A First Step Towards the Development of Yoruba Named Entity Recognition System

Abstract

The NER task can be considered solved for English and a few other European languages given the available research outputs, tools, resources and applications involving NER for these languages. The scenario is sharply different for Nigerian and most of African languages and hence the motivation for the research reported in this paper. The paper presents an exploration of the potency of some language independent features in the recognition of the mentions of persons, locations and organizations in Yor’ub’a text in a supervised machine learning set-up. The results are promising but as further investigations revealed, the size of the training corpus is yet an issue that needs to be addressed.

References


17. D. Seyler, T. Dembelova, L. Del Corro, J. Hoffart, and G. Weikum. A study of the


Index Terms

Computer Science  Information Systems

Keywords

Named entities, NER, Yoruba language, Natural language processing