

# PERSONAL CONCEPT HIERARCHY CONSTRUCTION

Hui Yang  
Carnegie Mellon University  
*huiyang@cs.georgetown.edu*

April 6, 2012

## Abstract

A concept hierarchy is a set of concepts and relations between those concepts. Since ancient times, concept hierarchies have been used to organize and access information. In some situations, task-specific and user-specific concept hierarchies are necessary to allow an overview and easy access a large set of documents. For example, in regulatory reforms, rule-makers in government regulatory agencies must quickly identify and respond to issues raised in public comments. A concept hierarchy constructed for a set of public comments hierarchically organizes the comments and a user is able to easily “drill down” into documents that discuss a specific topic.

Particularly, this dissertation addresses how to construct concept hierarchies from text collections automatically or with a-human-in-the-loop. The novel metric-based concept hierarchy construction framework transforms concept hierarchy construction into a multi-criterion optimization problem. It incrementally clusters concepts based on minimum evolution of hierarchy structure, as well as optimization derived from the modeling of concept abstractness and concept coherence. Moreover, this dissertation represents the semantic distance between concepts as a wide range of features, each of which corresponds to a state-of-the-art concept hierarchy construction technique, such as lexico-syntactic pattern, contextual information, and co-occurrence. The use of multiple features allows a further study of the interaction between features and different types of semantic relations as well as the interaction between features and concepts at different abstraction levels.

Besides the automatic framework for concept hierarchy construction, this dissertation also proposes an effective human-guided concept hierarchy construction framework to address personalization by learning from periodic manual guidance and directing the learned models towards personal preferences. Through human-computer interactions, the human and the machine work together to organize concepts into hierarchies. The machine’s predictions not only save the user’s effort but also make sensible suggestions to assist the user. This is one of the first works of real-time machine learning for organizing personalized information in an interactive paradigm.

This dissertation also studies user behaviors during concept hierarchy construction. It explores whether people create concept hierarchies more quickly or more consistently using

---

the proposed frameworks, whether there are consistent dataset-specific or user-specific differences in the hierarchies that people construct, whether people are self-consistent, and how these factors interact with different construction methods. The user study elaborates that dataset difficulty is a major factor affecting how people organize information into concept hierarchies. It also reveals that people are quite self-consistent in building hierarchies. This novel finding provides foundations to study the differences in concept hierarchy construction behaviors between individuals.

Last but not least, the dissertation proposes a novel similarity metric for measuring hierarchy similarity. Fragment-based Similarity (FBS) employs a unique bag-of-word representation for hierarchies and takes a fragment-based view to calculate hierarchy similarity. FBS well approximates tree edit distance and greatly improves tree edit distance's efficiency from NP-hard to only  $O(n^3)$  and  $O(n)$  if pairwise node similarities are pre-calculated.

The research in this dissertation is an important step forward of concept hierarchy construction. It addresses important problems of concept hierarchy construction, especially considers how to better model these problems with good theoretical foundations, to study these problems via extensive empirical experiments and user studies, and to solve these problems by developing practical applications for constructing personal concept hierarchies. Available at <http://www.cs.georgetown.edu/~huiyang/publication/dissertation.pdf>.