Named Entity Boundary Detection for Religious Unhealthy Statements in Social Media

Author: Prasan Yapa

Abstract

Named entity recognition (NER) can be introduced as one of the fundamental tasks in natural language processing. The type and boundary are the two components of a named entity (NE) in text. There are various research have been applied to detect the NE types and boundary as two separate tasks and there is no existing mechanism to consider both of these tasks together. NER in social media analysis considering expressions, disputes etc. can also be considered as something which carries a huge demand. Furthermore, most of these mechanisms have been followed considering English as the testing language corpus and there is an overwhelming demand for such a system which supports for multiple languages including Sinhala. So, the intention is to implement a system for NE boundary detection for Sinhala language considering religious unhealthy statements in social media. Detecting both NE boundary and NE type as an aggregate mechanism will tune up the accuracy and performance of NE linking to knowledge bases. The approach will be determined by some of the aspects such as identifying the existing mechanisms of NE type detection and NE boundary detection, identifying the complexity indexes, matrices and relationships of religious unhealthy statements in Sinhala, identifying the novelty in implementing NE boundary detection considering NE types of religious unhealthy statements and finally enhancing NE linking considering NE boundary detection. The ultimate target is to implement a prototype which out forms the stateof-the-art existing baselines.