Semantic Digital Libraries: Improving Usability of Information Discovery with Semantic and Social Services

Sebastian Ryszard Kruk
Presentation Outline

- Navigation
- Problem statement and hypothesis
- Ontologies
- The prototype
- Evaluation
Problem and Hypothesis
Problem Statement

Digital library users are
- **missing a librarian** => problems with information discovery and understanding complex metadata
- **missing peers** => cannot share experience with other users visiting the library
- **missing connection with other sources** => library resources cannot become a part of the information processing workflow

Digital Library system
- **knowledge organization systems** => islands of highly organized information
- **poor information discovery** => loosing their position to other sources
- **incompatible taxonomies and schemata** => loosing potential of rich metadata
Hypothesis

Semantic and social technologies in digital libraries improve information discovery compared to classic approaches:

- Users find information **more easily**
- **Precision** in searching is improved
- Users’ **satisfaction** is increased
- Users **retain** more information
Semantic Digital Libraries: Architecture & Ontologies
SemDL Architecture

Existing reference digital library architectures
- Alexandria DL architecture (Frew et al, 1998)
- DELOS reference model (actors) (Candela et al, 2007a)
- Interaction Triptych Model (Fuhr et al, 2007)

Missing:
- Object Model: integration of metadata, reuse of library resources
- Digital Library Services: interoperability, user annotation, advanced search and browsing

Published in: Kruk et al., 2005 (DEXA); Kruk and McDaniel, 2008 (Springer); Kruk et al., 2009 (accepted to TEL)
Ontologies for SemDL

Requirements:

- Support for a complex and dynamic structure of information objects; reuse, aggregation; scientific publications workflow
- Support for reach, interconnected and interoperable bibliographic metadata; align existing concepts, e.g., MARC21, BibTeX, Dublin Core, SKOS, Address Ontology
- Support for communities of library users: FOAF, SIOC, Tom Gruber’s Tagging Ontology
- Support for rights management; model based on ODRL and XACML

Published in: Kruk et al., 2005 (DEXA); Kruk and Haslhofer, 2006 (NKOS, ECDL); Kruk and McDaniel, 2008 (Springer); Kruk et al., 2009 (accepted to TEL)
SemDL Ontologies Example

- **Ontologies designed**: JeromeDL structure ontology, MarcOnt bibliographic ontology, FOAFRealm/SSCF ontology, Extensible Access Control ontology, S3B Tagging Ontology

- **Ontologies used**: FOAF, SKOS, SIOC, Address ontology

Published in: Kruk et al., 2005 (DEXA); Kruk and Haslhofer, 2006 (NKOS, ECDL); Kruk and McDaniel, 2008 (Springer); Kruk et al., 2009 (accepted to TEL)
Semantic Digital Libraries: Navigation
Social Semantic Collaborative Filtering

Motivation

- support identifying and finding experts, and propagating their expertise
- allow to express users’ interests and filtering knowledge base using disambiguation mechanisms
- feature security mechanisms for efficient and secure information gathering and dissemination

Model

- graph of quantified social relations
- graph of inclusions of collections annotated with KOS concepts social relations
- access control based on the position in the social network

Published in: Grzönkowski, Gzella, Kruk, et al., 2009 (Journal of Web Based Communities); Choi, Kruk, et al., 2006 (IRW, WWW); Kruk, et al, 2006 (ASWC); Kruk and Decker, 2005 (Semantic Desktop Workshop, ISWC)
Published in: Grzonkowski, Gzella, Kruk, et al., 2009 (Journal of Web Based Communities); Choi, Kruk, et al., 2006 (IRW, WWW); Kruk, et al, 2006 (ASWC); Kruk and Decker, 2005 (Semantic Desktop Workshop, ISWC)
Evaluation of SSCF Model

Question for evaluation:

Is the social network better informed with SSCF?

Assumptions for evaluation model:

- The quality of the information provided by a user on a certain collection is proportional to the expertise level of the user on the topic of the collection.
- It is possible to find a user with a high expertise on the given topic within the social network.

