

# Key question 3: How to balance automated processing and human interpretation?

Key findings from work assignment 2 of Cedefop's Comparing VET-qualifications project

Marye Hudepohl  
Ockham-IPS, Netherlands



Claudia Plaimauer  
3s, Austria



# Overview

- \_\_Goals & Research questions
- \_\_Approach
- \_\_Workflow
- \_\_Conditions on the workflow
- \_\_Operationalising the workflow
- \_\_Testing an automated workflow
- \_\_Results of the testing exercise
- \_\_Conclusions & Recommendations
- \_\_Discussion



# Goals & Research questions

## **Goal:**

Design and test an automated workflow (prototype) for comparing learning outcomes (LOs) descriptions included in national qualification with the ESCO KSC skills pillar.

## **Research questions:**

- \_\_How can digital technologies support the gathering, structuring and analysis of VET qualifications?
- \_\_How can digital technologies address the linguistic challenges involved?
- \_\_What role could ESCO play in an automated comparison of VET qualifications?



# Approach

## **Desk research, expert consultations**

- \_\_Propose a workflow;
- \_\_Identify existing/emerging digital technologies for potential use;

## **Prototype development**

- \_\_Select the most promising technologies;
- \_\_Identify and tackle encountered challenges, tailoring existing applications in a trial and error approach;

## **Prototype testing**

- \_\_Compare LOs of one selected VET qualification against full set of ESCO skills;



# Workflow (1/3)

## **Access national qualifications**

\_\_ Pre-process national qualifications as well as reference point or system;

## **Parse LO descriptions**

\_\_ Text segmentation;  
\_\_ Part-of-speech tagging ('chunking');

## **Normalise detected text segments**

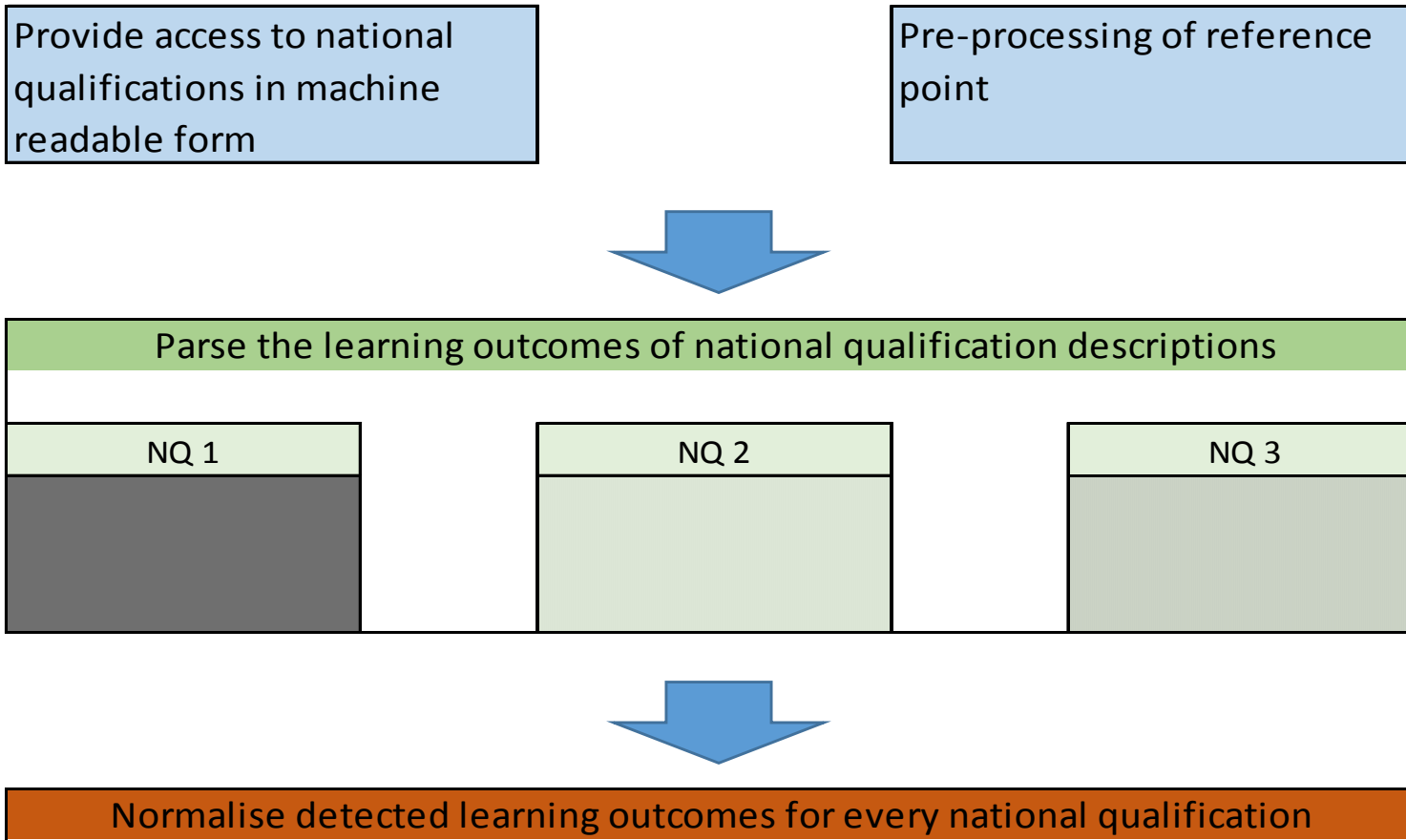
\_\_ "Translate" text segments into taxonomy concepts;

