem, certain classical IT questions arise about the apparent lack of process-based work. Development of any technology involving human-computer interaction faces risks with respect to efficiency, effectiveness, and potential applicability when the only people involved in its testing are those who developed the technology. Long-term marketability of these packages demands a focus on the needs of the user. To ensure a user-focused environment and appropriate package capabilities, process-descriptive input from users needs to be considered. Without sufficient foresight, time, and resources, developers may not produce a product of value to the intended users. With this in mind, a structural

examination of the environment in which DM tools are applied, namely knowledge discovery, seems well overdue.

The KD process itself represents the design and application of a dynamic approach to domain understanding through the specification of its underlying mechanisms [31]. A generalization of the process as a whole has been outlined by Brachman and corroborated by IBM in their overviews of KD/DM technologies [15, 29]. The elements of this process are outlined in Table 1. Since decision making in general plays an integral role throughout the process, Mintzberg’s classical three-phase decision-making process model serves as an ideal basis for framing the generalization and providing a way of comparing other popularized business processes [21].

Brachman’s Elements of the KD Process	Mintzberg’s 3 Phases
<i>Task Discovery, Data Discovery, Data Cleansing, Data Segmentation</i>	<i>Identification Phase</i>
<i>Model Selection, Parameter Selection, Model Specification, Model Fitting</i>	<i>Development Phase</i>
<i>Model Evaluation, Model Refinement Output Evaluation</i>	<i>Evaluation (Selection) Phase</i>

Table 1. Brachman’s KD Process Elements in the Context of Mintzberg’s 3 Phases

The KD Process can therefore be discussed as the interaction of three general phases of ⁽¹⁾ *Domain Identification*, ⁽²⁾ *Strategy Development / Application*, and ⁽³⁾ *Results Evaluation*. As shown in Figure 2, the robustness of this delineation can be illustrated in

its applicability to other classical processes, such as new product development, product experimentation, and IT implementation [25, 30, 32]. In the presence of the phase-iteration often associated with these processes, the comparability becomes evident.

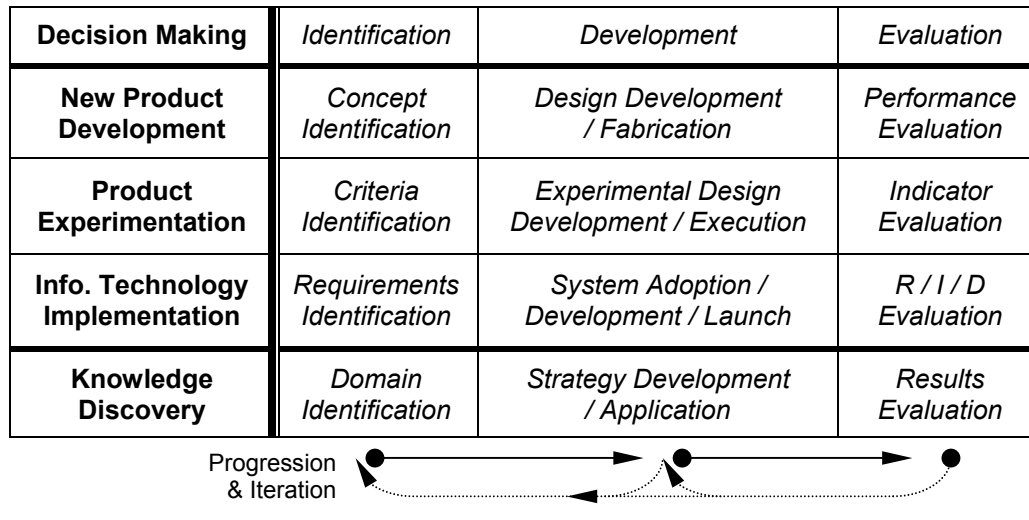


Figure 2. Three Phase Process Frameworks, Comparability and Iteration
 (Note: For IT Implementation, R / I / D = Routinization / Infusion / Diffusion)

Iterations within the KD process occur ⁽¹⁾ when search results are ambiguous, leading to either a re-specification of the domain and/or the search strategy, or ⁽²⁾ when the search strategy that has the greatest value requires further domain specification to be properly executed. In any event, the greatest challenge to the analyst is found within the *Strategy Development / Application* phase, where the actual process of data transformation occurs. The ambiguities increase the difficulty in finding useful, uncorrupted information.

