Using Artificial Intelligence for maintaining and improving ESCO

MSWG 13-03

European Skills, Competences, Qualifications and Occupations

1. Background

Artificial intelligence (AI) offers the potential to support, scale up, semi-automate or automate the various steps that are manually performed today to maintain and extend ESCO. It has become a common industry practice to define and maintain so-called knowledge graphs that merge taxonomies such as ESCO, O*Net, etc. The main reason for doing this is that bringing structure to employment related data is crucial to build accurate solutions for solving problems such as job market analysis, promotion of job mobility, courses recommendation and others.

2. Challenges

The main challenges for a classification like ESCO are:

- How to quantify and maintain the relations between the concepts (skills, occupations) in the graph or taxonomy?
- How to map real world data from qualifications (e.g. learning outcomes, course descriptions), candidates (e.g. CVs, profiles and search queries), jobs (e.g. vacancies) and employers (e.g. their products and services) to such a clean and structured taxonomy?

AI in the context of ESCO is about supporting the (nowadays manual) creation of structure within the space of employment data and extracting and visualising valuable insights from it. The AI challenges can be divided in the following categories:

1. *Develop algorithms to map raw, real world data to ESCO*
    a. Named entity recognition for real world data
    b. Semantic matching of real world data to ESCO

2. *Develop algorithms to quantify relations between concepts in ESCO*
    a. Quantify relations between skills
    b. Quantify relations between occupation and skill
    c. Quantify relations between occupations
    d. Quantify relations between course and occupation
    e. Quantify relations between course and skill
    f. Hierarchy building and completion

3. *Develop algorithms to quantify quality issues*
    a. Suspicious/missing relations between concepts
    b. Missing concepts
    c. Duplicate concepts
    d. Vague descriptions
    e. Translation issues

4. *Develop algorithms for employment data analytics*
    a. Identify skill and occupation trends
    b. Emerging skills and occupations

5. *Develop algorithms to visualise ESCO*

3. Use cases

This section presents a non-exhaustive list of practical AI use cases with reference to where they fit in this classification.

a. ESCO skill and occupation identification from free text such as work history and job description (1.a, 1.b)

Raw data from job descriptions, user interests, work history and job titles are usually present as free and unstructured text. It is crucial to develop Natural Language Processing[1] (NLP) algorithms to tokenise, parse and map such flat text to known ESCO concepts. In addition, a great source of information for ESCO is other classifications (both private and public, both national and international). Processing all these classifications manually is a very time consuming process. Automated mapping could make this process significantly more efficient.

b. Link learning outcomes to skills and occupations (1.a, 1.b, 2.e)

One of the objectives of ESCO is to bridge the world of education and the labour market. Just as with the previous use case, doing this in an entirely manual way can be quite time consuming. Therefore, NLP algorithms should be developed to map raw learning outcomes to ESCO skills. The methodology for this is likely going to be similar to the one in the previous use case, but algorithms will need to be retrained given that text from qualifications has a different nature than text from work histories or vacancies. This use case is currently being addressed in the pilot for linking qualifications to ESCO skills.

c. Find close occupations and skills (1.a, 1.b, 2.b, 3.a)

Changes in society are having a big impact on the labour market (e.g. digitisation, greening economy, labour market shocks). These changes require a lot of people to find jobs in different areas or sectors. To that end, it's important to get a better understanding about what occupations are similar to each other and to what extent. This can help people to think about the next step in their career. One way to start this is by looking at the skills that occur in several occupations. Using Machine Learning and NLP can take us further into including semantic meaning when comparing occupations or skills.

d. Detect missing skills and occupations (1.a, 1.b, 3.b)

By analysing large amounts of CVs and vacancies, new terms and skills should emerge. Currently this is done through a manual, lengthy process of gathering feedback from different stakeholders. AI can help identifying new occupations and skills through big data analytics or by automated knowledge graph building.

[1] Natural Language Processing is the application of computational techniques to the analysis and synthesis of natural language and speech. A natural or ordinary language is any language that has evolved naturally in humans.

e. Detect ESCO quality issues (3.a – 3.e)

The current manual management of ESCO requires a lot of effort and resources. AI can help to find inconsistencies in the classification such as duplicates, discrepancies between the term of a skill and its description, etc. NLP can be used to detect quality issues, by using a more evidence-based approach with actual real-world statistical data, so that manual effort is spent more wisely. In addition, at the moment a heuristic approach is used to clean up the data. For example, a maximum number of optional skills per occupation is targeted. Doing this using a more evidence based approach based on actual real world statistical data will improve quality.

4. Proposals by the ESCO Maintenance Committee

During the meeting of 22 October 2020 with the ESCO Maintenance Committee (MAI), the Commission presented the goals and methodology adopted in the rollout of AI models for maintaining and improving ESCO. The MAI confirmed that there is an important role for AI to play in the future maintenance of ESCO. The main points discussed concern the complexity of building a representative data set to train the different AI models. More specifically, MAI members highlighted the following points:

- Building such a model requires an extensive understanding of the different economic sectors of the labour market and the ability to compare similar occupations and similar skills. This means that for a first training of the model more general knowledge could suffice, while additional effort might be required for later developments.
- It is crucial to use the most appropriate data. Data collected from online vacancies might have different levels of detail compared to the ESCO data. Similarly there might be imbalances in the data regarding representation of different Member States. Only homogeneous data can be compared and the AI models should compensate for imbalances in the data.
- The importance of having transparent AI models was mentioned. It would be helpful to indicate from which kind of data the individual AI models are obtained.
- Language could be a barrier in different ways. Even if the model is currently trained using only the English language, the interest of the Commission is to build a model using all the 27 ESCO languages. To respond to this need, a large amount of data in every language is needed. Moreover, MAI members with experience using AI in different languages have expressed the need to continuously support the models in the translations of terms with multiple meanings.

Overall, when sharing experiences with colleagues who have already worked with AI models, similarities in methodologies have been spotted. Further collaborations in data sharing are expected.

5. Input from the ESCO Member States Working Group

The Commission will invite the Member States' input on technical issues related to the usage of AI in ESCO, in particular on types of useful data, possible contributions by Member States and possible other efforts for using AI in the labour market. The

Member States' representatives at the MSWG will receive such questions in the coming days with an EU Survey, for replying after further elaboration at the national level. The same questions will be presented at the MSWG meeting of 19 November 2020.