

# Cluster-Based Visualization of Web Search Results

Matthias Tilsner

March 11, 2009

1 Introduction

2 Analysis

3 Evaluation

4 Approach

5 Conclusion

# Project Domain

## Web Documents

- URL
- header data
  - title
  - meta-tags
- body data
  - textual content
  - links

## Search Parameters

- query string
  - search terms

# Project Domain

## Web Documents

- URL
- header data
  - title
  - meta-tags
- body data
  - textual content
  - links

## Search Parameters

- query string
  - search terms

# Visualization Techniques

## Relevance to Search Terms/Query

- position [2, 3, 5, 8]
- color [4, 6, 8]
- size [7, 9]

**Information Visualization** - [ Diese Seite übersetzen ]  
**Information Visualization** is a central forum for all aspects of **informa**  
**visualization** and its applications.  
[www.palgrave-journals.com/ivd/](http://www.palgrave-journals.com/ivd/) - Ähnliche Seiten

**Information visualization** - Wikipedia, the free encyclopedia  
**Information visualization** the interdisciplinary study of the visual  
representation of large-scale collections of non-numerical **informatio**  
such as files ...  
[en.wikipedia.org/wiki/Information\\_visualization](http://en.wikipedia.org/wiki/Information_visualization) - 68k -  
[im Cache](#) - Ähnliche Seiten

**Visualization (computer graphics)** - Wikipedia, the free  
**Information visualization** concentrates on the use of computer-  
tools to ... Practical application of **information visualization** in c  
programs ...  
[en.wikipedia.org/wiki/Visualization\\_\(graphic\)](http://en.wikipedia.org/wiki/Visualization_(graphic)) - 79k -

20. ■ 6.2.2.6 TMC: Microsoft's Small Private Network Turned Into A VC  
21. ■ Visualization and Analysis: Information For Data Network Services ...  
22. ■ A Uniform Resource Identifier (URI) Scheme for the Semantic ...  
23. ■ Windows 97: Overview ... General Overview Section for 98 ...  
24. ■ Network's Configuration: Questions Have Come To Focus Here ...  
25. ■ General Overview: Details on the Components of the Network ...  
26. ■ HTTP: Overview: Message and Protocol Details ...  
27. ■ General Overview: Details on the Components of the Network ...  
28. ■ Web: Overview: Details on the Components of the Network ...  
29. ■ Web: Overview: Details on the Components of the Network ...  
30. ■ Web: Overview: Details on the Components of the Network ...  
31. ■ Web: Overview: Details on the Components of the Network ...  
32. ■ Web: Overview: Details on the Components of the Network ...  
33. ■ Web: Overview: Details on the Components of the Network ...  
34. ■ Web: Overview: Details on the Components of the Network ...  
35. ■ Web: Overview: Details on the Components of the Network ...  
36. ■ Web: Overview: Details on the Components of the Network ...  
37. ■ Web: Overview: Details on the Components of the Network ...  
38. ■ Web: Overview: Details on the Components of the Network ...  
39. ■ Web: Overview: Details on the Components of the Network ...  
40. ■ Web: Overview: Details on the Components of the Network ...



(a) position [3]

(b) color [4]

(c) size [7]

# Visualization Techniques

## Relationship of Search Results

- position [8]
- connection [2, 6, 9, 16]

Bookstein, B.	
Leuk, M. E.	27
Karo, M. E.	
Ozkarahan, E.	36
Salton, G.	37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100

(d) position [8]

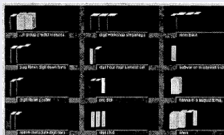


(e) connection [16]

# Visualization Techniques

## Similarity of Search Results

- position [1, 2, 7, 6, 9, 10, 11, 12, 13, 14]



(f) position [10]

# Interaction Techniques

## Details on Demand

- highlighting [15]
- popup [2]
- zoom [5, 7, 9, 10]

## Data Refinement

- redefine search terms [11]
- search result list [2, 6]



# Interaction Techniques

## Details on Demand

- highlighting [15]
- popup [2]
- zoom [5, 7, 9, 10]

## Data Refinement

- redefine search terms [11]
- search result list [2, 6]

# Current Status

## Summary of Existing Research

- either extending or abandoning conventional result lists
- display of all result items or a subset
  - zoom in for structural details
  - popup metadata