Evaluation setup:

- a model of the social network - 1000 users
- distribution of relationships: bell-curved ($\mu = 25$, $\sigma = 12.5$) and zipfian ($\theta = 1.9$)

Measuring:

- Average Maximal Expertise ($R$) - average value of the highest expertise level found within given degree of separation ($R$)
Evaluation of SSCF Model

Q1: Can a user access information gathered by the domain experts?

- For Zipf’s distribution maximal average expertise for R=6 is 91% - answer: very probable
- For Bell-curved distribution maximal average expertise for R>3 is above 96% - answer: even more probable

Q2: Is the average expertise level higher in the social network?

- For both types of distributions the average expertise of a single member (R = 0) is much lower than in the social network.

Published in: Grzonkowski, Gzella, Kruk, et al., 2009 (Journal of Web Based Communities); Choi, Kruk, et al., 2006 (IRW, WWW); Kruk, et al, 2006 (ASWC); Kruk and Decker, 2005 (Semantic Desktop Workshop, ISWC)
Shortcomings of faceted navigation

- RDF is not a homogeneous information space
- Join operator is **unintuitive** to the end user (Oren, 2006)
- No filtering based on only given value
- No union and difference operators
- Most of solutions are monolithic (no MVC)
- Poor accessibility: information overload

Extended model

- Extensions to inverted and existential operators
- Browse and similarity operators
- New combination operators: union, difference, binding

Published in: Kruk et al., 2007 (ODBASE)
MultiBeeBrowse

- Zoomable User Interface: basic, structured, browsing and complete history view
- Collaborative Browsing (using SSCF and RSS)
- Adaptable Browsing Interface (incl. concepts suggestions, facets labeling, results presentation)
- Services for Accessible Faceted Navigation
- Model of meta-operations:

![Diagram of decision trees](attachment:Diagram.png)

Published in: Kruk et al., 2007 (ODBASE)
Evaluation

Comparing three solutions:
- **MultiBeeBrowse**
- **BrowseRDF**
- **Longwell**

Features comparison:

<table>
<thead>
<tr>
<th>Operator</th>
<th>MBB</th>
<th>Browse RDF</th>
<th>Longwell</th>
<th>Other</th>
</tr>
</thead>
<tbody>
<tr>
<td>search</td>
<td>+</td>
<td>±</td>
<td>±</td>
<td>-</td>
</tr>
<tr>
<td>selection</td>
<td>+</td>
<td>±</td>
<td>±</td>
<td>±</td>
</tr>
<tr>
<td>exist. property</td>
<td>+</td>
<td>+</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>browse</td>
<td>+</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>combine</td>
<td>+</td>
<td>±</td>
<td>±</td>
<td>±</td>
</tr>
</tbody>
</table>

Published in: Kruk et al., 2007 (ODBASE)
Semantic Digital Libraries: JeromeDL - the prototype
JeromeDL

Semantic digital library project based on cooperation of

- Gdańsk University of Technology
- DERI, National University of Ireland, Galway

Distributed under Open Source (BSD) license

10+ instances worldwide:

- DERI, Ireland: Library, Books, EastWeb DL
- GUT, Poland: WBSS, Kashebian, PMR Journal
- INEGI, Mexico: internal digital library
- dContentWare, Italy: core of the project
- Bosco Inc., India: 1000+ resources
- WKU, KY, USA: learning materials repository

Published in: Kruk, Decker and Zieborak, 2005 (DEXA); Kruk et al., 2007 (Semantic Web Challenge, ISWC), Kruk et al., 2008 (ECDL), Kruk and McDaniel, 2008 (Springer)
Differentiators of JeromeDL

- combining semantic bibliographic descriptions and social media
- advanced, personalized search solutions
- social networking platform integrated with user profiling component
- extensible access control system based on social network relations
- collaborative filtering and browsing
- dynamic collections
- integration with other Web 2.0 services

Published in: Kruk, Decker and Zieborak, 2005 (DEXA); Kruk et al., 2007 (Semantic Web Challenge, ISWC), Kruk et al., 2008 (ECDL), Kruk and McDaniel, 2008 (Springer)
3-layered Architecture

Published in: Kruk, Decker and Zieborak, 2005 (DEXA); Kruk et al., 2007 (Semantic Web Challenge, ISWC), Kruk et al., 2008 (ECDL), Kruk and McDaniel, 2008 (Springer)
Search and Browsing

