ICWL Workshop 2007

Web-Based Learning in Cultural Science Communities:
Information System Support for Transcription, Localization and Addressing

Marc Spaniol
Jinhua, PR China, 28th of August 2007
Agenda

• Motivation
  - Communities of Practice in the Cultural Sciences
  - Knowledge Organization and Creation in the Cultural Sciences

• Media-centric Knowledge Organization and Creation in Communities
  - Theory of Transcriptivity
  - SECI-Model (Socialization – Externalization – Combination – Internalization)

• Information System Support for Media-centric Working Practices
  - A media theory for computer science
  - Interoperable multimedia metadata
  - LAS (Learning as a Service)

• ATLAS-based Community Information Systems
  - MECCA (Movie E-learning Combination and Categorization Application)
  - MIST (Media Integrated Story-Telling)

• Conclusions and Outlook
Communities of Practice (CoPs) in the Cultural Sciences

- Characteristics of CoPs [Wenger 1998] in the Cultural Sciences
  - Collaboration in the Scope of a (Research)-Project
  - Legitimate Peripheral Participation
  - Limited financial Resources for Computer Scientific Support

- Typical Working Practices
  - Recontextualization of Artifacts based on the underlying Theory
    - Comparison, Discussion und Commenting von Archives
    - Applicable Opinion to be discussed in a specific Context
    - Keeping Discourse going on, not deciding true/false
  - Work on Terminology based on evolutionary Categorizations
  - Individual Modifications of Multimedia Contents (Variations)
  - Creation of Reference Collections

 Fluid Archive
Knowledge Organization and Creation in the Cultural Sciences

• How does Learning take place?
  - Partition of scientific Disciplines based on Snow: “The Two Cultures“ [1959]
  - Linear: Model „Natural Sciences and Engineering Sciences“
    ▪ Old Knowledge is replaces by new Knowledge
    ▪ Medial Artifacts disappear in the Archive, but remain quotable
    ⇒ Knowledge is „transported“
  - Non-linear: Model „Cultural Sciences and Humanities“
    ▪ New Knowledge emerges from the Discourses about old Knowledge
    ▪ Medial Artifacts are kept in dynamic Archives
    ⇒ Knowledge is medial „illustrated“

• How to support the non-linear Model of Knowledge Organization and Creation by means of Information Systems?
  ⇒ Combining:
    Operational Media Theory and Knowledge Management Theory
Theory of Transcriptivity

• Media Operations
  - Transcription: Improving readability and making sense of Pre-Texts
  - Addressing: Transforming "unaddressed" Information in "addressed" Messages
  - Localization: Adopting Media to local Practices

• “Interleaved” Media Operations

• Interdependencies
  Multimedia Content ↔ Community Context not systematized

Transkribieren – Medien/Lektüre
[Jäger, Stanitzek 02]
### SECI-Model

- Combination of western and asian Knowledge Theories
- Continuous Knowledge Transformation: Individual ↔ Community
- Media Operations and Security Aspects disregarded

The SECI-Model is based on the combination of western and asian Knowledge Theories and focuses on the continuous transformation of knowledge between individual and community. Media operations and security aspects are disregarded.

<table>
<thead>
<tr>
<th>Transformation from</th>
<th>Transformation into</th>
</tr>
</thead>
<tbody>
<tr>
<td>implicit Knowledge</td>
<td>explicit Knowledge</td>
</tr>
<tr>
<td>Socialization</td>
<td>Externalization</td>
</tr>
<tr>
<td>Internalization</td>
<td>Combination</td>
</tr>
</tbody>
</table>

**The Knowledge Creating Company**  
[Nonaka & Takeuchi 95]
Information System Support for Media-centric Working Practices

- Synthesis of Theory of Transcriptivity and SECI-Model
- Focus on Media-theoretic Aspects in Knowledge Organization and Creation

[Diagram showing the cycle of Information System Support for Media-centric Working Practices with steps: Pre-Texts, Transcription, practical Localization, formal Localization, Addressing, Transcript in Context, Transcript, and EC-TEL 2007 [Spaniol et al.]]
Transcription in Information Systems

- Ongoing Processes and Implications for Information Systems:
  - Creation and Management of (Multimedia)-Metadata
  - Illustration of (Multi)-Media Dependencies
  - Extension of the Archive by additional Multimedia Contents

⇒ Implications for Metadata Management:
  1. Multimedia Descriptors
  2. Graph Representations
  3. Extensibility of Schema and Descriptors
Transcription

