Knowledge Organization - an introduction

by Dino Karabeg
and Alexander Sigel
Motivation 1

Yr. 1950

The Computer

Yr. 2000

The Web
Motivation 2
Our goal

Organize and study the background knowledge resources
Knowledge Organization
Lecture Plan

• What is KO? Approaches in Library and Information Science
• Survey of KO learning resources
• Computer-science Approach 1: Value Matrix
• Computer-science Approach 2: Methodology
What is Knowledge Organization?

Birger Hjørland
KO in the narrow sense

• In the narrow meaning Knowledge Organization (KO) is about activities such as document description, indexing and classification performed in libraries, bibliographical databases, archives and other kinds of “memory institutions” by librarians, archivists, information specialists, subject specialists, as well as by computer algorithms and laymen.

• Library and Information Science (LIS) is the central discipline of KO in this narrow sense (although seriously challenged by, among other fields, computer science).
KO in the broader sense

- In the broader meaning KO is about the social division of mental labor, i.e. the organization of universities and other institutions for research and higher education, the structure of disciplines and professions, the social organization of media, the production and dissemination of “knowledge” etc.
Central claim

KO in the narrow sense cannot develop a fruitful body of knowledge without considering KO in the broader perspective.
Knowledge Organization

KO as a field of study is concerned with the nature and quality of such knowledge organizing processes (KOP) as well as the knowledge organizing systems (KOS) used to organize documents, document representations, works and concepts.
KO theory

• KO has mainly been a practical activity without much theory

• Practical KO may have been seen as a syntactic rather than as a semantic activity

• The problem is not just to formulate a theory, but to uncover theoretical assumptions in different practices
Syntactic vs. Semantic

- Syntactic labor is determined by the form alone of symbols.
- Semantic labor is concerned with transformations motivated by the meaning of symbols.
The documentalists made a generic concept “document” to include not only books, articles, “records” and object such as globes, but any kind of material indexed to serve as some kind of documentation, including pictures, maps and globes. Even animals were considered documents (if captured and kept in a zoo).
Information

• Introduced by computer scientists as ‘information processing’

• Belief that Shannon’s “information theory” was a long-needed answer to a theory also about libraries and scholarly communication

• Documents are more related to the concept and theory of semiotics
Knowledge

- Knowledge Organization originated in library field ca. 1900
- Henry Bliss: *The organization of knowledge and the system of sciences*, 1929
- Bliss understood “knowledge” in the Platonic tradition as “verified, true belief”
A classification of books to be effective on the practical side must correspond to the relationships of subject-matters, and this correspondence can be secured only as the intellectual, or conceptual, organization is based upon the order inherent in the fields of knowledge, which in turn mirrors the order of nature.

(John Dewey, from preface to Bliss, 1929)
Pragmatist view of KO

Cherry trees will be differently grouped by woodworkers, orchardists, artists, scientists and merry-makers.

(John Dewey, 1920/1948)
Theoretical approaches to KO

1. Traditional approach, ca. 1876
2. Facet-analytic approach, ca. 1933
3. Information retrieval tradition (IR), 1950s
4. User oriented and cognitive views, 1970s
5. Bibliometric approaches, 1963
6. Domain analytic approach, 1994
7. Other approaches
1. Traditional approach

Dewey Decimal Code (DDC), 1876

Melvil Dewey (1851-1931)
The Lancashire cotton industry: a study in economic development

Assigned DDC Code: 338.4767721094276
Principles developed within traditional approach

- Principle of controlled vocabulary
- Cutter’s rule about specificity
- Hulme’s principle of literary warrant (1911)
- Principle of organizing from the general to the specific
Criticism of the traditional approach

- Traditional (early) KO systems lack theoretical foundation
- Dewey’s interest was not to find an optimal system to support users of libraries, but rather to find an efficient way to manage library collections
- Natural order $\rightarrow$ Scientific classification $\rightarrow$ Library classification
2. Facet-analytical approach

- **Personality** is the distinguishing characteristic of a subject.
- **Matter** is the physical material of which a subject may be composed.
- **Energy** is any action that occurs with respect to the subject.
- **Space** is the geographic component of the location of a subject.
- **Time** is the period associated with a subject.
Criticism of the facet-analytical approach

• The underlying philosophical assumption that elements do not change their meaning in different contexts, according to modern theories of meaning, is problematic.
3. Information retrieval tradition (IR)

- 1950s
- System driven (system makes decision what to present to user)
- Query transformation
- Google
- Cranfield experiments (evaluating recall and precision) showed superiority of simple retrieval systems
Criticism of the IR approach

• Ignores users’ cognitive behavior and context.

• Assumption that texts contain all necessary information needed to retrieve them.

• Relevance feedback is based on unverified premises about the users’ ability to evaluate relevance.

• Positivist assumptions: It has mainly been based on statistical averages, and has neglected to investigate how different kinds of representation and algorithms may serve different views and interests.
4. User-oriented views

- Based on empirical studies of users
- Market-oriented
- User determines the relevance
- Knowledge organization done by users (ex. Folksonomies)
Criticism of the user-oriented approach

• Pre-scientific form of knowledge organization

• Positivist averaging (‘one size fits all’): What has been neglected is to develop different representations of the same documents to serve different users.
5. Bibliometric approaches

- Based on using bibliographical references to organize networks of papers, mainly by bibliographic coupling

- Science Citation Index, Kessler, 1963

- Google

- Citations provided by highly qualified subject specialists

- Dynamic, self-organizing
Criticism of the bibliometric approach

- Relation between citations and subject relatedness is indirect and unclear (social vs. intellectual organization of knowledge)
- Does not provide logical structure with mutually exclusive and collectively exhaustive classes
- Explicit semantic relations are not provided
- Bibliometric maps show networks of cooperating authors, while thesauri show ontological links
6. Domain analytic approach (DA)

- Sociological-epistemological standpoint: The Indexing of a given document should reflect the needs of a given group of users or a given ideal purpose.

- Ex. KVINFO, Nynne Koch

- Different points of view need different systems of organization

- Important: Collective views shared by many users. Different paradigms exist in domains of knowledge and need to be identified
Criticism of the domain analytic approach

• The kind of information which is judged relevant for a given task depends on the theory of the person doing the judgment. Ex. schizophrenia - problematic communication between mother and child or genetic factors?

• Terminology of a field as point of departure. Dilemma – how to select a terminology?
7. Other approaches

- Semiotic
- Critical-hermeneutical
- discourse-analytic
- genre-based
- document typology
- mark up languages
- document architectures
Selected issues in KO

• How to integrate socio-cultural differences in KO?
• Multilingual issue in KO
• Ethics in KO
• How to represent work-oriented and organizational environments in KO?
• How to integrate different structures on the Web?
• How to organize multidimensional knowledge?
• The creation of interdisciplinary ontologies
• Who should do KO?
Challenges in transition from book knowledge to networked electronic media

- Learn from the past
- Free ourselves from the past
- Learn from computer science
What KO approaches are represented on TED?
Computerized knowledge organization
Example: Quanta
THANKS!