

Using Concept Maps in NDLTD as a Cross-Language Summarization Tool for Computing–Related ETDs

Ryan Richardson (ryan@vt.edu), Edward A. Fox

Department of Computer Science, 660 McBryde Hall 0106, Virginia Tech, Blacksburg, VA 24061

Abstract

Concept maps, introduced by Novak, aid learners' understanding. We hypothesize that concept maps also can function as a summary of large documents (e.g., ETDs). We are building a system that will automatically generate concept maps from English-language ETDs in the computing field. The system also will provide Spanish translations of these concept maps for native Spanish speakers. Using these machine translation techniques, we believe concept maps could allow researchers to discover pertinent dissertations in languages they cannot read, helping them to decide if they want a potentially relevant dissertation translated.

We are using a state-of-the-art natural language processing system, called Relex, to extract noun phrases and noun-verb-noun relations from ETDs, and then produce concept maps automatically. We also have incorporated information from the table of contents of ETDs to create novel styles of concept maps. Currently we are producing concept maps for the Virginia Tech CS collection (175 ETDs), which covers a broad range of computer science. We intend to automatically produce concept maps for computing-related ETDs for a larger segment of the NDLTD holdings. We have recently conducted two user studies, to evaluate user perceptions about these different map styles.

We are using several methods to translate node and link text in concept maps from English to Spanish. Nodes labeled with single words from a given technical area can be translated using wordlists, but phrases in specific technical fields can be difficult to translate. Thus we have amassed a collection of about 580 Spanish-language ETDs from Scirus and two Mexican universities and we are using this corpus to mine phrase translations that we could not find otherwise.

We plan to test the usefulness of the automatically-generated and translated concept maps in a user experiment to be conducted at Universidad de las Americas (UDLA) in Puebla, Mexico. This experiment will determine if concept maps can augment abstracts (translated using a standard machine translation package) in helping Spanish speaking users find ETDs of interest.