MedioVis

Visual Interface for Searching and Exploring Multimedia Libraries

Prof. Dr. Harald Reiterer
Harald.Reiterer@uni-konstanz.de

Prof. Dr. Marc Scholl
Marc.Scholl@uni-konstanz.de

Mathias Heilig
Mathias.Heilig@uni-konstanz.de
Library of the University of Konstanz
(Director Petra Hätscher)
Application Domain Mediothek

Human-Computer Interaction Group
(Prof. Harald Reiterer)
Interaction Design & Visualization

Database & Information System Group
(Prof. Marc Scholl)
Data Management & Retrieval
Overview

Motivation & Goals

Results

New Interaction & Visualization Techniques

HyperGrid & HyperScatter & Bargrams & Networks

Lessons Learned: Design Principles

New Data Management & Retrieval Techniques

XML Database System

XQuery/XPath with Fulltext Retrieval & Updates
Motivation I Poor Usability
Content: Users want more than pure Catalog Content because a lot of interesting information are available in the Web.

Services: A lot of useful Web Services are available which should be integrated (Web 2.0).
Motivation

Why novel interaction, navigation and visualization techniques?

Traditional Ulks are cognitively demanding because …

• often search function and result visualization are separated and distributed over several pages
• inflexible and static result lists (no filtering, no sorting, …) without possibilities to control the level of detail or to browse for details on-the-spot
• lots of unstructured content in numerous pop-up windows with demanding window management and without possibilities to visually compare results

Effects of the growing information space:

• searching, finding and analyzing information are becoming the essential tasks,
• therefore users need analytic and goal-directed searching, filtering and comparison …
• … as well as browsing-oriented and opportunistic search possibilities.

Our alternative: mediovis with zoomable user interfaces (ZUlks)
**MedioVis Client Application**

- **Visible frontend**
  - Integration of functionality of web interface

- **Invisible backend**
  - Integration of web services and web interfaces of the "Web 2.0"
  - XML-Retrieval with XPath
  - XML-Database
  - Integration of metadata with web crawlers
  - Data Pumps

**Server**

- Operational library system with Web OPAC functionality
- Operational data, e.g. lending, reservation
- Library catalog data

**Library**

- "Web 1.0"
- "Web 2.0"
# The Table

- well-known and popular in “real life”
- perfect for sorting, filtering and comparing multidimensional (meta)data
- trade-off between visible amount of information and usability of large/wide tables

→ displays only "the tip of the iceberg"!

## The Browser

- popular interaction paradigm: point & click
- perfect for presenting loosely structured and heterogeneous content on multiple interlinked pages

→ only loosely structured and spread over multiple pages → no comparisons and no global overview

→ „drills down” into the iceberg!

