

OAI6: Breakout Group 6
Access Data Mining: A new foundation for Added-value services in full text repositories

Björn Mittelsdorf (b.mittelsdorf@sulb.uni-saarland.de)
Ulrich Herb (u.herb@sulb.uni-saarland.de)

Brainstorming Ideas (roughly ordered)

Authority/Standardisation

- Central Unique Author Identification
 - Author
 - Identification/Profile
 - Picture
 - Projects
 - Competence
- Network of Authors
 - Social
 - Professional
 - Expertise
 - Field of interest

Visualisations/Indexing Dimensions

- Paper's Context
- Visual Social Graph
- show development of ideas (network graph displaying publication times for a document set)
- Visualisation of publications geo-location
- Position publication in the „landscape of science“
- Project's social map
- Visualize Data and Connections
- Semantic classification
- Numerical Semantics (Speech Independence)

Barrier Reduction

- Connect to the world (link between science and application)
- Publication-News-Binding
- Solicit further research
 - Need stack
 - Wish list
 - Request notification
- Practicable access to repositories not only via modern PC capabilities and resolution (e.g. mobile phones, hand helds, OLPC etc.)

Reception Tracking

- Consistent access statistics
- Re-use tracking
- Enhanced (complex) metrics (for better evaluation)

Assistance

Both:

- Automatic update/linking pre-print + new version
- Thumbnail/snapshot creation (first page display)
- Integrate everything (integrate information and processes and results) seamless working
- Modification of the document catalogues structures

Authors:

- Real-Time assistance in form fill in

- Automatic Metadata Creation/Lookup
- Reduce Redundant Work (intelligent submission)
- Dynamic publication (version management of production and reception; collaborative production)
- Easy submission process
- Dynamic publication list (exportable)
- Bonus Point System
- Easy feedback from authors to repository
- Repository as workspace
- Repository as research/production environment
- Educational assistance/encouragement for new authors (How-tos)
- Automatic/easy classification/positioning of new publication
- Automatic citation generation

Users:

- Track/pursue other searchers' way through the repository
- User recommendations as part of repository
- Graph/Image extraction from papers
- Dataset extraction
- Assign personalised searchable attributes
 - Personal comments
 - Pictures as bookmarks
 - Memory aids
 - Relevance statement
- Transparent result display relevance criteria

Brainstorming Summary

Many of the ideas presented above are about optimising and augmenting repository usage. There is a strong trend towards facilitation of the article import not only by enhancing the software but also by establishing human communication through repository functions.

Usage should become more “seamless” serving even readers with restricted access in an acceptable manner. In addition the repository shall not remain a rather old fashioned file server but become part of a decentralised and location-independent work space. This does not restrict itself to article creation but also to the reading work asking for tools that help us to “keep” the information gathered for further use. Repositories should provide useful every day work services instead of being content with being used.

The everyday and the scientific world should rejoin: Connecting research results with news reports and strengthening the communication between scientists and interest groups can aid the former in funding discussions and supply the latter with actual and precise information.

Communication can be further reformed by establishing a set of professional information.

Authors report these information about themselves, ranging from expertise and project lists to photographs for better recognition.

Many ideas revolved about network graphs and visual browsing structures that emerge from “social” network characteristics inherent to publications or authors.

Discussion Summary

The discussion following the brainstorming centred on more abstract problems.

The heterogeneity of services creates all kind of barriers.

Technical barriers emerge from a very strong researcher stereotype consisting of a broadband internet connection, powerful hardware and a sumptuous display. This stays true even for the ideas generated in the brainstorming which have a strong tendency towards graphical services.

Regardless of the efforts of DCMI metadata sets are not equivalent across disciplines, countries, cultures, and languages. The creation and optimisation of overall searches and interoperable services thus have to solve many questions aside from their actual developments.

Among these questions, two became salient in the limited time:

- The lack of interoperability in indexing systems (especially semantic ones) resulting from language and culture barriers and the independence/autonomy of standards
- The synchronisation problem concerning uniform identifiers, e.g. the Uniform Author Identifier.

The workshop group formulated the possibility of a central authority to manage the identifiers as well as the possibility to create a distributed network of identification institutions synchronising themselves in a network of equals. Sadly, nobody had any idea who this authority could be.

It was pointed out that the actual creation of services would be more effective if such questions were addressed in advance to the implementation. This demands communication with the scientific organisations as well as the publishers and the individual multipliers, emphasising the benefits and advantages of norms and standards for evaluation and development.

Conclusion

Repositories and thus repository softwares are not perfect like most services which we summarise as “the internet”. There are lots of functionalities but also behaviour patterns that must and will evolve in, hopefully the near, future. Well-documented standards have to be part of a development such as this, but it can be argued that there is also a need for one entity or a group of organisations monitoring standards and allocating identifiers.

In combination with the strong tendency towards making repositories more “social” it is obvious that privacy, informational self-determination, internationally equal access and equal scope of design have to be respected.

We, as moderators, are very glad to have our project objectives, i.e. usage statistics and enhanced metrics, mentioned in the brainstorming.