Enhancing Access to Knowledge Management Literature – Proposal for Domain-Based Classification Scheme and Thesauri

Denise A. D. Bedford
Goodyear Professor of Knowledge Management
Kent State University, Kent Ohio

Communication Culture and Technology, Georgetown University
Research Context


• However, a recent survey of knowledge management practitioners and professionals raised awareness of the challenges involved in finding, accessing and using the knowledge management literature (Bedford 2012).

• The survey responses also highlighted the need for knowledge organization systems and services to facilitate discovery and access in particular.

• This research presents a prototype classification scheme and thesaurus to support the field of knowledge sciences.
Value of Knowledge Organization Systems

- Knowledge organization systems and services fulfill a critical role in any discipline.

- Provide a consensus-based description of the discipline.
- Provide a guide to the literature for those who are somewhat familiar with the field or are new to the field.

- Provide a mechanism for drawing into a single view the historical and the current literature, and they provide an extensible framework for the literature of the future.

- Help us to understand how the discipline relates to and overlaps with other disciplines.

- They support discovery, organization and management of the information in the discipline.
Review of Current Practice

- Knowledge management is a field which is not currently supported by any discipline-specific classification schemes, thesauri, or authority control systems.

- The characterization of knowledge management by knowledge organization systems in other disciplines is suboptimal and needs to be addressed.

- By suboptimal we infer that it is both under-represented and misrepresented in other domain systems.
IEEE Thesaurus

Knowledge management
UF: Intellectual capital
BT: Computer applications
BT: Management
NT: Knowledge transfer
RT: Competitive intelligence
RT: Information management
RT: Knowledge discovery
RT: Knowledge engineering
RT: Management information systems
RT: Semantic Web

Library of Congress

Knowledge management
BT: Management
NT: Information Technology
NT: Intellectual Capital
NT: Organizational Learning

ERIC Thesaurus

Knowledge management
BT: Information management
RT: Access to information
RT: Communities of practice
RT: Human capital
RT: Information dissemination
RT: Information networks
RT: Information science
RT: Information scientists
RT: Information systems
RT: Information technology
RT: Information transfer
RT: Information utilization
RT: Knowledge economy
RT: Organizational communication
RT: Organizational effectiveness
RT: Total quality management
RT: Training
RT: Workplace learning
Challenges and Opportunities

• We should be able to define our own discipline and make KOS available for others to use – publishers, authors, search system vendors, semantic analysis developers

• What we see in these thesaurus representations, though, reminds us of how important it is to begin with a strong knowledge architecture, to involve the discipline in development, and to align with good design principles

• Let me explain how we approached this task and the steps we took along the way

• We hope to make this available to knowledge professionals for collaborative development in the future
The research explores and proposes answers to four questions.

- Question 1: How do we represent the discipline of knowledge sciences?
- Question 2: What does a classification scheme for the discipline look like, and how do we construct it?
- Question 3: What does a thesaurus for the discipline look like, and how do we construct it?
- Question 4: What types of authority control sources are needed to support the discipline, and how can we construct them?

Each question is a research project in its own right with a distinct product.
Research Question 1 – Scope and Coverage

- Innovation
- Intellectual Capital
- Knowledge Technology
- Knowledge Strategy
- Knowledge Architecture
- Knowledge Operations
- Knowledge Asset Mgmt.
- Organizational Learning
- Organizational Culture
- Collaboration/Communities
A classification scheme is a taxonomy that is comprised of a structured set or group of classes.

Classes define “sets” of things that fit the definition or the description of the class.

Traditionally, classification schemes either have no class level definitions, a scope definition or they default to a de facto definition based on assigned members.

This doesn’t work well for classification schemes which are intended to be used for automated classification – in this case a class and its subclasses must be grounded on a deep and explicit definition – at the concept or semantic rule level.
The proposed classification scheme aligns with the definition of the discipline.

The initial version of the scheme is limited to two levels - Primary Classes and Secondary Classes.

- Primary classes align with the facets of the discipline
- Secondary classes represent subcategories within each primary domain.