As a final note, it should be emphasized that the three-phase depiction of such processes is only one of several ways that may prove appropriate. At the same time,

experience has shown that attempts to expand process depictions need not involve increased accountability [26, 33]. Perhaps the greatest benefit of a framework for KD process consideration is in its ability to distinguish the prominent phase in which data mining tools and techniques are often applied; specifically that of *Strategy Development / Application*.

3. KD Process Dynamics

In examining the dynamics of KD processes, the experience of representatives from different facets of the data mining community was necessary; the three prominent roles were: Client, Developer, and 3rd-party Consultant. Three firms (one for each role) provided examples for this study of their extensive experience with KD tasks, including data drawn from a range of ERP system types. For purposes of confidentiality, they will not be identified.

The role of **Client** was provided by a manufacturer and supplier to hospitals throughout North America of medical products and services; this company has one of the most comprehensive catalogs in their field and has been an innovator in the implementation and use of modern information technologies, including enterprise and data mining packages. The role of package **Developer** was provided by a specialty software firm; creator of award winning data mining and text mining suites. In contrast

to these, the **3rd-party Consulting** group has earned itself a niche in the field of data mining consulting with an ever-expanding portfolio in the use of such packages.

To elicit information about the experiences of these firms, semi-structured interviews were conducted with three representatives from each firm (a total of 9 interviews). Although opinions among them varied, only those held in common by representatives in an individual firm were reported here. Overall, the experiences of these firms added support to the general process framework of contemporary researchers and contributed insights to the mechanisms specific to each phase of knowledge discovery.

3.1. Domain Identification

The specification of task domain is fundamental to the effectiveness of all later phases. In particular, the designation of managerial performance measures and output requirements are needed to understand and limit the task progression risks that cause excessive analyst bias and exorbitant and unnecessary analytical effort. Such caution also applies to the specification of data sources and of desired output characteristics that occur in this phase. Numerous features of such specifications may lead to increased task complexity and non-stationarity [6]. Speculation and qualitative observation have suggested that prior domain-specific knowledge about the dimensions, relationships and patterns may be essential in reducing these and other complicating issues [2].

Since the time invested in any KD process is indicative of both the direct and opportunity costs of a knowledge-seeking firm, the starting conditions are important. Such a view was shared by representatives of the Client firm, faced with continued pressures to innovate but also continuously to adjust its strategy to meet the demands of an ever-shifting market. The Client firm representatives suggested that they try to avoid wasting time investigating “ill-defined” problems and, in fact, prefer to spend additional time “early-on defining and specifying” the scope of each task and its related data requirements. Ultimately, if the analyst finds the task to be intractable or inappropriate, the Client would prefer time spent “redefining the problem rather than investing in a futile endeavor”.

The experiences of the Consulting firm and Developer suggested that some users might not be as focused as the interviewed Client firm with respect to domain identification. While firms at the forefront in business intelligence typically have very well organized data sets, consistent data-dictionaries and strong understandings of their requirements, the data sets initially used or provided by “starting firms” may be incomplete or inappropriate at first. To identify such potential issues, the Consulting firm reviewed the initial data through the use of baseline summary statistics and checked on redundancy. If necessary, cleansing and aggregation techniques were applied. When issues could not be resolved through transformations of the available data, consultants requested additional or more appropriately organized material from the client. This can lead to an early “redefinition” of the analytical goals. Yet the task is ultimately limited

by the capabilities of the client: “Our packages can only give answers. (Client) firms need to come up with the questions.”

Several noteworthy ways are suggested to facilitate domain identification. The efforts of Reinartz and Wirth [24], Hunter [14], Kuhlthau [17], and Engels [7] emphasized the need to provide a formal and accurate description of the tasks at hand, prior to the development of a strategy. They stress that even if users (data analysts) are domain experts, they typically will not have extensive knowledge about KD and DM techniques when first charged with a task. The user-guidance supported framework of Engels consisted of two levels, the first facilitating the translation of ill-defined tasks into logical formalizations, the second requiring the specification of sets of procedures needed to attain the desired knowledge level.

