Bring Order to the Web: Automatically Categorizing Search Results

by Hao Chen, Susan Dumais

cse 450: Web Mining Seminar
Jian Wang
A user interface that organizes Web search results into hierarchical categories.

Two main components:
- A text classifier that categorizes web pages on-the-fly
- A user interface that presents the web pages within the category structure and allows the user to manipulate the structured view
RELATED WORK

- Generating structure
  - Structural information
  - Clustering
  - Classification

- Using structure to support search
  - Web search services use category information to organize the search results
  - Few studies evaluated the effectiveness
TEXT CLASSIFICATION

- Data Set
  - A collection of web pages from LookSmart’s Web directory

- Pre-processing
  - A short summary of information about each result

- Classification
  - Support Vector Machine (SVM)
USER INTERFACE

- Information Overlays
  - Percentage of documents falling into the category
  - Additional category information
  - Summaries of the web pages

- Distilled Information Display
  - How many categories
  - How many pages
  - How to rank pages
  - How to rank categories
USER STUDY

- Methods
- Subjects
- Procedure
- Search Tasks
USER STUDY (continued)

- Results
  - Subjective questionnaire measures
  - Accuracy/GiveUp
  - Search Time
  - Interaction Style – Hovering, Page Views, ShowMore, SubCategory
CONCLUSION

- The results convincingly demonstrate that the category interface is superior to the list interface in both subjective and objective measures.

- Further research
  - How the results generalize to other domains and task scenarios
  - How best to present concise views of search results in their category contexts