

Wyniki wstępnych prac badawczo-rozwojowych w zakresie oceny podobieństwa kwalifikacji

Michał Marcińczuk

Politechnika Wrocławska

Grzegorz Wojdyga, Łukasz Kobyliński

IPI PAN

Marcin Będkowski

Uniwersytet Warszawski, IBE

Agenda

1. Informacje wstępne
2. Zbiór ZRK i ESCO
3. Odnoszenie ZRK do baz zewnętrznych
4. Grupowanie kompetencji z bazy ZRK
5. Ocena i rozwinięcie wyników uzyskanych w ramach ekspertyzy
6. Podsumowanie

Informacje wstępne

- Cel: stan wiedzy dotyczący grupowania, a także sprawdzania podobieństw pomiędzy kwalifikacjami
- Okres realizacji: maj–wrzesień 2019,

Zbiór ZRK i ESCO

Baza ESCO (European Skills, Competences, Qualifications and Occupations) dzieli się na 3 części:

- 942 zawodów
- 13485 umiejętności
- 8155 kwalifikacji wymaganych do tych zawodów

Search

OCCUPATIONS

- 0 Siły zbrojne
- 1 Kierownicy
- 2 Specjaliści
- 3 Technicy i inny średni personel
- 4 Pracownicy biurów
- 5 Pracownicy usług i sprzedawcy
- 6 Wykwalifikowani rolnicy, robotnicy leś...
- 7 Robotnicy przemysłowi i rzemieślnicy
- 8 Operatorzy i monterzy maszyn i urzą...
- 9 Pracownicy przy pracach prostych

Polish (pl)

inżynier języka naturalnego

[Discuss this topic in the Online Forum](#)

Description

Language engineers work within the field of computing science, and more specifically in the field of natural language processing. They aim to close the gap in translation between accurate human translations to machine-operated translators. They parse texts, compare and map translations, and improve the linguistics of translations through programming and code.

Regulatory aspect

W bazie zawodów regulowanych prowadzonej przez Komisję można sprawdzić, czy i w jaki sposób dany zawód jest regulowany w państwach członkowskich UE, państwach EOG lub Szwajcarii. Baza zawodów regulowanych: http://ec.europa.eu/growth/single-market/services/free-movement-professionals/qualifications-recognition_pl

Hierarchy

- 2 Specjaliści
 - 21 Specjaliści nauk fizycznych, matematycznych i technicznych
 - 215 Inżynierowie elektrotechnologii
 - 2152 Inżynierowie elektroniki
 - inżynier elektronik
 - inżynier języka naturalnego

Essential skills and competences

- [interpretować wymogi techniczne](#)
- [korzystać z oprogramowania do rysunków technicznych](#)
- [oceniać technologie tłumaczeniowe](#)
- [określać wymagania techniczne](#)
- [prowadzić badania naukowe](#)

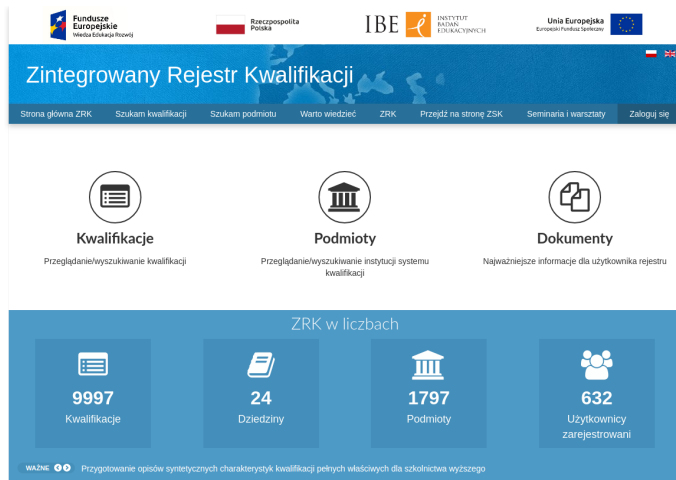
Rysunek 1: Przykładowe kwalifikacje z katalogu IBE

Cechy charakterystyczne

- prawie wszystkie pola są przetłumaczone na języki Unii Europejskiej
- bazy zawodów i kwalifikacji są w relacji wiele do wielu
- kwalifikacje (dyplomy) nie zostały uwzględnione
- jest w niej zastosowana hierarchia zgodna ze standardami
- wygodne API

Baza ZRK:

- **wersja** — maj 2019,
- **liczba kompetencji** — 9682,
- **kluczowe pola:**
 - id,
 - nazwa,
 - charakterystyka,
 - efekty uczenia.



The image shows the homepage of the Zintegrated Register of Qualifications (ZRK). At the top, there are logos for the European Union, the Polish Republic, and the Institute of Educational Research (IBE). The main header features the text 'Zintegrowany Rejestr Kwalifikacji' and a navigation menu with links: 'Strona główna ZRK', 'Szukam kwalifikacji', 'Szukam podmiotu', 'Warto wiedzieć', 'ZRK', 'Przejdź na stronę ZSK', 'Seminarium i warsztaty', and 'Zaloguj się'. Below the navigation, there are three main sections: 'Kwalifikacje' (with a calendar icon and description 'Przeglądanie/wyszukiwanie kwalifikacji'), 'Podmioty' (with a building icon and description 'Przeglądanie/wyszukiwanie instytucji systemu kwalifikacji'), and 'Dokumenty' (with a document icon and description 'Najważniejsze informacje dla użytkownika rejestru'). At the bottom, a blue section titled 'ZRK w liczbach' displays four statistics: 9997 Kwalifikacje, 24 Dziedziny, 1797 Podmioty, and 632 Użytkownicy zarejestrowani. A footer note states: 'WAŻNE: Przygotowanie opisów syntetycznych charakterystyk kwalifikacji pełnych właściwych dla szkolnictwa wyższego'.