In associated work, Ali and Wallace [1] proposed a paradigm in which managerial goals are mapped out and linked to performance measures. Within such a framework, managers specify, either directly or in conjunction with formal analytical techniques and rule structures, ⁽¹⁾ generic data items to be used, ⁽²⁾ appropriate data mining methods, and ⁽³⁾ algorithm performance measures connotative of managerial goals. When formal analytical tools and structures are required for the translation of managerial goals (e.g., WIP inventory reduction) into recognizable performance measures (e.g., stability of product defect reduction tactics), analysts may require support from any number of auxiliary systems. This may be particularly true when awareness of overall task performance must be maintained by the analyst. In relation to the theoretical nature of

the process, the hypothesized variations in technique specification, the presence of continuous time performance feedback has been proposed to reduce analyst reliance on undirected activities. In fact, it has been suggested that the absence of such mechanisms threatens the efficiency and effectiveness of time spent.

3.2. *Strategy Development / Application*

Once the task domain has been partially structured, the specification and application of effective KD strategies can be considered. This requires analysts to weigh tradeoffs of DM application alternatives. These dictate, at every event of the analysis, such issues as ⁽¹⁾ what prior domain knowledge is applied and partially exhausted? ⁽²⁾ which available data mining tools are to be utilized? and ⁽³⁾ what is the *orientation* of the search event?

Data mining techniques described as either *directed* or *undirected* searches, are performed with an analogy to confirmatory and exploratory analyses. Fully directed techniques required the *a priori* specification of inputs, outputs, and models. Less directed techniques, often utilizing step-wise and self-organizing approaches, searching for optimal subsets of inputs, internal model characteristics and output structures, such as the number of classifications to be considered [11]. Such analysis makes possible the generation of relational information of potentially novel form, unrestrained by analyst bias. These novel gains are offset by the risk of detecting spurious, nonsensical, or irrelevant information, as well as typical extensions to the amount of time required per

search event. An additional tradeoff lays in an accompanied loss of statistical power, understanding, and search expediency. A matrix of these tradeoffs is presented in Figure 3.

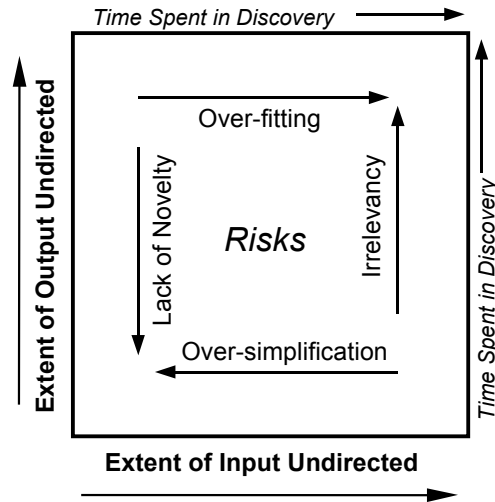


Figure 3. Tradeoffs Between Directed and Undirected Searches

Face validity for this tradeoff matrix was provided by post-interview follow-ups with the firm representatives: all nine concurred on its general appropriateness.

Small models or overly constrained searches, while often easy to comprehend, may simply be inadequate depictions of the active relationships, and hence provide poor information. Large or complex models, while providing a means of improving predictive capabilities, suffer from difficulties in understanding and face an increased risk of over-fitting to trial data. A similar problem arises when the size of the trial data set is limited. Furthermore, difficulties associated with over-fitting may stem from searches that require the specification of a large number of potential models of a single dependent variable: at a significance level of α , a system will, on average, find α of the total number of models

to be significant, from purely random data alone. As the number of models investigated approaches α^{-1} , unless additional considerations such as Bonferroni adjustments are made, the probability of detecting a spurious and biased relationship is no longer slight. However, standard statistical corrections have their drawbacks, since they penalize adequate models for the indecisiveness of the analyst. These considerations, along with other issues such as time devoted to search, represent a number of critical tactics faced by the analyst prior to each iteration. Once specified, the task is delegated to the data mining system.

Similar statements can be made with regards to analysis characterized by undirected input specifications. They represent only the endpoints of a continuum of semi-directed approaches. The robustness permitted by such options may contribute to the efficacy of a KD process, and the selection of specific orientations may be highly dependent on the level of unspent analyst knowledge throughout the discovery process. No formal study has been proposed to compare the usefulness of these approaches, nor the time dependency of the extent to which these techniques are used. However robust undirected strategies are often applied when, as a representative of the Consulting firm remarked, “well defined problems are paired with ill-defined (or non-intuitive) solutions.” This common sentiment was shared by the Client firm, though its representatives made no comment on its specific tactics. At the same time, the Client firm did suggest that the presence of time constraints and a “desire to provide manageable (and believable) results” tends to force a “streamlining of the analysis whenever possible.”