- **TagsTreeMaps** - filtering with hierarchical tags
- **MultiBeeBrowse** - social browsing
- **Dynamic collections** - defined based on triple filtering and SPARQL queries
- **Recommendations of related resources** based on semantic resource description
- **Query templates** in natural language
- **Semantic Query Expansion** based on user’s context and semantic annotations
- **Social Semantic Collaborative Filtering**
- **flexible API** for integration of external services, e.g., Exhibit (SIMILE, MIT)

Published in: Kruk, Decker and Zieborak, 2005 (DEXA); Kruk et al., 2007 (Semantic Web Challenge, ISWC), Kruk et al., 2008 (ECDL), Kruk and McDaniel, 2008 (Springer)
Semantic Digital Libraries: Evaluation
Evaluation Procedure

Evaluating usability (system, user)

Two digital libraries in their basic (vanilla) setup
- JeromeDL - semantic digital library
- DSpace - classic digital library (control group)

Database:
- noise: 529 articles from DERI JeromeDL instances
- reference set: 35 articles on Internet psychology

Participants: 59 commenced evaluation, 26 completed

Published in: Kruk et al., 2008 (ECDL), Kruk and McDaniel, 2008 (Springer)
Questions for Evaluation (1)

Do semantic and social services improve the quality of answers?
- Slightly better results for JeromeDL group, improving significantly over time (results statistical significance close to acceptance threshold)

Do semantic and social service increase the quality of references provided by the participants?
- Slightly better results for JeromeDL group, improving significantly over time (could not confirm statistical significance)

Do semantic and social service increase the satisfaction from using a digital library? (statistical significance significance)

Published in: Kruk et al., 2008 (ECDL), Kruk and McDaniel, 2008 (Springer)
Questions for Evaluation (2)

Which services are found to be most useful?
- recommendations and social filtering (results statistically significant)

Do semantic and social services increase information retention? (results statistically significant)

Quality of answers: JeromeDL - 2.78, DSpace - 2.44

Accuracy of references: JeromeDL - 6, DSpace - 1

Satisfaction:

<table>
<thead>
<tr>
<th></th>
<th>JeromeDL</th>
<th>DSpace</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding</td>
<td>29.11</td>
<td>10.89</td>
</tr>
<tr>
<td>Easy of execution</td>
<td>2.00</td>
<td>-17.22</td>
</tr>
<tr>
<td>Intuitiveness</td>
<td>21.11</td>
<td>-1.00</td>
</tr>
</tbody>
</table>

Would you like to continue using this library?

Published in: Kruk et al., 2008 (ECDL), Kruk and McDaniel, 2008 (Springer)
Conclusions
I have presented

- Architecture and ontologies for Semantic Digital Libraries
- Examples of search and browsing services:
  - Social Semantic Collaborative Filtering
  - MultiBeeBrowse
- JeromeDL - the prototype
- Evaluation of semantic and social services
What about hypothesis?

Semantic and social technologies in digital libraries improve information discovery compared to classic approaches:

- Users find information more easily ✓
- Precision in searching is improved ✓
- Users' satisfaction is increased ✓
- Users retain more information ✓
The Impact


30+ Papers (excluding 9 chapters in the book):


- **MultiBeeBrowse** - ODBASE 2007, Conference on Teaching and Learning 2007, CHI 2008

- **Social Semantic Collaborative Filtering** - Semantic Desktop at ISWC, 2005

- **NLQ** - IADIS International Conference WWW/Internet 2006


- **HyperCuP** - ESWC Demo Session 2006


3 Invited talks: EPFL, UCD, Polish Information Processing Society

3 workshops: Irish DL Summit, Web Archiving, Special Session at NKOS 2006

10+ open source projects - corrib.org, opensource.knowledgehives.com

17 MSc Theses supervised at GUT

Startup company (Knowledge Hives) continuing R&D efforts initiated in SemDL domain
Semantic and Social Services Improve Usability of Information Discovery in Semantic Digital Libraries

Sebastian Ryszard Kruk
http://www.sebastiankruk.com/