<table>
<thead>
<tr>
<th>Film</th>
<th>Content</th>
<th>People Involved</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flags of Our Fathers</td>
<td>Drama, War</td>
<td>Clint Eastwood</td>
<td></td>
</tr>
<tr>
<td>Medley</td>
<td>Adventure, History, Sci-Fi, Action</td>
<td>Guillermo del Toro</td>
<td></td>
</tr>
<tr>
<td>Wild Bill, Billy</td>
<td>Drama</td>
<td>Clint Eastwood</td>
<td></td>
</tr>
<tr>
<td>Welcome to Woopport</td>
<td>Comedy</td>
<td>Doris Petrie</td>
<td></td>
</tr>
<tr>
<td>Elvis, Elvis</td>
<td>Documentary</td>
<td>Samuel Fuller, Gillo Jacques</td>
<td></td>
</tr>
<tr>
<td>A Decade Under the Influence</td>
<td>Documentary</td>
<td>Ted Demme, Richard LaGravenese</td>
<td></td>
</tr>
<tr>
<td>How to Get Away with a Dead Guy</td>
<td>Comedy, Drama</td>
<td>Antonio Gomora</td>
<td></td>
</tr>
<tr>
<td>The Big Fat Greek Life</td>
<td>Comedy</td>
<td>Frazier Phipps, Peter Bonerz, Carl Aiello</td>
<td></td>
</tr>
<tr>
<td>Mystic River</td>
<td>Drama, Crime, Mystery</td>
<td>Clint Eastwood</td>
<td></td>
</tr>
<tr>
<td>The Blues</td>
<td>Drama, Musical, Documentary</td>
<td>Charles Burnett, Clint Eastwood, Mike O'Keefe</td>
<td></td>
</tr>
<tr>
<td>Bloodwork</td>
<td>Crime, Thriller, Drama, Mystery</td>
<td>Clint Eastwood</td>
<td></td>
</tr>
<tr>
<td>Chicago</td>
<td>Romantic, Crime, Drama</td>
<td>Faye Dunaway</td>
<td></td>
</tr>
<tr>
<td>Mom's Boyfriend</td>
<td>Documentary</td>
<td>Adam Low</td>
<td></td>
</tr>
<tr>
<td>Where the Money Is</td>
<td>Drama</td>
<td>Alan Rudolph</td>
<td></td>
</tr>
<tr>
<td>Breakfast at Tiffany</td>
<td>Comedy</td>
<td>Tom Hiddleston</td>
<td></td>
</tr>
<tr>
<td>相近的电影：</td>
<td>Drama</td>
<td>Thea Thane</td>
<td></td>
</tr>
<tr>
<td>True Crime</td>
<td>Crime, Thriller, Drama</td>
<td>Clint Eastwood</td>
<td></td>
</tr>
<tr>
<td>Affair; 100 Years...100 Movie...</td>
<td>Documentary</td>
<td>Tom Donchess</td>
<td></td>
</tr>
<tr>
<td>Satellite Shore</td>
<td>Documentary</td>
<td>Debbie Newcom</td>
<td></td>
</tr>
<tr>
<td>Absolute Power</td>
<td>Thriller, Crime, Drama</td>
<td>Clint Eastwood</td>
<td></td>
</tr>
<tr>
<td>Wrong Side of the Street: Boy, The</td>
<td>Crime, Drama</td>
<td>Clint Eastwood</td>
<td></td>
</tr>
<tr>
<td>Wild Bill, Hollywood Avarick</td>
<td>Documentary</td>
<td>Todd Robinson</td>
<td></td>
</tr>
</tbody>
</table>
Hyperlinked content in a grid structure = HyperGrid


<table>
<thead>
<tr>
<th>Film Title</th>
<th>Content</th>
<th>People Involved</th>
</tr>
</thead>
<tbody>
<tr>
<td>A Century of Cinema</td>
<td>Documentary</td>
<td>Caroline Thomas</td>
</tr>
<tr>
<td>A Decade Under the Influence</td>
<td>Documentary</td>
<td>Ted Barnes, Michael LaGravinese</td>
</tr>
<tr>
<td>A Perfect World</td>
<td>Action, Drama, Crime</td>
<td>1997</td>
</tr>
<tr>
<td>A Perfect World</td>
<td>Action, Drama, Crime</td>
<td>1997</td>
</tr>
</tbody>
</table>

Columns = Aspects of Interest (AOI)

Rows = individual entries
Integration of Multimedia Metadata and Web Services

Evaluation Study in the Lab

- Quantitative Usability Test – KOALA vs. MedioVis
- Test Null-Hypothesis:
  - No significant difference in terms of task completion time (efficiency)
  - No significant difference in terms of user satisfaction (subjective user rating)

<table>
<thead>
<tr>
<th>Time frame</th>
<th>Dec ’04/Jan ’05</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participants</td>
<td>24 (16 female, 8 male)</td>
</tr>
<tr>
<td>Age</td>
<td>~23 years, StD.: 2.69</td>
</tr>
<tr>
<td>Occupation</td>
<td>Students at the University of Konstanz</td>
</tr>
<tr>
<td>Field of study</td>
<td>11 x Literature, Art &amp; Media Science (main target group)</td>
</tr>
<tr>
<td></td>
<td>13 x very heterogenous fields but no computer science students</td>
</tr>
<tr>
<td>Computer experience</td>
<td>Beginner to advanced user</td>
</tr>
<tr>
<td>System knowledge</td>
<td>Koala: unexperienced to very experienced</td>
</tr>
<tr>
<td></td>
<td>MedioVis: no subject had used MedioVis before</td>
</tr>
</tbody>
</table>

