Chapter 6
User–Centered Maintenance of Concept Hierarchies

Kai Eckert
University of Mannheim, Germany

Robert Meusel
University of Mannheim, Germany

Heiner Stuckenschmidt
University of Mannheim, Germany

ABSTRACT
Taxonomies are hierarchical concept representations that have numerous important applications from improved indexing of document collections to faceted browsing and semantic search applications. The maintenance of taxonomies includes the dynamic extension, analysis, and visualization of these representations. Instead of focusing on the construction of taxonomies from scratch, however, the authors describe several successful approaches to the semi-automatic maintenance of taxonomies. These approaches have in common that they incorporate the human expert as a central part of the system.

1 INTRODUCTION
Beside the full-fledged ontologies that are supposed to form the basis of the Semantic Web, other constructs exist to represent background knowledge. Many of them have a long tradition and history that goes back well before the birth of the Internet, including glossaries, dictionaries, vocabularies, gazetteers, taxonomies and thesauri, just to name a few. They all, including ontologies, have in common that they are used to represent and organize knowledge in a structured way to fulfill specific functions for different purposes, thus we refer to all of them as Knowledge Organization Systems (KOS). A subset of them have in common that they organize the concepts representing single units of knowledge in a hierarchical way, which we call concept hierarchies.

Concept hierarchies have numerous important applications from improved indexing of document collections to faceted browsing and semantic search applications. But the creation and maintenance of concept hierarchies is cumbersome and very time consuming. On the other hand, many
concept hierarchies already exist and more and more of them become publicly available, ideally as linked open data, and can be reused for different purposes. If an existing concept hierarchy is to be reused, several tasks have to be performed, reaching from the proper selection of the source to start with to the adaptation for the desired purpose which includes deletion of unnecessary concepts, merging and splitting of concepts and especially the addition of missing concepts.

In this chapter, we describe several successful approaches to the semi-automatic creation and maintenance of different types of concept hierarchies. These approaches have in common that they incorporate the human expert as a central part of the system.

### 1.1 Concept Hierarchies

Contrary to full-fledged ontologies, concept hierarchies use only a limited set of semantic relations. Typically there is one main relation that leads to the hierarchical structure of the concepts. This relation is usually referred to as “broader than”, respectively “narrower than”, but can also have a very specific meaning, like “is a”/“has subclass” or “part of”/“has part”. A common property of these semantic relations is transitivity which is essential for building a hierarchy: if A is broader than B and B is broader than C, A has also to be broader than C or the organization in a hierarchy would be counter-intuitive.

The most abstract and widely used representative of concept hierarchies are taxonomies. Taxonomies basically organize concepts by supertype-subtype or parent-child relations and according to this all concepts have to be as disjunct as possible. An example for a taxonomy is the Linnaean taxonomy:

**Linnaean taxonomy** is one of the first appearance in science of taxonomies, which was used to classify concepts. In this special taxonomy Carolus Linnaeus classified the known and still existing concepts belonging into the imperium naturea into three different kingdoms: mineral, animal and vegetable. All in all his structure included four other classification levels beneath kingdom: class, order, genus and species. The work of Linnaeus was the starting point for the botanical nomenclature as known today and is used to classify all kinds of animals and plants by biologists.

In comparison to taxonomies, thesauri basically extend the core functionality of taxonomies with additional accepted relations like synonym, antonym or related term to improve their ability to describe the world or a specific domain. The concepts in a thesaurus are not required to be disjunct. An example would be WordNet:

**WordNet** is one of the largest living general English thesaurus. The concepts are called synsets and include nouns, verbs, adjectives and adverbs which are cognitive synonyms describing the concept.

The approaches in this chapter can be used in principle for both thesauri and taxonomies, as well as for other concept hierarchies. To avoid the long-winded term concept hierarchy, we refer to it often as thesaurus or taxonomy, usually depending on the current example dataset: the Medical Subject Headings (Section 3) are usually considered to be a thesaurus, while the Indiana Philosophy Ontology (Section 4) - despite the ontology in the name - is referred to as taxonomy on the project website. Here, we describe both concept hierarchies briefly:

**Medical Subject Headings (MeSH)** is a medical thesaurus created and maintained by the National Library of Medicine (NLM). It includes over 25,000 concepts which are arranged in an alphabetic and a hierarchical structure. This thesaurus is used by the NLM to index medical