The Role of Human Generated and Automatically Extracted Lexico-Semantic Resources in Web Search

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

ABSTRACT

The tutorial examines the role of knowledge from lexico-semantic resources, whether human-generated or automatically-extracted, in information retrieval in general, and Web search in particular. A better understanding of the structure and meaning of queries enables a better match of queries against documents, and better ranking of search results. The generation of alternative search results of finer granularity (e.g., quote for a stock symbol, weather forecast for a location, contact info for a business), which more directly answer the user's query, can increase the search effectiveness and the time to result. Similarly, suggesting relevant query completions reduce the time needed to type the entire query, and therefore the time to result. In general, enhancements of the search experience, in the form of spell checking the queries or offering alternative query refinements, represent valuable aids to users of information retrieval systems. Concretely, the tutorial teaches the audience about characteristics of existing, human-generated resources; methods for extracting open-domain classes, instances and relations from the Web; the role of human-generated vs. automatically-extracted knowledge resources, in enhancing information retrieval; implications in semantic annotation of queries, understanding query intent, and information access and retrieval in general.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing; H.3.1 [**Information Storage and Retrieval**]: Information Search and Retrieval; I.2.7 [**Artificial Intelligence**]: Natural Language Processing; I.2.6 [**Artificial Intelligence**]: Learning

General Terms

Algorithms, Experimentation

1. TUTORIAL PRESENTER

Marius Paşca is a research scientist at Google. He graduated with a Ph.D. degree in Computer Science from Southern Methodist University, Dallas, Texas and an M.Sc. degree in Computer Science from Joseph Fourier University, Grenoble, France. Current research interests include factual information extraction from unstructured text and natural-language matching functions for information retrieval.