Visual interactive search and exploration of a digital library

Syed Sheece Raza Gardezi
Supervised by Hanna Suominen
u5752631
Contents

- Problem background
- Related Work
- Dataset
- Game Plan
- Overview of system
- Interactive Visualizations: Timeline, Sunburst Area and Graph
- Evaluation
- Conclusion
- Question Answer
Problem Background

Appx. 20,000 new documents are added IEEE Xplore each month.

18,000+ New articles added to ACM DL each year.

Elsevier had 400,000 articles published in 2015.

Active researchers globally at 7.8 million.
Related Work

Evetwatch - NICTA

2013- 340 million tweets per day
Multi-task learning for cross-platform siRNA efficacy prediction: an in-silico study
Liu, Q. · Xu, Q. · Zheng, V., et al. in BMC Bioinformatics (2010)

Deep sparse multi-task learning for feature selection in Alzheimer 's disease diagnosis

Improving prediction accuracy of tumor classification by reusing genes discarded during gene selection
Yang, J. · Li, G. · Meng, H., et al. in BMC Genomics (2008)

Computational reconstruction of proteome-wide protein interaction networks between HTLV retroviruses and Homo sapiens
Mei, S. · Zhu, H. in BMC Bioinformatics (2014)

Multi-modal microblog classification via multi-task learning
Zhao, S. · Yao, H., et al. in Multimedia Tools and Applications (2014)
- Range adjustable
- Java based
- Restricts general usage
German trenches of the Vimy area during the Great War simulation with time and location (A), the scanned period map (B) is being used both as a virtual “prop” map (and trench geometries) and (C) a live mini-map tracks the users movement in part of the virtual world.
Friday 1 March 1918
Recco - Amman. Pilot Lt. Taylor. In air 2 h. 20 m. Did a good recco.

Saturday 2 March 1918
Data

- NICTA - 6000 papers - 120,489 keywords - 6941 authors
- Open data available to public
Game Plan

- Information seeking and discovery
  - Discovering underline trends
  - Extending search capabilities

- Collection and artefacts
  - Choosing compromise btw vantage points
    - entire digital library
    - Individual paper

- Visualization for exploration
  - Visuals that add value to search
  - Interactions that are natural
Overview of the system

Digital Library

<table>
<thead>
<tr>
<th>Title 1</th>
<th>Title 2</th>
<th>Title 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
</tr>
<tr>
<td>Value 4</td>
<td>Value 5</td>
<td>Value 6</td>
</tr>
<tr>
<td>Value 7</td>
<td>Value 8</td>
<td>Value 9</td>
</tr>
<tr>
<td>Value 10</td>
<td>Value 11</td>
<td>Value 12</td>
</tr>
</tbody>
</table>

Database

<table>
<thead>
<tr>
<th>Title 1</th>
<th>Title 2</th>
<th>Title 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
</tr>
<tr>
<td>Value 4</td>
<td>Value 5</td>
<td>Value 6</td>
</tr>
<tr>
<td>Value 7</td>
<td>Value 8</td>
<td>Value 9</td>
</tr>
<tr>
<td>Value 10</td>
<td>Value 11</td>
<td>Value 12</td>
</tr>
</tbody>
</table>

Information Processing

Documents (text fields: title, abstract) (name: document ID)

Potential Keyword List

Information Analysis System

Keywords Extraction

Inverse Index (Keyword: document of occurrence)

Keyword List (keyword, score)

Document Vectors (name: documentID, Content: keywords)

Clustering

5 top topics assigned to each document

Topics (topWords, topDocuments)

Interactive Visualizations

Data Wrapper

TimeLine

Graph

Area
Sunburst
Area

Open computer science library!
Evaluation

- Survey was conducting after taking ethics approval from NICTA

- Related questions deal with the study of effectiveness, usefulness and engagement of such interfaces for exploring digital Libraries.

- 75% of the participants said it helped them understand the data better.

- All of the participants agreed that such tools will support them in search task.
Conclusion

- A way of exploring digital libraries was implemented.
- More is required. Giving user ability to choose topic, categories and timeline to adjust visualization.
- System is scalable.
- Computation occurs at user end.
Question/ Answers