Media in general:

\[ M \in \{ M_{\text{Text}}, M_{\text{Image}}, M_{\text{Audio}}, M_{\text{Video}} \}^n \cup \{ P \}, \ n \in \mathbb{N}_0 \]

Media in Information Systems:

\[ M \setminus P \]

Transcripts:

\[ t_p := (\tau(p), \mu(p), \iota(p)) \]

Transcription:

\[ \tau : P \rightarrow M \setminus P \]

Meta-Transcription (manually):

\[ \mu : P \rightarrow M_{\text{Text}} \cup \emptyset \]

Meta-Transcription (in Information Systems):

\[ \iota : P \rightarrow M_{\text{Text}} \]
Formal Localization in Information Systems

- Ongoing Processes and Implications for Information Systems:
  - Individual Media Categorizations
  - Creation and Management of Reference Collections
  - Fine-grained Access Rights

⇒ Implications for Metadata Management:
  1. Classification Schemes and Vocabularies
  2. Multimedia-Variations
  3. Copyright-protection and -management
Formal Localization: Keyword Index / Categorization

Keyword Index:
\[ S := \{s\} \]

Access Relation:
\[ \lambda_i : S \rightarrow \mathcal{P}(M) \]

Categorization:
\[ K := \{k\} \quad k := (id, s) \]

(Tree)-Order:
\[ o := (k, k', \kappa(k_{id}, k'_{id})) \]

Access Relation:
\[ \lambda_{k_{id}} : ID \rightarrow \mathcal{P}(M) \]
Formal Localization: Local Access Relations

Global Access Relation:
\[ I_{\text{global}} := \{ K_{\text{global}}, O_{\text{global}}, \lambda_{\text{global}} \} \]

Local Access Relation:
\[ I_{\text{local}} := \{ K_{\text{local}}, O_{\text{local}}, \lambda_{\text{local}} \} \]

Operations in Categorizations:
- Insert
- Delete
- Rename

Operations in Media Relations:
- Add
- Delete
Addressing in Information Systems

• Ongoing Processes and Implications for Information Systems:
  - Illustration of (Multi)-Media Contents in Context
  - Opportunities for Legitimate Peripheral
  - (Asynchronous) Cooperation Support

⇒ Implications for Metadata Management:
  1. End Device Adaptation
  2. Profiles
  3. User Preferences und Usage History
<table>
<thead>
<tr>
<th>Addressing</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Discussions:</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Hypermedia Documents:</strong></td>
</tr>
<tr>
<td><strong>Discussions on hypermedia Documents:</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Context-sensitive Addressing of Multimedia Contents:</strong></td>
</tr>
</tbody>
</table>
Practical Localization

- Ongoing Processes:
  - Process of Perceptions
  - Understanding of Interrelations
    “Multimedia Contents ↔ underlying Theory”
  - Opportunities to Comment on
    multimedia Comment and/or underlying Theory

⇒ “Outside” of the Information System
⇒ No direct Implications on Metadata Management
Practical Localization as a social Process:

\[ \sigma_n : T \times H \rightarrow P \]

Cognition / Interpretation of Multimedia Artifacts by Community Members

\[ \sigma_n (t,h) := p \]

⇒ Not measurable within the Information System
   (but, e.g. implicitly measurable via Usage History)
⇒ New, “lived” Experiences originates
⇒ Process might by repeated infinitely often
## Interoperable Multimedia Metadata

<table>
<thead>
<tr>
<th>Multimedia Descriptors</th>
<th>Transcription</th>
<th>formal Localization</th>
<th>Addressing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dublin Core</td>
<td>0/+</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>SMIL</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>IEEE LOM</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CDWA (Lite)</td>
<td>0</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>RDF</td>
<td>0/+</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>MCF</td>
<td>+</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>PICS</td>
<td>0/+</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>OWL</td>
<td>-</td>
<td>0</td>
<td>-/0</td>
</tr>
<tr>
<td>MPEG-7/21</td>
<td>+</td>
<td>0/+</td>
<td>+</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Multimedia Descriptors</th>
<th>Transcription</th>
<th>formal Localization</th>
<th>Addressing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dublin Core</td>
<td>0/+</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>SMIL</td>
<td>+</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>IEEE LOM</td>
<td>-</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CDWA (Lite)</td>
<td>0</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>RDF</td>
<td>0/+</td>
<td>+</td>
<td>0</td>
</tr>
<tr>
<td>MCF</td>
<td>+</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>PICS</td>
<td>0/+</td>
<td>0</td>
<td>-</td>
</tr>
<tr>
<td>OWL</td>
<td>-</td>
<td>0</td>
<td>-/0</td>
</tr>
<tr>
<td>MPEG-7/21</td>
<td>+</td>
<td>0/+</td>
<td>+</td>
</tr>
</tbody>
</table>

**Legend:**
- **+** qualified
- **o** neutral
- **-** unqualified
ATLAS
Architecture for Transcription-, Addressing and Localization Systems

How to develop Community Information Systems that meet the requirements of Working Practices in the Cultural Sciences?

