Functional Parsers For Indian Languages

The construction of tools to parse natural languages is a multi-layered challenge for programmers. The question we pursue here is: are functional languages suitable for building tools to parse Indian languages?

The process of parsing a language involves several steps, and in this study we mainly look towards syntactic parsing, i.e: parsing a sentence in terms of the syntactic roles of the words in a sentence.

Functional, type-theoretic languages are uniquely suited for describing natural languages, and for building NLP tools. An example of such is given in the paper Dependency Grammars as Haskell Programs, by Tomasz Obrebski. We demonstrate this by implementing the dependency parser engine described in his paper, and running it with a small sample English grammar.

Another, more extensive example of functional type-theoretic languages used for NLP applications is Grammatical Framework – a programming language used to write multilingual grammars. Grammatical Framework on both a syntactic and semantic level- semantic relations are defined using a language-independent Abstract Syntax Tree, and “linearised” i.e: converted into strings, using a concrete constituent-based grammar.

Paninian grammar is a type of dependency grammar that is uniquely suited for Indian languages, which are usually free-word order. As part of further work in this project, we attempt to implement a limited-domain basic Paninian Grammar structure, using Grammatical Framework.

Intern Name: Atreyee Ghosal
Mentor Name: Venkatesh Choppella