MyCoRe Distributed Search and Standard Library Interfaces

Mathias Zarick, University of Rostock
Overview

- Distributed Search and Integration Service
- Occurrences (replication / „on-the-fly“-processing)
- Demands to the service
- Distributed MyCoRe
- OAI – Cross Archive Searching Services
- Summary and Conclusion
Extension of the Architecture of Digital Libraries for Distributed Search

- **Export Interfaces**
  - Remote access
  - Certain protocols (for this XML-based protocols are presenting themselves)
  - Standards (Z39.50, OAI, ...)

- **Distributed Search and Integration Service**
  - Access and investigation on remote data sources
  - Homogenization / Consolidation of query results from the single heterogeneous remote data sources
Digital Library with Integration Capability

Export Interfaces

User Management and Access Services
Search Services and Metadata Management
Document Management
Order Services
Delivery Services
Accountancy and Payment Services
Deliverer Management and Document Procurement

Distributed Search and Integration Services
Distributed Search and Integration Service: Occurrences

- Replication
  - Replication of metadata and/or documents to a local instance
  - Investigation on the replicated data
  - Example:
    - Cross Archive Searching Services (OAI)
  - Advantage: Performance, Homogenization already during replication → one data model
  - Disadvantage: enormous need of storage volume, redundancy, uncertain Validity, Access rights are still there?
Distributed Search and Integration Service: Occurrences (2)

- On-the-Fly-Processing
  - Real distributed search „On-the-Fly“
  - direct forwarding of queries to the remote repositories
  - Conversion to other models, query languages and concepts by a federation service / wrapper
  - Advantage: Results are up-to-date, no consistency problems, low or no need of storage volume „near“ the distributed search and integration service
  - Disadvantage: more time-consuming when performing a distributed investigation, due to response times and the expenditure for the homogenization
Important Demands To A Distributed Search and Integration Service

- parallel search
- result merging and preparation for a global presentation (at the same time homogenization)
- duplicate recognition
- global ranking
- ...

05.09.2002 Mathias Zarick, University of Rostock
Result Merging

- Merge the results from the single repositories to a global result
- problem: many heterogeneous data formats
- Transformation of these formats
- For global result is XML very useful, Transformation of XML e.g. via XSL
Result Merging – An Example
Duplicate Recognition

- Recognition of identical object in several heterogeneous repositories
- present those only once in global result; show only list of repositories, where to find such object
- problem: not or only hard decidable; What is identical?
- Example: Maxim Gorki = Maksim Gorkii = Aleksei Maksimovich Peshkov
- Desirable: global Identification mechanism like for example ISBN
- There are: DOI, NBN - Persistent Identifier of Deutsche Bibliothek: Project CARMEN-AP4
MyCoRe – Architecture

- Command Line Interface
- Web Interface

Searching \(\Rightarrow\) Editing

MyCoRe

Backend:
Content Manager
DB2
Each MyCoRe instance offers a service for remote querying its content.

Other Instances can use this service, to query remote installations.

Therefore 2 basic components:
1. a remote query server (RQS)
2. a remote query client (RQC) (integrated in MyCoRe)

Communication between RQS and RQC via a chosen protocol.
1. should recognize and avoid communication errors
2. security demands → rights / encryption
Distributed MyCoRe

- Implementation in Java
  - mycore.communication.*;

- HTTP-Connection
- HTTPS-Connection
- Socket-Connection

- Data stream based on XML
OAI – Open Archives Searching Service

- [http://www.openarchives.org](http://www.openarchives.org)
- Metadata access on remote repositories
- Permits only replication and investigation on replicated data, no search via OAI-protocol possible
- based on HTTP
- HTTP-request, response in XML
- Metadata records in several metadata formats (at least Dublin Core)
Cross Archive Searching Service

User Interface

Search Component

Manage Database Component

Harvest Component

Scheduler

Indexing Component

DB2

OAI Repository

OAI

OAI Repository

OAI

OAI Repository

...
Conclusion

- Digital Library needs export interfaces for integration for distributed search
- 2 main occurrences for a realisation: „On-the-Fly“ & Replication
- Duplicate recognition very difficult, global identification mechanisms might bring a solution
- Implementation of MyCoRe Distributed Search
- Implementation of a Cross Archive Searching Service (OAI)
Thank you for your Attention!!!

THE END