## **Map normalised text segments onto reference point or system**

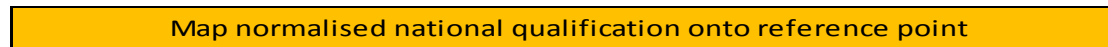
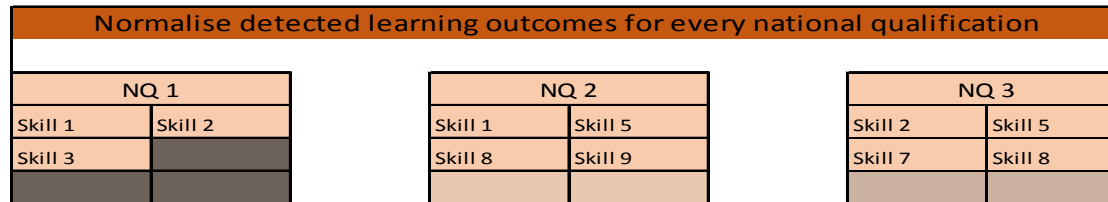
\_\_ Registering overlap and divergence.



# Workflow (2/3)



# Workflow (3/3)



RP	NQ1	NQ2	NQ3
Skill 1	yes	yes	no
Skill 2	yes	no	yes
Skill 3	yes	no	no
Skill 4	no	no	no
Skill 5	no	yes	yes
Skill 6	no	no	no
Skill 7	no	no	yes
Skill 8	no	no	yes
Skill 9	no	yes	no
Skill 10	no	no	no
Skill 11	no	no	no
Skill 12	no	no	no

shared and diverging LOs, expressed by skills  
concepts of the reference point

unmappable LOs,  
expressed in free text



# Conditions on the workflow

Ideally the digital tool should have the ability to

- \_\_process different text formats;
- \_\_identify vertical dimension of learning outcomes;
- \_\_parse natural language;
- \_\_process different languages;

Furthermore, the digital tool should

- \_\_use open-access software;
- \_\_not be too demanding to operate.





# Operationalising the workflow (1/2)

## **Selection of already available tools:**

- \_\_Natural Language Toolkit (NLTK) for working with text data;
- \_\_Python (Anaconda Navigator) as programming language;
- \_\_Pandas for working with tabular data ('dataframes');

## **Trial-and-error approach to tailoring these tools identified several challenges:**

- \_\_It is very time consuming to make already existing coding examples fit the needs of our workflow;
- \_\_Proper feature detection is impossible without supplementing ESCO with additional synonym dictionaries;
- \_\_Development of automated classifier needs manually pre-processed training material.



# Operationalising the workflow (2/2)

## **Incorporate suggestions and feedback from experts**

- \_\_ Aim at supporting rather than fully automating the process;
- \_\_ Feasibility of integrating machine learning depends on amount and diversity of data to be processed;
- \_\_ Implementing a machine learning approach should require significantly less efforts than doing the comparison manually;

## **Adapted approach to developing a prototype**

- \_\_ Narrow focus on certain processing subtasks;
- \_\_ Use whole reference system (ESCO skills) rather than pre-selected reference points (occupational skills profiles) only.