Fundusze Europejskie
Wiedza. Edukacja. Praca


Rzeczpospolita
Polska


IBE
INSTYTUT
BADAŃ
EDUKACYJNYCH


Unia Europejska
Europejski Fundusz Społeczny

Zintegrowany Rejestr Kwalifikacji





Strona główna ZRK Szukam kwalifikacji Szukam podmiotu Warto wiedzieć ZRK Przejdź na stronę ZSK Seminarium i warsztaty Zaloguj się


Kwalifikacje
Przeglądanie/wyszukiwanie kwalifikacji


Podmioty
Przeglądanie/wyszukiwanie instytucji systemu kwalifikacji


Dokumenty
Najważniejsze informacje dla użytkownika rejestru

ZRK w liczbach

 9997 Kwalifikacje	 24 Dziedziny	 1797 Podmioty	 632 Użytkownicy zarejestrowani
--	---	--	---

WAŻNE: Przygotowanie opisów syntetycznych charakterystyk kwalifikacji pełnych właściwych dla szkolnictwa wyższego

Opis zbioru ZRK

Strona główna / Kwalifikacje

Kwalifikacje Eksport

Wyswietlone 21-30 z 9997. Wyszukiwanie zaawansowane

Lp.	Nazwa kwalifikacji	Kategoria kwalifikacji	Rodzaj kwalifikacji	Poziom PRK/ERK	
	<input type="text"/>	-	-	-	
21	Aktor scen muzycznych (343601)	ze szkolnictwa zawodowego	pełna	5	
22	Aktywne prowadzenie sprzedaży skierowanej do klientów biznesowych – przedstawiciel handlowy	rynkowe	cząstkowa	4	
23	Animator kultury (343901)	ze szkolnictwa zawodowego	pełna	6	
24	Asystent Fryzjera (932919) od 1 września 2017r.	ze szkolnictwa zawodowego	pełna	2	
25	Asystent kierownika produkcji filmowej / telewizyjnej (343902)	ze szkolnictwa zawodowego	pełna	4	
26	Asystent kierownika produkcji filmowej i telewizyjnej (343902) od 1 września 2019r.	ze szkolnictwa zawodowego	pełna	4	
27	Asystent osoby niepełnosprawnej (341201)	ze szkolnictwa zawodowego	pełna	4	
28	Asystent osoby niepełnosprawnej (341201) od 1 września 2019r.	ze szkolnictwa zawodowego	pełna	4	
29	Asystentka stomatologiczna (325101)	ze szkolnictwa zawodowego	pełna	4	
30	Asystentka stomatologiczna (325101) od 1 września 2019r.	ze szkolnictwa zawodowego	pełna	4	

« 1 2 3 4 5 6 7 8 9 10 »

Rysunek 2: Przykładowe kwalifikacje z bazy ZRK

<https://rejestr.kwalifikacje.gov.pl/frontend/index.php?r=kwalifikacja%2Fview&id=12545>

[Strona główna](#) / [Kwalifikacje](#) / Technik hutnik (311704) od 1 września 2017r.

Kwalifikacja - podgląd

Informacje ogólne

Efekty uczenia się

Podmioty

Pozostałe informacje

Nazwa kwalifikacji Technik hutnik (311704) od 1 września 2017r.

Skrót nazwy

Rodzaj kwalifikacji kwalifikacja pełna

Poziom PRK/ERK 4

Krótką charakterystyką kwalifikacji, obejmującą informacje o działaniach lub zadaniach, które potrafi wykonywać osoba posiadająca tę kwalifikację

Absolwent szkoły kształcącej w zawodzie Technik hutnik powinien być przygotowany do wykonywania następujących zadań zawodowych:

- 1) obsługiwanie maszyn i urządzeń w procesach technologicznych;
- 2) konserwacji i przeglądów bieżących maszyn i urządzeń hutniczych;
- 3) prowadzenia kontroli wyrobu;
- 4) organizowania i prowadzenia procesów hutniczych;
- 5) wykonywania rozruchu urządzeń i systemów mechatronicznych;
- 6) organizowania i prowadzenie konserwacji oraz przeglądów bieżących maszyn i urządzeń hutniczych;
- 7) nadzorowania procesów hutniczych zgodnie z systemem zarządzania jakością.

Rysunek 3: Przykładowa kwalifikacje z bazy ZRK

Statystyka:

- ok. 10 000 kwalifikacji
- ok. 700 kwalifikacji zawiera pełne opisy
- około 9 tysięcy czyli ponad 93% to opisy dyplomów.
- pozostałe rekordy można podzielić na:
 - zawody (np. piekarz, technik samochodowy)
 - kompetencje (np. montaż układów i urządzeń elektronicznych)
 - rzemiosło (np. krawiec - dyplom mistrzowski)

Odnoszenie ZRK do baz zewnętrznych

W naszych testach sprawdzaliśmy czy można przy pomocy najbardziej popularnych metryk wyznaczyć analogiczne kwalifikacje w innej bazie danych (ESCO).

