

Knowledge Organization in the Domain of Digital Libraries: Using Metadata as a Case Study

Shu-Jiun (Sophy) Chen

Computing Centre of Academia Sinica

2007.10.18

Outline

- Introduction
- The Concept of Knowledge Organization
- Metadata Studies
- Process of the Study
- Preliminary Results
- Conclusions and Future Work

Introduction

- Metadata is an emerging domain which comes from cross disciplinary traditions to support digital libraries.
- However, the concepts of metadata are in the shaping.
 - When we identify what metadata is, for example,
 - we all agree that [DC \(Dublin Core\)](#) is one of metadata standards (*data structure standards*).
 - How about [DACS \(Describing Archives: A Content Standard\)](#)? Do we count it as a kind of metadata standards (*data content standards*) ?
 - How about [AAT \(The Art & Architecture Thesaurus\)](#) related to metadata concept? Is it a kind of metadata standards (*data value standards*) ?

Introduction

- Research Questions:
 - What is **the scope of metadata studies** in the context of digital libraries?
 - How to systematically **construct a knowledge map** for an emerging domain such as metadata studies?
 - Present structure and context association of a domain knowledge
- Purpose of the Study
 - To construct systematically a knowledge organization structure for metadata domain

Key Concepts of Knowledge Organization

- Knowledge Organization (KO) is about **activities** such as **document description, indexing and classification** performed in libraries, in bibliographical databases, in archives and in other kinds of “memory institutions” and on the Internet. (Hjørland, 2007)
 - **In the narrow sense** (within LIS) is KO concerned with information in **bibliographical records**.
 - **In the broader sense** is KO about how knowledge is organized in, for example, **scientific disciplines, higher educational institutions, encyclopedias, dictionaries, languages, genres, theories, minds etc.**

Key Concepts of 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 and concepts (Hjørland, 2007). Examples of KOS are (Hodge, 2000)
 - **Term Lists**
 - Authority files
 - Glossaries
 - Dictionaries
 - Gazetteers
 - **Classification and Categories**
 - Subject headings
 - Classification Schemes, Taxonomies, and Categorization Schemes
 - **Relationship Lists**
 - Thesauri
 - Semantic Networks
 - Ontologies

Approaches to Knowledge Organization

- “Traditional approaches”
 - eg. LCC (Library of Congress Classification)
- Management oriented approaches
- Logical and facet-analytic approaches
- Computer based approaches
- Bibliometric approaches
- User oriented and cognitive approaches
- **Domain analytic approaches**
 - Domain analysis (DA) is an approach to information science (IS) that **emphasizes the social, historical, and cultural dimensions of information**.
 - DA’s focus on domains stands in **contrast to the alternative metatheories of cognitivism and information systems**, which direct attention to psychological processes and technological processes, respectively.
- Other approaches (Hjørland and Hartel, 2003; Hjørland, Apr. 2007)

A Pilot Study

Domain analytic approach

- Methodology
 - Qualitative method
 - Grounded Theory-based approach
 - A general methodology for developing theory that is grounded in data systematically gathered and analyzed (Strauss and Corbin 1994 , p. 273).
 - Hermeneutics approach to documents (epistemological stance of this study)
 - Meaning is constructed by the researcher through reading extant metadata texts.

Data Collection

- JCDL 2001~2007 (22 papers)
 - Joint Conference on Digital Libraries (JCDL)
 - About 60 papers related to metadata topic per year
- Keyword: metadata
 - 2001 3/61 papers
 - 2002 4/63 papers
 - 2003 3/56 papers
 - 2004 4/64 papers
 - 2005 3/63 papers
 - 2006 3/59 papers
 - 2007 2/61 papers



Data Analysis

- Grounded Theory-based approach
 - Set of techniques
 - Identify categories and concepts that emerge from text
 - Link concepts into substantive and formal theories
- Open Coding
 - Process of breaking down, examining, comparing, conceptualizing and categorizing data (*Labeling phenomena, Discovering categories*)
- Axial Coding
 - Data assembled in new ways after open coding, by making connections between categories (taking categories and identifying, *expanding out our knowledge of the categories*)
- Selective Coding
 - The process of selecting the core category
 - Then relating it to other categories

Selected Examples of Codes Used in Open Coding

- Metadata generation
- Natural language process (NLP)
- Machine-Learning (ML)
- Automatic metadata-generation
- Human metadata-generation
- Bibliometrics
- Research projects
- Core technologies
- Preservation metadata
- Semantic metadata
- Digital libraries
- Protocols
- Interoperability
- OAI
- Dissemination of digital content
- Digital video library
- Metasearching
- Semantic web
- RDF
- Descriptive metadata
- Collection descriptions...

