



Linking learning outcomes of qualifications to ESCO skills A pilot for testing

Francesco Losappio and Jan Lutz

ETF - Training programme on Big Data for LMI – 22-24 November 2021

Why linking qualifications to skills

Enrich information on qualifications: linking their learning outcomes to skills allows to directly identify relevant qualifications for missing skills.

Promote transparency of information on qualifications: employers grasp qualifications' labour market value in a cross-border context, thus supporting labour mobility.

Individuals improve chances on labour market through better skills-based job matching.

Support upskilling-reskilling by matching individuals with the right training.

Background

2018 study recommended testing an automated approach based on Natural Language Processing with an initial degree of human intervention.

COM presented the study outcomes during the 2019 EQF AG/MSWG joint meeting and proposed to pilot the proposed approach.

A pilot project with 5 Member States (LV, NL, PL, RO and SI) was conducted in 2019. COM developed a first proof of concept for an IT tool supporting the linking of LOs to skills.

A second pilot was organised in 2020. COM provided an improved version of the IT tool.

Objectives of the pilot

Testing the linking in different EU languages and for qualifications of different type and levels.

Develop an enhanced matching algorithm based on artificial intelligence to support the linking process in all EU languages.

Assess the required effort from awarding bodies/competent authorities to create these links.

Test the usability of the hierarchical structure of ESCO skills and introduce rules and matching relations between skills and learning outcomes of qualifications.

Getting feedback on the ESCO skills pillar.

Pilot features

13 participants: 9 EU MS (BE, CZ, FR, IT, LV, NL, PL, RO, SI), ETUC, ETF, EURASHE and the IDB.

Examine high number (min 25) of diverse qualifications in different national databases.

Different description styles in different languages.

VET and Higher Education qualifications with different EQF levels.

Link with occupations: skills from ESCO OCCS can be used to validate or improve matching results.

Visualisation of key performance indicators (data per qualification and per project).

Technical Solution

Machine Learning Algorithm

Expected Output:

Suggestions of ESCO skills based on the input text

Challenges:

- Input texts in several languages
- Input text variations: broad vs. detailed descriptions
- Relevant segmentation of the text
- Word ambiguity
- Concepts “compatibility”

Qualifications

Participant	# qualifications	# qualifications with mappings	% mapped
Belgium-Flanders	25	24	0.75
Belgium-Fr	47	2	0.03
ETF-Cameroon	1	1	0.99
ETF-Cape-Verde	14	14	0.34
ETF-Kenya	15	7	0.28
ETF-Mozambique	14	13	0.7
ETF-South-Africa	20	11	0.32
France	24	22	0.44
IDB	42	14	0.19
Italy	32	32	0.97
Latvia	24	24	0.32
Netherlands	19	9	0.11
Poland	26	25	0.28
Romania	25	4	0.02
Slovenia	39	39	0.92
	367	241	0.444

Mapped Learning Outcomes: Languages

Language	#Mapped LO (sum)
EN	885
NL	836
SL	702
PL	384
FR	383
PT	330
IT	156
ES	44
RO	4
	3724

- The current data set can already be used for evaluation/comparison of different candidate algorithms. More data would be beneficial to improve the algorithms.
- Mapping is done best in the native language. This allows the algorithm to learn more about the specifics of a language and results in better matches.

Mappings

Participant	#Mappings (sum)	#Mappings /LO (min)	#Mappings /LO (median)	#Mappings /LO (max)
Belgium-Flanders	2486	1	2	26
Belgium-Fr	55	1	1.5	14
ETF-Cameroon	30	1	2.5	6
ETF-Cape-Verde	444	1	5	26
ETF-Kenya	395	1	3	11
ETF-Mozambique	851	1	3	10
ETF-South-Africa	285	1	3	13
France	552	1	1	10
IDB	246	1	5	22
Italy	1883	1	10	45
Latvia	1675	1	2	18
Netherlands	153	1	2	8
Poland	548	1	1	7
Romania	7	1	1	4
Slovenia	2388	1	3	13
	11998			

1-3 mappings per learning outcome is the expected number

Possible reasons for larger numbers:

- Formulation of the learning outcomes
- Sentence splitter needs improvement
- User friendliness of the tool (e.g. scrolling up and down)

Mapping Algorithm

	% mapped skills in topX	
	API Search	ML Suggestions
Top1	17.05	30.41
Top3	30.17	48.42
Top5	36.70	56.43
Top10	45.30	65.97
Top20	52.94	73.46
Top50	63.43	81.62

- Methodology: can be improved but performs better than the basic API search
- Amount of available training data plays an important role
- It is important to have non-English data to take language specific context into account.

