Information Retrieval Systems as Integration Platforms for Language Technologies

Douglas W. Oard
University of Maryland

Tutorial Outline

1. An interactive IR framework
    a. Ranked retrieval with a bag of terms
    b. Supporting document selection
    c. Supporting query reformulation
2. Web search
    a. Finding useful terms: anchor text
    b. Beyond content: link analysis
    c. Supporting selection: indicative summaries
3. CLIR
    a. What to translate: blind relevance feedback and backoff translation
    b. Crossing the language barrier: evidence mapping
    c. Supporting query reformulation: interactive translation selection
4. Speech retrieval
    a. Finding useful terms: document expansion
    b. Beyond content: dialog patterns
    c. Supporting selection: topic segmentation
5. Evaluation
    a. Inexpensive replication: TREC-like evaluations
    b. Focus on interaction: iCLEF

Abstract

At one time it might have been possible to think of natural language processing, speech processing and information retrieval as separate fields, but an increasing degree of interdependence is now clearly evident. This tutorial will explore those connections from the perspective of information retrieval system design. An overarching framework for interactive retrieval will be introduced, and then specialized to describe Web search, cross-language retrieval, and retrieval from spoken word collections. These applications will then be used to illustrate the critical dependence on component technologies such as computational morphology, acquisition of translation knowledge from corpora, speaker identification, automatic speech recognition, and summarization. Information retrieval systems can offer a useful environment for extrinsic evaluation of new component capabilities, so the tutorial will conclude with a review of evaluation techniques that can help to reveal the contribution of specific components. Attendees will receive a copy of the presentation slides and recommendations for further reading on each of the major topics of the tutorial.

This tutorial is designed for participants that bring expertise in one or more human language technologies. No prior exposure to information retrieval research methods is assumed.