ELSEVIER

Information & Management 40 (2003) 533–539

INFORMATION & MANAGEMENT

www.elsevier.com/locate/dsw

Technology and knowledge: bridging a "generating" gap

Israel Spiegler*

Technology and Information Systems Department, The Leon Recanati Graduate School of Business Administration, Tel Aviv University, Tel Aviv 69978, Israel

Received 22 August 2001; accepted 27 May 2002

Abstract

Refuting the notion of technology as a replacement of knowledge, this paper focuses on a gap between them that needs to be bridged. The idea is that *technology* represents the means, and *knowledge* the end of a process that includes many explicit and implicit methods for generating knowledge by using technology. Among these methods is data mining (DM), the leading thrust in the effort to gain actionable information from operational databases of organizations; this is particularly evident in direct marketing, customer relationship management (CRM), user profiling, and e-commerce applications.

Two models of knowledge are reviewed. The first follows a conventional hierarchy of data, information and knowledge with a spiral and recursive way of generating knowledge. The other presents a reverse hierarchy where knowledge precedes the data-to-information process. The models are compared and discussed in the context of knowledge management (KM), using DM as an example.

Keywords: Knowledge; Technology; Knowledge management; Data mining; Knowledge life cycle

1. Introduction

Much confusion exists today about the relationship between technology and knowledge. To many in managerial roles, technology is conceived of as a replacement for knowledge. The computer, database management software, data warehouses, data marts, and particularly, the Internet are often equated with information and knowledge. We agree that such technological means are undeniably vehicles for fast and efficient access to data, and that software packages and communication facilities enable accessing, retrieving, transformations, visualizations, and other operations on data. Still, they are in the realm of information and not knowledge.

Before trying to capture the relationship between knowledge and technology, the terms *data*, *information*, and *knowledge* must be clarified. Spiegler [18] suggests a recursive and spiral model of linking the three, where "yesterday's data are today's information, and tomorrow's knowledge, which in turn recycles back through the value chain into information and then into data". This model, depicted in Fig. 1, defines a cycle, almost a life cycle, of knowledge generation. An interesting and diametrically opposed view is proposed by Tuomi [20] who argues that knowledge is needed before data are collected and indeed it determines what data to store. This paper compares the two views and discusses their relevance to the generation of knowledge.

* Tel.: +972-3-640-7488; fax: +972-3-640-7741.
E-mail address: spiegler@post.tau.ac.il (I. Spiegler).

0378-7206/02/$ – see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0378-7206(02)00069-1

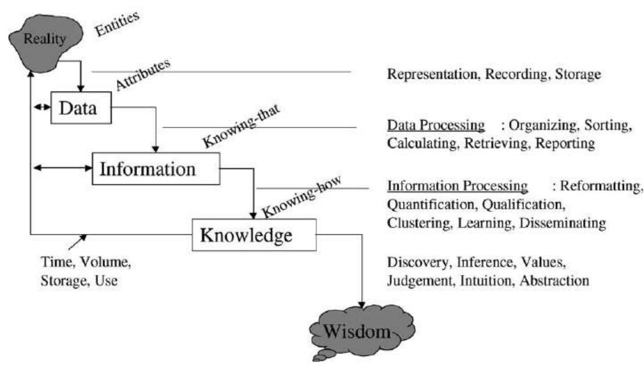

Fig. 1. Data, information and knowledge.

Knowledge, information, and data are often used interchangeably. In the earliest days of computing, the terms *data* and *information* were used with *data processing* turning the former into the latter. Then, *data management* and *information management* were introduced, and now we have *knowledge management* (KM) and the coming of knowledge based business [8]. Serious attempts are being made to distinguish these concepts [1] but definitions of KM are still notably similar to those given in the past for MIS, DSS, EIS and related systems.

Many approaches are used to generate knowledge. Among them is abstract thought, data mining (DM), and the commercial practice of managing and utilizing the organization's data resources. The objective of DM is to detect, interpret, and predict qualitative and quantitative patterns in data, leading to *information* and *knowledge*. A wide variety of models and algorithms are employed, from statistics, artificial intelligence, neural nets and databases, to machine learning [2,5]. The core mining techniques applied by researchers are clustering, classification, association, and time series [10].