Przy porównaniu braliśmy pod uwagę nazwy kwalifikacji oraz ich charakterystyki.

Przy testach chcieliśmy porównać działanie 4 metryk:

- Odległość Jaccarda
- TF-IDF
- *Embeddings* cosine distance
- Word mover's distance

Word Mover's distance



Rysunek 4: Przykład WMD

Testy na bazie ESCO

ZRK	ESCO
Opiekun osoby starszej (341202)	opiekun osób starszych w placówce całodobowej
Technik inżynierii środowiska i melioracji (311208)	inżynier inżynierii środowiska
Technik mechanik (311504)	mechanik pojazdów samochodowych
Operator obrabiarek skrawających (722307)	operator maszyn do obróbki skrawaniem
Monter budownictwa wodnego (711701)	inżynier budownictwa wodnego
Murarz-tylnik (711204)	murarz
Stolarz meblowy - dyplom mistrzowski Stolarz meblowy - świadectwo czeladnicze	cieśla

Dane wymagały preprocessingu, który polegał na:

- zamianie wszystkich dużych liter na małe
- usunięciu numerów i znaków interpunkcyjnych, a także wszystkich wyrazów, które nie są rzeczownikami, przymiotnikami oraz czasownikami
- zamianie wszystkich pozostałych wyrazów na ich formę bazową.

Odległość Jaccarda - nazwy kwalifikacji

ESCO/ZRK	opiekun osoba starsza	technik mechanik	murarz	monter budownictwo wodny	operator obrabiarka	stolarz meblowy
opiekun osoba starsza placówka całodobowa	0.6	0	0	0	0	0
mechanik pojazd samochodowy	0	0.25	0	0	0	0
murarz	0	0	1	0	0	0
inżynier wodny budownictwo	0	0	0	0.5	0	0
operator maszyna obróbka skrawanie	0	0	0	0	0.2	0
cieśla	0	0	0	0	0	0

Odległość Jaccarda - charakterystyki kwalifikacji

ESCO/ZRK	opiekun osoba starsza	technik mechanik	murarz	monter budownictwo wodny	operator obrabiarka	stolarz meblowy
opiekun osoba starsza placówka całodobowa	0.14	0.05	0.01	0.01	0.01	0.06
mechanik pojazd samochodowy	0.05	0.11	0.08	0.04	0.11	0.07
murarz	0.02	0.05	0.1	0.02	0.06	0.08
inżynier wodny budownictwo	0.05	0.07	0.03	0.07	0.06	0.06
operator maszyna obróbka skrawanie	0.04	0.07	0.01	0.01	0.08	0.02
cieśla	0.01	0.06	0.12	0.05	0.07	0.1

Embeddings cosine similarity - nazwy kwalifikacji

ESCO/ZRK	opiekun osoba starsza	technik mechanik	murarz	monter budownictwo wodny	operator obrabiarka	stolarz meblowy
opiekun osoba starsza placówka całodobowa	0.83	0.22	0.11	0.06	0.17	0.15
mechanik pojazd samochodowy	0.11	0.65	0.3	0.42	0.49	0.35
murarz	0.14	0.52	1	0.41	0.29	0.53
inżynier wodny budownictwo	-0.02	0.46	0.31	0.87	0.33	0.24
operator maszyna obróbka skrawanie	-0.02	0.5	0.3	0.39	0.82	0.37
cieśla	0.15	0.42	0.61	0.35	0.2	0.46

Embeddings cosine similarity - charakterystyki kwalifikacji

ESCO/ZRK	opiekun osoba starsza	technik mechanik	murarz	monter budownictwo wodny	operator obrabiarka	stolarz meblowy
opiekun osoba starsza placówka całodobowa	0.7	0,27	0.19	0.32	0.29	0.43
mechanik pojazd samochodowy	0.27	0.71	0.49	0.52	0.52	0.56
murarz	0.31	0.51	0.7	0.53	0.52	0.57
inżynier wodny budownictwo	0.45	0.64	0.56	0.75	0.61	0.52
operator maszyna obróbka skrawanie	0.35	0.73	0.52	0.54	0.66	0.61
cieśla	0.28	0.58	0.72	0.56	0.51	0.65

Word mover's distance - nazwy kwalifikacji

Im mniejsza miara tym większe podobieństwo.

ESCO/ZRK	opiekun osoba starsza	technik mechanik	murarz	monter budownictwo wodny	operator obrabiarka	stolarz meblowy
opiekun osoba starsza placówka całodobowa	0.52	1.31	1.37	1.35	1.33	1.32
mechanik pojazd samochodowy	1.3	0.83	1.23	1.13	1.16	1.17
murarz	1.35	1.03	0	1.19	1.23	1.14
inżynier wodny budownictwo	1.37	1.15	1.25	0.34	1.26	1.28
operator maszyna obróbka skrawanie	1.4	1.14	1.26	1.25	0.76	1.26
cieśla	1.34	1.11	0.89	1.23	1.29	1.17

Word mover's distance - charakterystyki kwalifikacji

Im mniejsza miara tym większe podobieństwo.

