Using big data & AI for identifying Labour market information in Austria

Claudia Plaimauer 3s Unternehmensberatung GmbH www.3s.co.at

ILO Workshop Geneva, 19-20/09/2019

Our incentives for using big data & AI

- Save time and money;
- Gain a little more independence from traditional data collections;
- Expand our possibilities of data analysis e.g. when
 - Monitoring national labour market demand;
 - Analysing occupational skills profiles;
 - Generating input for the maintenance of labour market taxonomies.

Preliminary project experience, e.g.

- Test AI methods for
 - normalising free text survey results;
 - processing online vacancies;
 - validating occupational skills profiles;
 - maintaining labour market taxonomies.
- Support Textkernel in setting up Jobfeed AT (<u>www.jobfeed.com/at/home.php</u>), a big data platform for systematically querying Austria's online job market.
- Conduct annual analysis of Austrian online vacancy market since survey year 2016.
- Use Jobfeed-based analysis to inform curriculum development;
- Investigate potential of automation for a comparison of European VET qualifications.

Testing AI methods and big data in taxonomy management

Goal 1: Validation of taxonomy terms

- Are the taxonomy terms chosen to represent occupational requirements (,skills') actually used in vacancies?
- Which formats occur frequently in vacancies, which rarely, or never?

Goal 2: Amendment of taxonomy

- Are skills terms commonly used in vacancies missing?
- Is there any indication that skills concepts are missing?

Goal 3: Does the current format of the taxonomy in any way impede its application in NLP?

The Austrian PES' central LM taxonomies



Method used for validating 'skills' terms

Text string matching Frequency counts

29,000+ ,skills' designations from PES' taxonomy Requirement descriptions taken from 850,000+ vacancies

Insights from validating 'skills' terms

- 1. Clear, descriptive, consistent and unambiguous taxonomy terms have a format that doesn't fully correspond to the language used in vacancies: 56% of our ,skills' terms never appeared in job ads.
- 2. The longer the taxonomy term the less frequently it is used in vacancies.
- 3. Preferred (PT) and non-preferred (NPT) taxonomy terms mostly show the same frequency distribution.
- 4. Some naming strategies are a hindrance in NLP and should be avoided, e.g. excessive contextualisation (,Use 3-dimensional printing in car construction'), or explanatory adjuncts in parentheses (,determination of optimum selling price (basics)').

Mix of methods used for supplementing the ,skills' thesaurus

Automated methods:

- Key word extraction;
- Frequency counts;
- Data cleansing (detection of spelling variants, declensions and typing errors);
- Key word classification;
- Text string matching;
- Co-occurrence analysis.

Editorial methods:

- Supplemental investigations to clarify content and use;
- Semantic analysis;
- Terminology control;
- Conceptual modeling.

textkernel



Insights from supplementing 'skills' terms

- The Austrian 'skills' taxonomy is fairly comprehensive: only 1900+ words/phrases/text segments were identified as missing (of these only approx. 900 specialist skills); in the end 97 new specialist skill concepts and a total of 635 new specialist skill terms were supplemented.
- 2. Text mining is a highly effective method for identifying evidencebased amendment needs for thesauri, but it comes at a price.
- 3. Automated skills mining delivers amendment candidates only; considerable additional human processing is necessary to fully integrate new terms, or even concepts.
- 4. In order to improve its applicability in NLP our 'skills' taxonomy should also include formats predominately found in vacancy text, at least as hidden search terms.
- 5. Certain naming strategies should be avoided to improve the taxonomies applicability in NLP.

Thank you for your attention!

Claudia Plaimauer 3s Unternehmensberatung Wiedner Hauptstraße 18 1040 Vienna, Austria Tel +43-1-5850915/33 plaimauer@3s.co.at

