

Project Update

CS533c Information Visualization

Chia-Ning Chiang

Ron Jung-Rung Han

An Analysis Project

- Using VxInsight
- On the InfoVis Contest Dataset
- “History of InfoVis” A collection of InfoVis papers for the past 10 years

Input files of VxInsight

- 1. A file containing the data, typically a Microsoft Access file or Microsoft Excel file.
- 2. A database alias file with extension <dataset>.db.
- 3. A VxInsight configuration file with extension <dataset>.config.
- 4. A stopword file with extension <dataset>.stop.
- 5. A local cache file with extension <dataset>.cache.

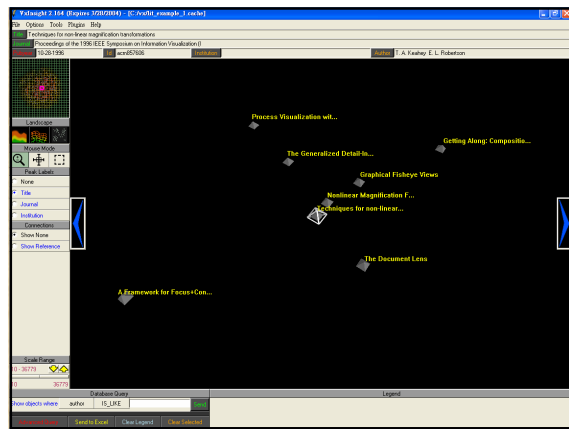
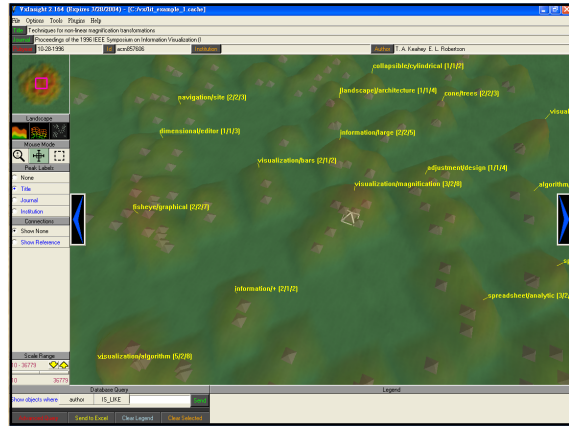
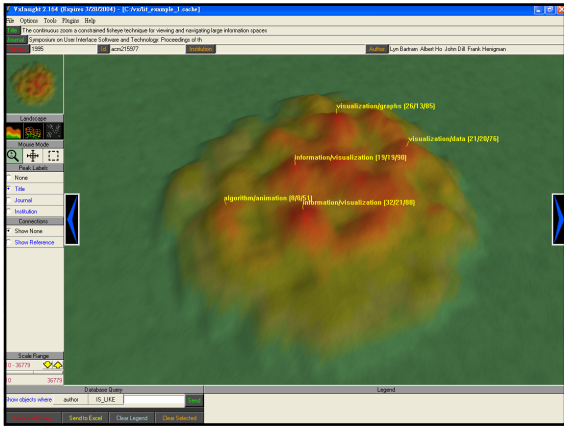
Data conversion

- Source format XML => Tab delimited text
- Tab delimited text => MS Excel / Access

Data conversion

- Tried XML parsing packages
- Wrote my own parser
- Problems with data: extra “<“
- New line / return characters in the Abstract field
- Incomplete data:

```
<article id="acm673478">
<title>MineSet(tm) A System for High-End Data
Mining and Visualization</title>
<source ref="acm645922">Proceedings of the
22th International Conference on Very Large
Data Bases</source>
<pages from=""/>
<authors>
</authors>
<date from="1996"/>
<references>
</references>
</article>
```



What's next?

- Different types of similarity functions
- Co-citation, bibliographical coupling
- Words in the title