

Intelligent Sorting System for Curriculum Vitae using Natural Language Processing

Tharindu Weerasooriya^{1*}

Natural language Processing (NLP) has undergone tremendous development over the past few decades. The logic behind sentence analysis plays a vital role in NLP applications. The present study makes use of Stanford CoreNLP, an NLP tool that enables Parts-of-Speech (POS) tagging and NamedEntity Tagging to extract the essential information from a curriculum vitae (CV), followed by ranking the best candidates according to the information included in the CV. The system design is as follows: the proposed system initially categorizes the candidates according to the post applied. The second step checks for the basic qualifications required by the company. If the basic requirements are not met, the CV is rejected. The third step uses POS tagging to interpret and assign marks for each section in the CV. The extracurricular activities section is grammatically ambiguous as it contains achievements in sports, clubs and societies. The research was aimed at classifying the extracurricular activities using a mix of rule based parsers and the NamedEntity Tagger. Firstly, the sentence is passed through the rule based parser, which classifies it as a sport or a club activity (using a word match specific to each group). The category which has the highest match is given $\frac{3}{4}$ mark of the decision. The NamedEntity tagger searches the sentence for any sports or organizations, and the classification is given a $\frac{1}{4}$ point in the decision. The sentence is categorized into the relevant category depending on the highest score. During testing, in a CV which contained 28 extracurricular activities, the system classified 14 achievements in Sports and 14 achievements in Clubs and Societies. However, the correct classification should be 17 in Sports and 11 achievements in Clubs and Societies. The methodology would succeed in sorting ambiguous sentences, where a corpus based method would fail (i.e. "Compered at Kelani Hockey 6's"). The keyword of the sentence is Hockey, but it is not an achievement in sports). Being an adaptable system using NLP, it could be customized to assign a weighted score for specific keywords depending on the requirement of the organization. The fourth step is to assign a total score to the CV. At the end of the cycle, the system would output the list of the top 50 CVs qualified for the post. This system was tested with a sample data set from the CV bank of the Career Fair 2015 (CF) of the University of Kelaniya. The manual CV sorting process of the CF required at least 2 minutes per CV and each CV was sorted individually. The system was less time consuming, more organized and efficient.

Keywords: *Natural Language Processing, Parts-of-Speech Tagging, NamedEntity Tagger, Curriculum Vitae, Keyword Classification*

¹ University of Kelaniya, Sri Lanka *cyriltcw@gmail.com
