

2007 Student Research Symposium

Language Technologies Institute



Generalizing from Relevance Feedback using Named Entity Wildcards

Abhimanyu Lad

Abstract:

Traditional adaptive filtering systems learn the user's interests in a rather simple way -- words from relevant documents are favored in the query model, while words from irrelevant documents are down-weighted. This biases the query model towards specific words seen in the past, causing the system to favor documents containing relevant but redundant information over documents that use previously unseen words to denote new facts about the same news event.

We propose new ways of generalizing from relevance feedback by augmenting the traditional bag-of-words query model with named entity wildcards that are anchored in context. The use of wildcards allows generalization beyond specific words, while contextual restrictions limit the wildcard-matching to entities related to the user's query.

We test our new approach in a nugget-level adaptive filtering system and evaluate it in terms of both relevance and novelty of the presented information. Our results indicate that higher recall is obtained when lexical terms are generalized using wildcards. However, such wildcards must be anchored to their context to maintain good precision. How the context of a wildcard is represented and matched against a given document also plays a crucial role in the performance of the retrieval system.