

## Abstract

Unified Language System for Engineering Design (ULSED): A Framework and  
Automation Tools for Better Design Information Retrieval

by

Jia-Long Wu

Doctor of Philosophy in Mechanical Engineering

University of California, Berkeley

Professor Alice M. Agogino, Chair

In the current fast-paced global market, it is critical for companies to have shorter and more effective product design processes. Information management plays an important role in assisting the design teams to achieve that. However, with the heterogeneous nature of the information sources design teams draw upon, traditional information management systems are not able to provide required information effectively. One major reason is that different information sources use different terminologies. The National Library of Medicine (NLM) recognized this problem and initiated the Unified Medical Language System (UMLS) project that aims at providing a platform for medical information exchange and facilitating electronic biomedical information retrieval from multiple sources. This research expands the UMLS model and proposes the Unified Language System for Engineering Design (ULSED) to allow more effective access and exchange of engineering design information.

A knowledge repository of the concepts in engineering design is needed in ULSED. This research introduces a method for extracting keyphrases from engineering design

documents to suggest terms to represent key concepts. The keyphrase extraction task is modelled as a multi-objective optimization problem and solved with a genetic algorithm. One of the objectives is to maximize the total statistical significance of a set of phrases, which is based on the clumping properties of content-bearing phrases proposed by Bookstein. Another objective is to minimize the number of keyphrases while maintaining the integrity of the phrase set. Results from a set of conference proceedings show promising results when evaluated by a group of experts in engineering design.

ULSED provides another knowledge source as a semantic network to connect concepts. This dissertation applies Latent Semantic Analysis (LSA) to discover the semantic relationships between keyphrases. A prototype of the query expansion mechanism was implemented in an engineering educational digital library to assist users in improving their search queries.

Federating search over and exchanging information between multiple repositories are important in order to achieve effective collaborations between organizations. One obstacle is that these repositories often use different data models. This dissertation develops a framework for translating data models and implements a web search service for federating multiple repositories.