

ClusteringWiki: Personalized and Collaborative Clustering of Search Results

David C. Anastasiu
Texas State University-San
Marcos
San Marcos, TX, USA
da1143@txstate.edu

Byron J. Gao
Texas State University-San
Marcos
San Marcos, TX, USA
bgao@txstate.edu

David Buttler
Lawrence Livermore National
Laboratory
Livermore, CA, USA
buttler1@llnl.gov

Categories and Subject Descriptors: H.3.3 [Information Storage and Retrieval]: Information Search and Retrieval – *Clustering*

General Terms: Algorithms, Design, Human Factors

Keywords: Personalized clustering, Search result clustering, Wiki, Mass collaboration, Web 2.0

1. EXTENDED ABSTRACT

How to organize and present search results plays a critical role in the utility of search engines. Queries are inherently ambiguous and search results are often diverse with multiple senses. The user has to sift through many irrelevant results to locate those relevant ones. Clustering divides a long list of disparate search results into a few topic-coherent clusters, allowing the user to quickly locate relevant results by topic navigation. There is good evidence that clustering improves user experience and search result quality [2].

Search result clustering is a highly user-centric task. Automatic algorithms often fail to fulfill the human factors in the objectives of search result clustering, generating meaningless, awkward or nonsense cluster labels [1]. In this demo, we explore a completely different direction in tackling the problem, utilizing the power of direct user intervention and mass-collaboration. In **ClusteringWiki** (dmlab.cs.txstate.edu/ClusteringWiki/), the user can edit and annotate the membership, structure and labels of clusters through a Wiki interface to personalize her search result presentation. Such edits and annotations can be implicitly shared among users as a mass-collaborative way of improving search result organization and search engine utility.

In collaborative tagging, tags can be used to collectively classify and find information. **ClusteringWiki** extends conventional tagging and allows tagging of structured objects, which are clusters of search results organized in a hierarchy.

Figure 1 shows a snapshot of **ClusteringWiki**. The left-hand *label panel* presents a hierarchy of cluster labels. The right-hand *result panel* presents search results for a chosen cluster label. A logged-in user can edit the current clusters by creating, deleting, modifying, moving or copying nodes in the cluster tree. Each edit will be validated against a set of predefined consistency constraints before being stored. Detailed descriptions of the framework and implementation can be found at dmlab.cs.txstate.edu/ClusteringWiki/pdf/cw.pdf.

Demonstration scenarios. After login, you are able to

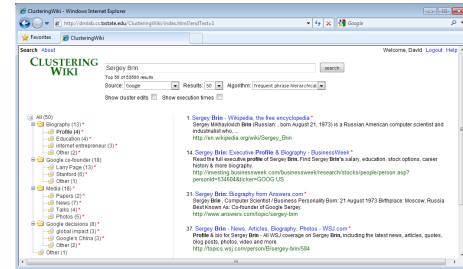


Figure 1: Snapshot of ClusteringWiki.

personalize a cluster tree for a query, e.g. “Sergey Brin”. You can edit the tree labels in the label panel by renaming, creating, copying, moving and deleting labels. You can also re-assign the cluster membership of search results by copy, move and delete operations. To reduce user editing effort, cluster edits in **ClusteringWiki** are available through context menus with pre-validated operations attached to labels and results. To perform a move/copy operation you can use cut/copy and paste or simply drag and drop a result or label.

To verify that **ClusteringWiki** retains personal preferences, you can log out, log in, and issue the same query. The resulting personalized cluster tree should allow you to explore the search results more effectively. To demonstrate aggregated clustering, **ClusteringWiki** lists the top 10 queries edited by the most users. You (without login) can choose some of these queries and observe how the aggregated trees differ from the initial trees. Aggregated trees reflect the collaborative effort and common preferences of many users.

Cluster editing takes user effort. **ClusteringWiki** attempts to reuse such effort. While preference aggregation can be considered sharing among users, preference transfer is sharing among queries. In **ClusteringWiki**, preference transfer is executed regardless of your login status. For example, you can issue a query “Sergey Brin” and edit the cluster tree. Then you can issue a similar query “Sergey M. Brin” and observe how those stored preferences for “Sergey Brin” are enforced in producing the cluster tree for “Sergey M. Brin”.

Acknowledgments: This work (LLNL-CONF-461652) was performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory (contract DE-AC52-07NA27344).

2. REFERENCES

- [1] C. Carpineto, S. Osiński, G. Romano, and D. Weiss. A survey of web clustering engines. *ACM Comput. Surv.*, 41(3):1–38, 2009.
- [2] C. D. Manning, P. Raghavan, and H. Schtze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.