Proposal for a Provisional Knowledge Management Taxonomy

Michael J.D. Sutton, Ph.D. Student
Graduate School of Library and Information Studies
McGill University, Montréal, Québec, Canada
michael.sutton2@sympatico.ca

ABSTRACT

The evolving field of discourse called "knowledge management" possesses an implicitly structured, ill-defined set of concepts and terms. This deficiency has confused both academics and practitioners. This paper suggests that any systematic investigation into phenomena would significantly benefit from a well-defined theoretical framework as a foundation for study of the phenomena. Since the literature associated with a knowledge management taxonomy is exceptionally thin, this paper will propose a provisional taxonomy for further discussion and research.

RÉSUMÉ

Le champ évoluant du discours appelé knowledge management possède une série mal définie des notions et des termes implicitement structurée. Ce manque a embrouillé les universitaires et les praticien-nes. Cette communication suggère que n’importe quel examen sur le phénomène serait avantage considérablement par un cadre théorique bien défini comme une base d’étude du phénomène. Etant donné que la littérature associée à la taxonomie de connaissance de gestion est exceptionnellement médiocre, cette étude proposera une taxonomie provisoire pour de la discussion et de la recherche d’avantage.

INTRODUCTION

During the past two decades a field of discourse has evolved that evokes mystery, respect, ridicule, and sometimes even laughter. That emergent field has been labeled "knowledge management." Additional broader, related and narrower concepts have sprung up alongside "knowledge management." Regrettably, for those who are trying to conduct research on the subject of knowledge management, a provisional taxonomy of concepts and terms has not yet been established or promulgated.

This lack of a provisional taxonomy has consequently confused academics and practitioners alike because of the significant ambiguity surrounding the concepts. Associated concepts, such as, competitive intelligence, knowledge engineering, text mining, continuous learning, or communities of practice also display an annoying level of ambivalence. This confusion of terms may contribute to a hesitation and slower than

---

28 The author would like to acknowledge support for this research from a number of corporate and university sources: Denis Lefebvre, CEO, COVANSYS Canada; Ralph Wood, President, Meridian Systems; Diane Crouse, Director and James Roy, Senior Consultant, Knowledge Management Practice, Consulting & Audit Canada; Danny Boulanger, CEO, Interdoc Corporation; Dr. John Willinsky, Professor, University of British Columbia; Paul Hession, CIO, Fisheries & Oceans Canada; Alain Godbout, Partner, GMG & assoc.; Sandy Foote, Vice-President, Sierra Systems Inc.; Alex Beraskow, Partner, IT/Net.

29 The author would like to further acknowledge the integral assistance, feedback and encouragement of three of his dissertation team members: Dr., Lorna Rees-Potter, Dr. Jamshid Beheshti, and Dr. France Bouthillier.
necessary adoption of knowledge management as a potential academic discipline.

One of the initial challenges of a provisional taxonomy was the proposed location of knowledge management in terms of other scientific fields, disciplines and departments (Schlick 1925). Indeed, a critical question that arises from the analysis is:

"Does Knowledge Management encompass the qualities of an Art as well as empirical traits of a Science, or is it within the exclusive the domain of one or the other?"

This question may seem trite, yet, when we reflect upon the potential interaction and contributions of epistemology, linguistics, metaphysics, philosophy, sociology, and theology with foundational characteristics of knowledge management, we are presented with some unusual paradoxes (Brook 2000). For example, is business knowledge obtained a priori or a posteriori? Is this a valid question that should be answered only within the context of knowledge management or within the wider spectrum of epistemology (Feyerabend 1991, Lehrer 2000, & Maritain 1995)? Though an interesting research question to explore, these issues would warrant a separate research initiative and paper for more in depth discussion.

The broader exploration of interest may be whether knowledge management is a component of a higher-level science, such as knowledge science (Fichte 1982), in the same sense that information management is a set of methods and methodologies within the context of information science, and classification theory exists within the realm of library science.

Consequently, a number of assumptions were made by the researcher in order to attempt to cluster knowledge management terms and concepts within an initial conceptual framework. For the purposes of this study, the researcher has assumed a context for the organization and management of knowledge that exists within a continuum encompassing something called knowledge science. Thus, an hierarchy has been proposed which moves within a spectrum from the stratum of data, through an information layer, and eventually into a knowledge layer (Bellinger 1997, Knowledge Management Forum 1996, Liebowitz 1999, & Wiig 1995). The proposed provisional taxonomy parallels this conceptual framework.