ESCO/ZRK	opiekun osoba starsza	technik mechanik	murarz	monter budownictwo wodny	operator obrabiarka	stolarz meblowy
opiekun osoba starsza placówka całodobowa	0.94	1.12	1.2	1.2	1.15	1.13
mechanik pojazd samochodowy	1.08	1	1.04	1.13	1	1.03
murarz	1.15	1.09	1.02	1.17	1.08	1.07
inżynier wodny budownictwo	1.06	1.03	1.1	1.01	1.04	1.09
operator maszyna obróbka skrawanie	1.16	1	1.13	1.15	0.99	1.14
cieśla	1.21	1.09	1	1.12	1.07	1

Testy pokazały, że przy porównaniach dłuższych tekstów (charakterystyk czy zestawów uczenia się) metody te potrafią wskazywać na błędną kwalifikację, jednak zawsze docelowa kwalifikacja jest punktowana dość wysoko.

Docelowo najlepiej, gdyby program proponował kilka najbliższych wyników, z których użytkownik mógłby wybrać jego zdaniem najwłaściwszą.

Grupowanie kompetencji z bazy ZRK

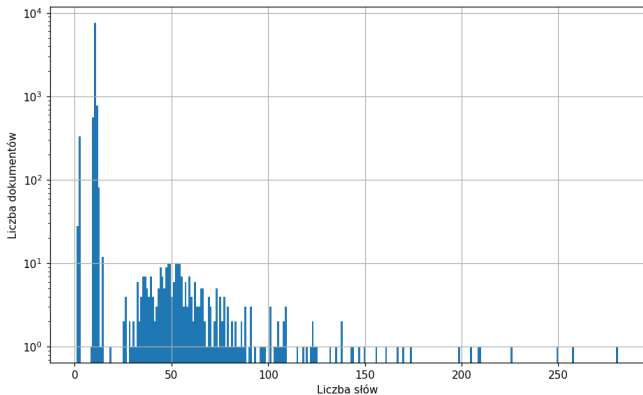
- podobieństwo kompetencji przez analizę podobieństwa charakterystyk — pole tekstowe, opisowe,
- metody porównywania tekstów (charakterystyk kompetencji):
 - model wektorowy z ważeniem tf-idf,
 - modele językowe,

I. Wstępna analiza z wykorzystaniem systemu WebSty

- **WebSty** (ws.clarin-pl.eu/websty.shtml) — webowy system do analizy podobieństwa treści,
- predefiniowane metody porównywania:
 - Analiza autorstwa,
 - Analiza stylu gramatycznego,
 - **Grupy podobieństwa treści**,
 - Klasyczna analiza autorstwa
- konfiguracja dla *Grupy podobieństwa treści*:
 - tokenizacja i analiza morfologiczna przy użyciu WCRFT2,
 - cechy: 1000 najczęstszych lematów rzeczowników,
 - lematy występujące tylko w jednym dokumencie były pomijane.
 - ważenie cech przy użyciu tf-idf.
- **Efekt: brak wyników grupowania ze względu na zbyt dużą liczbę podobnych tekstów**

Analiza zbioru

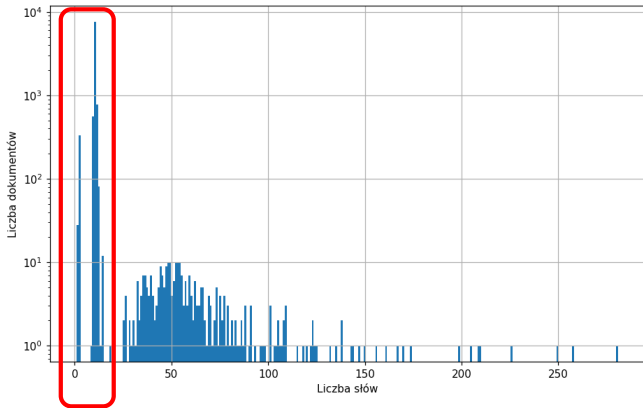
- krótkie (do 12 słów) — prawie 97% całego zbioru,
- długie (powyżej 12 słów) — ok 3% całego zbioru (ok 320 tekstów).



Rysunek 5: Rozkład długości charakterystyk kompetencji

Analiza zbioru

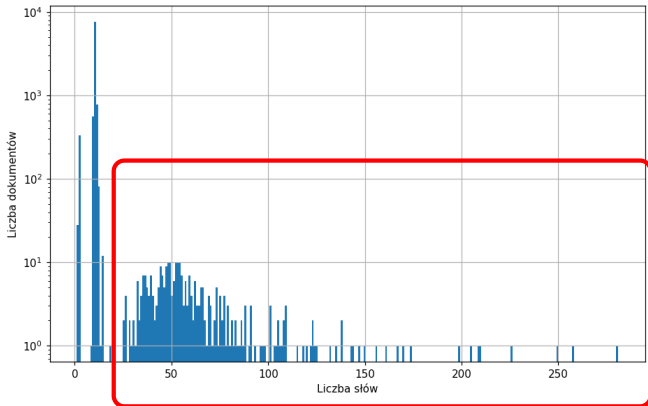
- krótkie (do 12 słów) — prawie 97% całego zbioru,
- długie (powyżej 12 słów) — ok 3% całego zbioru (ok 320 tekstów).



Rysunek 5: Rozkład długości charakterystyk kompetencji

Analiza zbioru

- krótkie (do 12 słów) — prawie 97% całego zbioru,
- długie (powyżej 12 słów) — ok 3% całego zbioru (ok 320 tekstów).



