ENABLING SEMANTIC SEARCH IN STRUCTURED P2P NETWORKS VIA DISTRIBUTED DATABASES AND WEB SERVICES

Maria Teresa Andrade
FEUP / INESC Porto
mandrade@fe.up.pt; maria.andrade@inescporto.pt
Introduction:
- current situation and challenges in p2p networks

Context of the work
- MOSAICA
  - goals, system architecture, functionality

The proposed approach
- concept
- tools used, modular functionality, layered architecture

P2P Content Management System
- use cases, uploading and searching content, deployed testbed

Conclusions
Introduction

Current situation and challenges in p2p networks
Current situation in p2p

• **unstructured architectures**

  • the network is flooded with query messages to locate streams and files (all nodes receive the query)

• **structured architectures**

  • based on Distributed Hash Tables (DHT) that maintain an index to the available resources, facilitating location

  • query messages are forwarded to only some nodes, which are more likely to have the location of the desired resource
Challenges in p2p for multimedia

• generally, users when searching for multimedia resources
don not know the exact name of the file they are looking for
often want to receive a useful set of results that are clearly related
to the subject they are interested in

• while unstructured designs allow this kind of proximity behavior
  at the cost of larger search times and increased network traffic

• common structured, DHT-based, designs do not
  search is performed using a hash key and only exact matches are returned
Challenges in p2p - semantic queries

- Current research initiatives include
  - Locality Sensitive Hashing (LSH) techniques (nearest-neighbor search problem in a Euclidean space)
  - Hierarchical multiple indexes, allowing query-to-query mapping and thus the implementation of a recursively query process
  - The CUBIT project at Cornell University using an overlay where peers are organized in concentric rings
  - Though promising good results, require complex management and/or processing of intermediate results
Context of the work

- MOSAICA
- goals, system architecture, functionality
MOSAICA goals

• Leveraging Web2.0-based technologies to enhance critical thinking and open-mindedness

• tool-box of Web based technologies for the preservation and sharing of cultural heritage resources (photos, documents, video, sound, etc.)

• sharing of cultural resources owned by private people or organizations such as museums.

• motivate users to create innovative experiences, multimedia virtual expeditions, telling their personal stories, and sharing them with remote peers, via a rich interactive environment

• reach as many people as possible

• low cost, high availability, simplicity
MOSAICA architecture

Online Utilities
- Virtual Expedition Maker
- Semantic Annotator
- Ontology Editor

Multifaceted Interfaces
- Repository of Educational Resources
  - GIS
  - Semantic Directory
  - Semantic Search

Logical Reasoning and Inference Engine

Distributed (p2p) Content Management Services
- Distributed Ontology
  - Distributed resources - content annotations

WWW
MOSAICA functionality

• **Explorative usage**
  • visit places of interest by zooming in on an area on MOSAICA's geographical interface, obtaining related multimedia info,
  • or by exploring MOSAICA semantic directory, or by submitting a query.

• **Collaborative usage**
  • annotate digital cultural objects with free-text comments or with semantic inter-related concepts using the MOSAICA Ontology
  • submit and exposed to the public annotated resources
  • design own Virtual Expeditions, using digital resources available in MOSAICA, and suggest them to other visitors by storing them in the MOSAICA P2P repository

• **Guided usage**
  • select ready-made, thematically-oriented Virtual Expeditions, from within the MOSAICA distributed repository
MOSAICA functionality

- **Explorative usage**
  - visit places of interest by zooming in on an area on MOSAICA's geographical interface, obtaining related multimedia info,
  - or by exploring MOSAICA semantic directory, or by submitting a query.

- **Collaborative usage**
  - annotate digital cultural objects with free-text comments or with semantic inter-related concepts using the MOSAICA Ontology
  - submit and expose to the public annotated resources
  - design own Virtual Expeditions, using digital resources available in MOSAICA, and suggest them to other visitors by storing them in the MOSAICA P2P repository

- **Guided usage**
  - select ready-made, thematically-oriented Virtual Expeditions, from within the MOSAICA distributed repository
MOSAICA functionality

- **Explorative usage**
  - Visit places of interest by zooming in on an area on MOSAICA’s geographical interface, obtaining related multimedia info, or by exploring MOSAICA semantic directory, or by submitting a query.

