A New Visual Search Interface for Web Browsing

Songhua Xu$^{1,2}$ Tao Jin$^2$ Francis C.M. Lau$^2$

1 Department of Computer Science, Yale University
2 Department of Computer Science, The University of Hong Kong

WSDM ‘09
Barcelona, Spain
Feb. 11, 2009
Motivation

• Search engines: one of the most useful tools computer scientists offered to the modern society.
• A great deal of research on search engine design and backend retrieval algorithms.
• Search interface is also important, yet much less studied.
A new visual search interface

- Group the search results according to their contents
- Present selected search results in summary form
- Quickly grasp the key topics and contents without being overwhelmed with details
- Digest over search results
Organization of the talk

- Overview of our visual search interface
- Architecture of our visual interface
- Related work
- Search result organization
- Displaying region allocation for topic groups
- Selective presentation of representative search results
- Demo
- Conclusion
- Q & A
Novelty of our new search interface

- Traditional search interface
  - List view
  - Linearly organized

- Our new search interface
  - Summarized view
  - Top-down style
  - Analogous to e-map
  - Selected texts and images of search results
Features of our new search interface

- Multi-level view of search results
- High-level overview of all the topics
  - Each topic mapped to a certain screen region
  - Each topic contains a few top ranked search results
- Low-level detailed view of the topics
  - Zoom-in view of the topic
  - Sub-topics or abstracts
  - Four levels in our prototype
A visual search example
A visual search example
Organization of the talk

- Overview of our visual search interface
- Architecture of our visual interface
- Related work
- Search result organization
- Displaying region allocation for topic groups
- Selective presentation of representative search results
- Demo
- Conclusion
- Q & A
Architecture of our search interface

1. Query
2. Top N relevant search results [via commercial search engine: Google, Yahoo, etc.]
3. Topic detection
4. Search result classification
5. Topic naming
6. Groups view
7. Search result view
8. Hierarchic presentation of search results
Organization of the talk

• Overview of our visual search interface
• Architecture of our visual interface
• Related work
• Search result organization
• Displaying region allocation for topic groups
• Selective presentation of representative search results
• Demo
• Conclusion
• Q & A
Related work: cluster-based document collection browsing


P. Pirolli, P. Schank, M. Hearst, and C. Diehl. Scatter/gather browsing communicates the topic structure of a very large text collection. CHI '96.


Related work: visual search interfaces


Related work: visual data exploration


Organization of the talk

• Overview of our visual search interface
• Architecture of our visual interface
• Related work
• **Search result organization**
• Displaying region allocation for topic groups
• Selective presentation of representative search results
• Demo
• Conclusion
• Q & A
Search result organization

• Semantically meaningful document clustering

• Nonnegative matrix factorization analysis
  – Identifying topics present in a document corpus
  – Correspond well to semantically meaningful topics in the document content space


• Hierarchical search result clustering
Organization of the talk

• Overview of our visual search interface
• Architecture of our visual interface
• Related work
• Search result organization
• Displaying region allocation for topic groups
• Selective presentation of representative search results
• Demo
• Conclusion
• Q & A
Displaying region allocation: determining size of a topic group

• Utilize the webpage ranks, compute the average webpage rank for each topic group
  – \( dc_1, \ldots, dc_n \) (n topic groups)
  – \( r_1, \ldots, r_n \) (n average webpage ranks)
  – Assuming S is the whole screen space:

\[
s_i \triangleq \frac{1}{\sum_{i=1}^{n} \frac{1}{r_i}} S
\]

• Open source hierarchical document clustering software package (http://demo.carrot2.org)
Displaying region allocation: using rectangular regions

- Using rectangular topic regions
  - Easy to compute

  Layout allocation problem

  Rectangle-in-rectangle packing problem

- Simple and visual elegance
Displaying region allocation: Layout design

• Automatic yellow-page pagination and layout algorithm:

Displaying region allocation: Layout design (cont.)

- A modification: introducing a bias term

\[ \psi \triangleq \sum_{i=1}^{K} (GS_i(x) + GS_i(y)) \]

- \( GS_i(x) \) the horizontal distances of the center for the i-th topic displaying region
- \( GS_i(y) \) the vertical distances of the center of the i-th topic displaying region
- \( K \) topics detected
Organization of the talk

• Overview of our visual search interface
• Architecture of our visual interface
• Related work
• Search result organization
• Displaying region allocation for topic groups
• Selective presentation of representative search results
• Demo
• Conclusion
• Q & A
Selective presentation of representative search results

- Key information in the search results:
  - Text
  - Images
- A summarized version of this information
- Three modes provided in our system:
  - Text only
  - Image only
  - Mixed text and image
- Mixed text summaries and thumbnails
  - Better performance
  - Study conducted by Woodruff et al., UIST 98.
Selective presentation of search results: text only mode

- Extractive summarization

- Display region allocate
  - $D_{x_j}$ the screen area
  - Extracted key sentences length
Selective presentation of search results: image only mode

• Select key image of webpage
  – No images from advertisements
  – No website navigation images

• Image Classification techniques
  – Sophisticated methods are available
  – Keep it simple
  • Image size
  • Image position
  • Image show pattern
Selective presentation of search results: image only mode

- Content images
- Pick from content images
  - Less than N
    - pick the first image
  - More than N
    - Pick first a x b images to combine a rectangle image
    - Use open-source image mosaic package (jimage-mosaic.sourceforge.net)
Selective presentation of search results: image only mode

Two examples
Selective presentation of search results: mixed image & text mode

- Position of text and image
  - Left and Right
  - Upper and Bottom
  - Around
To recap

- A user-friendly and informative graphical front-end for search engines
- Content oriented search result organization & presentation
- Topic detection and search result classification
- Mixed image and text summary of representative search results
Organization of the talk

- Overview of our visual search interface
- Architecture of our visual interface
- Related work
- Search result organization
- Displaying region allocation for topic groups
- Selective presentation of representative search results
- Demo
- Conclusion
- Q & A
Demo

• Our prototype visual search interface site:
  http://www.cs.hku.hk/~songhua/visualsearch/
  or
  http://www.gceel.com

• Special thanks for:
  – Google App Engine Service
  – The Carrt2 Project
Organization of the talk

- Overview of our visual search interface
- Architecture of our visual interface
- Related work
- Search result organization
- Displaying region allocation for topic groups
- Selective presentation of representative search results
- Demo
- Conclusion
- Q & A
Conclusion

• A visual search interface for efficient navigation of search results, facilitating users to
  – quickly grasp latent topics
  – quickly locate the interested/desired search results
• Content oriented search result organization & presentation
• Questions & comments are welcome!

songhua.xu@yale.edu

Thank you for your time and attention!