⇒ Application Server for discourse and media-centric Working-practices:
  - Semantic Freedom in Annotating and Categorization of multimedia Artifacts
  - Context-sensitive Media Archives
  - Community Participation during Design and Implementation

socio-technical Information System Development

Self-Monitoring Tools

Measurement, Analysis & Simulation

Community-specific IS

ICWL 2003: [Spaniol et al.]
J.UKM 2006: [Klamma et al.]
EC-TEL 2006: [Spaniol]
**Key Features:**

- 3-Tier Architecture
- Connectors
- Components (internal)
- Services (internal & external)
- Java based Implementation
- Access Control on Service- and Object-Level
- MPEG-7/21 compliant Multimedia-Services
- Extension and Replacement of Services at Runtime
Access Control in LAS

Relevant Data
- Users / Agents
- Groups (hierarchical ordered)
- Security Objects (e.g. Multimedia Artifacts)

Access Right Management
- Permissions / Prohibitions
- Roles (Method Level)
- Access Control Lists (Object Level)
MPEG-7-Services in LAS

Key Features:
- MPEG-7 compliant Creation, Retrieval, Update, Deletion and Validation of Contents
- Persistency based on a MPEG-7 XML-Database (IBM, Oracle, eXist)
- Automatic MPEG-7-Binding via Apache XMLBeans
XML-Databases and MPEG-7

- Open-Source Database eXist
  - XQuery- and XUpdate Support
  - Indexing
  - Lack of Definition and Validation according to a given Schema
  - Lack of Transaction Control

- Oracle 10g
  - XQuery Support "encapsulated" in SQL
  - Transaction Control
  - Validation according to Schema only for simple ones (e.g. without Inheritance)
  - Lack of Indexing for complex Schema (e.g. MPEG-7) \(\Rightarrow\) very slow

- IBM DB2 Version 9 (Viper)
  - XQuery Support "encapsulated" in SQL
  - Transaction Control
  - Validation according to Schema only for small ones (e.g. not MPEG-7)
  - Despite lack of Indexing for complex Schema (e.g. MPEG-7) very fast
Multimedia Management in LAS

Key Features:
- Media Management via associated Community-FTP-Server
- LAS Access Control
LAS-Summary

• Community Support
  - Java Platform Independence
  - Easy to Maintain and Extend at Runtime
  - Core Services for Community-Administration
  - Fine-grained Access Control
    - Method Level
    - Object Level
      ⇒ „Security-First“-Concept
      ⇒ Platform for Community-Hosting

• Multimedia Support
  - Interoperable (Multimedia) Metadata
  - Cross-Community Support by MPEG-7/21 Compliance
  - Automatic Conversion to/from Dublin Core
    ⇒ Schema-compliant MPEG-7/21 Modifications via XMLBeans

⇒ Better Community and Multimedia Support than comparable Application Server such as Tomcat or Zope/Plone
Support of Film Study Communities in Research and Education by:

- Exploration of Multimedia Artifacts
- Illustration of Multimedia Dependencies
- Creation of Reference Collections
- Comprehensive Discourse Support
- Blended Learning
- Media Screenings 24/7 via Internet
MECCA – Transcription

- Creation and Management of
  - Media Collections
  - Technical and semantic Metadata
  - Multimedia Variations

- Automatic DC ↔ MPEG-7 Conversion

- Recall of already stored Metadata Descriptions
MECCA – formal Localization

Media Frame

- Information System Support for „fluid Archives“
  - Work on Terminology via evolutionary Categorizations
  - Creation and Management of Context-sensitive Access Relations
  - Multimedia Management via associated FTP-Server

- Contrast: Community Context ↔ Individual Context
MECCA – Addressing

- Media Screenings via Internet 24/7
- "Semantic Browsing" of Multimedia Contents
- Context Information via
  - Categorization
  - Technical Metadata
  - Semantic Metadata
- Recall of Discourses about
  - Media
  - Categorizations

"activating" Relations

"noticing" Relations
MECCA – Experiences

• Evaluated by
  - Project „Face in Film“ (SFB 427 „Media und cultural Communication“)
  - Students at Ruhr-University of Bochum as Blended Learning Materials in the Seminar „Pathos & Passion“

