TOWARDS SMART STATISTICS IN LABOUR MARKET DOMAIN

Inna Novalija, Marko Grobelnik
Artificial Intelligence Laboratory
Jožef Stefan Institute
Introduction

• Major initiatives related to digitalization of labor market - training of new skills and meeting the labour demand.

• Traditional way/standard measures of the labour demand and labour supply based on surveys – job vacancy surveys, wage survey, labour force surveys.

• The data-driven smart labour market statistics intends to:
  o use the available historical job vacancies data,
  o use the available real-time job vacancies data,
  o use the available real-time and historical dataset of additional data (described below),
  o align data sources,
  o construct models and obtain novel smart labour market indicators that will complement existing labour market statistics,
  o provide a system for delivering results to the users.
Background

Policy questions:
- job creation,
- education and training systems,
- labour market segmentation,
- improving skill supply and productivity.

Stakeholders:
- Statisticians from National and European statistical.
- Individual persons who are searching for new employment opportunities.
- Public and private employment.
- Education and training institutions from different levels and forms of education.
- Ministries of labour/manpower, economy/industry/trade, education, finance, etc.
- Standards development organizations.
- Academic and research institutes.
Data Sources (1)

- **The main data sources** available for the development of smart labour market statistics are historical and current data about job vacancies in the area of digital technologies and data science around Europe (~5,000,000 job vacancies 2015-2018).

- **Additional data sources** may include:
  - Labour supply data (based on user profile analysis).
  - Social media data, such as news, Twitter data that might be relevant for labour market.
Data Sources (2)

Example of Job Vacancies Crawled and Processed

<table>
<thead>
<tr>
<th>JOB LIST</th>
</tr>
</thead>
<tbody>
<tr>
<td>10770 JOBS FOUND OUT OF 4444880</td>
</tr>
<tr>
<td>TIME INTERVAL: 12/11/2017 - TODAY</td>
</tr>
<tr>
<td>BIOSTATISTICIAN - OBSERVATIONAL STUDIES HEOR</td>
</tr>
<tr>
<td>Quintiles, Barcelona, Spain</td>
</tr>
<tr>
<td>PUBLISHED ON JANUARY 7, 2018</td>
</tr>
<tr>
<td>DESCRIPTION</td>
</tr>
<tr>
<td>- Analyze plan, statistical analysis and final statistical reports using the appropriate methodology. Principal responsibilities: Osteoarthritis and quality of life. Language: English.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ANALISTA DE DATOS - R AVANZADO, MADRID</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFI Informática, Madrid, Spain</td>
</tr>
<tr>
<td>PUBLISHED ON JANUARY 7, 2018</td>
</tr>
<tr>
<td>DESCRIPTION</td>
</tr>
<tr>
<td>- Describir diferentes productos y mostrarlos mismos como distribuidor de seguros de salud, tickets de comidas, guía del transporte, ADSL, etc.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SOFTWARE QUALITY ASSURANCE INTERN FOR DATA SERVICES JOB</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spain</td>
</tr>
<tr>
<td>PUBLISHED ON JANUARY 7, 2018</td>
</tr>
<tr>
<td>DESCRIPTION</td>
</tr>
<tr>
<td>- Support for the development of new technology and innovation platforms. Business Analytics &amp; Technologies. Enterprise Information Management. Hecho +10 años en SAP</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FULLSTACK PHP DEVELOPER, MADRID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open Sistemas, Madrid, Spain</td>
</tr>
<tr>
<td>PUBLISHED ON JANUARY 7, 2018</td>
</tr>
<tr>
<td>DESCRIPTION</td>
</tr>
<tr>
<td>- Desarrollador de aplicaciones. Integración en equipo de trabajo en compañía discreta y liderar en productos y servicios relacionados con integración no. Big Data... Hasta 12 días en Teamexample.com</td>
</tr>
</tbody>
</table>
Data Sources (3)

Map of Recent Job Vacancies

**MAP**

Shows the number of locations that offer job positions for the given query. Clicking on the pin gives the location name and number of jobs found for that location.
Conceptual Architecture

Data Sources
- Web-scraped job vacancies
- Labor supply
- Additional data sources

Modeling
- Demand analysis
- Ontology development
- Ontology evolution

Target users
- Individual persons (residents, non-residents)
- Policy makers
- Educational and training centers
- Academia, research
- Statisticians
- Other

User cases/Applications
- Cross-country level
- Country level
- City/area level
- Conceptual level
Scenarios

• Demand Analysis
Demand analysis scenario suggests production of statistical indicators based on the available job vacancies using techniques for data preprocessing, semantic annotation, cross-linguality, location identification and aggregation.

• Ontology development
In the ontology development scenario an automatic (or semi-automatic) bottom-up process of creating ontology from available job vacancies will be suggested.

• Ontology evolution
Scenario 3 will suggest an automatic (or semi-automatic) ontology evolution process based on the real-time job vacancy stream.
Statistical Indicators

• Up-to date job vacancies statistics on a cross-country/country/city(area) level. Example: job vacancies in UK and France in the last month

• Up-to date skills statistics on a cross-country/country/city(area) level. Example: top 10 skills in UK in the last month

• Up-to date location statistics. Example: top locations for specific skill

• Ontology development statistics. Example: number of concepts in the ontology

• Ontology evolution statistics. Example: emerging skills in the ontology in the last 3 months
Conclusion

- We presented a proposal for developing smart labour market statistics based on streams of enriched textual data, such as job vacancies from European countries.

- We defined smart statistics scenarios, such as demand analysis scenario, skills ontology development scenario and skills ontology evolution scenario.

- The future work would include the implementation of the smart labour market scenarios, quality assessment and evaluation of the produced statistical outcomes.
Questions?