Dumais, S., Cutrell, E., & Chen, H. (2000). Classified displays of web search results.. 11th ASIS SIG/CR Classification Research Workshop, 87-90. doi:10.7152/acro.v11i1.12771

Classified Displays of Web Search Results

Susan Dumais[1], Ed Cutrell[1] and Hao Chen[2]
[1]Microsoft Research and University of [2]California, Berkeley
sdumais@microsoft.com

Abstract

Most text retrieval systems return a ranked list of results in response to a user's search request. Such lists can be long and overwhelming. Furthermore, results on different topics or different aspects of the same topic are intermixed in the list requiring users to sift through a long undifferentiated list to find items of interest. We have been exploring the use of automatic text classification techniques combined with novel interface ideas to allow users to quickly focus in on results of interest. Our approach combines the advantages of human knowledge in an initial classification stage with the broad coverage available with text retrieval systems. In a series of user studies we developed and evaluated several interfaces for structuring search results in order to better understand the cognitive processes that lead to effective analysis of search results. There are two key aspects to our work that we describe in more detail: 1) automatic text classification algorithms for quickly and accurately tagging new content, and 2) novel interface to support structured search.

1 Automatic Classification of Web Content

Web directories like Yahoo!, LookSmart, Snap, Open Directory employ human editors to categorize or tag Web pages. A similar approach has been used for centuries by librarians in systems like Dewey Decimal classification, Library of Congress classification, NLM Classification, etc. The resulting category structures can be browsed directly, or they can be used to facilitate search. Since this approach depends on human tagging, coverage is limited. Human classification can be augmented by automatic text classification techniques. A wide variety of text retrieval and machine learning techniques can be used to learn a model for each category based on a labeled set of training documents. The model can then be applied to new unlabeled documents to determine their categories. Chakrabarti et al.[1], Dumais and Chen[3], Stata et al.[4] and others have developed automatic classifiers for web pages using content from Web directories as training data. For present purposes, the important aspect of this work is that new web content can be tagged, thus dramatically extending the reach of directory services. How best to present the resulting category information to help users winnow through a large set of search results is the focus of our interface experiments.

2 User Interfaces for Integrating Search Results and Category Information

Figures 1 and 2 show the two basic interfaces for presenting search results explored by Chen and Dumais[2]. In the *Category* interface (Figure 1), search results are organized into hierarchical categories. The best matching web pages within each category are shown initially, and additional pages can be seen on demand by category expansion. In order to show both category context and individual search results in limited screen space, only the title of each page is shown

ISSN: 2324-9773

Dumais, S., Cutrell, E., & Chen, H. (2000). Classified displays of web search results.. 11th ASIS SIG/CR Classification Research Workshop, 87-90. doi:10.7152/acro.v11i1.12771

initially. The summary of each page is available as hover text. The *List* interface (Figure 2) is similar to current systems for displaying web search results. For comparability to the Category condition, only titles are shown initially with summaries available on demand as hover text. Chen and Dumais found large and reliable advantages for the category interface -- participants preferred the category interface and they were 50% faster at finding relevant information.

In order to better understand this category advantage, we developed and evaluated several new interfaces. We explored two methods for adding contextual information to the List interface. The first approach presented summaries inline instead of as hover text; the second approach added category names to the inline summaries in the list interface much as Web directories like Yahoo!, Snap, LookSmart and Open Directory do in presenting their results. We also explored methods for removing aspects of the context from the Category interface. The first approach removed the category names but the results were still grouped by category; the second approach removed the page titles, leaving only category names as browsing interfaces do.

Three key findings emerged from a series of four experiments using seven different interfaces. 1) In all cases, Category interfaces were faster than List interfaces. This was true even when we added category names and inline summaries to the Lists, and when we degraded the Category organization by removing category names or page titles. Interestingly, the List interface augmented with category names contains the same information as the Category interface, but search times and user preferences support the Category approach. Grouping results from the same category together visually appears to be the key. 2) The best performance in the Category interfaces was achieved when both category names and page titles are available. Either alone worked better than any list presentation, but the combination of specific results in the category context was the most effective for quickly analyzing search results. 3) Inline summaries were more effective than summaries presented as hover text for both the List and Category interfaces. This is surprising since more scrolling is required, but apparently the cognitive costs of deciding which title to examine in more detail outweigh the additional scrolling required.

These results provide interface design guidelines for effectively combining category context with search results. In addition, we have shown how automatic text classification can extend the reach of existing directories.

