Interoperability for Repository Federations
November 21st, 2007, Graz, Austria

Fridolin Wild
Vienna University of Economics and Business Administration

Agenda
- The iCamp Project
- Interoperability
- Retrieval in Repository Federations (ObjectSpot)
- (Preliminary) Retrieval Evaluation
- Future Work
- Outlook: Networking Feed-Based Microcontent

The iCamp Project
- heterogeneous tools
- persons in distributed locations
- content as "social currency"
- Common model for competence acquisition
Augmented Landscapes

Individuals use subsets of tools and services provided by institution actors can choose from a growing variety of options
actions span a range of services gradually transcend institutional boundaries actors appear as emigrants or immigrants leave and join institutional landscape for particular purposes

Interoperability (WP3)

Interoperability is a property that emerges, when distinctive information systems (subsystems) cooperatively exchange data in such a way that they facilitate the successful accomplishment of an overarching task.

(Wild & Sobernig, 2006)

The Concept of Interoperability

(modified from Kosanke, 2005; IEC, 2005)
Potential Levels of Integration

Presentation Integration
- portlets
- webapp mash-ups

Remoting: Service Orientation
- Information Retrieval
- Information Filtering

Information Integration & Dissemination
- Data Integration
- Data Exchange

Repository Interoperability

SQI
Mediator Service
The Net: SQI Targets
Typical Integration Patterns
Portlet & Portlet Configurator
Federated Ranking

Wild, 2007
Retrieval via SQI

Mediator
- Mediates
- Encapsulates complexity
- Eases consumer access
- Enables performance optimisation (e.g. caching, aggregation, …)
- Enables query routing
- Basis for an overlay structure (if necessary)

Typical Integration Patterns
All together significantly more than 10,000,000 artefacts!
Trend: Even more Scattered Collections

- Survey in (mainly) Central Europe (100 universities) showed:
  - Universities like sharing with selected partners (42%) and publicly for free (29%)
  - But: they don’t manage their assets
    - Only 18% have a repository accessible to outside
    - Only 16% have a cross-organisational repository
- Institutional Repositories emerge (LCMS, Video-Lectures, Fides, …)
- Microformats are popping up (enabling for embedded, even individual repositories)
- Trading platforms emerge and grow (e.g. slideshare, EducaNext, preprint servers, …)

Federated Ranking

- Federated search: multiple search engines
- Central mediator (also back-end)
- Mediator facilitates access to multiple repositories
- Repositories may contain different data and metadata
- Sorting and ranking algorithms ensure relevance
- User interaction provides feedback for ranking improvement

Federated Ranking (Wild, Sigurdarson, Kreuzinger & Chvatal, 2007)
Federated Ranking Algorithm

(Preliminary) Evaluation Results

RHL vs. Precision & Recall

Preliminary Results

Ranked Half-Life (Borlund 1998)

- Median of grouped continuous data
- Result position where half of the relevant documents are found before (lower value = better)
- e.g. RHL value 5: Half of all relevant documents in the answer set within top 5 results.

\[ M_r = L_m + \left( \frac{n/2 - \sum f^2}{F(\text{med})} \times CI \right) \]

- \( L_m \) = lower real limit of the median class
- \( n \) = number of observations
- \( \sum f^2 \) = cumulative relevance values up to and including the class preceding the median class
- \( F(\text{med}) \) = relevance value of the median class;
- \( CI \) = class interval, commonly in IR = 1.
Results

- Ranked: squared weights for titles
- Modified: equal weights
- Precision: only in the current subsample of the 1.000 documents big corpus

Results (2)

Future Work: Retrieval

- Implicit Feedback
- Collaborative Foraging Strategies
- Thorough Evaluation
Future Work

- Investigate implicit feedback means:
  - harvest behaviour traces
  - e.g. local feedback from detail views
  - e.g. track query modifications to adjust weights for disambiguation
- Investigate Collaborative Foraging Strategies
  - e.g. through investigating tagging
  - e.g. through investigating digest sharing
- Repeat Evaluation with Reuters Corpus
  - Simulation of specific situations (bad fast repos vs. good slow reps; varying noise levels to simulate varying topicality of a repository federation)

Thank You

www.icamp.eu

Outlook: Distributed Feed Networks

Feedback Spec
Wordpress Plug-In
Interaction Standards

The Missing Link: FeedBack

Example: WordPress (1)
Example: WordPress (2)

Example: WordPress (3)

Future Work

- Study interaction behaviour with a trial
#eof.
BEWARE, THE END IS NEAR.