Selected Examples of Codes Used in Open Coding – Properties and Dimensions

Digital Content _____ Category

Resource Types _____ ***Property***

- video(CPJC0103) _____ ***Dimensions***
- geospatial(CPJC0303)
- image(CPJC0401)
- historical newspapers(CPJC0503)
- scientific literature (CPJC0601)

Selected Examples of Codes Used in Axial Coding — Causal Conditions

- Automatic metadata-generation
- What causes Automatic metadata-generation?
 - For timely access to textual resources in digital libraries

Selected Examples of Codes Used in Axial Coding — Context

- Automatic metadata-generation
 - Quality of automatic assignment of metadata:
high
 - Scalability: large

Selected Examples of Codes Used in Axial Coding — Intervening condition — Consequences

- Rich Metadata (Phenomena)
 - (Causal Conditions)
To fulfill effective search of digital libraries
 - (Intervening condition)
entering metadata efficiently, accurately, concisely
 - (Consequences)
a web-based solution of cataloging tool

Some very Preliminary Results

Knowledge Organization Structure:

Attributes of conceptions of metadata studies

Digital Libraries (challenges, scalability, access)

Digital Content (dissemination, meaningful experience of digital information—semantic web, large collections-CPJC0403, CPJC0404, CPJC0502, CPJC0503, data management-CPJC0601, long-term preservation/web preservation-CPJC0705) (CPJC0201)

(by resource types) Digital Video Library(CPJC0103)

(by resource types) Geospatial / Georeferenced (CPJC0303)

(by resource types) Image Digital Libraries(CPJC0401)/Image Collections (CPJC0404)

(by resource types) Historical newspapers (CPJC0503)

(by resource types) Scientific Literature (CPJC0601)

Purposes (CPJC0503)

Support of Research and Education (CPJC0503)

Services

browsing-CPJC0702) (CPJC0204)

Search (effective-CPJC0204)(quality-CPJC0502)

Discovery (quality-CPJC0502)

Retrieval (Traditional, Collection Understand-CPJC0404)

Collection Understand 【Properties】 view artifacts 【Dimensions】 specific, comprehensive(CPJC0404)

Visualization Tools (CPJC0404)

Semantic Web

Standards

Metadata Studies

Metadata terms (Surrogates)

Metadata Types

Descriptive Metadata (CPJC0502) (CPJC0604)

Issues : descriptive granularity, loss of context (CPJC0502)

Structural Metadata (CPJC0201)

Preservation Metadata (CPJC0705)

Metadata Generation

Automatic metadata-generation (CPJC0101, CPJC0301, CPJC0401) / Automatic Metadata

Assignment (CPJC0402, CPJC0501)

Extraction (CPJC0402)

Core Technologies

Natural Language Processing (NLP) (CPJC0101) (CPJC0402)

Machine-Learning (ML) (CPJC0101)