Results from the pilot (1)

- Linking learning outcomes to skills is done best in the native language.
- Human intervention is an important component for ensuring quality of linking.
- Further improvements of the AI algorithm will reduce the time and resources needed to perform the work.
- Level of granularity and sectoral coverage of ESCO skills can lead to different results according to type of qualifications tested.
- The structure of learning outcomes is a key factor for the functioning of the AI algorithm.

Results from the pilot (2)

- The ESCO skills hierarchy supports the search of relevant skills within the ESCO dataset.
- Linking to higher levels of the hierarchy and an improved transversal skills structure could further support qualitative linking, in particular for HE qualifications.
- On average, 1 to 3 ESCO skills or knowledge concepts are mapped to each learning outcome.
- The performance of the AI algorithm showed promising results, however more data are needed to further improve quality of suggestions.

Next steps – Third phase of the pilot project

Continue the linking exercise to obtain larger amount of validated data in order to develop and compare the performance of different NLP algorithms.

Test improvements of ESCO v.1.1 and the application of the revised transversal skills structure in this domain.

Further improve the IT tool to link learning outcomes to ESCO skills, with the goal of offering it as a standalone service available to Member States and other organisations.

Qualification splitting

Master of Arts in het tolken

In progress

Concept identifier

ADV-0383

Description

1. Eén of twee vreemde talen beheersen, op het niveau C2 of minstens C1 van het Gemeenschappelijk Europees Referentiekader voor Moderne Vreemde Talen voor begrijpen (luisteren en lezen), spreken (productie en interactie) en schrijven. 2. In het Nederlands en de gekozen vreemde talen/vreemde taal effectieve schriftelijke en mondelinge communicatie met een hoge graad van specificiteit verzorgen in diverse professionele contexten als meertalig communicatiespecialist, als vertaler, of als tolk. 3. Een geavanceerde taalkennis, inclusief contrastieve kennis toepassen in uiteenlopende schriftelijke en mondelinge communicatieve situaties. 4. Verworven inzichten uit de vertaal- en tolkwetenschap, of uit de studie van communicatie in organisaties en van meertalige en interculturele contexten kritisch toepassen in uiteenlopende communicatieve situaties. 5. Een geavanceerde kennis van de culturele en institutionele context van de betrokken taalgebieden en de ontwikkeling ervan inzichtelijk toepassen in uiteenlopende communicatieve situaties. 6. Vanuit deze kennis en inzichten zelfstandig en kritisch reflecteren over diverse aspecten van communicatie. Het daarbij gehanteerde probleemoplossend vermogen ook inzetten in een ruimere context. 7. Bij het voeren van mondelinge en schriftelijke professionele communicatie, het vertaal- of het tolkproces specifieke technologische hulpmiddelen efficiënt en adequaat gebruiken en evalueren. 8. De specificiteit van de professionele context van het vertalen of tolken, of de communicatie in organisaties en in meertalige of interculturele contexten, kennen; vertrouwd zijn met de bijhorende normen en ernaar handelen. 9. In het domein van de tekst-, vertaal- of tolkwetenschap of de meertalige, interculturele of organisatiecommunicatie zelfstandig onderzoek concipiëren, plannen en uitvoeren op het niveau van een beginnend onderzoeker en er schriftelijk en mondeling over communiceren zowel met een breed als met een gespecialiseerd publiek. 10. Nieuwe ontwikkelingen zowel binnen de discipline van het vertalen of het tolken of de studie van communicatie in organisaties en meertalige, interculturele contexten, als binnen de algemene context van taal en cultuur, actief opvolgen, er kritisch over reflecteren en ze plaatsen in een breed maatschappelijk en intercultureel kader. 11. Deontologische en ethische aspecten van onderzoek onderkennen en in overeenstemming daarmee handelen. 12. Taal- en cultuurgevoeligheid en respect voor diversiteit integreren in het wetenschappelijk werk, de professionele omgeving of het maatschappelijk debat.

