

Nom et Prénom : Mme BLANCHETE ILHAM

Date de soutenance : 28/09/2024

Directeur de Thèse :MOURCHID MOHAMMED

Sujet de thèse :

Formalizing linguistic resources based on the root and pattern approach to develop an Arabic NLP application using NooJ linguistic platform

Abstract:

This thesis provides a comprehensive linguistic resource for the Arabic language based on the root and pattern approach. We have started by implementing the verb resource and linking it with a Masdar resource. Then, we implemented a noun resource and linked it with a broken plural resource. We have also implemented a transformational grammar that converts a verbal sentence to a nominal one using verb and Masdar resources. We have implemented the linguistic resources using the NooJ linguistic platform; each implemented resource contains a comprehensive dictionary that defines words and associates them with the possible morph-phonemic, morph syntactic, and semantic features. The resource also includes the inflectional and derivational grammars, which generate the possible inflectional forms of each entry and the derived forms, respectively. To implement our resources, we have conducted a linguistic study from which we have extracted the verbs' classification and the rules to generate the broken plural resource. We have implemented a representative model for each verb class, which generates the possible inflectional forms and links the entry with the possible linguistic features. Each representative model has been assigned to verbs with the same linguistic features. We have also conducted a linguistic study to extract rules to restrict the broken plural generation process. We have developed a natural language processing application to exploit our implemented resource; the application performs three main tasks: the first is to perform a complete/semi conjugation process of the chosen verb. The user can also explore the linguistic features of a particular verb, list the possible meaning of the verb, and explore the possible Masdar forms linked to the chosen verb. The application also provides a task where the user can explore different nouns that share a specific root and explore the possible associated linguistic feature to list the possible broken plural linked to the chosen noun. The third task is to extract all Masdar forms that share a specific root and list the associated linguistic features of any Masdar form.