Evaluation Results

F(1, 22) = 48.265; p<0.01 (significant)
Standard Deviation: 15.27s (Koala);
11.08s (MedioVis)

F(1, 22) = 107.768; p<0.01 (significant)
Standard Deviation: 15.06s (Koala); 11.40s (MedioVis)

Attrakdiff
F(1, 22) = 43.89; p<0.01 (significant)
Standard Deviation: 0.91 (Koala);
0.76 (MedioVis)

Attrakdiff
F(1, 22) = 38.18; p<0.01 (significant)
Standard Deviation: 0.84 (Koala);
0.54 (MedioVis)
Field Study Questionnaire - System Usability Scale

Average Scale

Answered questionnaires in proportion to the number of all sessions (1‰)
Outlook Multiple Coordinated Views & new Visualizations

HyperScatter
(semantic) zoomable 2D-Scatterplot Overview & correlation

Parallel Bargrams
Bargrams+ Parallel Coordinates correlation, distribution & filtering

Network
Zoomable network Visualize relations

HyperGrid
Zoomable table relating, sorting, Comparing & browsing

Lessons Learned  Design Principles

Support various ways of formulating an information need

Integrate analytical and browsing oriented ways of exploration

Provide views to different dimensions of an information space

Make search a pleasurable experience
Let us have a look inside ...
Large datasets need a DBMS

Complex datasets are tedious in an RDBMS

Text data need full-text retrieval

"Everything" is in XML, anyway

3 criteria for DBMS selection:

→ Performance, performance, performance

... BaseX is your friend
History

- DBIS Group: Research on XML since 2000
- XPath Accelerator<sup>2001</sup>, Staircase Join<sup>2003</sup>, LoopLifting<sup>2004</sup>: early, frequently referenced works on XML storage and querying
- Results have influenced BaseX:
  - flat table storage of tree-structured input
  - elegant and efficient query algorithms
  - Content/Index-based query evaluation
  - XQuery Full-Text: bringing the DB and IR community together
  - XQuery Update: fulfilling classical database requirements
Indexing

- Straightforward in relational databases:
  
  ```sql
  SELECT year FROM media WHERE title = "Crime and Punishment"
  → access title index, return year values of matching tuple
  ```

- Tricky query optimization in XML, different queries → same results:
  
  - `//*[ title = "Crime and Punishment" ]/year`
  - `/path/to/media[where/title = "Crime and Punishment"]/year`
  - `/media/year[../title = "Crime and Punishment"]`
  
  → access title index, check specified descendants/ancestors, return year

- But: flexibility of XML pays out...
Querying Hierarchical Data

<Medium id="303-357-556857">
<Title>Learning Italian</Title>
<Type>Multi-Part</Type>
<Section>Teaching/Languages</Section>
<Medium>
<Title>Master Book</Title>
<Type>Book</Type>
<Description>Lections 1 - 7</Description>
<Signature>6 fsi 470/l24-nb1</Signature>
</Medium>
<Medium>
<Title>Listening Examples</Title>
<Type>CD</Type>
<Description>Lections 1 - 7</Description>
<Signature>6 fsi 470/l24-nb2</Signature>
</Medium>
…

Find Media with "Italy" in the Title and CDs attached...

//Medium[Title ftcontains "Italy" with stemming][./Type = "CD"]
23,000 Library queries, U Konstanz

- Year: 1900-2000 (1.53 mio hits)
- Type: DVD, Genre: Drama (823 hits)
- Title: Shrek (1 hit)

Execution times:
- 35% < 1 ms
- 75% < 10 ms

Queries, sorted by execution time
Application Fields

• Focus on library catalogs
  → many query optimizations inspired by library data

• Downloads @ Sourceforge since first official release:

...many other application fields, different user groups world wide:
  Computer Scientists, Linguists, Historians, Biologists
BaseX – Querying the Library Data
BaseX – Frontend

Auction Data
Thankyouforyourattention!

http://hci.uni-konstanz.de/MedioVis

http://sourceforge.net/projects/mediovis

http://basex.org