The subsequent miscommunication between cross-disciplinary fields that are now drawing upon the emergent knowledge management stream is presenting challenges for discussion and dialogue. For example, the term "information" has numerous, ambiguous definitions across the fields of communications theory, information and library science, computer science, software engineering and knowledge engineering. This paper suggests that any systematic investigation into phenomena would significantly benefit from a well-defined theoretical framework as a foundation to study the phenomena. Since the literature associated with a knowledge management taxonomy is exceptionally thin and there is only one known published model for a knowledge management thesaurus (Fiddler 2000), this paper will propose a provisional taxonomy for further discussion and research.
If quantitative and qualitative research is to be investigated and scrutinized, the structure and content of the underlying research needs to be reported without ambiguity or contradiction. When research is undertaken, there is an imperative that calls for an acceptable level of clarity in the description of theories, hypotheses, research components, measurements, methodologies and metrics. Additionally, researchers cannot easily generalize findings when the experiment or lesson learned contains poorly scoped or defined terms. An agreed upon taxonomy could provide a common linguistic and semantic framework by articulating the relationships between words and concepts for those interested in knowledge management (Wiig 1993).

**APPROACHES TO A PROVISIONAL TAXONOMY AND POTENTIAL BENEFITS**

A taxonomy consists of a framework of concepts based upon similarities as well as classification differences of identified concepts, terms and phrases. A taxonomy is composed of two parts (Chrisman 1988):

1) a classification scheme representing relationships that permit a researcher to arrange entities into groups (called taxa), and
2) methods used to specify categories within the classification system—the self-differentiating capability of a group of entities to cluster together into a single 'taxon' based upon similarities or satisfactory differences from other groups of entities.

A rigorous taxonomy can form the basis for a concise and accurate conceptual framework that represents the subject nodes in a given knowledge domain. Such a taxonomy can be employed to organize information in the chosen domain for eventual retrieval, for example as in Godin, (1999). A taxonomy could be used to introduce a level of terminological control. Thus, researchers and practitioners would be able to identify subject categories in advance they could use to search repositories. Alternatively, a taxonomy could be used to pre-select keywords for new published material to be facilitate its precise retrieval while allowing for creative and critical need to push semantic boundaries and "bound-ed-ness."

The objective of this paper is not to evaluate or assess the strengths or weaknesses of different taxonomy development methods. The objective is to select and employ a method that will yield results and can satisfy the goal of proposing a provisional taxonomy for further discussion. Brief summaries of the numerous approaches that furnish the basis for the creation of a taxonomy are described below:

- Biological cladistics
- Formal Concept Analysis (FCA)
- Grounded theory
- Library science's theory of the organization of information
- Ontological terminology systems
• Organizational systematics.

Biological cladistics is a formalized process for grouping biological organisms that are descended from common ancestors. It is based upon the theories of evolution and natural selection and the grouping of similar organisms into a family, genus, and species, which has had its greatest application within the biological sciences (Mayr 1969). In this approach attributes considered essential are identified and selected, and used to group organisms, such as mammary glands for the genus *mammals*. This approach has also been the preferred method for taxonomy development in the manufacturing sector, and elsewhere (Bolden 1997, McCarthy 1995, & McCarthy 2000).

FCA is a method for identifying, structuring, and displaying graphical relationships between concepts within domains through underlying mathematical formula (Ganter 1999). Rudolf Wille founded FCA in the early 1980's at the TH Darmstadt, Germany. FCA is an approach based upon the lattice theory of Garret Birkoff (Birkoff 1993), and establishes a formal context based upon a set of objects, a set of attributes, and a binary relation between the objects and attributes. The context is formalized into a mathematical notation and can be represented as a table or a line diagram consisting of nodes. It has proven useful for understanding given knowledge domains using a mathematical framework of evaluation (Burmeister 2000, Darmstadt Formal Concept Analysis Research Group 2000, & Erdmann 1998).

Grounded theory is a qualitative social science research method first formulated in Glaser (1967). The research is neither statistical nor quantitative (Strauss 1990); but instead proposes an inductivist approach which gathers data from a variety of sources, analyzes and codes it, theoretically samples it, and finally generates hypotheses and theories from interpretive procedures in order to understand a particular phenomenon. Although different from the hypothetico-deductive line of inquiry (Glaser 1978), the two approaches share the same capability to predict facts from observed data. In Grounded Theory the method of 'constant comparison' imposes a rigor to the observation of data that permits the researcher to interpret diverse patterns in the data and create higher levels of abstraction from the underlying detailed data observed (Strauss 1987).