We discuss here how to generate knowledge by means of technology. Our point of departure is that technology is the prerequisite and means, and knowledge the end result of the generation cycle. The current literature on KM and DM makes generous use of the term *knowledge* but seldom makes a good definition of it [4]. Some writers prefer to concentrate on KM, leaving knowledge as a black box or a commodity, and refer to it using managerial terms like "markets", "buying", and "renting". As one source puts it "since epistemologists spend their lives trying to understand what it means to know something, we will not pretend to provide a definitive account ourselves . . . we offer . . . a pragmatic description that helps us communicate what we mean when we talk about knowledge in organizations" [7].

2. Technology

The term *technology*, or information technology (IT), is used in this paper to include hardware, software, communications and other means of processing data. In this sense, technology is no more a substitute for knowledge than the brain is a substitute for the mind. While knowledge is an ongoing process, technology is a means or vehicle for processing data, and producing and disseminating information. IT in itself hardly creates knowledge or guarantees the generation of knowledge. The assumption that technology can replace human knowledge or create its equivalent has been proven false time and again.

Perhaps the big divide between humans and machines lies in this very issue of knowledge possession. Even the most liberal of artificial intelligence proponents are cautious in attributing knowledge to machines [23]. Humans deal with and possess knowledge, whereas machines handle the manifestations or representations of knowledge. These representations are at least one step lower in the chain of abstraction from reality. Thus, we ordinarily find data and information at the machine level.

Computers are called data processors, information processors and even knowledge generators. But, as already pointed out, though they can help store and access many facts, they cannot replace expert know-how simply by stockpiling more and more of these facts into their databases [9].

The same holds for most software systems. In a case study of implementing Lotus Notes, regarded by some as knowledge-enhancing software, it was observed that its introduction into an organization did not, in itself, produce a change of information sharing or communication patterns—two basic enablers of organizational knowledge.

From a managerial point of view, the only too common practice of relying on technology as a strategic asset is, at least, risky if not futile. Technology in such a broad sense has a "self-canceling advantage" since it quickly becomes freely available [22]. Internet sites, rare and expensive less than a decade ago, are now commonly downloaded by ever-increasing numbers of users, providing them with technological advantages. Hence, organizations are now beginning to view knowledge as the true long-run strategic advantage in competitive markets.

3. Knowledge

Any definition of knowledge must start from data and information. Following Peter Drucker, who calls information "data endowed with relevance and purpose", and accepting that the value of information is determined by the receiver and not by the sender, we can make the parallel inference that:

> If data becomes information when they add value, then information becomes knowledge when it adds insight, abstraction, and better understanding.

The distinction between data and information appears already in the CODASYL report of 1971 when generalized features of database systems were made [6]. The following is a summary of the distinction between data, information, and knowledge:

- *Data* are symbols inscribed by human hands or by instruments.
- *Information* is a judgment, by an individual or groups, that given data resolve questions, disclose or reveal distinctions, or enable new action. Information, thus, exists in the eyes of the beholder; the same data can become nonsense to one person and gold to another.
- *Knowledge* is the capacity for effective action in a domain of human actions.

Using a restaurant simile, "data are the symbols on the menu, information is the understanding of the restaurant's offerings, knowledge is the dinner. You don't go to the restaurant to lick the ink or eat the menu" (by Lewis Perelman).

Bourdreau and Couillard [3] see information as the result of analyzing and interpreting data—phrases or images that carry meaning. Such assigning of meaning to information is an example of borrowing and enhancing terms that are found in many areas, particularly in computers and high-tech markets.

In setting forth the original notion of information, Shannon hardly ascribed meaning to it at all in the ordinary sense. Information theory is indeed a non-semantic mathematical theory of communication defining a channel's capacity to transmit data. As Shanon explicitly states, "Information, in this theory must not be confused with meaning" [17].

