Acquisition, Representation and Usage of Conceptual Hierarchies

Marius Paşca
Google Inc.
Mountain View, California 94043
mars@google.com

1 Tutorial Description

Through subsumption and instantiation, individual instances (*"artificial intelligence"*, *"the spotted pig"*) otherwise spanning a wide range of domains can be brought together and organized under conceptual hierarchies. The hierarchies connect more specific concepts (*"computer science subfields"*, *"gastropubs"*) to more general concepts (*"academic disciplines"*, *"restaurants"*) through IsA relations. Explicit or implicit properties applicable to, and defining, more general concepts are inherited by their more specific concepts, down to the instances connected to the lower parts of the hierarchies. Subsumption represents a crisp, universally-applicable principle towards consistently representing IsA relations in any knowledge resource. Yet knowledge resources often exhibit significant differences in their scope, representation choices and intended usage, to cause significant differences in their expected usage and impact on various tasks.

This tutorial examines the theoretical foundations of subsumption, and its practical embodiment through IsA relations compiled manually or extracted automatically. It addresses IsA relations from their formal definition; through practical choices made in their representation within the larger and more widely-used of the available knowledge resources; to their automatic acquisition from document repositories, as opposed to their manual compilation by human contributors; to their impact in text analysis and information retrieval. As search engines move away from returning a set of links and closer to returning results that more directly answer queries, IsA relations play an increasingly important role towards a better understanding of documents and queries. The tutorial teaches the audience about definitions, assumptions and practical choices related to modeling and representing IsA relations in existing, human-compiled resources of instances, concepts and resulting conceptual hierarchies; methods for automatically extracting sets of instances within unlabeled or labeled concepts, where the concepts may be considered as a flat set or organized hierarchically; and applications of IsA relations in information retrieval.

2 Tutorial Outline

- Subsumption, inheritance, instantiation;
- Resources of IsA relations;
- Acquisition of IsA relations from text;
- Applications.

3 Instructor

Marius Paşca is a research scientist at Google. Current research interests include factual information extraction from unstructured text within documents and queries and its applications to Web search.