References

1. Chakrabarti, S., Dom, B., Agrawal, R. & Raghavan, P. Scalable feature selection, classification and signature generation for organizing large text databases into hierarchical topic taxonomies. *The VLDB Journal 7*, 1998, 163-178.

2. Chen, H. & Dumais, S. T. Bringing order to the Web: Automatically categorizing search results. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems,* 2000, 145-152.

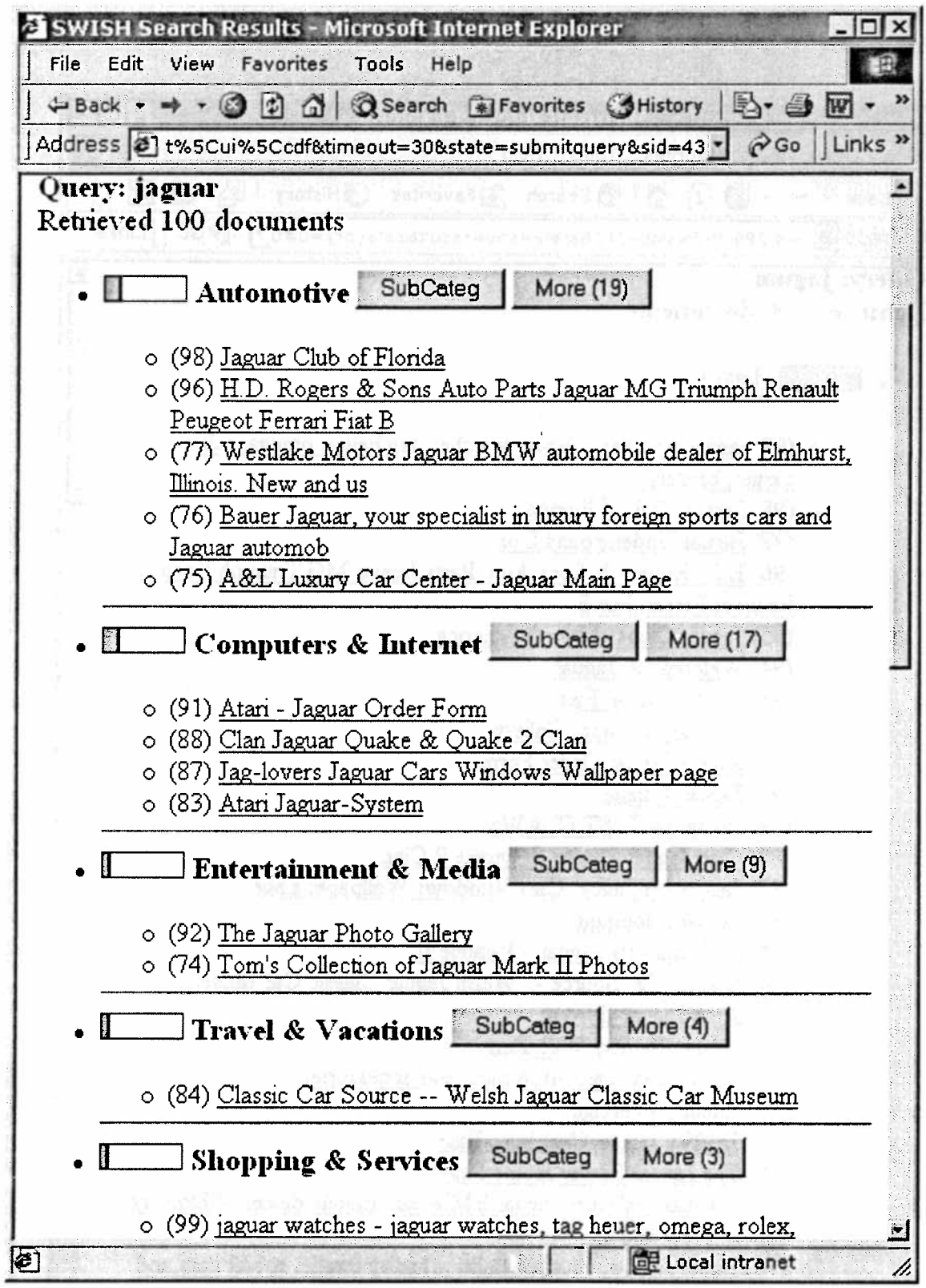
3. Dumais, S. T. & Chen, H. Hierarchical classification of web content. In *Proceedings of SIGIR'2000*, 256-263.