Concept Uri

<http://data.europa.eu/esco/concept/lo-linking/belgium-flanders/ADV-0383>

Learning Outcome Entities 14

1

Eén of twee vreemde talen beheersen, op het niveau C2 of minstens C1 van het Gemeenschappelijk Europees Referentiekader voor Moderne Vreemde Talen voor begrijpen (luisteren en lezen), spreken (productie en interactie) en schrijven

In het Nederlands en de gekozen vreemde talen/vreemde taal effectieve schriftelijke en mondelinge communicatie met een hoge graad van specificiteit verzorgen in diverse professionele contexten als meertalig communicatiespecialist, als vertaler, of als tolk

Een geavanceerde taalkennis, inclusief contrastieve kennis toepassen in uiteenlopende schriftelijke en mondelinge communicatieve situaties

Verworven inzichten uit de vertaal- en tolkwetenschap, of uit de studie van communicatie in organisaties en van meertalige en interculturele contexten kritisch toepassen in uiteenlopende communicatieve situaties 5

Een geavanceerde kennis van de culturele en institutionele context van de betrokken taalgebieden en de ontwikkeling ervan inzichtelijk toepassen in uiteenlopende communicatieve situaties

Vanuit deze kennis en inzichten zelfstandig en kritisch reflecteren over diverse aspecten van communicatie

Het daarbij gehanteerde probleemoplossend vermogen ook inzetten in een ruimere context

Bij het voeren van mondelinge en schriftelijke professionele communicatie, het vertaal- of het tolkproces specifieke technologische hulpmiddelen efficiënt en adequaat

Qualification splitting

```
>>> inputString
1.Een of twee vreemde talen beheersen, op het niveau C2 of minstens C1 van het Gemeenschappelijk Europees Referentiekader voor Moderne Vreemde Talen voor begrijpen (luisteren en lezen), spreken (productie en interactie) en schrijven. 2. In het Nederlands en de gekozen vreemde talen/vreemde taal effectieve schriftelijke en mondelinge communicatie met een hoge graad van specificiteit verzorgen in diverse professionele contexten als meertalig communicatiespecialist, als vertaler, of als tolk. 3. Een geavanceerde taalkennis, inclusief contrastieve kennis toepassen in uiteenlopende schriftelijke en mondelinge communicatieve situaties. 4. Verworven inzichten uit de vertaal- en tolkwetenschap, of uit de studie van communicatie in organisaties en van meertalige en interculturele contexten kritisch toepassen in uiteenlopende communicatieve situaties 5. Een geavanceerde kennis van de culturele en institutionele context van de betrokken taalgebieden en de ontwikkeling ervan inzichtelijk toepassen in uiteenlopende communicatieve situaties. 6. Vanuit deze kennis en inzichten zelfstandig en kritisch reflecteren over diverse aspecten van communicatie. Het daarbij gehanteerde probleemoplossend vermogen ook inzetten in een ruimere context. 7. Bij het voeren van mondelinge en schriftelijke professionele communicatie, het vertaal- of het tolkproces specifieke technologische hulpmiddelen efficiënt en adequaat gebruiken en evalueren. 8. De specificiteit van de professionele context van het vertalen of tolken, of de communicatie in organisaties en in meertalige of interculturele contexten, kennen; vertrouwd zijn met de bijhorende normen en ernaar handelen 9. In het domein van de tekst-, vertaal- of tolkwetenschap of de meertalige, interculturele of organisatiecommunicatie zelfstandig onderzoek concipiëren, plannen en uitvoeren op het niveau van een beginnend onderzoeker en er schriftelijk en mondeling over communiceren zowel met een breed als met een gespecialiseerd publiek. 10. Nieuwe ontwikkelingen zowel binnen de discipline van het vertalen of het tolken of de studie van communicatie in organisaties en meertalige, interculturele contexten, als binnen de algemene context van taal en cultuur, actief opvolgen, er kritisch over reflecteren en ze plaatsen in een breed maatschappelijk en intercultureel kader. 11. Deontologische en ethische aspecten van onderzoek onderkennen en in overeenstemming daarmee handelen. 12. Taal- en cultuurgevoeligheid en respect voor diversiteit integreren in het wetenschappelijk werk, de professionele omgeving of het maatschappelijk debat.'
```