Library science's theory of the organization of information is an approach based upon a robust theoretical framework (Dewey 1916). Three basic concepts within the theory of the organization of information relevant to this paper are: indexing, classification and thesaurus construction (Soergel 1974). The concept of abstracting has specifically been left out because it is a distillation and summarization technique, not a method of pointing to or locating a concept.

Indexing is a method of creating signs that point to a smaller subset of a collection of objects, such as the index in a monograph (O'Connor 1996). Classification is a method for grouping entities with a common attribute or a set of attributes (Austin 1981). Thesaurus construction is a method incorporating the basic principles of indexing and classification (Aitchison 1997). The thesaurus relates concepts, (or terms), within a controlled terminology. A thesaurus is comprised of a list of terms and their relationships
expressed in an indexing language. A proposed thesaurus may be evaluated on degree of conceptual completeness, terminological completeness, and thesaurus display (Soergel 1974). A provisional thesaurus may also be used as a taxonomy for describing a knowledge domain.

Ontological terminology systems are representational systems for modeling domains of knowledge through a semantic network. It is often used in the fields of Artificial Intelligence and Medicine (Jouis 1998 & Gangemi 1998). Terms, classification and definitions are described within a formal expressive language. Often, such ontologies have axioms and definitions associated within them as some form of predicate logic or a computer language. A simple ontology is equivalent to a taxonomy (Fischer 1998).

Organizational systematics is a classification method that permits the identification of useful scientific groupings (McElvey 1978). For example, taxa are in common use for unions, associations, industry sectors, manufacturers, businesses, business strategies, and organization types (Amaro 1999).

**STRUCTURE OF A TAXONOMY**

An initial literature search was carried out to determine how taxonomies in other disciplines were created and presented. The literature search was restricted to knowledge organization and knowledge management journals, library and information science journals, management and organizational design journals. The results helped to direct the Grounded Theory approach to positing a Provisional Knowledge Management Taxonomy.

Like a thesaurus and ontology, a taxonomy is constructed within a model of hierarchical relationships (Aitchison 1972). An hierarchy distinguishes a taxonomy from a simple list of terms, and is based upon (National Information Standards Organization 1993):

"degrees or levels of super ordination and subordination, where the super ordinate descriptor represents a class or a whole, and subordinate descriptors refer to its members or parts. The following relationship indicators may express reciprocity:

**BT** (Broader Term), a label for the super ordinate descriptor
**NT** (Narrower Term), a label for the subordinate descriptor."

The hierarchical relationship encompasses three situations that are logically different and mutually exclusive:

1. generic relationship -
   a. a link between a class and its members or species,
   b. an is-a relationship as in *whale is-a mammal*;
2. whole-part relationship -
   a. a link where a concept is essentially included in another concept,
   b. an is-part-of relationship as in *province of Ontario is-part-of the nation of Canada*;
3. instance relationship –
   a. a link between a category of things or events and specific instances of the
      thing or event,
   b. an is-an-instance-of relationship as in Canadian Rockies is-an-instance-
      of mountain ranges.

Each of these types of relationships can be logically tested and validated (ISO Technical
Committee 46 – Documentation 1986).

The proposed taxonomy of this research paper subsumed these into one hierarchical
relationship template, (as specified below). The three relationships employed in this
provisional taxonomy were:

- Hierarchical Relationships:
  - BT - Broader Term
  - NT - Narrower Term

- Associative Relationships:
  - RT - Related Term

- Equivalence Relationships:
  - USE - Use
  - UF - Use for
  - EXP - Abbreviation Expansion
  - ABB - Abbreviation.

In the proposed taxonomy a number of instructions were not used. Their value and utility
will only become evident once the Provisional Knowledge Management Taxonomy has
been distributed for review by domain experts and more in-depth research work begun.
The purposefully omitted instructions included:

- Classification Relationships (facet or node indicators),
- SN - Scope Notes,
- SEE - See Reference,
- SEEN - See from Reference, and
- HN - History Notes.

Software products were acquired to make the task of denoting concepts and relationships
easier to identify and maintain. Additionally, these products provided reporting
mechanisms for listing the terms and relationships. Two firms30, Multisystems from the
USA and Questans from the UK, donated research copies of their respective products,
MultiTes Ver. 7, (http://www.multites.com) and STRIDE Ver. 6.2.3,
(http://www.questans.co.uk) to assist the researcher to collect and display the taxonomy
of terms. The software products furnished a number of views of the taxonomy:

1) Alphabetical—a report of each term followed by all of its relationships with other
terms;
2) Top Term—a report of families of descriptors based on hierarchical relationships
   (BT/NT);

---

30 The researcher is indebted to these firms; and acknowledges their generous assistance in supporting the research goal of this project.
CONSTRUCTION OF A KNOWLEDGE MANAGEMENT TAXONOMY

A taxonomy would normally be generated by either of two methods:

1) consultation with subject authorities or
2) examination of a sample set of the documents related to the domain under consideration (ISO Technical Committee 46 – Documentation 1986).

