

# 7<sup>th</sup> OJA Forum | 25 June 2024

Online Event

All times stated in CEST

## Agenda

### Session 1 | Tooling Up: Visualizing and Utilizing Online Job Ad Data

11:00 am **Welcome**

11:15 am **From data to knowledge on skills**

Jiří Braňka - European Centre for the Development of Vocational Training (Cedefop)

Building on over a decade of experience in developing and disseminating skills intelligence, Cedefop explores various ways how to present the online job advertisement data, so it gives detailed and value-added information, while respecting the specifics and biases of the source. With new indicators in development, Cedefop attempts to measure better how skills are important and relevant across jobs.

11:45 am **The Future of Work Initiative in the Arab Region: ESCWA's Skills Monitor Pipeline**

Sama El Hage Sleiman – United Nations Economic and Social Commission for Western Asia (ESCWA)

The ESCWA is leading an innovative project to systematically analyze and capture the diverse array of skills from 22 Arab countries through the ESCWA Skills Monitor (ESM) tool. The presentation will illustrate the ESM pipeline from collecting and curating OJA data sets to their augmentation. Our discussion will include an in-depth look at the "Skills Forest" a unique model that illustrates the dynamic and intricate relationships between different jobs, illustrating mobility between jobs based on diversified skills. The session will conclude with a live demo of the ESM, highlighting its real-time analytical power and its transformative potential for policymakers, educators, and industry professionals.

12:15 pm **From data to action: developing a LMI model for employability in a Spanish Online University**

Federico Christmann – Open University of Catalonia

The "Labor Market Information for Educators" (LMI4E) shows how Higher Education Institutions can benefit from LMI to align teaching programs with labor market insights and improve students' employability. Our tool feeds from Lightcast's OJV data, ESCO, and other secondary sources to

provide meaningful guidance to reshape teaching programs and offer new curricula. This is implemented following a production process that consists of a) data cleaning and validation, b) identification of the relevant labor market for each degree, c) data storage according to the institutional framework, and d) data visualization and reporting.

12:45 pm **Break**

## Session 2 | Mapping Occupational Titles to OJAs

13:45 pm **Profession classification in the messy real world: methodology and challenges**

Kasper Kok – Textkernel

Job titles in online job ads come in many forms of wording and granularity. Therefore, standardization to a consistent catalogue or taxonomy is a prerequisite to effective analytics and matching. Most public taxonomies are created in a 'top-down' fashion and don't reflect the variety of job titles as seen in OJAs. Textkernel maintains a 'bottom-up' professions taxonomy instead, for which CV and job ad data is the starting point. In this presentation, we will discuss aspects of the methodology for creating this taxonomy and keeping it up to date and aligned across languages. We will also discuss some of the main challenges, such as dealing with differences in granularity, seniority distinctions, and noise removal.

14:15 pm **A Hybrid Methodology for Job Ad Title Normalization**

Ibrahim Rahhal – International University of Rabat

Occupation identification is an important step in structuring job ads and gaining insights from them. Using traditional approaches requires extensive labeled datasets, limiting their applicability. We propose a hybrid methodology using BERT for sector classification and similarity measures for job title identification. We introduce a novel document embedding strategy to enhance accuracy. Results show that incorporating document embedding-based methods, such as weighting and noise removal, boosts accuracy by 23.5% compared to Bag of Words models surpassing 85% accuracy in some sectors. Evaluation demonstrates superiority or parity with state-of-the-art methods, highlighting efficacy in identifying emerging occupations in the Moroccan job market.

14:45 pm **Break**

14:55 pm **Leveraging Semantic Search with BERT: Classifying German Job Titles using the KLDB**

Rahkakavee Baskaran – &effect

This session demonstrates a semantic search approach for classifying German job titles from online job advertisements using KIdB (german classification of occupations). The algorithm is based on a BERT model (Bidirectional Encoder Representations from Transformers). The classification has an accuracy of 0.86 and a macro f1 value of 0.70 on the five digits of the KIdB classification. In her presentation, Rahkakavee Baskaran explains the process from data pre-processing to model development and highlights the challenges encountered and insights gained.

15:25 pm **Demystifying Deep Learning: Unveiling the Mechanisms Behind a KIdB Job Requirement Level Classifier through Explainable AI**

Kai Krüger – Federal Institute for Vocational Education and Training

This presentation introduces a transformer-based deep learning model designed to predict the fifth digit of the KIdB (german classification of occupations), which represents the job requirement level, for a given job posting. Using the explainable AI technique known as Integrated Gradients, we demonstrate which parts of the input—specifically, which words—are most influential in the model's decision-making process. This analysis helps verify whether the patterns learned by the model represent plausible indicators for classifying the job requirement level, such as job titles, qualifications, or tasks, or if they merely reflect statistical artifacts in the training and evaluation data.

15:55 pm **Feedback and farewell**