Rysunek 5: Rozkład długości charakterystyk kompetencji

II. Grupowanie krótkich tekstów

Grupa 1 (85)

- tytuł zawodowy/naukowy: doktor nauk **biologicznych** w zakresie **biochemia**
- tytuł zawodowy/naukowy: doktor nauk **biologicznych** w zakresie **biotechnologia**
- tytuł zawodowy/naukowy: doktor nauk **chemicznych** w zakresie **biochemia**
- tytuł zawodowy/naukowy: doktor nauk **rolniczych** w zakresie **biotechnologia**
- tytuł zawodowy/naukowy: doktor nauk **rolniczych** w zakresie **inżynieria rolnicza**,
- (...)

Grupa 2 (9)

- ogólnoakademicki; tytuł zawodowy/naukowy: magister lub inny równorzędny;
- profil kształcenia: ogólnoakademicki; tytuł zawodowy/naukowy: inżynier lub inny równorzędny;
- profil kształcenia: praktyczny; tytuł zawodowy/naukowy: inżynier lub inny równorzędny;
- (...)

Grupa 3

- nie dotyczy
- nan

Podobieństwo z użyciem modeli językowych

- **przeznaczone dla tekstów z rozłącznymi zbiorami terminów istotnych** — głównie krótkie opisy, np. *Technik budowy jednostek pływających* i *Technik budownictwa okrętowego*
- **transformacja dokumentu do postaci wektora liczb rzeczywistych** z wykorzystaniem modeli językowych,
- **podobieństwo jako wartość cosinus kąta między wektorami.**
- modele **fastText**:
 - fasttext.cc/docs/en/crawl-vectors.html — modele dla 157 języków uczonych na Common Crawl,
 - fasttext.cc/docs/en/pretrained-vectors.html — modele dla 294 języków uczone na wersjach językowych Wikipedii
 - clarin-pl.eu/dspace/handle/11321/600 — model uczony na KGR10.

Tytuł zawodowy/naukowy: Doktor nauk społecznych w zakresie nauki o bezpieczeństwie

- tf-idf
 - 0.56 *Tytuł zawodowy/naukowy: Doktor nauk społecznych w zakresie nauki o polityce*
 - 0.54 *Tytuł zawodowy/naukowy: Doktor nauk społecznych w zakresie nauki o obronności*
 - 0.54 *Tytuł zawodowy/naukowy: Doktor nauk społecznych w zakresie nauki o mediach*
- fastText
 - 0.53 *Tytuł zawodowy/naukowy: Doktor nauk społecznych w zakresie nauki o obronności*
 - 0.45 *Tytuł zawodowy/naukowy: Doktor nauk o zdrowiu w zakresie nauk o zdrowiu*
 - 0.45 *Tytuł zawodowy/naukowy: Doktor nauk ekonomicznych w zakresie finanse*

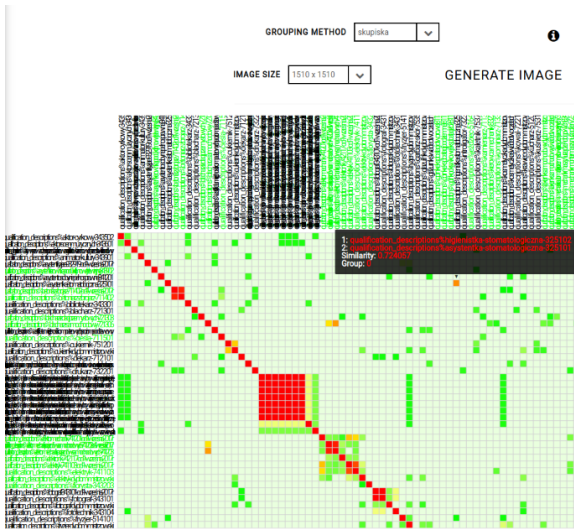
Obserwacje i wnioski

- **elementy strukturalne zawiązują podobieństwo opisów** — kluczowe jest odfiltrowanie terminów strukturalnych przed porównaniem,
- **podobieństwo między terminami z dziedziny jest niższe niż między terminami pokrewnymi**, np.:

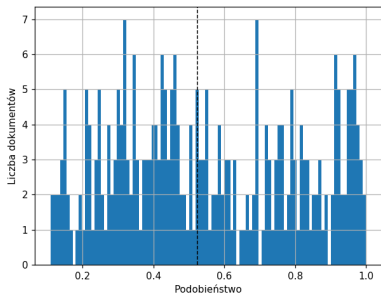
	1	2	3	4	5	6
1. dekarz	–	0.72	0.69	0.42	0.41	0.47
2. cieśla	–	–	0.60	0.32	0.28	0.34
3. mechanik	–	–	–	0.25	0.34	0.24
4. dach	–	–	–	–	0.49	0.53
5. montaż rynien	–	–	–	–	–	0.62
6. kładzenie dachówki	–	–	–	–	–	–

- **brak wektorów dla fraz wielowyrazowych** — średnia wektorów dla poszczególnych słów.