- **Collaborative usage**
  - Annotate digital cultural objects with free-text comments or with semantic inter-related concepts using the MOSAICA Ontology.
  - Submit and expose to the public annotated resources.
  - Design own Virtual Expeditions, using digital resources available in MOSAICA, and suggest them to other visitors by storing them in the MOSAICA P2P repository.

- **Guided usage**
  - Select ready-made, thematically-oriented Virtual Expeditions, from within the MOSAICA distributed repository.
MOSAICA functionality

• **Explorative usage**
  - visit places of interest by zooming in on an area on MOSAICA's geographical interface, obtaining related multimedia info,
  - or by exploring MOSAICA semantic directory, or by submitting a query.

• **Collaborative usage**
  - annotate digital cultural objects with free-text comments or with semantic inter-related concepts using the MOSAICA Ontology
  - submit and expose to the public annotated resources
  - design own Virtual Expeditions, using digital resources available in MOSAICA, and suggest them to other visitors by storing them in the MOSAICA P2P repository

• **Guided usage**
  - select ready-made, thematically-oriented Virtual Expeditions, from within the MOSAICA distributed repository
MOSAICA functionality

- **Explorative usage**
  - Visit places of interest by zooming in on an area on MOSAICA's geographical interface, obtaining related multimedia info,
  - or by exploring MOSAICA semantic directory, or by submitting a query.

- **Collaborative usage**
  - Annotate digital cultural objects with free-text comments or with semantic inter-related concepts using the MOSAICA Ontology
  - Submit and expose to the public annotated resources
  - Design own Virtual Expeditions, using digital resources available in MOSAICA, and suggest them to other visitors by storing them in the MOSAICA P2P repository

- **Guided usage**
  - Select ready-made, thematically-oriented Virtual Expeditions, from within the MOSAICA distributed repository

**Associating a concept to the resource**
MOSAICA functionality

- **Explorative usage**
  - Visit places of interest by zooming in on an area on MOSAICA's geographical interface, obtaining related multimedia info,
  - or by exploring MOSAICA semantic directory, or by submitting a query.

- **Collaborative usage**
  - Annotate digital cultural objects with free-text comments or with semantic inter-related concepts using the MOSAICA Ontology
  - Submit and expose to the public annotated resources
  - Design own Virtual Expeditions, using digital resources available in MOSAICA, and suggest them to other visitors by storing them in the MOSAICA P2P repository

- **Guided usage**
  - Select ready-made, thematically-oriented Virtual Expeditions, from within the MOSAICA distributed repository
MOSAICA functionality

Explorative usage
• visit places of interest by zooming in on an area on MOSAICA’s geographical interface, obtaining related multimedia info,
• or by exploring MOSAICA semantic directory, or by submitting a query.

Collaborative usage
• annotate digital cultural objects with free-text comments or with semantic inter-related concepts using the MOSAICA Ontology
• submit and exposed to the public annotated resources
• design own Virtual Expeditions, using digital resources available in MOSAICA, and suggest them to other visitors by storing them in the MOSAICA P2P repository

Guided usage
• select ready-made, thematically-oriented Virtual Expeditions, from within the MOSAICA distributed repository
The proposed approach for the distribution and management of resources

- concept, tools used
- modular functionality, layered architecture
Concept

• The challenge for the MOSAICA repository and content management system
  • to allow efficient semantic search
  • low-cost, usability, high-availability, simplicity

• The solution
  • p2p design adopting a two layer approach
    • upper layer handling semantic and keyword based queries
    • exposing distributed services through a Web Services interface
      • notably the Ontology service
    • built using the JXTA framework and distributed databases
  • lower layer handling media resources distribution and download
    • based on DHT p2p architecture, notably using BitTorrent protocols
The challenge for the MOSAICA repository and content management system to allow efficient semantic search.