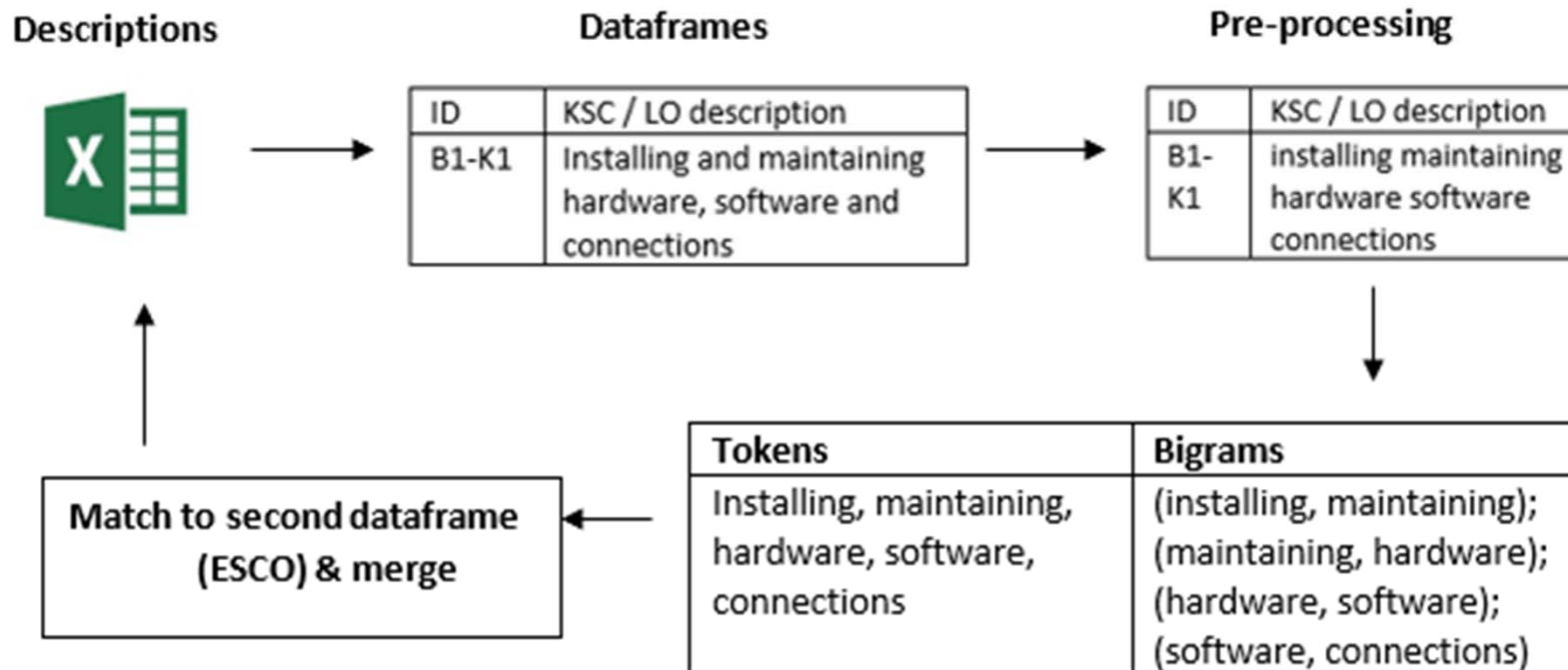


# Testing of the workflow

## Test files / source data:

\_\_Short LO descriptions and ID's for Dutch ICT service technician (core tasks / work processes);

\_\_Short KSC descriptions and ID's from (full) ESCO KSC skills pillar (v103).



# Results of the testing exercise (1/4)

## Frequency analysis (single tokens)

\_\_6,571 unique tokens in ESCO, 35 in qualification;

\_\_0.87% of ESCO tokens high-frequency, 46.2% only occur once;

\_\_most-common terms are generally action verbs, mid-frequency terms generally more context-defining.

High-frequency (top5)	Mid-frequency	Single-Occurence
manage	digital	herbicide
equipment	facilities	hormone
operate	communication	mistelle
maintain	compliance	monogramprinting
perform	conditions	habits



