Using vacancy mining for validating & supplementing labour market taxonomies

Challenges and lessons learnt

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

26 June 2018
5th Cedefop Brussels-based seminar
LMSI systems for VET policies

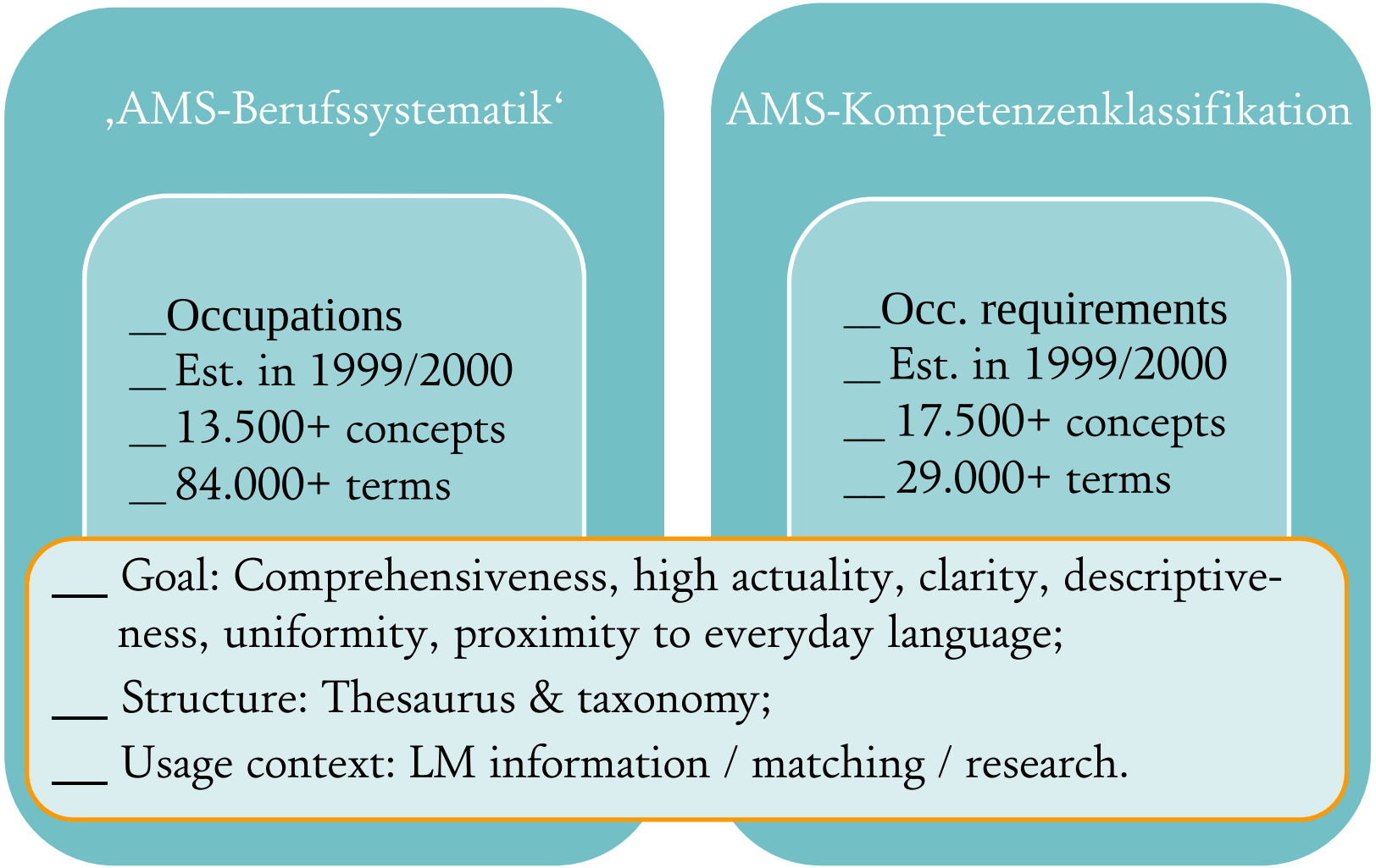
Content

- Use of AI and Big Data in LMSI at 3s;
- The Austrian PES' central LM taxonomies;
- Testing AI-based methods for taxonomy management:
  - Goals & expectations;
  - Research questions;
  - Expected significance of results;
- Validation of terms;
- Enrichment of vocabulary & conceptual content;
- Lessons learnt & outlook.