III. Grupowanie długich tekstów (WebSty)

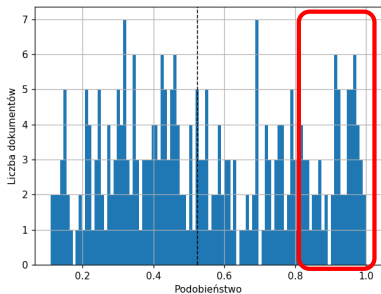


Rysunek 6: Mapa ciepła przedstawiająca podobieństwo tekstów



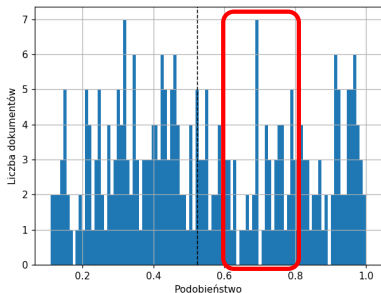
Rysunek 7: Rozkład podobieństwa dla najbardziej podobnych propozycji

- wyniki dla 286 z 320 tekstów,



Rysunek 7: Rozkład podobieństwa dla najbardziej podobnych propozycji

- wyniki dla 286 z 320 tekstów,
- 0,8–1,0 (87 par) — poprawnie wskazane jako podobne,



Rysunek 7: Rozkład podobieństwa dla najbardziej podobnych propozycji

- wyniki dla 286 z 320 tekstów,
- 0,8–1,0 (87 par) — poprawnie wskazane jako podobne,
- 0,6–0,8 (121 par) — niższa precyzja, **kryterium podobieństwa stało się podobieństwo struktury opisów**

Podobieństwo 0,65 dla pary:

[12543]: *Operator maszyn i urządzeń hutniczych (812121) od 1 września 2017r.*

[1]Absolwent szkoły kształcącej w zawodzie [2]Operator [3]maszyn [4]i urządzeń hutniczych [5]powinien być przygotowany do wykonywania następujących zadań zawodowych:

1) [7]obsługiwanie maszyn i urządzeń [9]w [10]procesach technologicznych[11];

[12]2[13]) [15]konserwacji [16]i przeglądów bieżących [17]maszyn i urządzeń [18]hutniczych;

3) prowadzenia kontroli wyrobu[19].

[12368] *Mechanik – operator maszyn do produkcji drzewnej (817212)*

[1]Absolwent szkoły kształcącej w zawodzie [2]mechanik-operator [3]maszyn [4]do produkcji drzewnej [5]powinien być przygotowany do wykonywania następujących zadań zawodowych:

1) [6]montowania maszyn i urządzeń do produkcji drzewnej;

2) kontrolowania i nadzorowania pracy maszyn i urządzeń do produkcji drzewnej;

3) [7]obsługiwanie maszyn i urządzeń[8], stosowanych [9]w [10]procesie produkcji materiałów drzewnych[11];

[12]4[13]) [14]prowadzenia bieżącej kontroli jakości surowców i produktów;

5) wykonywania przeglądów i [15]konserwacji [17]maszyn i urządzeń[19].

- **odcięcie listy cech do 1000 pozycji** — odcięcie po frekwencji powoduje, że terminy dziedzinowe są niewystarczająco częste, aby znalazły się na liście cech,
- **brak uwzględnienia przymiotników** — przymiotniki są dobrym wyznacznikiem dziedziny, np. *urządzenie **hutnicze***, a *urządzenie **rolnicze***,
- **brak uwzględnienia terminów wielowyrazowych** — wyrażenia identyfikujące określone zawody znajdowały się na liście terminów wskazanych przez TermoPL, np. *pojazd samochodowy*, *wyrób włókienniczy*, *obsługa ciągnika rolniczego*.

IV. Rozszerzona analiza długich tekstów (TfidfVectorizer)

- lokalny potok przetwarzania:
 - tager morfologiczny — WCRFT2 poprzez API NlpRest2 (CLARIN-PL),
 - TfidfVectorizer z pakietu Gensim,
- konfiguracja:
 - n-gramy lematów długości od 1 do 3 tokenów,
 - ignorowanie wartości pos: interp, num, conj i prep,
 - ograniczenie do 4 tys. najczęstszych cech,
 - wykorzystanie stoplisty,
 - ważenie cech przy pomocy tf-idf.

TermoPL

<http://zil.ipipan.waw.pl/TermoPL>

<https://ws.clarin-pl.eu/termopl.shtml>

#	Ranking	Pojęcie	Pojęcie zlematyzowane	C-value	Długość	Freq_s	Freq_in	Kontekst
5	5	wykonywać	[wykonywanie]	65	1	657	363	74
13	13	zadanie	[zadanie]	27	1	289	284	12
12	12	zawód	[zawód]	27	1	284	108	9
1	1	zadanie zawodowy	[zadanie zawodowe]	2.6e+2	2	265	3	1
26	26	absolwent	[absolwent]	17	1	262	262	3
20	20	szkoła	[szkoła]	20	1	261	261	4
14	14	urządzenie	[urządzenie]	26	1	259	197	60
2	2	absolwent szkoły	[absolwent szkoły]	2.6e+2	2	256	0	0
18	18	praca	[praca]	21	1	218	198	56
34	30	prowadzić	[prowadzenie]	15	1	158	142	41
35	31	wyrób	[wyrób]	14	1	146	143	57
36	32	osoba	[osoba]	14	1	143	55	25
39	34	maszyna	[maszyna]	13	1	133	104	24
46	40	organizować	[organizowanie]	11	1	116	37	12
54	44	technika	[technika]	9.8	1	100	99	62