- NO manually labelled data used
- Dutch was not in training set

```
1.Een of twee vreemde talen beheersen, op het niveau C2 of minstens C1 van het Gemeenschappelijk Europees Referentiekader voor Moderne Vreemde Talen voor begrijpen (luisteren en lezen), spreken (productie en interactie) en schrijven. 2. In het Nederlands en de gekozen vreemde talen/vreemde taal effectieve schriftelijke en mondelinge communicatie met een hoge graad van specificiteit verzorgen in diverse professionele contexten als meertalig communicatiespecialist, als vertaler, of als tolk. 3. Een geavanceerde taalkennis, inclusief contrastieve kennis toepassen in uiteenlopende schriftelijke en mondelinge communicatieve situaties. 4. Verworven inzichten uit de vertaal- en tolkwetenschap, of uit de studie van communicatie in organisaties en van meertalige en interculturele contexten kritisch toepassen in uiteenlopende communicatieve situaties 5. Een geavanceerde kennis van de culturele en institutionele context van de betrokken taalgebieden en de ontwikkeling ervan inzichtelijk toepassen in uiteenlopende communicatieve situaties. 6. Vanuit deze kennis en inzichten zelfstandig en kritisch reflecteren over diverse aspecten van communicatie. Het daarbij gehanteerde probleemoplossend vermogen ook inzetten in een ruimere context. 7. Bij het voeren van mondelinge en schriftelijke professionele communicatie, het vertaal- of het tolkproces specifieke technologische hulpmiddelen efficiënt en adequaat gebruiken en evalueren. 8. De specificiteit van de professionele context van het vertalen of tolken, of de communicatie in organisaties en in meertalige of interculturele contexten, kennen; vertrouwd zijn met de bijhorende normen en ernaar handelen 9. In het domein van de tekst-, vertaal- of tolkwetenschap of de meertalige, interculturele of organisatiecommunicatie zelfstandig onderzoek concipiëren, plannen en uitvoeren op het niveau van een beginnend onderzoeker en er schriftelijk en mondeling over communiceren zowel met een breed als met een gespecialiseerd publiek. 10. Nieuwe ontwikkelingen zowel binnen de discipline van het vertalen of het tolken of de studie van communicatie in organisaties en meertalige, interculturele contexten, als binnen de algemene context van taal en cultuur, actief opvolgen, er kritisch over reflecteren en ze plaatsen in een breed maatschappelijk en intercultureel kader. 11. Deontologische en ethische aspecten van onderzoek onderkennen en in overeenstemming daarmee handelen. 12. Taal- en cultuurgevoeligheid en respect voor diversiteit integreren in het wetenschappelijk werk, de professionele omgeving of het maatschappelijk debat.'
```

Qualification splitting

Technicien Geometre-Topographe

 In progress ▾

Concept identifier

GT-TO

Iscedf

<http://data.europa.eu/esco/isced-f/0731>

Description

Maitriser l'utilisation des instruments et les différentes méthodes d'acquisition des données en Topographie et Bâtiments et Travaux publics Appréhender les applications et méthodes particulières de saisie et représentation des données Maitriser les règles de base relatives aux traitements numériques des données et effectuer des calculs à l'aide de logiciel informatique Appréhender les procédés conventionnels et informatisés de traitement graphique Appréhender la réglementation professionnelle et d'aménagement fonciers ainsi que les notions de droit en général Réaliser des mesures et des levés topographiques et implanter des détails Exploiter et traiter numériquement les données / mesures de terrain Réaliser des documents graphiques Participer à des interventions spécifiques liées à la Géomatique

Concept Uri

<http://data.europa.eu/esco/concept/lo-linking/etf-cameroon/GT-TO>



```
>>> inputString = "Maitriser l'utilisation des instruments et les differentes methodes d'acquisition des donnees en Topographie et Batiments et Travaux publics Apprehender les applications et me  
thodes particulieres de saisie et representation des donnees Maitriser les regles de base relatives aux traitements numeriques des donnees et effectuer des calculs a l'aide de logiciel informati  
que Apprehender les procedes conventionnels et informatises de traitement graphique Apprehender la reglementation professionnelle et d'amenagement fonciers ainsi que les notions de droit en gene  
ral Realiser des mesures et des leviers topographiques et implanter des details Exploiter et traiter numeriquement les donnees / mesures de terrain Realiser des documents graphiques Participer a  
des interventions specifiques liees a la Geomatique"
```

```
>>>  
>>> splittedPhrases = splitInSentences(inputString, 0.15)  
>>> for phrase in splittedPhrases: print(phrase)
```

```
...  
Maitriser l'utilisation des instruments et les differentes methodes d'acquisition des donnees en Topographie et Batiments et Travaux publics  
Apprehender les applications et methodes particulieres de saisie et representation des donnees  
Maitriser les regles de base relatives aux traitements numeriques des donnees et effectuer des calculs a l'aide de logiciel informatique  
Apprehender les procedes conventionnels et informatises de traitement graphique  
Apprehender la reglementation professionnelle et d'amenagement fonciers ainsi que les notions de droit en general  
Realiser des mesures et des leviers topographiques et implanter des details  
Exploiter et traiter numeriquement les donnees / mesures de terrain  
Realiser des documents graphiques  
Participer a des interventions specifiques liees a la Geomatique
```


