

Computational Morphology of Kannada

Kavi Narayana Murthy, Srinivasu Badugu
School of Computer and Information Sciences
University of Hyderabad
email: knmuh@yahoo.com, srinivasucse@gmail.com

Abstract

In this paper we describe our model of computational morphology in detail. We have developed comprehensive morphological analyzers and generators for both Kannada and Telugu and in this paper we shall look at Kannada morphology in some detail. Our work differs from all the other works on morphology in several significant and fundamental ways. Although linguists generally understand the difference between language and script and the natural importance of spoken language over the written form, most practical work in Natural Language Processing (NLP) as also to some extent in modern linguistics seems to be degenerating into mere string manipulation based on the written form of words. There appears to be too much of importance given to the written form and too little to meaning. Words are taken to be sequences of characters separated by spaces. In this paper we highlight the weaknesses of this view point and argue for a more semantically oriented and hence more universal view of what constitutes a word. In this process we will also have some things to say about language, grammar and linguistics. We shall then develop our computational model of morphology involving a number of extensions to Finite State Automata. In the latter part, we shall deal with the morphology of Kannada in some detail within the framework our model of computational morphology. We shall also talk about the metrics for performance evaluation of morphology systems and provide some quantitative evaluations of our system.

Keywords: Morphology, Extended Finite State Models, Kannada