Semantic Annotation and Retrieval in E-Recruitment

E-recruitment processes prioritize matching between a job description and user queries to identify relevant candidates. Existing e-recruitment systems face challenges in extracting job descriptions due to unstructured nature of content, and text nomenclature differences for defining the same content. The systems are particularly unable to extract effectively contextual entities, such as job requirements and job responsibilities from job description. They also lack in finding effectively desired search results due to semantic differences in job descriptions and users English natural language queries. This thesis proposes a framework to cater for challenges in the existing e-recruitment systems.

The proposed framework extracts entities from job descriptions using a domain specific dictionary. The extraction process first performs linguistic analysis and then extracts entities and compound words. After the extraction of entities and compound words, it builds job context using a job description domain ontology. The ontology provides an underlying schema for defining how concepts are related to each other. Besides building a contextual relationship among entities, the entities are also enriched using Linked Open Data that improves search capability in finding suitable jobs. In the proposed framework, Web Ontology Language (OWL) is used to represent information for machine-understanding. The framework appropriately matches users queries and job descriptions.

The evaluation dataset has been collected from various jobs portals, such as Indeed, Personforce, DBWorld. A total of 860 jobs were collected that belong to multiple categories, such as technology, medical, management and others. The dataset was vetted and verified by HR experts. The evaluation has been performed using precision, recall, F-1 measure, accuracy and error rate. The proposed framework achieved an overall F-1 measure of 87.83% and accuracy of 94% for entities extraction. The application has a precision of 99.9% in representing and retrieving job descriptions from its knowledge base. The job description ontology has an overall coverage of 96%. The evaluation results show that the proposed framework performs well in extracting, modelling, enriching, and retrieving job description against queries. At current, the SExEnT framework is not able to automatically generate pattern/action rules, provide a complex ranked retrieval of job descriptions against user profile and automatically extending dictionary to increase extraction precision. In future, the framework can be extended to work on these limitations.

(Malik Nabeel Ahmed Awan)
PhD Candidate

(Dr. Sharifullah Khan)
PhD Supervisor

(Dr. Sharifullah Khan)
Associate Professor
NUST School of Electrical En...
and Computer Science...