Improvements on algorithm to suggest skills

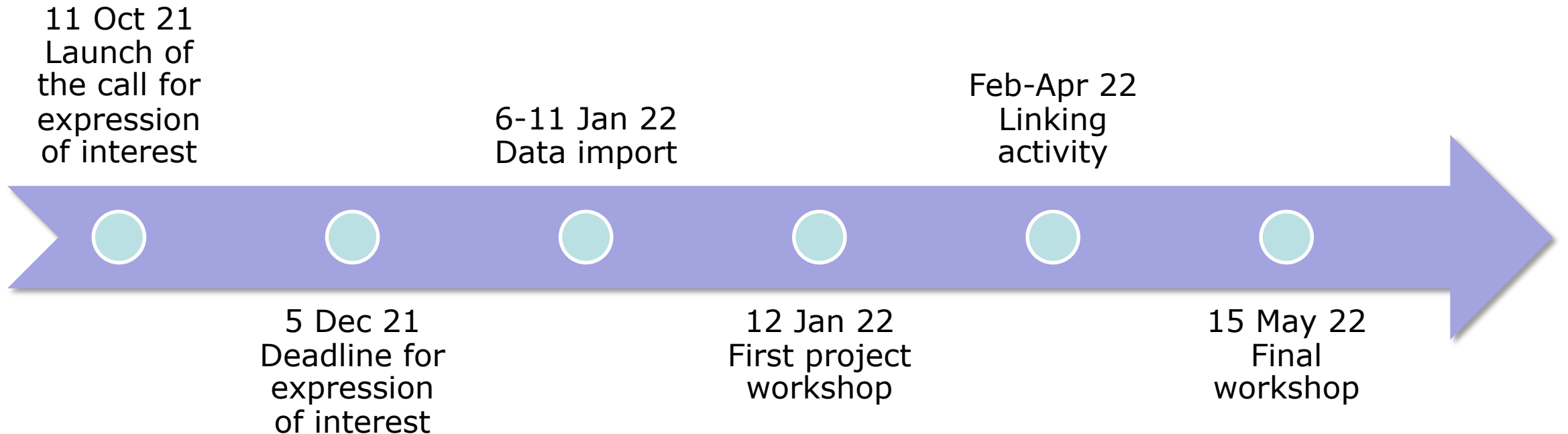
Skill suggestions based on three algorithms in the new version of the tool:

- Word2vec model: English
- Semantic Embedding Mapping model: English
- Sentence BERT: Multilingual

Available validation data allow to compare performance and iteratively improve:

	All_esco_api	All_w2v	All_sem11	All_sem13	All_sem17	All_sem18	SBert1	SBert6	SBert7	SBert8	Combined
Unique Total Records	3232	3259	3213	3213	3213	3213	3240	3244	3244	3244	3283
Unique Top1	0.170482673	0.304081	0.16651105	0.2159975	0.2296919	0.2284469	0.2148148	0.2096178	0.212392	0.204377	0.2564727
Unique Top3	0.301670792	0.484198	0.28913788	0.3709928	0.3856209	0.3918456	0.3632716	0.3757707	0.372996	0.348952	0.4794395
Unique Top5	0.366955446	0.564284	0.35792095	0.4431995	0.4662309	0.4721444	0.4469136	0.4627004	0.45561	0.433724	0.5653366
Unique Top10	0.452970297	0.659712	0.44475568	0.5496421	0.5739185	0.5689387	0.5469136	0.573058	0.565351	0.544698	0.679866
Unique Top20	0.529393564	0.734581	0.53283536	0.6277622	0.660442	0.6495487	0.645679	0.6747842	0.660604	0.646424	0.7675906

Project timeline



THANK YOU