Motivated by similar issues, the Consulting firm utilized formal “streamlining” practices upon each occasion in which the data set was revisited. Such tactics included the reevaluation of data under consideration through the creation of potentially more meaningful complex variables, such as ratios and binaries. Such data reduction is crucial, particularly in light of the time requirements of more complex analyses, which “typically increase with the cube of the number of independent variables simultaneously considered.” In order to gain through an entirely structured reduction approach however, the analyst must have an adequate level of understanding of the problem and available data. Thus a high level of communication between the analyst and decision makers is paramount.

When the potential subset of variable candidates appeared irreducible, the representative claimed that analysis proceeded in “a number of necessary recursions.” Due to excessive time requirements of rule and relationship convergence for large variable sets, a tactic that was shown to be successful was to observe interim progress in model specification. Initially, the analyst considered the entire irreducible variable subset, attempting to fit “a single encompassing model.” However, when interim reports provided by the engine suggested that “stagnation in fit improvement” seemed evident, even at relatively low levels of fitness, the analyst often stopped the engine and attempted to reduce the set by relatively comparing interim effect metrics (when available). The analyst then utilized the reduced set in subsequent attempts at developing descriptive

models, which (by their simplified nature) may be significantly less time intensive, often returning to these primary interim results when specifying each additional model.

While agreeing with the frequent need for data reduction, the Development firm proposed alternate approaches. The consideration of simple metrics, such as correlations, was joined with the consideration of more complex techniques, such as use of packaged tools that utilized hyper-cube distribution comparisons to remove less relevant variables. Due to time constraints, somewhat subjectively selected subsets could be considered on a piece-wise basis. In the representatives' experience, the separate analysis of the top eight variables, followed by the analysis of the second and subsequent subsets of eight independent variables typically provided a means of assessing their ultimate relevance while adding only a small affordable costs to analysis.

An additional tactic proposed by the representative was to fragment the data set through fully undirected analysis tools, a tactic analogous to cluster analysis. The observation clusters derived could then be analyzed separately, beginning once again with preliminary correlation and distribution comparison metrics with the hope of finding more definitive dependencies. The belief behind such a tactic is that various sub-populations of the data behave according to relatively distinct and dissimilar dependent models. This approach is characteristically unstructured in terms of both the relevant inputs and the nature of the outputs.

3.3 Results Evaluation and Process Iteration

The ultimate phase of the knowledge discovery process involved the interpretation of the results provided by analyst-specified algorithmic search. This may be revisited several times before the task is adequately completed. In fact, common among the three representative firms was the claim that the evaluation of information provided by data mining engines was the most likely step to result in a recursion to either steps associated with domain identification or further steps associated with strategy development and application. Also though an analyst's understanding of the domain generally increased through such iteration, the ability to "streamline" subsequent searches might not. A representative from the Client firm claimed that, as subsequent evaluations and iterations occurred, "the likelihood of gaining additional knowledge follows a U-shaped curve."

Such an observation is not unsupported by theory. A similar phenomenon has been proposed by Combs in his generalized inverted-bell depiction of search processes [4]. The initial analysis is facilitated by the application of prior knowledge, while search events close to process completion are facilitated by the application of newly acquired knowledge, gained during analysis. According to Combs, intermediate search events are representative of the trough in the inverted-bell (See Figure 4).

Further evidence of the relative losses and gains in knowledge applicability can be found in recent database search literature. Particularly relevant is Spink's [28] examination of the use of term relevant feedback (TRF) within the search process. TRF refers to the application of information acquired in past queries as inputs for subsequent ones, within a single search process. Similar methodologies were adopted by Hsei-Yee [13] and Jokic [16] in their investigations into the effects of prior task knowledge and the availability of search assistance mechanisms on overall search performance.