Porównanie wyników

- **wyższa liczba wyników** — wzrost z 286 do 309,
- **wyniki dla pozycji, dla których wcześniej było brak wyników:**
 - Monter sieci i instalacji sanitarnych (712618) od 1 września 2017r.*
 - (0.90) *Monter sieci, instalacji i urządzeń sanitarnych (712616)*
 - (0.78) *Technik inżynierii sanitarnej (311218) od 1 września 2017r.*
 - (0.71) *Technik urządzeń sanitarnych (311209)*
- **lepsze propozycje**, np. dla *Technik włókiennik (311932)* wcześniej było (0.77) *Piekarz (751204)*, a obecnie:
 - (0.88) *Technik włókiennik (311932) od 1 września 2017r.*
 - (0.61) *Operator maszyn w przemyśle włókienniczym (815204) od 1 września 2017r.*
 - (0.60) *Operator maszyn w przemyśle włókienniczym (815204)*
- **brak wyników** dla tekstów, dla których wcześniejsze wyniki były niepoprawne, np. dla *Kominiarz (713303)* było *Technik urządzeń dźwigowych (311940)*.

- zauważalny wpływ przymiotników oraz fraz wielowyrazowych, np. dla *Operator maszyn i urządzeń hutniczych (812121) od 1 września 2017r.* najbardziej istotnymi cechami były frazy:
 - 0.306 urządzenie hutniczy
 - 0.306 maszyna urządzenie hutniczy
 - 0.260 hutniczy
 - 0.181 urządzenie proces technologiczny
 - 0.181 urządzenie hutniczy prowadzenie
- dalsze pomysły:
 - lepsze frazowanie, np. z użyciem gramatyki NKJP i narzędzia Spejd,
 - uzupełnienie stoplisty o pełne frazy, a nie tylko pojedyncze słowa.

Ocena i rozwinięcie wyników uzyskanych w ramach ekspertyzy

Ocena i rozwinięcie wyników uzyskanych w ramach ekspertyzy

	Podejście 1 (IPI)	Podejście 2 (IBE)
1. Podstawa porównania	charakterystyki	efekty uczenia się
2. Cechy	n -gramy (1-,2-,3-)	rzeczowniki
3. Liczba cech	ok. 4000	ok. 3300
4. Ważenie cech	TFIDF	0–1

- Podejście 1. – „oszczędne”, niekiedy dla danej kwalifikacji nie znajdzie się żadna podobna.
- Podejście 2. – „rozrzutne”, dla każdej kwalifikacji mamy wiele kwalifikacji podobnych (choć stopień podobieństwa bywa niski).

Pytanie: Które podejście jest bardziej efektywne? Rozważmy kilka przykładów.

11948: Asystentka stomatologiczna (325101)

Najbardziej znaczące frazy:

0.265 lekarz dentysta

0.265 dentysta

0.254 gabinet

0.231 lekarz

0.165 stomatologiczny powinien przygotować

11948: Asystentka stomatologiczna (325101)

Najbardziej podobne dokumenty:

[1] 0.8071 item_11959 Higienistka stomatologiczna (325102)

[2] 0.2290 item_11972 Technik dentystyczny (321402)

[3] 0.2127 item_12299 Technik prac biurowych (411004)

Podejście 2

1. Opiekun medyczny (532102) 12310 1.0
2. Technik elektroradiolog (321103) 12271 0.989
3. Technik sterylizacji medycznej (321104) 12272 0.98
4. Opiekunka dziecięca (325905) 12276 0.98
5. Protetyk słuchu (321401) 12037 0.97
6. Higienistka stomatologiczna (325102) 11959 0.95
7. Ortoptystka (325906) 11962 0.95
8. Technik dentystyczny (321402) 11972 0.95
9. Technik farmaceutyczny (321301) 11999 0.95
10. Technik ortopeda (321403) 12021 0.95

Skąd rozbieżności?

Fundusze Europejskie
Wzrostka i Innowacja Regionalna

MINISTERSTWO
EDUKACJI
NARODOWEJ

IBE kwalifikacje dla każdego

Unia Europejska
Europejski Fundusz Społeczny

Zintegrowany Rejestr Kwalifikacji

Strona główna ZRK Szukam kwalifikacji Szukam podmiotu Warto wiedzieć ZRK Przejdź na stronę ZSK Zaloguj się

Kwalifikacja - podgląd

Informacje ogólne Efekty uczenia się Podmioty Pozostałe informacje

Nazwa kwalifikacji

Skrót nazwy

Rodzaj kwalifikacji Poziom PRK/ERK

Krótką charakterystyką kwalifikacji, obejmującą informacje o działaniach lub zadaniach, które potrafi wykonywać osoba posiadająca tę kwalifikację

Absolwent szkoły kształcącej w zawodzie asystentka stomatologiczna powinien być przygotowany do wykonywania następujących zadań zawodowych:

- 1) przygotowywania gabinetu i stanowiska pracy lekarza dentystry do pracy;
- 2) asystowania lekarzowi dentyście podczas wykonywania zabiegów;
- 3) przygotowywania i przechowywania leków, materiałów, narzędzi oraz konserwowania na bieżąco sprzętu stosowanego w stomatologii;
- 4) wykonywania czynności administracyjnych i prowadzenia dokumentacji związanej z funkcjonowaniem gabinetu dentystrycznego.

Orientacyjny nakład pracy potrzebny do uzyskania kwalifikacji [godz.]