The researcher chose to employ an abbreviated grounded theory approach to achieve the second method. Since this paper resulted from the development of a term research paper and not from a formal research initiative, the label abbreviated grounded theory approach is purposefully used. Therefore, the use of the grounded theory approach was abbreviated because of the objective of informally proposing a provisional taxonomy.

Grounded theory is an approach to qualitative analysis that is based upon a complex iterative process of data observation that begins with a research situation. The research situation was the lack of a taxonomy to express relationships between terms and concepts used in the emerging field of knowledge management.

The researcher begins by collecting data in the literature and identifying the core theoretical concepts through the selection of a number of sample cases from the literature—the citations being the principle units of data (Pandit 1996). Data collection takes place by means of theoretical sampling—where the researcher decides, during coding and analysis, where next to collect cases in order to purposefully increase the diversity of the sample and search for different concepts (Glaser 1967). The researcher then identifies and documents any proposed relationships between the core concepts and the underlying data. The key analytical strategies used in grounded theory (Trochim 2000) are:

1. Coding – the process of developing initial categories and later moving to higher-level categories and concepts;
2. Memoing – the process of recording thoughts, ideas, relationships throughout the analysis; and
3. Sorting and Reporting – the process of reviewing the details of the data from numerous perspectives and integrating the ideas into a coherent framework or argument.

A theoretical sample of 100 articles and 20 monographs was chosen from a private research collection of 500 journal articles and reports, and 100 monographs. The sample set represented 20% of the base collection. Then, candidate terms were identified by scanning the selected sample set and noting the occurrence of terms within titles, subtitles and section headings, abstracts, figures, tables and bibliographies. The extracted term was a representation of a concept in the form of a noun (term) or noun phrase (compound term). Descriptors (preferred terms) were distinguished from non-preferred terms (synonyms or quasi-synonyms) by applying the USE or ABBREV instructions (National
Information Standards Organization 1993).

The selected terms appeared to be structural landmarks within the documents and books. Candidate terms were selected during this analytical process, coded and memoed. From the coded candidate terms list, significant single and compound terms were chosen for entry into MultiTes or STRIDE. Terms were regularized according to guidelines in the different standards (American National Standards Institute 1974, National Information Standards Organization 1993, & ISO Technical Committee 46 – Documentation 1986). Synonyms and abbreviations were then cross-referenced.

Like the concept identification stage, there are two methods that may be used to determine the relationship of terms:
1) consultation with subject authorities or
2) an analysis of the sample documents.

Again, the researcher chose to rely upon the grounded theory approach to propose relationships and analyze the sample documents for implied or expressed relationships. Formal glossaries and dictionaries of knowledge management terms are unknown; so these could not be consulted.

The volume of terms and concepts, which arose across the theoretical sample, as well as the occurrence of unique terms within a document, were used as criteria when proposing candidate terms and their relationships. Terms were assessed for ambiguity and clarity (Holm 1998). The terms were then evaluated for their homogeneity as well as their heterogeneity. From these investigative activities, a set of tentative relationships were established and reviewed. These relationships have resulted in a provisional taxonomy for discussion, further refinement and research.

PRESENTATION OF A KNOWLEDGE MANAGEMENT TAXONOMY

A non-graphical presentation of a taxonomy presents a number of problems for comprehension by a reader/classifier and an expert team of reviewers. First, the textual view—in terms of reports and lists—relies heavily upon the reader's cognitive capabilities to knit together a tapestry of the overall model in terms of its components. In a large taxonomy this is exceptionally challenging. Second, circular displays or primitive networks displays with coordinate grids, as illustrated for example in Soergel (1974), require significant labor to create and maintain manually. Consequently, the generated reports are text and list bound, not graphical in orientation. Extracted samples of the three primary textual views of the Knowledge Management Taxonomy are presented below from each report to describe a particular situation.

