# Extending Faceted Search to the Open-Domain Web

Weize Kong

Center for Intelligent Information Retrieval

College of Information and Computer Sciences

University of Massachusetts Amherst

Amherst, MA 01003

*wkong@cs.umass.edu*

### Abstract

Faceted search enables users to navigate a multi-dimensional information space by combining keyword search with drill-down options in each facets. For example, when searching "computer monitor" in an e-commerce site, users can select brands and monitor types from the the provided facets {"Samsung", "Dell", "Acer", ...} and {"LET-Lit", "LCD", "OLED", ...}. It has been used successfully for many vertical applications, including e-commerce and digital libraries. However, this idea is not well explored for general web search in an open-domain setting, even though it holds great potential for assisting multi-faceted queries and exploratory search.

The goal of this work is to explore this potential by extending faceted search into the open-domain web setting, which we call Faceted Web Search. We address three fundamental issues in Faceted Web Search, namely: how to automatically generate facets (facet generation); how to re-organize search results with users' selections on facets (facet feedback); and how to evaluate generated facets and entire Faceted Web Search systems.

In conventional faceted search, facets are generated in advance for an entire corpus either manually or semi-automatically, and then recommended for particular queries in most of the previous work. However, this approach is difficult to extend to the entire web due to the web's large and heterogeneous nature. We instead propose a query-dependent approach, which extracts facets for queries from their web search results. We further improve our facet generation model under a more practical scenario, where users care more about precision of presented facets than recall.

The dominant facet feedback method in conventional faceted search is Boolean filtering, which filters search results by users' selections on facets. However, our investigation shows Boolean filtering is too strict when extended to the open-domain setting. Thus, we propose soft ranking models for Faceted Web Search, which expand original queries with users' selections on facets to re-rank search results. Our experiments show that the soft ranking models are more effective than Boolean filtering models for Faceted Web Search.

To evaluate Faceted Web Search, we propose both intrinsic evaluation, which evaluates facet generation on its own, and extrinsic evaluation, which evaluates an entire Faceted Web Search system by its utility in assisting search clarification. We also design a method for building reusable test collections for such evaluations. Our experiments show that using the Faceted Web Search interface can significantly improve the original ranking if allowed sufficient time for user feedback on facets.

**Supervisor**: James Allan
**Available at**: `http://cs.umass.edu/~wkong/thesis.pdf`