

Universal Networking Language: A Tool for Language Independent Semantics?

Amitabha Mukerjee, Achla M Raina, Kumar Kapil,
Pankaj Goyal, Pushpraj Shukla

Indian Institute of Technology, Kanpur, India
{amit,achla,kapil,pankajgo,praj}@iitk.ac.in

Abstract. Given source text in several languages, can one answer queries in some other language, without translating any of the sources into the language of the questioner? While this task seems extremely difficult at first sight, it is possible that the ongoing UN sponsored Universal Networking Language (UNL) proposal may hold some clues towards achieving this distant dream. In this paper we present a partially implemented solution which shows how UNL, though not designed with this as the primary objective, can be used as the predicate knowledge base on which inferences can be performed. Semantic processing is demonstrated by Question Answering. In our system as of now, both the text corpus and the questions are in English, but if UNL can deliver on its promise of a single homogeneous language-independent encoding, then it should be possible to achieve question answering and other semantic tasks in any language.

1 Semantics Models And UNL

Many organizations worldwide are grappling with problems like the following: Given source text in several European languages, would it be possible to demonstrate semantic understanding in some other language (like Hindi) without explicitly translating any of the sources into the language of the questioner? This is, of course, an extremely difficult task, perhaps even an impossibly difficult task. We trust the reader will realize that this paper is merely a very preliminary investigation as indicated by the hesitant “?” at the end of the paper’s title. The key insight driving this research is the realization that if there is a mechanism for mapping any language into a uniform language-independent predicate structure, then it would constitute an important tool in this direction. While no system worldwide is anywhere near succeeding in this effort, the ongoing work on Universal Networking Language (UNL) [2] appears to hold the highest promise in terms of delivering on this dream.

UNL was developed as a universal knowledge-encoding mechanism, and is being primarily driven by the needs of the MT community. UNL provides for a uniform concept vocabulary (called “universal words” or UW’s – the same concept in any language results in the same UW, which is written out using English orthography). These UW’s are connected by a small set of about thirty-eight binary relations to obtain a set of predicate expressions that can encode the linguistic content of any sentence in any language of the world. One of the philosophical issues of course, is that the same con-