## Results of the testing exercise (2/4)

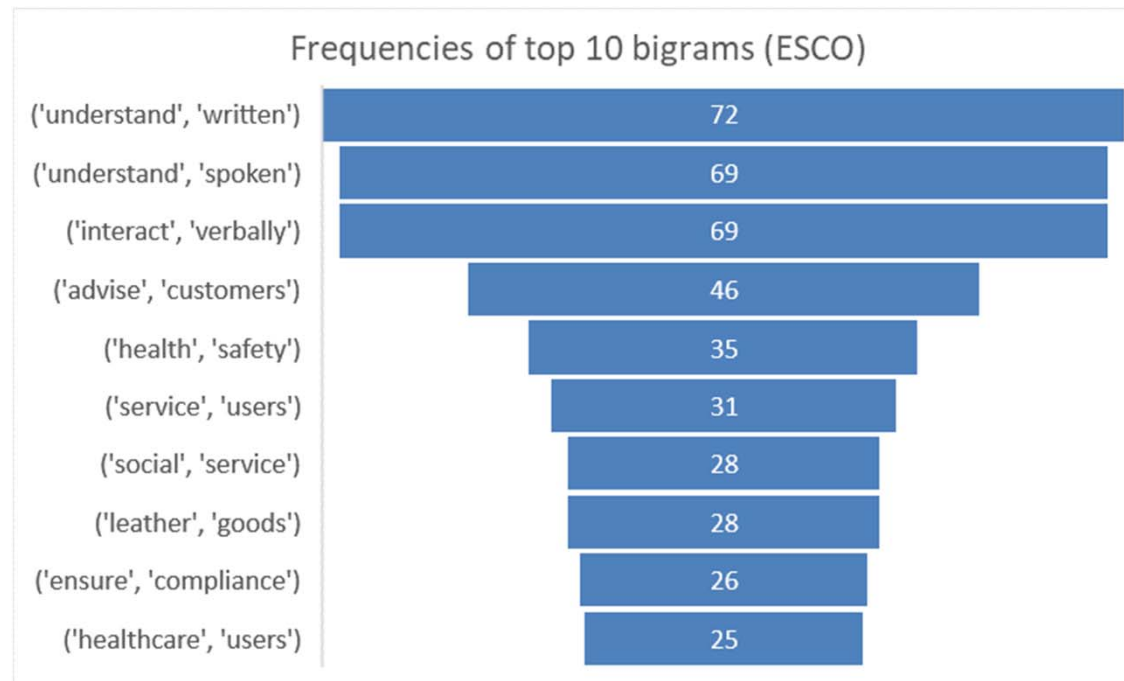
<b>Token / Term</b>	<b>ID</b>	<b>LO description</b>	<b>KSC description</b>	<b>Linked to ESCO profile(s)</b>
connections	B1-K1	Installing and maintaining hardware, software and connections	inspect for unauthorised connections	meter reader.
connections	B1-K1-W3	Realize connections	inspect for unauthorised connections	meter reader.
installing	B1-K1	Installing and maintaining hardware, software and connections	estimate costs of installing telecommunication devices	telecommunications engineer; specialised seller; telecommunications equipment specialised seller.
maintaining	B1-K1	Installing and maintaining hardware, software and connections	assume responsibility for maintaining a safe ship environment	fleet commander; marine surveyor.

# Results of the testing exercise (3/4)

## Frequency analysis (bigrams)

\_\_23,852 bigrams in ESCO, 28 in qualification;

\_\_Less than 1% (ESCO) occur more than 10x, 87.6% only occur once.



# Results of the testing exercise (4/4)

<b>Bigram</b>	<b>LO description</b>	<b>KSC description</b>	<b>Linked to ESCO profile(s)</b>
incident reports	Handling of incident reports	process incident reports for prevention	chemical metallurgist; process metallurgist; metal furnace operator; coking furnace operator; heat treatment furnace operator; metallurgist; mine rescue officer.
peripheral equipment	'Making systems, (peripheral) equipment and applications ready for use'	set up audiovisual peripheral equipment	camera operator; audio-visual technician; recording studio technician; broadcast technician.
peripheral equipment	'Replacement, repair and / or (dis) assembly of (parts of) systems and (peripheral) equipment'	explain characteristics of computer peripheral equipment	specialised seller; computer and accessories specialised seller.

# Conclusions & Recommendations

## **Conclusions:**

- \_\_ At this moment, limited possibilities in terms of automated comparison of qualifications (machine learning);
- \_\_ Proper feature detection (& matching) is impossible without supplementing ESCO with additional synonym dictionaries;

## **Recommendations with regard to ESCO**

- \_\_ Enriching vocabulary with stemmed versions of skills phrases;
- \_\_ Supplementing semantic structure to enable aggregations;
- \_\_ Dissecting complex skills into enabling skills components;
- \_\_ Consolidating terminology (e.g. summarising skills expressing the same meaning with different words under one concept).





# Discussion

- \_\_Is it possible and also economically feasible to fully automate the comparison of VET qualifications?
- \_\_How accurate / comprehensive / reliable would the results of an automated comparison of qualifications be?
- \_\_If automation is only used to support a manual comparison of qualifications, how should tasks be distributed between man and machine?
- \_\_What needs to be done to improve ESCO's suitability for NLP?



# Thank you!

**Marye Hudepohl**

Ockham-IPS, Netherlands

[m.hudepohl@ockham-ips.nl](mailto:m.hudepohl@ockham-ips.nl)

[ockham-ips.nl/](http://ockham-ips.nl/)



**Claudia Plaimauer**

3s Unternehmensberatung GmbH

[plaimauer@3s.co.at](mailto:plaimauer@3s.co.at)

[www.3s.co.at](http://www.3s.co.at)