Alphabetical Report
The Alphabetical Report presents each term followed by all of its relationships with other terms. The term knowledge management itself encompasses a wide range of narrower and related terms:
Proposal for a Provisional Knowledge Management Taxonomy

knowledge management
  ABB: EM
  BT: knowledge
  NT: cognition function
      competitive intelligence
      knowledge management institution
      knowledge management life cycle
      activity
      knowledge management methodologies
      knowledge management practice
      knowledge management strategy

knowledge management system
knowledge management tool
knowledge networking
measurement function
art of knowledge
artificial intelligence
cognitive science
content analysis
library science
management science
systems science

Extract 1: Knowledge Management in Alphabetical Report

Top Term Report
The Top Term Report presents families of descriptors based on hierarchical relationships (BT/NT). Here we see the need to differentiate between knowledge form and knowledge type:

knowledge form
  . behavioral knowledge
  . business knowledge
  . collective-individual knowledge
  . best practices
  . lessons learned
  . complex knowledge
  . customer knowledge
  . deep knowledge
  . embedded knowledge
  . empirical-conceptual knowledge
  . institutional knowledge
  . best practices
  . lessons learned
  . knowledge calculus
  . knowledge vector
  . knowledge domain
  . knowledge geometry

knowledge superiority
lost knowledge
operative knowledge
personal knowledge
planning knowledge
positive-speculative knowledge
private knowledge
process knowledge
public knowledge
rational knowledge
religious knowledge
scholarly knowledge
simple knowledge
stigmatized knowledge
symbolic-concrete knowledge
technological knowledge
working knowledge

Extract 2: Knowledge Form in Top Term Report

knowledge type
  . common knowledge
  . common-sense knowledge
  . conceptual knowledge
  . cultural knowledge
  . domain knowledge
  . expert knowledge
  . explicit knowledge
  . formal knowledge
  . hidden knowledge
  . human knowledge
  . implicit knowledge
  . informal knowledge

knowledge-based decision making
complete knowledge
incomplete knowledge
organized knowledge
philosophical knowledge
political knowledge
rudimentary knowledge
scientific knowledge
self-knowledge
social knowledge
specialized knowledge	
tacit knowledge
technical knowledge
traditional knowledge

Extract 3: Knowledge Type in Top Term Report

Hierarchical Report
The Hierarchical Report is similar to the Alphabetical Report; but presents the full hierarchy for all hierarchical relationships. Nesting of terms generally does not exceed 5 levels:
CONCLUSION AND FUTURE RESEARCH

The exercise of identifying, collecting, and relating terms from the broad field of knowledge management provided increased understanding of the knowledge domain the researcher will be working with for his dissertation. Often concepts in this area have inherently ambiguous definitions and use, i.e., synthesis and unity. At the same time, many of the existing concepts are recognized from their origins in other disciplines and fields, i.e., behaviour, or cognition. Researchers would benefit further from the capability to articulate key concepts and terms.

An abbreviated experience with the grounded theory approach provided the researcher with a brief, but insightful, exposure to the power, depth, rigour, discipline and benefits of this qualitative research method. Because of the qualitative nature of the researcher's proposed thesis, the grounded theory approach will be integrated as the foundation of his anticipated contribution to the body of knowledge in the knowledge management field of discourse.

Since the provisional taxonomy has not benefited from a review by other expert authorities, a vetting and consultation process with identified experts, (as a Community of Interest), should be sought as a desirable next step. The publication of this paper could form the initial basis for establishing a forum of informed participants as well as the potential publication of a formal dictionary or glossary from the taxonomy.

The potential conversion of a provisional taxonomy to a graphical presentation tool, such as Inxight Hyperbolic Viewer, (http://www.inxight.com), or The Brain, (http://www.thebrain.com), could provide further capability for constructive criticism of terms and relationships that were not evident in the textual view provided by the text-based reports. Graphical representations have always had an advantage over text in
conveying complex conceptual material.

This graphical version of the taxonomy could be published to the web and promulgated to journal authors and academics in the knowledge management field. The journal authors and academics could experiment with the graphical interface to choose terms and phrases as keywords for their scholarly submissions. Such a use could help market and sell the increased application of this type of keyword authority for publications in the field. Subsequently, precision of retrieval could be tested to see if it has increased since the controlled vocabulary was applied.

Finally, consideration might be given to establishing a micro-thesaurus for the emerging knowledge management discipline. The Provisional Knowledge Management Taxonomy could be used as the starting point for formalizing the relationships and terms into a thesaurus. In such an endeavor, classification relationships, (i.e., facet or node indicators from formal thesaurus construction standards), may take on a more pragmatic value for potential retrieval.

© Copyright 2001, Michael J.D. Sutton

7. REFERENCES


Schlick, M. General Theory of Knowledge. La Salle, IL: Open Court, 1925.


