HIERARCHICAL ANNOTATOR SYSTEM FOR KANNADA LANGUAGE

BHUVANESHWARI C. MELINAMATH

Department of Computer and Information Sciences, University of Hyderabad, Hyderabad, Andhra Pradesh, India

ABSTRACT

We have developed a wide coverage Hierarchical morpho-syntactic annotator system for Kannada language using hierarchical tag set. Developing annotator using hierarchical tag set is another new attempt, as there are no such systems exists for Kannada or any Indian Language. Attempts tried so far are using flat tag set. Our annotating system relay on five resources 1) Tag set 2) Dictionary 3) Morphological system 4) Named entity recognizer. Morphological system is developed using well defined saMdhi rules and using finite state transducer (FST) transition file shows the order of suffixation. The architecture is general and can be adopted for other language families just by replacing morph relevant information files. There is no hard coding. The system takes a Kannada sentence as input and gives POS tag/tags for each word of the sentence as output. There is not much work is done in automatic processing of Kannada language.

The major types of morphological process like inflection, derivation, and compounding are handled in this system. The goal of our work is to create a new computational model within the framework of finite state technology that will account for word formation processes in Kannada language. Annotation plays an important role in Natural Language Processing applications such as Parsing, Information Extraction, Information Retrieval and Machine Translation. Results are encouraging with respect to noun as compared to verbs. More than 90% results are obtained for nouns and around 85% results are obtained for verbs.

KEYWORDS: Part of Speech (POS), Natural Language Processing (NLP), Finite State Transducers (FST)