

Real-time Labour Market Information on Skill Requirements

Setting up the infrastructure for EU system

Overview of methodological and technical approach – Key Points

Mario Mezzanzanica
Expert Workshop
Milan, 20-21 March 2018

Overall Data Flow - Recap



Data
Ingestion

Pre-Processing

Information
Extraction

ETL

Presentation
Area



CEDEFOP

European Centre
for the Development
of Vocational Training

Data Ingestion

- **Goal:**
 - Collect massive amount of heterogeneous data from several unpredictable number of sources
- **Challenges:**
 - Guarantee data completeness and consistency
- **Approach:**
 - Develop a multi-technique framework (crawling, scraping, API) to fit different website characteristics: volume of vacancies, different languages, technology structure, non-invasive approach and policy agreements
 - Prevent data sources losses, via redundancy policies
 - Detect and collect metadata to improve information value
- **Features:**
 - To grant process governance in data Ingestion phase, we developed a monitoring system with a scheduling tool and an alerting module



CEDEFOP

European Centre
for the Development
of Vocational Training

Data Pre Processing

- **Goal:**
 - Feed information extraction phase with proper data
- **Challenges:**
 - Measure, monitor and increase Data Quality, to maximize completeness, consistency, complexity, timeliness and periodicity
- **Approach:**
 - Develop a multi-phase pipeline, focused on:
 - Vacancy Detection: analyze website page to select only content referred to vacancies
 - Deduplication: detect duplicated vacancy posts to obtain a single vacancy entity
 - Date detection: identify release and expire dates through vacancy description analysis
 - Vacancy duration: method to define expire date, when not explicitly available
- **Features:**
 - Guarantee Data Quality during all processing phases



CEDEFOP

European Centre
for the Development
of Vocational Training

Information Extraction

- **Goal:**
 - Extract and structure information from data, to be provided to the presentation layer
- **Challenges:**
 - Handle massive amount of heterogeneous data written in different languages
- **Approach:**
 - Develop an adaptable framework, tailored on different information features. Some relevant challenges:
 - **Occupation** feature classification: combined methods such as Machine Learning, Topic Modeling and Unsupervised Learning
 - **Skill** feature classification: another different combined methods, such as Text Analysis with corpus based or Knowledge based similarity
- **Features:**
 - Guarantee Explainable information extraction, logging classification methods and relevant features.



CEDEFOP

European Centre
for the Development
of Vocational Training

Presentation

- **Goal:**
 - Provide data in an Analytic environment, to support users across their different analysis needs
- **Challenges:**
 - Personalized data access to different types of stakeholders
- **Approach:**
 - Develop different navigation paths, for different users:
 - **Data scientists/analysts:** access data via a Discovery Lab environment, to ensure free data analysis needs
 - **Decision makers:** access information through storytelling and dynamic dashboards



CEDEFOP

European Centre
for the Development
of Vocational Training

Some references of our research group

(take a picture on QRcode to get the paper
Notice. PDF are for personal use only)



1. Mezzanzanica & Mercorio: [Big Data enables Labour Market Intelligence](#). In Encyclopedia of Big Data Technologies. Springer, 2018.
2. Mezzanzanica et al. [WoLMIS: a labor market intelligence system for classifying web job vacancies](#). Journal of Intelligent Information Systems, 2017.
3. Mezzanzanica et al. [Using Machine Learning for Labour Market Intelligence](#). In ECML PKDD 2017, LNCS. Springer, 2017
4. Mezzanzanica et al. [A Language Modelling Approach for Discovering Novel Labour Market Occupations from the Web](#). In 2017 IEEE/WIC/ACM International Conference on Web Intelligence, 2017.
5. Colombo, Mezzanzanica & Mercorio. [Applying ML Tools on Web Vacancies for Labour Markets and Skill Analysis](#). “Terminator or the Jetsons? The Economics and Policy Implications of Artificial Intelligence”, 2018



CEDEFOP

European Centre
for the Development
of Vocational Training