So, what is knowledge? Knowledge has an elusive and fragile manifestation; we know it when we use it. This notion of knowledge may partially explain why, when we attempt to capture, record or store knowledge, it becomes information or data. Nonaka defines knowledge as "a justified belief that increases an entity's capacity for effective action" [13,14].

The difficulty of defining knowledge is also due to the paradox that knowledge resides in a person's mind and at the same time has to be captured, stored, and reported. A wide range of characteristics is attributed to knowledge. Consider the following sample of definitions of knowledge:

1. Knowledge is the power to act and to make value-producing decisions [12].
2. Knowledge is information made actionable in a way that adds value to the enterprise [21].
3. Knowledge is mission specific professional expertise.
4. Knowledge is things that are held to be true in a given context and that drive people to action.

A recent source outlines the following knowledge perspectives: state of mind, object, process, condition of access to information, and capability.

Philosophers classify knowledge into *knowing-that* and *knowing-how.* Knowing-that is factual while knowing-how is actionable, a distinction that parallels data and DM. Data, stored in databases, are facts that can be recalled, processed, and disseminated. Once given relevance and purpose, data are turned into information and in turn into knowledge—the knowing-how to do something. This is the conventional view of the knowledge hierarchy. Indeed, one goal of DM is to gain new insights and knowledge from large databases.

Another way to classify knowledge is by skill levels: beginner, early user, professional, expert, virtuoso, and master. Such classification is applicable at the practical level, generally following training and practice.

Knowledge ranges from a mere recalling of facts (this some people think can be stored), to action and expertise, and thus, to a potential and an ability. We can carry the delineation a step further and propose that knowledge is the *production of new facts*, or even the production of new knowledge, a recursive or reflexive process that is, in fact, infinite. Such a model may be a partial melding of two knowledge models: the conventional data → information → knowledge, and the reverse hierarchy of knowledge preceding information and data.

The goal of any KM is to generate knowledge, which is done by representing *knowing-how* in terms of *knowing-that* and is not always proportional to the volume of accumulated or presented facts.

3.1. Levels of knowledge

In his seminal work on personal knowledge, Polanyi [15] outlines a three-level model of knowledge:

1. Skill: acting according to rules.
2. Know-how: skill plus acting in a social context.
3. Expertise: know-how plus the ability to influence rules and domain of knowledge.

Here again, the expertise level is recursive or reflexive. In fact, Polanyi defines knowledge as "an activity which would be better described as *a process of knowing*". Two types of knowledge are gaining general acceptance in the field of KM:

- *Tacit knowledge*: implicit, mental models, and experiences of individuals.
- *Explicit knowledge*: formal models, rules, and procedures.

These are used throughout Nonaka and Takeuchi's pioneering work on the knowledge creating company. These types are reviewed and extended further into individual, social, declarative, procedural, casual, conditional, relational, and pragmatic types of knowledge.

3.2. Related components

Many terms and components are used in conjunction with knowledge generation and appear in the literature. Among them are: experience, judgment, common sense, rules of thumb, values and beliefs, basic truths, context, best practices, emotions, desires, and socializing into a culture.

As a working definition, we can state that knowledge is *the process of knowing*, *a reflexive process that takes data and information*, *in a social context*, *mixes the ingredients and factors listed above*, *to generate new data*, *information*, *and/or knowledge*. Thus, knowledge constantly evolves or else reverts to its raw material.

Alavi and Leidner define four processes in KM: creation, storage and retrieval, transfer, and application. Our discussion here focuses on the first and the role that technology plays in this aspect.

4. Data mining

Bridging the gap between technology and knowledge, the objective of DM is to identify valid, novel and useful patterns, and associations in existing data in order to gain insights that add to organizational knowledge. DM and KD in databases (KDD) came about as a reaction to the exponential growth of databases in all facets of life, the idea being to extract high level information from an abundance of raw data. Tools and techniques are continually being developed to cope with mountains of data, usually operational, in order to glean some insight and knowledge from them. The terms DM and KDD are sometimes used interchangeably, and at times are distinguished, e.g. KDD refers to the overall process of discovering useful knowledge from data, while DM refers to the application of algorithms for extracting patterns from data. While DM is surely part of "technology" in the broad sense, we use it here to illustrate an approach to knowledge generation.

