Abstract

The field of natural language processing (aka NLP) is an intersection of the study of linguistics, computation and statistics. The primary goal of NLP is automated understanding of the semi-structured language that humans use. This study stems application in diverse fields like semantic analysis, summarization, text classification and the like. The domain natural language processing is a fledgling domain with no concrete indication of when it will mature. Compared to well established domains like “Study of Algorithms”, NLP is yet in its emerging period and hence there’s dearth of a concise piece of literature that elaborates on the phases of NLP and lists different techniques that can be adapted. NLP borrows heavily from foundational subjects of study like statistics, probability theory and theory of computation. In this paper, we describe three phases of natural language processing namely, language modelling, parts-of-speech tagging and parsing, outlining the approaches used that can be used.
1. Adwait Ratnaparkhi, A Maximum Entropy Model for Part-Of-Speech Tagging
2. DJurafsky, JH Martin, Speech and Language Processing.
3. Michael Collins, Head-Driven Statistical Models for Natural Language Parsing
4. Bill Wilson, University of New South Wales.
5. Roni Rosenfeld, Two decades of statistical language modeling: where do we go from here?

Index Terms

Computer Science Information Sciences

Keywords

NLP, Language Modelling, Parsing, POS tagging, HMM