Each subdomain is a logical subset of a primary knowledge domain. This design allows for the scheme’s scope and coverage as the discipline evolves.
Top Level KM Classification Scheme

- Knowledge Sciences Classification Scheme – Primary Domains
  - 01.00.00 Knowledge Theories & Models
  - 02.00.00 Knowledge Strategy & Leadership
  - 03.00.00 Intellectual Capital
  - 04.00.00 Knowledge & Information Economics
  - 05.00.00 Organizational Culture & Communication
  - 06.00.00 Collaboration and Communities
  - 07.00.00 Knowledge Asset Management
  - 08.00.00 Knowledge Architecture
  - 09.00.00 Knowledge Technologies
  - 10.00.00 Knowledge Assessment and Evaluation
  - 11.00.00 Innovation
  - 12.00.00 Organizational Learning
Build Out of Class Scheme – Step 1

• The secondary classes were identified following a three step methodology.

• Step 1 involved creating a test corpus

• The test corpus was comprised of 500 knowledge management peer reviewed published articles.

• Each article was manually assigned to one primary class.

• In addition, a set of 71 published knowledge management books were selected and classed to at least one primary class.
Build Out of Class Scheme – Step 2

- Step 2 involved converting each article and each back-of-the-book index into a machine readable format for semantic analysis.

- Each article was scanned and converted to text using optical character recognition software. The OCR’d files were then converted to text format.

- Noun phrases were extracted from each of the OCR’d journal articles using the SAS Content Categorization technologies.

- The back-of-the-book index of each book was scanned and converted to optical character recognition, and converted to text.

- Each index entry was cleaned up and converted to noun phrases.
Build Out of Class Scheme – Step 3

- Step 3 involved the creation of extended lists of concepts for each primary class in the scheme.

- On average, each list began as a raw unedited list of between 40,000 and 50,000 concepts.

- Each concept list was manually reviewed and reduced to between 5,000 and 10,000 concepts per primary class.

- The final concept lists were sufficiently rich to identify secondary class clusters.

- These clusters discovered within each concept list were then used to identify secondary classes.
Sample Secondary Classes

03.00.00 Intellectual Capital 07.00.00 Knowledge Asset Management
03.01.00 Human Capital 07.01.00 Knowledge Typologies and Topologies
03.02.00 Structural Capital 07.02.00 Knowledge Organization Systems and Services
03.03.00 Relational Capital 07.03.00 Knowledge Representation
03.00.00 Intellectual Capital 07.04.00 Knowledge Life Cycles
97.05.00 Knowledge Governance Models and Processes
07.06.00 Knowledge Security and Privacy
07.07.00 Knowledge Retention and Preservation
A thesaurus is different from a classification scheme.

We use a thesaurus to support concept level indexing - the thesaurus contains concepts that can be assigned to content either through human or machine assisted indexing.

Thesauri can also be an aid to searching where there are strong relationships among concepts and where the search engine can exploit these relationships.

In order to support both of these goals, a thesaurus must have a rich set of terms, must have a clear description of its warrant, and must have rigorous relationships established among concepts.
Knowledge Sciences Thesaurus

- The Knowledge Sciences Thesaurus must provide an accurate and inclusive description of the discipline.

- At the concept level, a knowledge sciences thesaurus must describe all of the ‘what we do’ and ‘what we know’ aspects of the discipline.

- The Knowledge Sciences Thesaurus should provide a domain-wide, application-independent framework for describing all of the domain’s areas of expertise and knowledge domains, current as well as historical, representing the vocabularies used by domain experts and domain novices, for information and knowledge created within the domain and in related domains.

- The architecture of the thesaurus is critical – the thesaurus and the classification scheme are distinct structures but they must work together.
Thesaurus and Classification Scheme Alignment

Thesaurus Record

Innovative culture
- TOP: 10.00.00 Innovation
- TOP: 05.00.00 Culture and Communications
- UF: innovation culture
- UF: culture of innovation
- BT: culture
- RT: disruptive innovation
- RT: incremental innovation

Primary Classes Assigned to Term

Other Related Concepts
Secondary Classes and Mini-Thesauri

• Each secondary class is supported by a deep definition of concepts - these become controlled vocabularies in the thesaurus.

• Each term is classified to one or more subclasses to ensure that its membership is tracked and its relationships are managed.

• In this way, we can relation terms across vocabularies to represent the multidisciplinary nature of the field.

• The design allows us to maintain subclass definitions and to extract subclass vocabularies and thesauri.
Thesaurus

- We anticipate that a fully elaborated thesaurus will consist of more than 200,000 terms.

- The depth and extent of the thesaurus is necessary because we need to include terms from across multiple disciplines and perspectives.