## Assumptions

- users: untrained, unwilling, impatient, used to conventional result lists
- search must be intuitive
- no existing solution has been widely accepted

# Current Status

## Summary of Existing Research

- either extending or abandoning conventional result lists
- display of all result items or a subset
  - zoom in for structural details
  - popup metadata

## Assumptions

- users: untrained, unwilling, impatient, used to conventional result lists
- search must be intuitive
- no existing solution has been widely accepted

# Open Issues

## Interface Requirements

- intuitive to use
- enables fast search success

## Beneficial Principles

- similarity to conventional result lists
- usable without training  
⇒ training provides additional functionality
- overview-zoom-detail interaction

# Open Issues

## Interface Requirements

- intuitive to use
- enables fast search success

## Beneficial Principles

- similarity to conventional result lists
- usable without training
  - ⇒ training provides additional functionality
- overview-zoom-detail interaction

# Idea

## Visualization

- fuzzy clustering
- “Concept Highlighter”-like visualization

## Interaction

- result filtering
- interactive creation of customized result set

# Idea

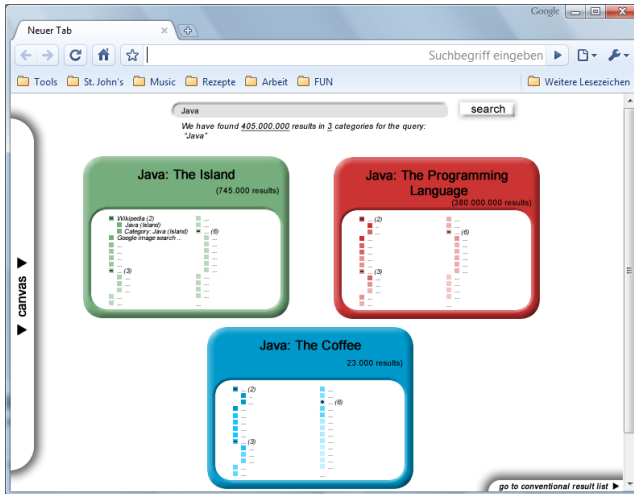
## Visualization

- fuzzy clustering
- “Concept Highlighter”-like visualization

## Interaction

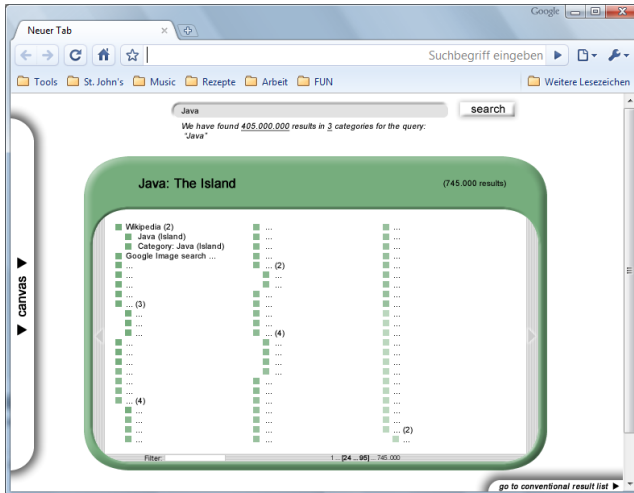
- result filtering
- interactive creation of customized result set

# Fuzzy Clustering

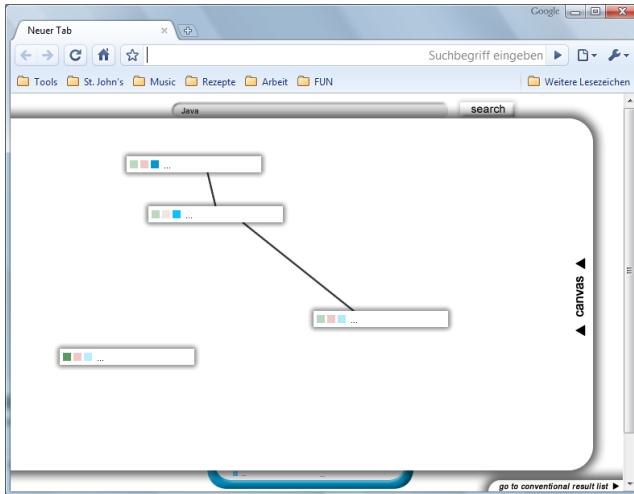




# “Concept Highlighter” -Like Visualization



# Interactive Creation of Customized Result Sets



# Risks and Potentials

## Risks

- performance
- intuitivity / usability
- quality (clustering & topic identification)

