European Commission

Information Note

Using Artificial Intelligence for maintaining and improving ESCO

MAI 34-03

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 to solve problems such as job market analysis, promote job mobility, recommend courses, 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?

Artificial intelligence 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 artificial intelligence 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 relation between skills
    b. Quantify relation between occupation and skill
    c. Quantify relation between occupations
    d. Quantify relation between course and occupation
    e. Quantify relation 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 algorithms to tokenise, parse and map such flat text to known ESCO concepts. The mapping method should be probabilistic in order to compute a confidence metric for each prediction. In this way, it is possible to minimise the number of wrong matches by only allowing predictions that have high confidence. Ideally the method should take the context into account when assigning input to an ESCO concept. Obtaining such a mapping algorithm can also be used to generate mapping files and to identify where there is a lot of variability in real world data (e.g. a large amount of different job titles mapping to the same occupation). 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, natural language processing 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. The pre-processing pipeline is also going to be tailored to the use case. 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 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 which skills occur in several occupations. However, using machine learning and natural language processing, we can go further to include 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.

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

Since the ESCO classification is managed manually, this requires a lot of effort and resources. Effort that would be better spent by evolving ESCO and preparing it for the future. We believe AI can help to find inconsistencies in the classification such as duplicates, discrepancies between the term of a skill and its description, etc. Natural language processing can be used to detect quality issues 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.