### Topic 2: Ontology-guided Job Market Demand Analysis

#### Introduction

This study aims to conduct identification for data science skills by analyzing job postings. job seekers and applications are the first target users for supplying in-demand while educators are the second target users for training and defining courses to shape their potential in high demand. As a quantitative method, Co-Word analysis is used to analyze the occurrence and co-occurrence frequency of skill pairs in the job adverts. The OBIE(Ontology-based Information Extraction) pipeline is utilized by excerpting pre-defined concepts and interpreting text using the ontology concepts. The SARO(the Skills and Recruitment Ontology) are ontology that expends the ’JobPosting’ taxonomy from schema.org1 and the’ Skills and Competences’ with more than 800 concepts from the ESCO ontology. Both methodologies are used in this research. Moreover, LIS (library and information science)and IS(information system) are used in review research to analyze and highlight job skills, changing with providing insight for researchers by supporting to improve the methodology of job advert.

#### Related Work

OBEI is a subfield of IE(information extraction) ontology that provides and definite concept specifically in the IE process. By consideration of Co-word analysis, the difference between present research and previous research is compared to provide:

* + The SARO ontology provides a structured representation for skill analysis and recruiting.
  + The OBIE approach automates keyword extraction from job postings, reducing the "indexer effect" and ensuring consistency in code, saving time and time.
  + Co-word analysis, when combined with inferential statistical analysis, provides a comprehensive examination of skill demands and their relationships within and outside the studied domain.

#### Co-work analysis for job advert

The Paris/Keele co-word approach is used to determine the abilities required for UK data scientists and the job posting format, with a Java implementation that supports longitudinal study.

* + ontology development and enrichment: in this case, SARO allows for the most effective analysis and reuse of job advertisements, which necessitates the right interpretation of abilities and competencies, as well as the formalization of ideas, connections, and examples.
  + Ontology-based information extraction and pre-processing: OBIE effectively pulls job posting and skill concepts from XML, transforms the results to RDF using the SILK framework, loads the recovered RDF triples, and creates a symmetric co-word matrix.
  + Co-word analysis: for cluster analysis, they use the Paris/Keels method to define the central skill, relationship skill, and current skill to provide deeply for dynamic analysis. IGraph and Plotly library are used to visualize the co-word by using Python.

#### User Study and Evaluation

The study's goal is to evaluate the OBIE method's usefulness and performance in extracting important information when compared to manual human extraction, as well as its potential for complementing skills analysis.

* + Data collection and preprocessing: For our study, we rely on 872 job adverts between August to November 2015, crawled from Adzuna .com using the Adzuna API10.
  + Evaluation of the OBIE method:
    - Gold Standard and Inter-Annotator Agreement: Two annotators separately examined a sample of 50 job advertisements to establish the greatest F-score possible for the OBIE pipeline, including IAA findings on subjectivity and aversion.
    - OBIE Evaluation & Discussion: GATE's Corpus Quality Assurance tool was used to calculate accuracy, recall, and F1-scores comparing two annotation sets, with severe and lenient F-measures provided. but in this evaluation, “Skill tools” got only lower F1-(59% strict and 60% lenient) and “Skill product” got only lower F1-(68% strict and 73% lenient).so we need improve performance.

#### Co-work analysis and discussion

The OBIE method's results can be utilized for exploratory data analysis, which are already available in the European Data Science Academy (EDSA) dashboard. The Paris/Keele method is used to abstract OBIE results by reducing data into workable clusters, varying permission values for co-occurrence and the number of permission links.

* + Quadrant I: Clusters 1 and 3 have high centrality and density, indicating they are crucial and core data science skills. Cluster 5 is also close to quadrant I, suggesting it has the potential to become part of the core skills.
  + Quadrant II: Clusters 5, 6, and 9 have high centrality but low density, requiring more investment and preparation to become core data science skills.
  + Quadrant III: Clusters 2 and 4 have low centrality but high density, indicating a close internal connection. Further insights can be gained through time-series analysis or network comparison.
  + Quadrant IV: Clusters 7 and 8 have low density and centrality, suggesting they are marginal and less developed skills. Time-series analysis would provide deeper insights into their contribution to the field.
* The evaluation of demand-derived technical abilities can help job searchers and candidates learn new skills, improve current ones, or pursue technical education. This information may encourage educators and training providers to develop new courses or update existing ones. The constellation of associated skills identifies important technical talents required in data science, such as EDI, Customer Support, SAS, and SQL.

#### Conclusion

The automated extraction approach surpasses the human method by 79% to 83% F-measure, allowing for co-word analysis and third-party querying. Future work will improve the extraction method, maximise demand data value, and detect employment sector trends.