## Potentials

- fast reduction of search space
- easy identification of relevant results
- topic distinction

# Risks and Potentials

## Risks

- performance
- intuitivity / usability
- quality (clustering & topic identification)

## Potentials

- fast reduction of search space
- easy identification of relevant results
- topic distinction

# Outlook

## Future Issues

- customized / optimized ranking
- result retaining & sharing
- Web Service API
- visualization customizing
- search term weighting

Thank you for your attention

“If you type ‘Google’ into Google, you can break the Internet”  
- Jen (*The IT Crowd*)

# References I

- [1] C. Benjamin, W. Woon, and K. Wong.  
A graphical and convenient tool for document comparison and visualization.  
In *Proceedings of the International Conference on Computer and Communication Engineering*, pages 362–367, May 2008.
- [2] K. Einsfeld, S. Agne, M. Deller, A. Ebert, B. Klein, and C. Reuschling.  
Dynamic visualization and navigation of semantic virtual environments.  
In *Proceedings of the Tenth International Conference on Information Visualization*, pages 569–574, July 2006.
- [3] Google Inc.  
Google search, 2009.  
<http://www.google.com>.
- [4] O. Hoeber and X. D. Yang.  
Visually exploring concept-based fuzzy clusters in web search results.  
In *Proceedings of the 4th International Atlantic Web Intelligence Conference*, pages 81–90, 2006.
- [5] M. Konchady, R. D’Amore, and G. Valley.  
A web based visualization for documents.  
In *Proceedings of the 1998 Workshop on New Paradigms in Information Visualization and Manipulation*, pages 13–19, New York, NY, USA, 1998. ACM.
- [6] S. Mukherjea and Y. Hara.  
Visualizing world-wide web search engine results.  
In *Proceedings of the 1999 IEEE International Conference on Information Visualization*, pages 400–405, 1999.

## References II

- [7] S. Mukherjea, K. Hirata, and Y. Hara.  
Visualizing the results of multimedia web search engines.  
In *Proceedings of the 1996 IEEE Symposium on Information Visualization*, pages 64–65, 122, Oct 1996.
- [8] L. T. Nowell, R. K. France, D. Hix, L. S. Heath, and E. A. Fox.  
Visualizing search results: Some alternatives to query-document similarity.  
In *Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 67–75, New York, NY, USA, 1996. ACM.
- [9] F. Paulovich, R. Pinho, C. Botha, A. Heijs, and R. Minghim.  
Pex-web: Content-based visualization of web search results.  
pages 208–214, July 2008.
- [10] A. Rauber and H. Bina.  
“andreas, rauber’? conference pages are over there, german documents on the lower left...”: An “old-fashioned” approach to web search results visualization.  
In *Proceedings of the 11th International Workshop on Database and Expert Systems Applications*, pages 615–619, 2000.
- [11] D. Roussinov.  
Internet search using adaptive visualization.  
In *Extended Abstracts of the 1999 Conference on Human Factors in Computing Systems*, pages 69–70, New York, NY, USA, 1999. ACM.
- [12] M. M. Sebrechts, J. V. Cugini, S. J. Laskowski, J. Vasilakis, and M. S. Miller.  
Visualization of search results: A comparative evaluation of text, 2d, and 3d interfaces.  
In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 3–10, New York, NY, USA, 1999. ACM.



## References III

- [13] M. Tvarozek and M. Bielikova.  
Personalized view-based search and visualization as a means for deep/semantic web data access.  
In *Proceeding of the 17th International Conference on World Wide Web*, pages 1023–1024, New York, NY, USA, 2008. ACM.
- [14] E. Weippl.  
Visualizing content-based relations in texts.  
In *Proceedings of the Second Australasian User Interface Conference*, pages 34–41, 2001.
- [15] M. Weiss-Lijn, J. McDonnell, and L. James.  
Visualising document content with metadata to facilitate goal-directed search.  
In *Proceedings of the Fifth International Conference on Information Visualisation*, pages 71–76, 2001.
- [16] C. M. Zaina and M. C. C. Baranauskas.  
Revealing relationships in search engine results.  
In *Proceedings of the 2005 Latin American Conference on Human-computer Interaction*, pages 120–127, New York, NY, USA, 2005. ACM.