4. Stata, R, Bharat, K. and Maghhout, F. The term vector database: Fast access to indexing terms for web pages. In *Proceedings of WWW9, 2000, 247-256.*

ISSN: 2324-9773

Dumais, S., Cutrell, E., & Chen, H. (2000). Classified displays of web search results.. 11th ASIS SIG/CR Classification Research Workshop, 87-90. doi:10.7152/acro.v11i1.12771

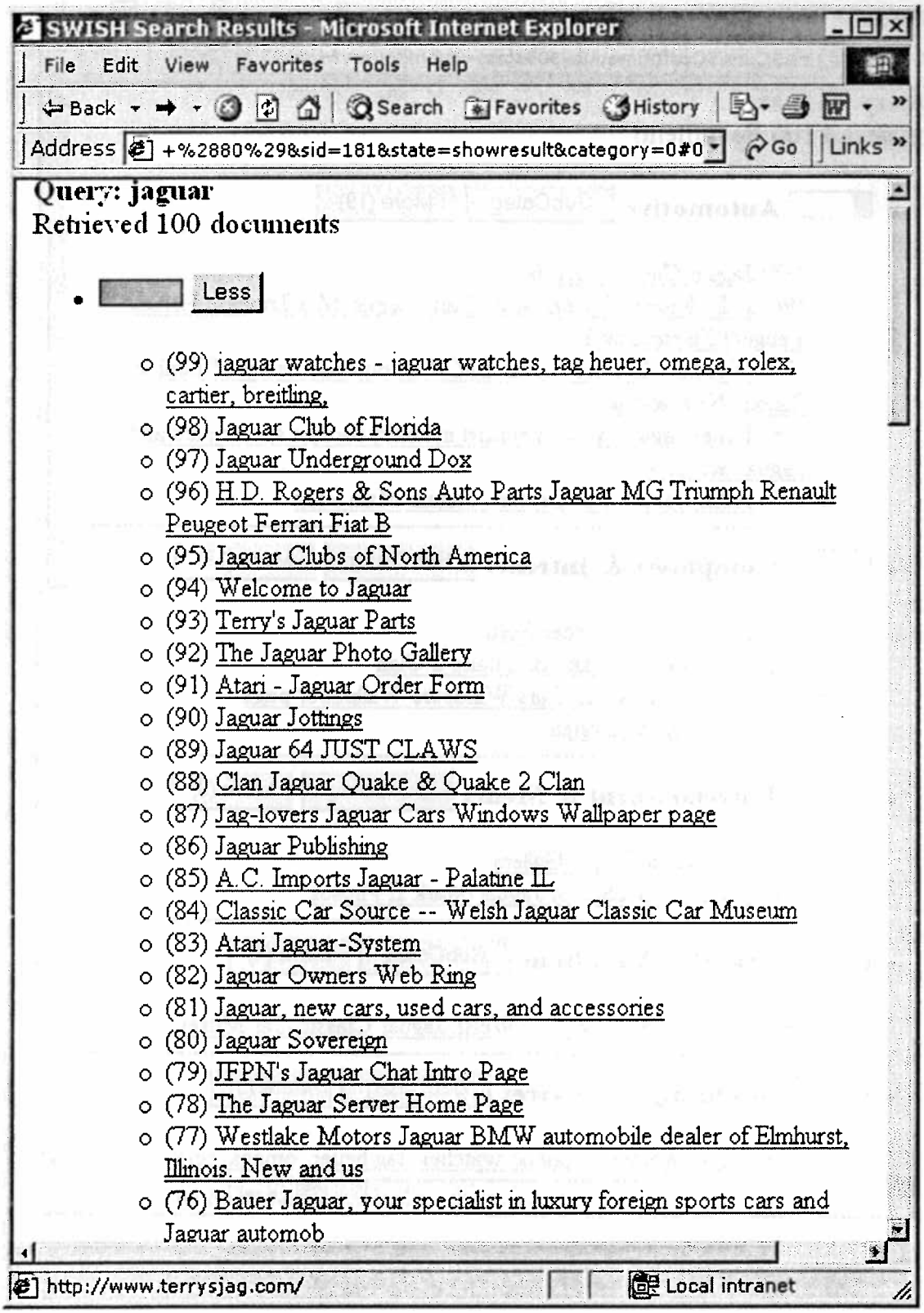
Figure 1. Category interface used by Chen and Dumais (2000).



ISSN: 2324-9773

Dumais, S., Cutrell, E., & Chen, H. (2000). Classified displays of web search results.. 11th ASIS SIG/CR Classification Research Workshop, 87-90. doi:10.7152/acro.v11i1.12771

Figure 2. List interface used by Chen and Dumais (2000)



ISSN: 2324-9773