Use of AI and Big Data at 3s

- 2013: 3s & Textkernel (www.textkernel.com) test automatised coding of free text survey results (occupations, occ. requirements, training needs);
- 2014/15: 3s tests semantic technologies for validating occupational skills profiles (in the context of Cedefop's mid-term skills supply and demand forecasts);
- 2015: Jobfeed AT (www.jobfeed.com/at/home.php) goes online (big data platform for systematically querying the Austrian online job market);
- 2017: Pilot project to test potential of semantic technologies for taxonomy maintenance tasks;
- 2017 & 2018: Analysis of Austrian online vacancy market (based on data from Jobfeed; results implemented in AMS Skills Barometer (bis.ams.or.at/qualibarometer).

The Austrian PES' central LM taxonomies



...and their maintenance

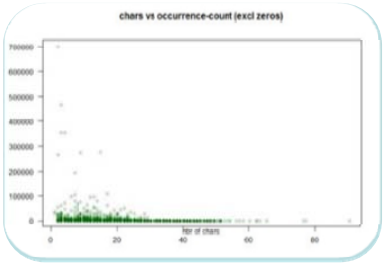
Impulse for amendments comes from

- Expert and non-expert users of these taxonomies;
- Guided, but also spontaneous feedback;
- User-independent quality checks;

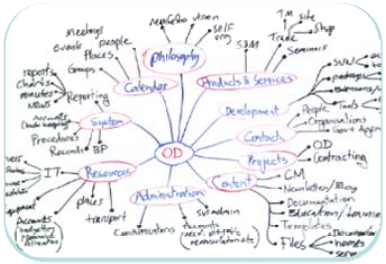
Techniques used in maintenance

- Editorial evaluation of user input/feedback;
- Functional analysis;
- Gap analysis;
- Semantic analysis;
- Terminology control;
- Computer-assisted evaluation of vacancy text. 

Testing AI-based methods for taxonomy management: Goals and expectations



Validation



Enrichment

Insights

Savings

Testing AI-based methods for taxonomy management: Research questions

Research question	Goal
Do taxonomy terms for occupational requirements actually occur in vacancies - and if yes, with which frequency?	➢ Validation of ‚skills' designations
Are terms which are frequently used in vacancies missing in the Austrian PES' taxonomies?	➢ Enrichment of vocabulary ➢ Detection of new concepts

Testing AI-based methods for taxonomy management: Expected significance of results

Lacking frequency of occurrence cannot be considered as incontrovertible evidence for a thesaurus term's futility because

- Taxonomies do not duplicate but interpret and structure reality; they aim at building a comprehensive model of a specific section of our world – and thus also contain 'structuring elements' without any observable LM relevance.
- It cannot be taken for granted that vacancy text always contains perfectly balanced occupational skills profiles (e.g. concealment of tacitly expected requirements, inflationary use of soft skills);
- Job titles (vacancy headings) and professional titles ('occupations') have different linguistic functions and context which is reflected in their wording.

Testing AI-based methods for taxonomy management: Expected significance of results - continued

Words/phrases extracted from vacancy text must always undergo terminological control prior to inclusion into a taxonomy, because

- Taxonomy terms aim at clarity, descriptiveness and consistency; taxonomy concepts are given unique preferred names that follow specific terminological rules, whereas the 'language of the labour market' at the most only follows conventions;
- Vacancy text exhibits the usual characteristics of naturally occurring language: misleading or vague wording, orthographic and grammatic errors, discriminatory practices, stylistic blunders, etc.

Testing AI-based methods for taxonomy management: Expected significance of results - examples

Job title → professional title /occupation:

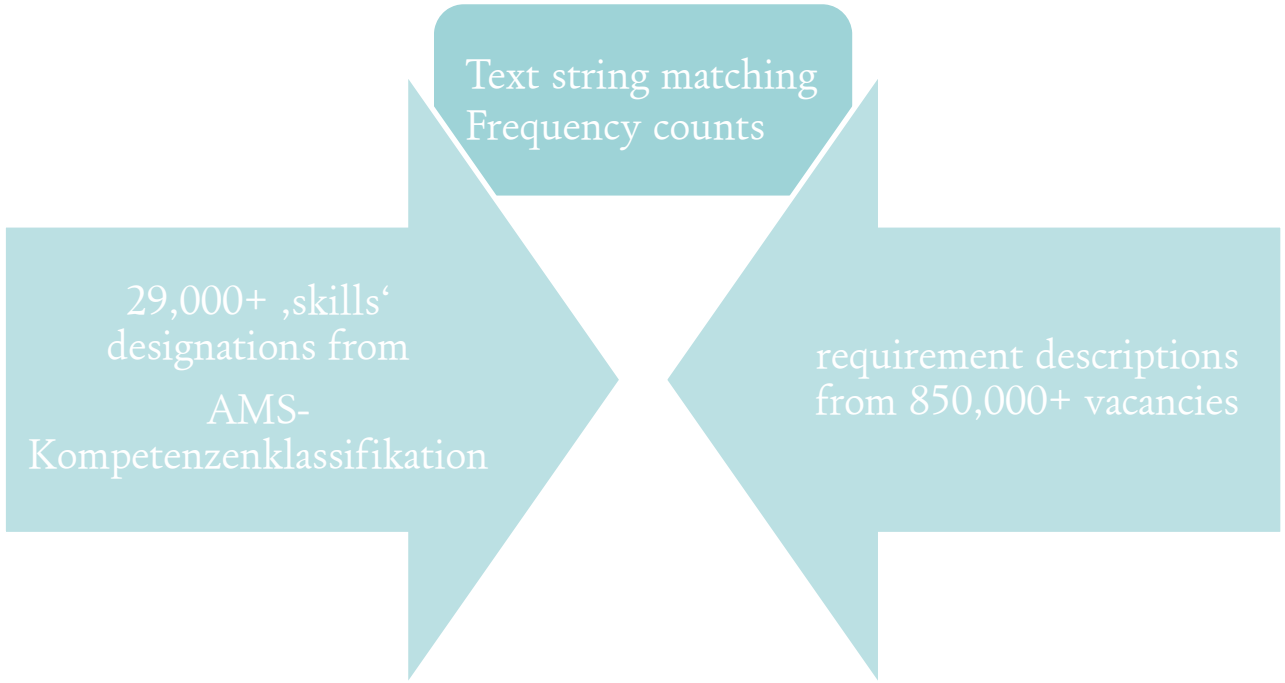
- ~~Senior~~ Projektmanager~~/i~~In
- Interviewer~~/i~~In ~~auf Werkvertragsbasis~~
- CATIA-KonstrukteurIn ~~im Flächendesign mit V5 / NX (w/m)~~
- BilanzbuchhalterIn ~~(m/w) mit Konzernerfahrung~~
- Brand ManagerIn ~~(w/m) für erfolgreiche Top-Marke im Food-Bereich~~;
- Customer Care Agent~~s~~ (m/w) ~~(TZ 30 Std./Wo)~~

Textstrings from vacancies → ‚skills' concepts:

- kommissionieren mit einem Handscanner im Kühlhaus
- Arbeiten mit Handscanner
- Scanntätigkeiten mit dem Handscanner
- Erfahrung mit Handscanner
- Kommissionierung und Umgang mit Handscanner
- Kommissionieren mit Handscanner
- Buchungen mittels Handscanner

Umgang mit Handscannern

Validation of ‘skills’ terms: Method used by Textkernel



Validation of ‘skills’ terms: Results

- 56% of AMS-Kompetenzenklassifikation‘s ‚skills‘ terms never appeared in vacancies;
- Negative correlation between term length and frequency of occurrence;
- Frequency distribution of +/- appearance in vacancies barely differed between preferred and non-preferred terms...
- ...but some non-preferred terms (NPTs) occurred much more frequently than their affiliated preferred term (PT);
- Some sub-sections of the ‚skills‘ taxonomy are closer aligned with the language of recruiters than others.

Validation of 'skills' terms: Results - some examples

No occurrence in vacancies, e.g.

- *Baugeräte warten und reparieren*
- *Verkaufspreis ermitteln (Grundkenntnisse)*
- *Tierbälge gegen Schädlingsbefall imprägnieren*

Frequency > 60.000:

- *Berufserfahrung*
- *Deutschkenntnisse*
- *Reisebereitschaft*

NPT more frequent than PT:

- NPT *Neukundengewinnung* (F=1.778) - PT *NeukundInnen-akquisition* (F=50)
- NPT *Raumplanung* (F=239) – PT *Raumplanungskenntnisse* (F=0)

Enrichment of vocabulary & conceptual content of the ‚skills' thesaurus: Mix of methods

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:

- Identification of additional open source data of related content;
- Exclusion of spelling variants, declensions, typing errors;
- Interpretation of quantitative output;
- Analogous & supplementary searches;
- Semantic analysis;
- Terminology control.

Enrichment of vocabulary & conceptual content of the ‚skills' thesaurus: subsequent editorial processing

- Editorial evaluation of amendment candidates;
- Supplementory enquiry in Jobfeed and other web ressources to clarify content, context and relevance of amendment candidates;
- Terminological adjustment of automatically detected terms to fit prescribed thesaurus format;
- Addition of NPTs, hidden search words, definitions and scope notes;
- Integration of new terms/concepts into semantic structure of taxonomy.

