

# Abstract

## **Translating Dialects in Search: Mapping between Specialized Languages of Discourse and Documentary Languages**

by

Vivien Petras

Doctor of Philosophy in Information Management and Systems

University of California, Berkeley

Professor Michael K. Buckland, Chair

The biggest problem in searching an information system is to find the appropriate search terms that not only represent the searcher's information need but also match the language used in the information system. This is a translation problem between a specialized dialect of discourse and the documentary language of the information system.

Discourse dialects evolve within specialized communities. They differ from general language and other communities' dialects in terminology (e.g. terms of art, jargon) and grammar patterns. A documentary language is the language used for document representation in an information system. A bibliographic database and its documentary language usually cover more than one domain of discourse.

This dissertation describes a mechanism that will provide a translation aid between specialized languages and the documentary language by suggesting appropriate search terms for a searcher's query in relation to the searcher's domain of discourse. With this kind of vocabulary support in the search process, the different specialized vocabularies can be disambiguated within the information system. Different perspectives on a topic can be represented to the searcher (based on the different discourses of the topic in the collection), which will help in navigating and exploring this information space more effectively.

The search term recommender system, based on statistical associations between specialized language terms and controlled vocabulary terms, is introduced and its applications for automatic text categorization, query expansion and reformulation, and terminology mapping are described. The search term recommender methodology is tested on three specialties in the Inspec bibliographical database and 33 specialties in the Ohsumed database in a text classification application. It is demonstrated that search term recommender systems are more effective when specialty-based.