



## Занятие 9

# Увязка результатов обучения из квалификаций с профессиональными умениями из ESCO Пилотное тестирование

Франческо Лозаппио и Ян Лутц

ЕФО — Обучающая программа по использованию больших данных в АРТ — 22—24 ноября 2021

# Зачем увязывать квалификации и умения

Обогащение информации о квалификациях: увязывание их результатов обучения с профессиональными умениями позволяет непосредственно определять подходящие квалификации для недостающих умений.

Содействие прозрачности информации о квалификациях: работодатели понимают ценность квалификаций для рынка труда в трансграничном контексте, тем самым способствуя трудовой мобильности.

Повышаются шансы граждан на рынке труда, благодаря более точному подбору рабочих мест на основе профессиональных умений.

Содействие повышению квалификации и переподготовке путем индивидуального подбора подходящих обучающих курсов.



## Общая информация

Обследование 2018 года рекомендовало протестировать автоматизированный подход на основе обработки естественного языка с определенной степенью человеческого вмешательства.

Комиссия представила результаты обследования во время совместного заседания Консультативной группы по вопросам ЕРК/ Рабочей группы государств-членов и предложила провести пилотное тестирование предложенного подхода.

Пилотный проект с участием 5 стран-членов (Латвия, Нидерланды, Польша, Румыния и Словения) был реализован в 2019 году. Комиссия разработала первую апробацию концепции технологического инструмента для увязывания результатов обучения с профессиональными умениями.

В 2020 году был организован второй пилотный проект. Комиссия предоставила улучшенную версию инструмента.



## Цели пилотного проекта

Протестировать увязывание на разных языках ЕС и на квалификациях разных типов и уровней.

Разработать улучшенный алгоритм соотнесения на основе искусственного интеллекта для содействия в процессе увязывания на всех языках ЕС.

Оценить, какой объем усилий требуется для такой увязки со стороны компетентных органов/ органов, присваивающих квалификации.

Протестировать удобство работы с иерархической структурой профессиональных умений ESCO и представить правила и отношения соответствия между умениями и результатами обучения из квалификаций.

Получить обратную связь о столпе профессиональных умений ESCO.

# Характеристики пилотного проекта

13 участников: 9 стран-членов ЕС (Бельгия, Чехия, Франция, Италия, Латвия, Нидерланды, Польша, Румыния, Словения), Европейская конфедерация профсоюзов (ETUC), ЕФО, Европейская ассоциация высших учебных заведений (EURASHE) и Межамериканский банк развития (IDB).

Изучить большое количество (не менее 25) различных квалификаций в разных национальных базах данных.

Разные стили описания и разные языки.

Квалификации профессионального и высшего образования с разными уровнями ЕРК.

Связь с профессиями: для валидации или улучшения результатов соотнесения можно использовать профессиональные умения из профессий ESCO.

Визуализация ключевых показателей результативности (данные по квалификациям и по проектам).

# Техническое решение

## Алгоритм машинного обучения

### Ожидаемый результат:

рекомендации профессиональных умений ESCO на основе введенного текста.

### Трудности:

- Вводимые тексты на нескольких языках
- Вариации вводимых текстов: широкие и подробные описания
- Подходящая сегментация текста
- Неоднозначность слов
- Концепция «совместимости»

# Квалификации

Участник	К-во квалификаций	К-во сопоставленных квалификаций	% сопоставленных
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

## Сопоставление результатов обучения: языки

Язык	К-во сопоставленных РО (сумма)
EN	885
NL	836
SL	702
PL	384
FR	383
PT	330
IT	156
ES	44
RO	4
	3724

- Текущий набор данных уже можно использовать для оценки/ сравнения разных алгоритмов-кандидатов. Для улучшения алгоритмов было бы полезно иметь больше данных.
- Сопоставление лучше всего делать на родном языке. Это позволяет алгоритму больше узнать об особенностях языка и способствует улучшению сопоставлений.