Surveys of techniques and models that have been devised for DM and knowledge generation are found in [11,16] and others. The relevant techniques are: classification, clustering, machine learning, and association.

Classification and clustering are prime targets of most empirical research of the real world and of DM. These aim to group entities or objects into classes so that there is maximum intra-class proximity between members, and maximum inter-class distinction among groups. Classification is a function that maps an entity or case into one of several classes. Two types of classification are distinguished: when the mapping is done to a predefined set of classes we have supervised classification, and if the process also determines the classes themselves we have unsupervised classification. Clustering and classification are among the core mining techniques, together with associations and time series analysis. Commonly, a clustering model provides:

(1) a representation scheme in the form of a data structure;
(2) an index to calculate similarity;
(3) a grouping technique.

Regression, perhaps the most common statistical method in social science studies, maps and fits an observed case to a predictive function calculated from the data.

Association is a relative newcomer to KDD. Association rules are a key DM tool. They take the form of $X \Rightarrow Y$, where X and Y are events, and when X occurs in the database so does Y, within a certain probability. An example is "98% of customers that purchase tires and automobile accessories also get automotive services". When applied to a database or data warehouse, such a rule may reveal new information hidden in the data. It is a technique found in direct marketing applications that attempt to infer possible and future patterns and modes of customer behavior.

Among the gamut of other methods and techniques used are: summarization, cross-tabulation, mean, standard deviation, etc. Knowledge is also generated by means of neural networks. Artificial neural networks emulate the inherent parallelism of circuits found in the brain to simulate a thinking and cognition process. They are used because of their flexibility in adjusting the number of nodes, connections, and layers to meet various problem definition needs. The models go through a stage of training and learning in the process of classification and clustering. The Appendix A lists a range of products and links found in the KDNuggets page, the official outlet of KDD.

5. Discussion

The two models of knowledge that are the basis of this paper deal essentially with relationships between data, information, and knowledge. A conventional hierarchy is depicted together with recursive loops. The reverse hierarchy proposed by Tuomi places knowledge ahead of the process. His premise is that seeing data as raw material for information and knowledge is misleading; raw data do not exist and human cognition cannot see simple facts without them being part of the current meaning structure. Such "thorny epistemological issues" led him to propose a reverse model. Applying articulation, structure, and verbalization to knowledge yields information. An addition of fixed representation and interpretation to the information produces data.

This idea is brought into the process of database and information system design [19]. So, in absolute, metaphysical terms Tuomi is right. However, attempting to capture the relevant environment in which a person or organization is required to act, the recursive model is more relevant. Moreover, it embodies the reverse model as it outlines ways in which knowledge reverts back to information and in turn transforms back to data. This moving spiral view of knowledge generation, which also appears in Nonaka and Takeuchi's work is a more comprehensive model and offers a life cycle framework for KM system design.

Tuomi rightly defines *yield* as the "intellectual dividend per effort invested", showing an exponential progression between data, information, knowledge, intelligence, and wisdom.

6. Summary and conclusions

Two models of knowledge were reviewed and compared. The first is a recursive and spiral model of linking data, information, and knowledge. A diametrically opposite view is that knowledge is needed before data are collected and indeed determines what data to store, thus treating the knowledge generation as a reverse hierarchy. The models have been compared and contrasted in terms of applicability as a framework and relevance to the generation of knowledge.

Both models acknowledge the role of technology in generating knowledge. We propose that one way that technology can be used for generating knowledge is DM.