- Each of the terms in the thesaurus will have deep relationing – particularly synonyms which will help to create links across those disciplines, and to bridge gaps between theory, design, development and practice.

- We have included some sample term relationing from just one of the secondary classes of the classification scheme – Organizational Culture.
Sample Related Terms

- Organizational culture
  - NT: business culture
  - NT: changing culture
  - NT: changeable culture
  - NT: corporate culture
  - RT: organizational culture analysis

- Corporate culture
  - UF: corporate cultures
  - UF: cultures in organizations
  - UF: cultures of firms
  - UF: cultures of organizations
  - UF: cultures of work organizations
  - UF: firm culture
  - UF: firm cultures
  - BT: culture
  - RT: corporate performance

- Culture
  - NT: abstract culture
  - NT: adaptive culture
  - NT: emerging culture
  - NT: folk cultures
  - NT: founding cultures
  - NT: efficient cultures
  - RT: cultural assessment
  - NT: organizational culture
  - NT: collectivist culture
  - NT: cultural artifacts
  - NT: cultural symbols
  - NT: cultural attributes
  - NT: cultural values
  - NT: cultural assumptions
  - RT: cultural ecologists
  - RT: cultural activities
  - RT: cultural assessments
  - RT: cultural antecedents
  - RT: cultural diffusion
• Another key KOS for the discipline is authority control.

• Authority control – in the form of personal name authority control and organizational name authority control – for published literature is well supported by national bibliographic sources.

• However, authority control for the informal literature including grey literature, internal documents, trade journals, conference and white papers, is not well supported.

• Authority control sources for the field of knowledge sciences would help us to identify individuals and organizations who actively engaged in practice.
Authority Control Tests

• We explored opportunities to support authority control for personal name authorities, organization name authorities, and geographic and administrative name authorities

• We leveraged two semantic technologies – the SAS Content Categorization Suite and the Open Calais engine to test the extraction of authorities for authority control systems

• While the technologies did a good job of extracting identifiers, it was clear that a significant amount of human knowledge is required to align the extracted forms with formal records and systems
Observation 1: Classification Scheme

• The prototyping and development of a Knowledge Sciences Classification Scheme is possible where semantic technologies are available to expedite the process.

• Both the Classification Scheme and the Thesaurus, though, rest on a consensus of the scope and coverage of the discipline.

• To ensure the classification scheme aligns with and supports the discipline, though, extensive review and feedback will be required from the discipline.

• Because knowledge sciences is an evolving discipline, there will need to be a strong governance model and an open system for collaborative development going forward.
Observation 2. Knowledge Sciences Thesaurus

• The prototyping and development of a Knowledge Sciences Thesaurus is also possible again leveraging semantic technologies.

• Semantic technologies can expedite the creation of a thesaurus, potentially reducing the time to delivery to under one year.

• Thereafter, the thesaurus would be collaboratively maintained by an academic institution with a knowledge science education program, and through a formal governance structure.

• The governance structure would be responsible for both the classification scheme and the thesaurus.
Observation 3. Knowledge Sciences Authority Control Sources

• While there is a need for authority control sources, particularly name authority and organizational authority sources, semantic technologies offer only a partial solution.

• They may, though, prove to be a starting point for working with national bibliographic sources to improve coverage of the field of knowledge sciences.

• Where identifiers are extracted, they should be added to the queue of national authority control sources and systems.
Observation 4. Custodianship, Support and Governance

- Knowledge organization systems supporting knowledge sciences should be governed as rigorously as are other disciplines such as chemistry, mathematics or psychology.

- However, there can be a single academic custodian, and the governance body can be comprised of members of the several professional societies, of organizations and of individual knowledge practitioners.

- The governance structure should also align with the warrants that undergird the knowledge organization systems.

- The development of new knowledge organization systems for the discipline present an opportunity to define open and collaborative approaches.
General Observation

• We believe this research demonstrates the feasibility of developing KOS tools for the discipline based on prototype development, engagement and ongoing collaboration.

• Prototype tools can be developed by academic institutions, made available as open source tools, and collaboratively built out and tested with the involvement of the existing professional and informal groups, of interested knowledge organizations, and general knowledge practitioners.

• We suggest that this research demonstrates it is possible for one organization or a small group of individuals to jump start the effort.
Thank you!

dbedfor3@kent.edu

db233@georgetown.edu