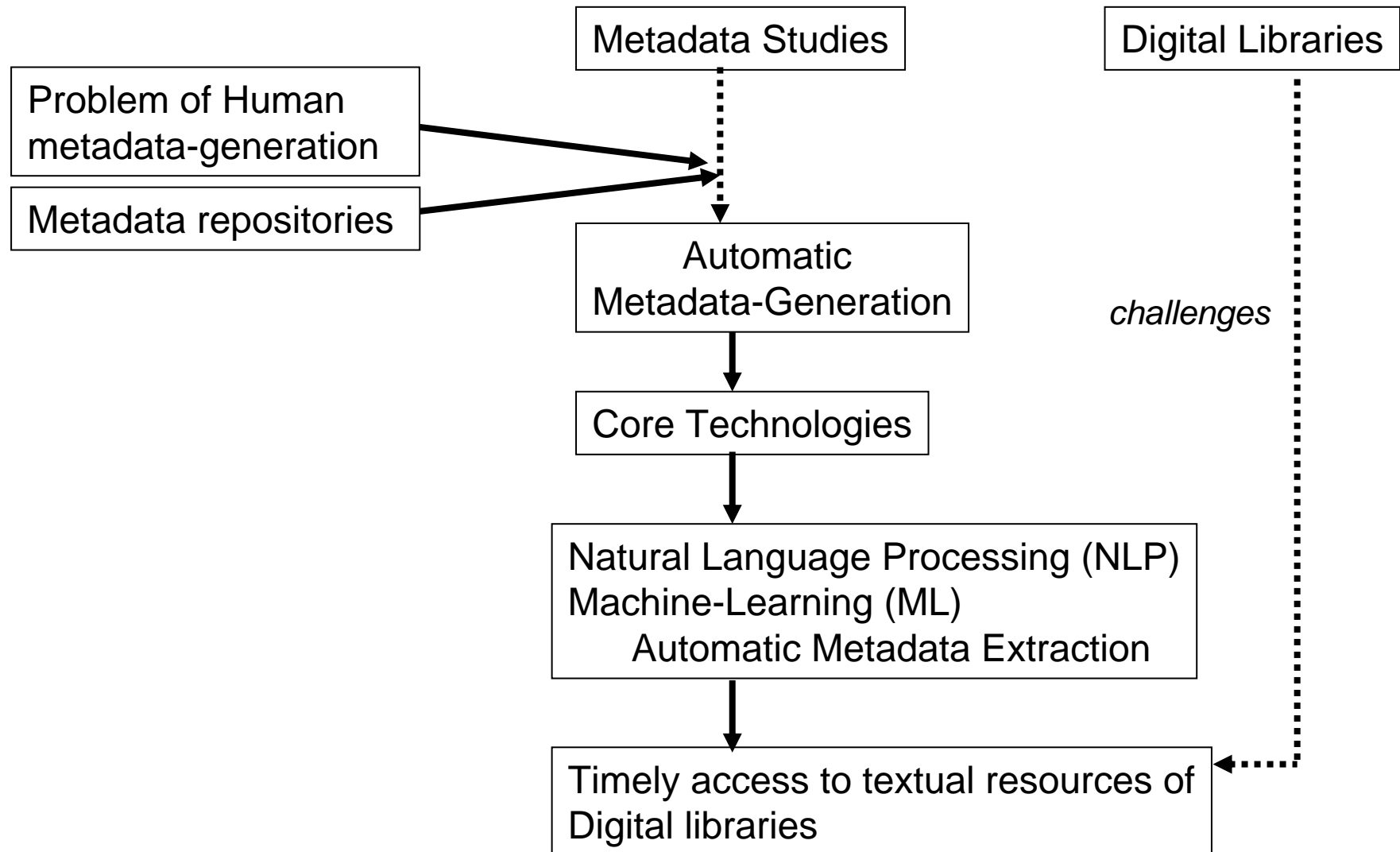
Tools (CPJC0501)

Human metadata-generation

Metadata Tools (Cataloging tools) (CPJC0204)

Some very Preliminary Results

Knowledge Organization Structure



Conclusions and the Future Work

- Applying grounded theory-based methodology applies is a way to explore knowledge organization structure in the domain of metadata studies.
- A more comprehensive investigation will develop a deeper understanding and gain a broader perspective of metadata theory.

Conclusions and the Future Work

- Triangulation
 - Data collection:
 - Include (JCDL), ECDL, ICADL and DC Proceedings
 - Journal papers
 - Monographs
 - Syllabus
 - Interview with lecturers/students who teach/take students who take metadata courses, professionals who plan or implement metadata for the digital libraries
 - Data analysis
 - Develop a refined grounded theory-based method as a reproducible methodology for other domains to construct the knowledge organization structure

Conclusions and the Future Work

- More interactive / friendly
 - How to visualize the knowledge organization structure?
 - Eg. Protégé (a ontology-based tool)
 - Present a storyline of metadata studies to show diversity of multiple disciplines

Thank you