## Соотнесения

Участник	К-во сопоставленных РО (сумма)	К-во сопоставлений/ РО (мин.)	К-во сопоставлений/ РО (медиан.)	К-во сопоставлений/ РО (макс.)
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			

Ожидаемое число: 1—3 соотнесения на каждый результат обучения

Возможные причины бóльших чисел:

- формулировка результатов обучения
- требуется улучшение разделителя предложений
- удобство использования инструмента (например, прокрутка вверх и вниз)

## Алгоритм соотнесения

	% соотнесенных проф. умений в топ-Х	
	Поиск при помощи API	Рекомендации на основе машинного обучения
Топ-1	17,05	30,41
Топ-3	30,17	48,42
Топ-5	36,70	56,43
Топ-10	45,30	65,97
Топ-20	52,94	73,46
Топ-50	63,43	81,62

- Методология: может быть усовершенствована, но работает лучше, чем обычный поиск при помощи API
- Важную роль играет количество доступных тренировочных данных
- Чтобы учесть специфический языковой контекст, важно иметь данные не на английском языке.

# Результаты пилотного проекта (1)

- Увязывание результатов обучения с профессиональными умениями лучше всего делать на родном языке.
- Вмешательство человека — важная составляющая обеспечения качества увязывания.
- Дальнейшее совершенствование алгоритма ИИ сократит время и ресурсы, которые требуются на выполнение работы.
- Уровень детализации и охвата отраслей профессиональными умениями ESCO может привести к различным результатам в зависимости от типа тестируемой квалификации.
- Структура результатов обучения является ключевым фактором функционирования алгоритма ИИ.

## Результаты пилотного проекта (2)

- Иерархия профессиональных умений ESCO способствует поиску нужных умений в наборе данных ESCO.
- Увязывание с более высокими уровнями иерархии и улучшенная структура сквозных профессиональных умений могут дополнительно способствовать количественному увязыванию, в частности что касается квалификаций ВО.
- В среднем с каждым результатом обучения соотнесено от 1 до 3 профессиональных умений или концепций знаний ESCO.
- Алгоритм ИИ продемонстрировал многообещающие результаты, но для дальнейшего улучшения качества его рекомендаций требуется больше данных.

## Дальнейшие действия — третья фаза пилотного проекта

Продолжать увязывание, чтобы получить больше подтвержденных данных для разработки и сравнения результативности разных алгоритмов обработки естественного языка.

Тестировать улучшения ESCO версии 1.1 и возможности применения пересмотренной структуры сквозных профессиональных умений в этой области.

Продолжать совершенствовать технологический инструмент для увязывания результатов обучения с профессиональными умениями ESCO. Цель — предлагать его странам-членам и другим организациям как отдельный сервис.

# Разделение квалификаций

## 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	
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	

# Разделение квалификаций

```
>>> 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.
```

- НЕ используются данные, размеченные вручную
- Нидерландский язык не использовался в тренировочном наборе

```
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.
```

# Разделение квалификаций

## Technicien Geometre-Topographe

In progress

Concept identifier  
GT-TO

Iscedf  
<http://data.europa.eu/esco/iscdf/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 levés topographiques et implanter des details Exploiter et traiter numériquement les donnees / mesures de terrain Realiser des documents graphiques Participer a  
des interventions spécifiques liées 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 levés topographiques et implanter des details  
Exploiter et traiter numériquement les donnees / mesures de terrain  
Realiser des documents graphiques  
Participer a des interventions spécifiques liées a la Geomatique
```



# Улучшения алгоритма в части предложения профессиональных умений

В новой версии инструмента рекомендации профессиональных умений базируются на трех алгоритмах:

- модель Word2vec: английский
- модель сопоставления на основе семантического встраивания: английский
- модель Sentence BERT: многоязычная

Наличие проверочных данных позволяет сравнивать результаты и улучшать алгоритмы с каждой итерацией:

	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

# Сроки реализации проекта

11 октября 2021.  
Приглашение к  
выражению  
заинтересованности

6—11 января 2022.  
Импорт данных

Февраль — апрель  
2022. Увязывание

5 декабря 2021.  
Конечный срок для  
выражения  
заинтересованности

12 января 2022.  
Первый семинар  
проекта

15 мая 2022.  
Заключительный  
семинар

THANK YOU