Enrichment of vocabulary & conceptual content of the ‚skills' thesaurus: from automatic detection to editorial integration

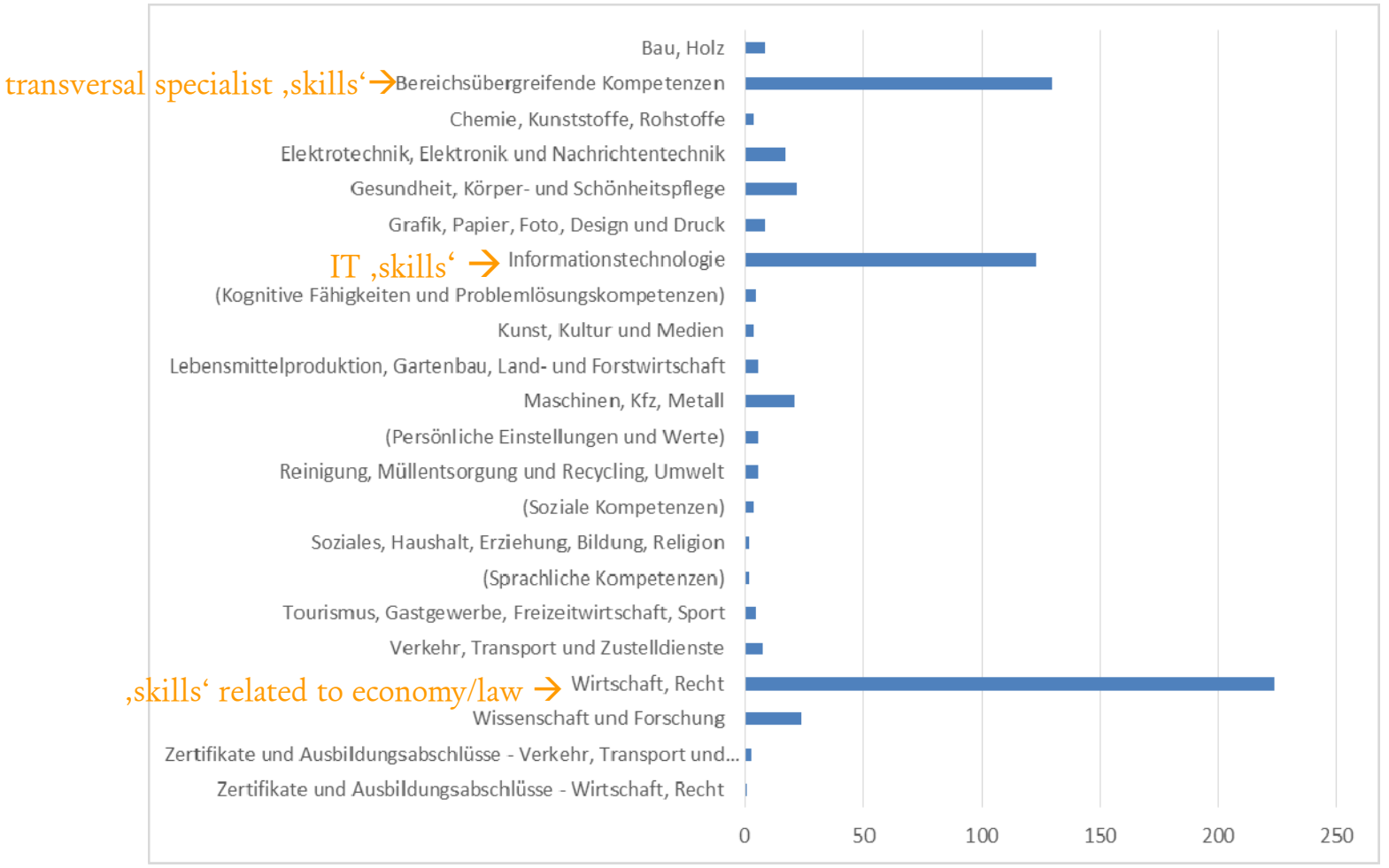
Output of 'skills'mining:

- 1.900+ potentially 'new' occupational requirements;
- approx. 900 of these resembled specialist 'skills';
- all 'skills' terms listed with frequency of occurrence and context (most frequently co-occurring occupation);

Result of subsequent editorial processing (focus on specialist 'skills'):

- Addition of 635 terms to the 'skills' thesaurus, of these
    - 366 NPTs;
    - 172 hidden search terms;
    - 97 PTs (= new concepts).

Enrichment of vocabulary & conceptual content of the ‚skills' thesaurus: Results



Lessons learnt & outlook

Text mining is a highly effective method for identifying evidence-based amendment needs for thesauri, but it comes at a price.

→ repeat text mining only at larger intervals.

There is a hard to reconcile tension between controlled vocabulary and natural language, especially

- pre-coordinated (e.g. *Schablonen herstellen und Dekorationstechniken kennen (Grundkenntnisse))* and
- formally disambiguated terms (e.g. *Hamster (Mail Transfer Agent))*

hamper the applicability of a taxonomy in automated vacancy coding.

→ Taxonomy should also include formats predominately found in vacancy text as NPTs or hidden search words to improve automated normalisation of requirement text.

Thank you for your attention!

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