Appendix A. KDNuggets page and products

- *Suites*, supporting classification, clustering, data preparation and additional discovery tasks
  - *agents*,
  - *association rules* and market basket analysis,
  - *bayesian and dependency networks*.
- *Classification* using: *multiple*, *decision trees/rules*, *neural nets*, *bayesian*, *other* (*SVM*, *genetic*, ...) approaches
  - *clustering*, for finding clusters or segments,
  - *data transformation and cleaning*,
  - *deviation and fraud detection*,
  - *estimation and forecasting*.
- *Libraries and developer kits* for creating embedded data mining applications
  - *OLAP and dimensional analysis*.
- *Sequential patterns*
  - *simulation of processes*.
- *Statistical analysis*
  - *text analysis* tools for searching and analyzing unstructured texts,
  - *visualization*, scientific and business-oriented,
- *Web Mining*, clickstream and session log analysis.

References

[1] M. Alavi, D. Leidner, Knowledge management and knowledge management systems: conceptual foundations and research issues, MIS Quarterly 25 (1), 2001, pp. 107–136.

[2] I. Bose, R.K. Mahapatra, Business data mining—a machine learning perspective, Information and Management 39, 2001, pp. 211–225.

[3] A. Bourdreau, G. Couillard, System integration and knowledge management, Information Systems Management (Fall) (1999) 24–32.

[4] L. Chen, T. Sakaguchi, M.N. Frolick, Data mining methods, applications, and tools, Information Systems Management (Spring) (2000) 65–70.

[5] H.M. Chung, P. Gray, Data mining, Journal of Management Information Systems 16 (1), 1999, pp. 11–16.

[6] CODASYL, Feature Analysis of Generalized Data Base Management Systems, ACM, May 1971.

[7] T.H. Davenport, L. Prusak, Working Knowledge: How Organizations Manage What They Know, Harvard Business School Press, Boston, Mass, 1998.

[8] S. Davis, J. Botkin, The coming of knowledge-based business, Harvard Business Review (September/October) (1994).

[9] H.L. Dreyfus, What Computers Still Can't Do, MIT Press, Cambridge, MA, 1997.

[10] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, R. Uthurusamy (Eds.), Advances in Knowledge Discovery and Data Mining, AAAI, MIT Press, Cambridge, MA, 1996.

[11] C.W. Holsapple, K.D. Joshi, Knowledge manipulation activities: results of a Delphi study, Information and Management 39, 2002, pp. 477–490.

[12] J., Kanter, Knowledge management, practically speaking, Information Systems Management (Fall) (1999) 7–15.

[13] I. Nonaka, K. Umemoto, D. Senoo, From information processing to knowledge creation: a paradigm shift in business management, Technology in Society 18 (2), 1996, pp. 203–218.

[14] I. Nonaka, H. Takeuchi, The Knowledge Creating Company, Oxford University Press, New York, 1995.

[15] M. Polanyi, Personal Knowledge, University of Chicago Press, Chicago, 1962.

[16] N. Ramakrishnan, A.Y. Grama, Data mining: from serendipity to science, IEEE Computer (August) (1999).

[17] C.E. Shanon, The Mathematical Theory of Communication, University of Illinois Press, Champaign, IL, 1962.

[18] I. Spiegler, Knowledge management: a new idea or a recycled concept, Communications of the AIS 3 (14), 2000, pp. 1–24.

[19] I. Spiegler, Information systems: a construct of transformations, IEEE Transactions on Systems, Man, and Cybernetics 25 (7), 1995, pp. 1121–1129.

[20] I. Tuomi, Data is more than knowledge: implications of the reversed knowledge hierarchy for knowledge management and organizational memory, Journal of Management Information Systems 16 (3), 1999/2000, pp. 103–117.

[21] E.F. Vail, Knowledge mapping: getting started with knowledge management, Information Systems Management (Fall) (1999) 16–23.
[22] A. Webber, What's so new about the new economy, Harvard Business Review (January/February) (1993).
[23] T. Winograd, The design of interaction, in: P. Denning, R. Metcalfe (Eds.), Beyond Calculations: The Next Fifty Years of Computing, Springer, New York, 1997.

Israel Spiegler is an associate professor and past chair of the Information Systems Department at Tel Aviv University. He holds an MSc and PhD degrees in computers and information from UCLA. He held faculty positions at Boston University, School of Information Science at Claremont Graduate University, and UCLA (visiting). His main areas of interest are data modeling, databases, artificial intelligence and knowledge management.