The solution:
- p2p design adopting a two layer approach
  - upper layer handling semantic and keyword based queries
    - exposing distributed services through a Web Services interface notably the Ontology service
      - built using the JXTA framework and distributed databases
  - lower layer handling media resources distribution and download
    - based on DHT p2p architecture, notably using BitTorrent protocols
p2p-cms modular functionality

MOSAICA Application Layer

MOSAICA p2pcms Layer

Local Services: 4
- Local Service #1: DB for metadata
- Local Service #2: Ontology server
- Local Service #3: GIS server
- Local Service #4: VE content server

Local Services: 2
- Local Service #1: DB for metadata
- Local Service #4: VE content server

Local Services: 0

MOSAICA peer
- WS API (JXTA layer)
- torrent seeder (Bit Torrent layer)

Search
Submit
GIS
VEs

Search
Submit
GIS
VEs

Search
Submit
GIS
VEs

“Local Services” of the p2pcms are exposed to MOSAICA clients
The **p2p-cms** implements a Web Services interface towards the semantic upper layer tools of MOSAICA:

- **semantic annotation**, semantic and GIS search, VE editor and player
- Offers functionality in a transparent way to the user
- Automatically invoked by the semantic tools of MOSAICA

### p2p-cms Functionality

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>login(LUser, password)</code></td>
<td>Login method. It returns a session hash if user credentials are correct.</td>
</tr>
<tr>
<td><code>registerContent(LSession, contentDat)</code></td>
<td>Content data XML is sent. ContentID is returned</td>
</tr>
<tr>
<td><code>requestContent(LSession, contentID, boolean wait)</code></td>
<td>XML describing content (title, torrent binary, etc.) is returned</td>
</tr>
<tr>
<td><code>searchContent(LSession, XMLquery, boolean wait)</code></td>
<td>Query contents using a XML query. XML list of found contents, for this user groups, is returned</td>
</tr>
<tr>
<td><code>removeContent(LSession, content ID)</code></td>
<td>Remove index to content</td>
</tr>
<tr>
<td><code>addMetadata(Lsession, content ID, indextype, property, value)</code></td>
<td>Submits new semantic descriptions to be propagated to the distributed databases</td>
</tr>
<tr>
<td><code>removeMetadata (Lsession, content ID, indextype, property, value)</code></td>
<td>Eliminates a metadata record</td>
</tr>
<tr>
<td><code>changePassword (Lsession, user, oldpassword, newPassword)</code></td>
<td>Updates password for user</td>
</tr>
</tbody>
</table>
p2p-cms architecture

- modular and layered architecture
P2P Content Management System

- use cases supported
- uploading and searching content
- testbed deployed
Some important use cases

XML File
[HTTP POST]

Call remote object
[JAVA RMI]

registerContent(sessionID, XML)
[JXTA Peer]

Read uploaded file and call WebService
(PHP)

registerContent(sessionID, XML)
[SOAP]

Generate content ID and return updated XML
[JXTA Peer]
Some important use cases

- register and submit content
Some important use cases

- search and obtain content
- play movie
testbed deployed

- Needed for operational testing, integration and data gathering for simulation and evaluation
- Includes JXTA and Bittorrent overlays
- A Rendezvous/Relay peer, normal peer, super seeder and RSS server can be installed in only one machine (independent applications)
- Needed software: P2P-cms application, Azureus, JXTA Shell, Web server (for RSS feeds), Mysql and AXIS 2 (with RSS and peer webservices)
testbed deployed

- XLAB
  - Peer
  - Rendez Vous / Relay
  - asterix.inescn.pt
- IDEA
  - Peer
  - Rendez Vous / Relay
- INESC
  - Peer
  - panix.inescn.pt
  - Peer
  - buz.fe.up.pt
- FEUP
  - Peer
- RSS
  - druix.inescn.pt
- STRAIGHT
  - Peer
  - Rendez Vous / Relay
  - Super Seeder
  - Super Seeder
  - ideafix.inescn.pt
Conclusions

- The two-layered approach
  - by which metadata is manipulated in one layer and content in the other layer
  - enables to obtain an optimal solution
  - on one side, it enables to take advantage of the great efficiency of DHT (Distributed Hash Tables) for the distribution of and access to media resources;
  - on the other hand, through the implementation of distributed databases for managing metadata, it enables to overcome the limitation of DHT-based P2P networks for performing semantic-based searches
  - additionally it enables the use of simple search mechanisms
  - however, semantic search is only possible through the use of an ontology server
  - XML based semantic queries are decomposed in a series of boolean expressions which are then used for indexing and searching
- Still, performance and scalability tests need to be performed
Thank you very much for your attention!