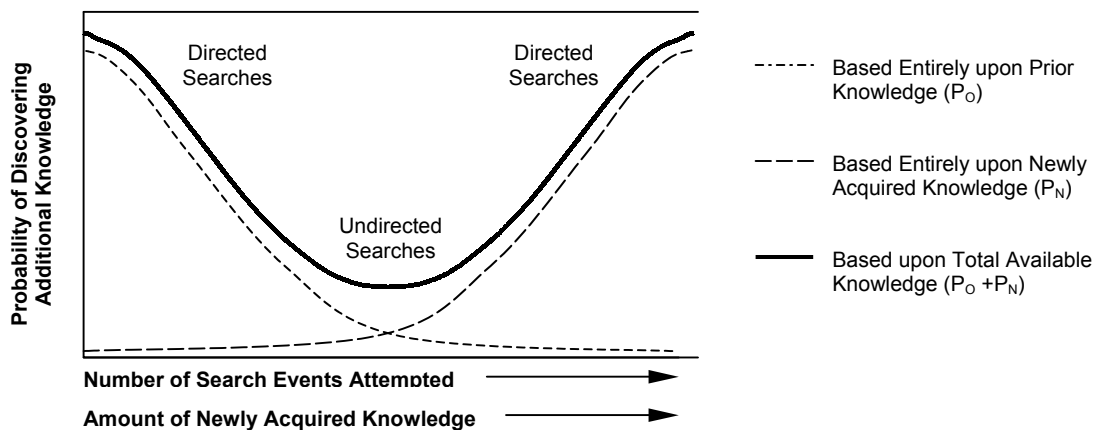


Figure 4. Theoretical Internal Dynamics of KD Processes

Once again a check to the face validity of this second extension to the basic decision framework was provided by *post hoc* assessments of those interviewed. Feedback showed unanimous support for this.

If a relative loss in the likelihood of discovery is associated with a greater utilization of undirected strategies, a number of interesting issues arise. In particular, since the use of undirected searches may be exceedingly costly, analysts must

theoretically weigh the net gains of increasing their potential for novel discovery against the net gains from pursuing relatively low cost, low potential directed techniques. As evident from the formality of alternation schemes proposed by the Consulting and Development firms, the relevancy of such issues seems apparent. As such, the timing of such alternations may have a profound impact on the efficacy of the process. However, the Development firm representatives insisted that the discovery process was a “cycle of trial and error.”

The analyst is ultimately charged with the responsibility of transferring as much of relevant analytical knowledge derived, or in the least the informational rules and relationships derived by the algorithm to the decision maker. Polanyi’s classical work on knowledge transfer categorizes knowledge into two distinguishable forms: explicit and tacit [23]. Explicit knowledge refers to that which is logically transferable, whereas tacit knowledge may be owner and context specific. In the realm of KD, explicit knowledge takes the form of formal models based upon the rules and relationships extracted throughout the process. Tacit knowledge on the other hand may refer to the reasoning behind the specific structure of the model developed, the rules sought out to constitute the model and the model’s perceived applicability. Nonaka posits that the organization as a whole must bear the responsibility of attempting to convert any such tacit knowledge into explicit forms [22]. Such efforts are indeed necessary if the role of KD is to support the overall organizational business intelligence rather than provide a black box internalization of consulting prowess.

References

- [1] F.O.G. Ali and W.A. Wallace. 1997. Bridging the Gap Between Business Objectives and Parameters of Data Mining Algorithms. *Decision Support Systems*. 21:3-15.
- [2] M.J.A. Berry and G. Linoff. 1997. *Data Mining Techniques for Marketing, Sales, and Customer Support*. New York: John Wiley & Sons Inc.
- [3] R.J. Brachman. 1996. The Process of Knowledge Discovery in Databases. In *Knowledge Discovery in Databases*, ed. Fayyad, U.M, et al., Menlo Park, CA: The AAI Press. pp. 1–31.
- [4] R.E. Combs and J.D. Moorhead. 1992. *The Competitive Intelligence Handbook*. Metuchen, NJ: The Scarecrow Press.
- [5] R. Davies. 1989. The Creation of New Knowledge by Information Retrieval and Classification. *Journal of Documentation*. 45: 273-301.
- [6] J.F. Elder and D. Pregibon. 1996. A Statistical Perspective on Knowledge Discovery in Databases. In *Knowledge Discovery in Databases*, ed. Fayyad, U.M, et al. Menlo Park, CA: The AAI Press. pp. 1–31.