• Positive Aspects
  - 24/7-Availability of Multimedia Contents in Research & Teaching
  - Common Terminology Work in spatial distributed Communities

• Usage of MPEG-7
  - Schema Size and Complexity induces Errors
  - Databases do not provide sufficient Schema Support
  \[\Rightarrow\] MPEG-7 Services required (without Protection of semantic Misinterpretation)

• Open Issues
  - Illustration of episodic Knowledge in Context
  - „Chaos“ (Folksonomies) vs. „Structure“ (Categories)
Support of an International Community of Researchers in order to Preserve the Afghan Cultural Heritage by:

- Exchange of Multimedia Artifacts
- Protection of geo-spatial Data of unauthorized access (e.g. Tomb Raiders)
- Recontextualization Multimedia Contents via non-linear Stories
- MPEG-7 based semantic Multimedia Annotations

ICWL 2006: [Spaniol et al.]
Knowledge Management Strategies 2007: [Spaniol et al.]
MIST – Transcription

• Creation & Management:
  - Media Collections
  - Metadata
  - Media Variations

• Collaborative Indexing based on:
  - Free Text Annotations (a lá Flickr)
  - Semantic MPEG-7 Basetypes
    ▪ Agent
    ▪ Event
    ▪ Concept
    ▪ Object
    ▪ Place
    ▪ Time
    ▪ State
**MIST – formal Localization**

- Illustration of episodic Knowledge
  - Modeling of non-linear Stories based on Movement Oriented Design (MOD) [Sharda 2005]; Cooperation with Victoria Universität Melbourne, Australia
  - Decomposition of Stories according to a Problem-Hierarchy
  - Recall of semantic high-quality Metadata on Multimedia Artifacts
MIST – Addressing

- Non-linear Access Sequences in multimedia Stories, i.e. „Success vs. Failure“ depending on Access Sequence
- Context Information via Multimedia Annotations
MIST – Experiences

• Evaluated by
  - Researchers at the Chair of Urban History, RWTH Aachen
  - Working Group of Prof. Nalin Sharda at Victoria University Melbourne, Australia

• Positive Aspects
  - MOD-Structure serves as guideline for the creation of “useful” stories
  - Web 2.0 alike participation

• Usage of MPEG-7
  - „Folksonomies“ based on MPEG-7 semantic Basetypes
  - MPEG-7 Basetypes used for geo-spatial and spatio-temporal Descriptions
  - Semantic Search by regular Expressions
  ⇒ Annotations with MPEG-7 Basetypes increase the Data Quality

• Open Issues
  - Application on mobile End Devices (e.g. Usage of MPEG-21)
  - Community-specific Folksonomies
Conclusions
Experiences in Using LAS

• Community-Information System Development with LAS
  - A media theory for computer science as a Model of Knowledge Organization and Creation
  - Allows an „evolutionary“ Service Development
    ▪ Easy to Maintain and Extend at Runtime
    ▪ Core Services for Community-Administration
  - Interoperability via
    ▪ MPEG-7/21 Compliance
    ▪ Dedicated MPEG-7-Services

⇒ Platform for Hosting of Multimedia-Communities

• Experiences in MPEG-7
  - Size of Schema and Complexity are problematic ⇒ MPEG-7-Profiles
  - MPEG-7 Validations on different Levels
    ▪ Syntax ⇒ MPEG-7 Services
    ▪ Semantic ⇒ „open Issue“
  - Insufficient Database support for XML and MPEG-7
Outlook

• Transfer of Results in different Application Domains
  - Entrepreneurial Training (VEL – Virtual Entrepreneurship Lab)
  - Plastics Technology (TRAMP – T.ool for R.epresentation & A.nnotation of M.ultimedia Content in P.lastics Engineering)

• Community Activities:
  - Multimedia Metadata Community (www.multimedia-metadata.info)
    - International Experts from Research and Industry
    - Biannual Workshop Series (7th in September 2007)
    - Joint Development of MPEG-7 based Applications
  - Web-based Learning
    - Participation, Promotion and Support of ICWL
    - Joint Research on Web-based Learning Applications

• Further Development of LAS towards mobile Community-Hosting
  - GPS-Embedding
  - Mobile End Devices

• Conceptual Enhancements
  - Graphical Editors for User Interface Design in Communities
  - Modeling of Media Usage based on the Actor-Network-Theory
  - Analysis of “Agency and Patienthal” Phenomena in Communities