- [7] R. Engels. 1996. Planning Tasks for Knowledge Discovery in Databases: Performing Task-Oriented User-Guidance. 1996 Annual Conference on Knowledge Discovery in Databases. pp. 170-175.
- [8] U.M. Fayyad, G. Piatetsky-Shapiro and P. Smyth. 1996. From Data Mining to Knowledge Discovery: An Overview. In Knowledge Discovery in Databases, ed. Fayyad, U.M, et al. Menlo Park, CA: The AAI Press, pp. 1–31.
- [9] D.B. Francis. 1997. Your Competitors: Who Will They Be?. *Competitive Intelligence Review*. 8:16-23.
- [10] S.H. Haeckel and R.L. Nolan. The Role of Technology in an Information Age: Transforming Symbols Into Action, In (eds.). *Annual Review of Institute for Information Studies*, 1994, pp. 1-24.
- [11] T.B. Ho. 1997. Discovering and Using Knowledge from Unsupervised Data, *Decision Support Systems*, 21:29-42.
- [12] X. Hu. 1993. *Conceptual Clustering and Concept Hierarchies in Knowledge Discovery*. Burnaby, BC: Simon Fraser University.

- [13] I. Hsieh-Yee. 1993. Effects of Search Experience and Subject Knowledge on the Search Tactics of Novice and Experienced Searchers. *Journal of the American Society for Information Science*. 44:161-174.
- [14] L. Hunter. 1995. Planning to Learn. In Ram, A. and Leake, D., eds., *Goal-Driven Learning*. London: MIT Press.
- [15] IBM Corporation. 1998. Data Mining: Extending the Information Warehouse Framework, www.almaden.ibm.com/cgi-bin/stss/get/data-mining.
- [16] M. Jokic. 1997. Analysis of Users' Searches of CD-ROM Databases in the National and University Library in Zagreb. *Information Processing & Management*. 33:785-802.
- [17] C.C. Kuhlthau. 1999. Accommodating the user's information search process: Challenges for information retrieval system designers, *American Society for Information Science*. *Bulletin of the American Society for Information Science*. 25:12-16.
- [18] C.C. Kuhlthau. 1999. The Role of Experience in the Information Search Process of an Early Career Information Worker Perceptions of Uncertainty, Complexity, Construction, and Sources. *Journal of the American Society for Information Science*. 50:339-412.

[19] G.M. Marakas. 1995. The Discovery-Learning DSS: Allowing for Discovery in the Decision Process. Proceedings of the 28th Annual Hawaii International Conference on System Sciences. pp. 72-81.

[20] H. Meyer. 1987. Real World Intelligence. New York: General Publishing Company Ltd.

[21] H. Mintzberg. 1973. The Nature of Managerial Work. Engelwood Cliffs, NJ: Prentice-Hall.

[22] I. Nonaka. 1994. A Dynamic Theory of Organizational Knowledge Creation. Organization Science. 5:14-37.

[23] M. Polanyi. 1966. The Tacit Dimension. London: Routedge & Kegan Paul.

[24] T. Reinartz and R. Wirth. 1995. Towards a Task Model for KDD-processes. MLNet Familiarisation Workshop. pp. 19-24.

[25] E. Rogers. 1995. Diffusion of Innovations, 4th Edition. New York: The Free Press.

[26] V.L. Sauter. 1999. Intuitive Decision-Making, Association for Computing Machinery. Communications of the ACM. 42:109-117.

[27] D. Sherry. 1997. Knowledge Discovery by Inspection. *Decision Support Systems*. 21:43-47.

[28] A. Spink. 1995. Term Relevance Feedback and Mediated Database Searching: Implications for Information Retrieval Practice and System Design. *Information Processing & Management*. 31:161-171.

[29] D.S. Tkach. 1998. Information Mining with the IBM Intelligent Miner Family. IBM Software Solutions White Paper.

[30] S.H. Thomke. 1998. Managing Experimentation in the Design of New Products. *Management Science*. 44:743-762.

[31] W.J. Trybula. 1997. Data Mining and Knowledge Discovery. *Annual Review of Information Science and Technology (ARIST)*. 32:197-229.

[32] S.C. Wheelwright and K.B. Clark. 1995. *Leading Product Development*. New York: The Free Press.

[33] R.W. Zmud and L.E. Apple. 1992. Measuring Technology Incorporation/Infusion. *The Journal of Product Innovation Management*. 9:148-156.