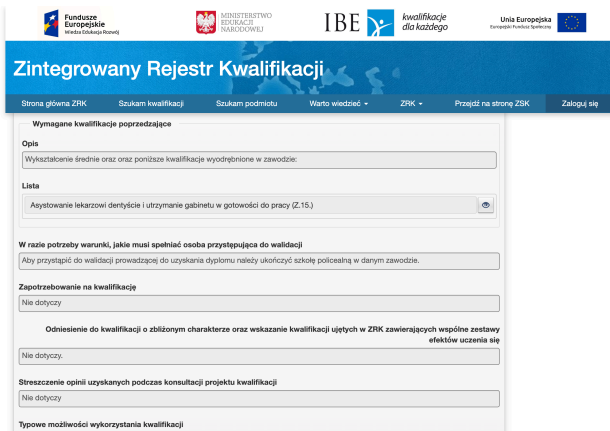
Grupa osób, które mogą być zainteresowane uzyskaniem kwalifikacji

Akcje

- Lista
- Eksport
- Wydruk

Rysunek 8: Struktura kwalifikacji


Skąd rozbieżności?



The screenshot displays the ZRK website interface. At the top, there are logos for Fundusze Europejskie, Ministerstwo Edukacji Narodowej, IBE, and Unia Europejska. The main header is blue with the text "Zintegrowany Rejestr Kwalifikacji". Below the header is a navigation bar with links: "Strona główna ZRK", "Szukam kwalifikacji", "Szukam podmiotu", "Warto wiedzieć", "ZRK", "Przejdź na stronę ZSK", and "Zaloguj się".

The main content area shows a qualification entry with the following sections:

- Wymagane kwalifikacje poprzedzające**
- Opis**: Wykształcenie średnie oraz oraz poniższe kwalifikacje wyodrębnione w zawodzie.
- Lista**: Asystowanie lekarzowi dentyście i utrzymanie gabinetu w gotowości do pracy (Z.15.)
- W razie potrzeby warunki, jakie musi spełniać osoba przystępująca do walidacji**: Aby przystąpić do walidacji prowadzącej do uzyskania dyplomu należy ukończyć szkołę policealną w danym zawodzie.
- Zapotrzebowanie na kwalifikację**: Nie dotyczy
- Odniesienie do kwalifikacji o zbliżonym charakterze oraz wskazanie kwalifikacji ujętych w ZRK zawierających wspólne zestawy efektów uczenia się**: Nie dotyczy.
- Streszczenie opinii uzyskanych podczas konsultacji projektu kwalifikacji**: Nie dotyczy
- Typowe możliwości wykorzystania kwalifikacji**



Rysunek 9: Kwalifikacje poprzedzające

Podejście 2a – po zsumowaniu kwalifikacji

1. Higienistka stomatologiczna (325102) 0.8
2. Ortoptystka (325906) 0.6
3. Technik sterylizacji medycznej (321104) 0.59
4. Technik elektroradiolog (321103) 0.59
5. Technik elektroniki i informatyki medycznej (311411) od 1 września 2017r. 0.58
6. Terapeuta zajęciowy (325907) 0.56
7. Opiekun medyczny (532102) 0.56
8. Opiekunka dziecięca (325905) 0.56
9. Technik ortopeda (321403) 0.55
10. Technik dentystyczny (321402) 0.53

Najbardziej znaczące frazy:

0.314 witryna

0.219 podstawowy zakres

0.219 modyfikacja

0.204 wprowadzać

0.204 witryna internetowy

Najbardziej podobne dokumenty

[1] 0.2468 item_12643 Projektowanie grafiki komputerowej

Podejście 2a – po zsumowaniu kwalifikacji

1. Programowanie, tworzenie i administrowanie stronami internetowymi i bazami danych (EE.09.) od 1 września 2017r. 0.21
2. Tworzenie aplikacji internetowych i baz danych oraz administrowanie bazami (E.14.) 0.16
3. Organizacja i prowadzenie procesu obsługi pojazdów samochodowych (M.42.) 0.09
4. Wykonywanie pomiarów sytuacyjnych, wysokościowych i realizacyjnych oraz opracowywanie wyników tych pomiarów (BD.31) od 1 września 2017r. 0.077
5. Organizacja i prowadzenie procesu obsługi pojazdów samochodowych (MG.43.) od 1 września 2017r. 0.077
6. Montaż i eksploatacja komputerów osobistych oraz urządzeń peryferyjnych (E.12.) 0.076

Najbardziej znaczące frazy:

0.379 potwierdzić

0.354 komputer

0.267 kompetencja

0.248 umiejętność

0.209 tekst

Najbardziej podobne dokumenty:

Podejście 2a – po zsumowaniu kwalifikacji

1. Tworzenie aplikacji internetowych i baz danych oraz administrowanie bazami (E.14.) 0.15
2. Obsługa oprogramowania i sprzętu informatycznego wspomagających użytkownika z niepełnosprawnością wzrokową (E.11.) 0.14
3. Obsługa oprogramowania i sprzętu informatycznego wspomagających użytkownika z niepełnosprawnością wzrokową (EE.07.) od 1 września 2017r. 0.14
4. Prowadzenie obsługi biura 0.14
5. Projektowanie lokalnych sieci komputerowych i administrowanie sieciami (E.13.) 0.10
6. Programowanie, tworzenie i administrowanie stronami internetowymi i bazami danych (EE.09.) od 1 września 2017r. 0.10

Które podejście jest lepsze?

To nie jest dobrze zadane pytanie. Nie na tym etapie.
Rozważmy podejście 1a – po zsumowaniu kwalifikacji.

Najbardziej znaczące frazy:

0.181 lekarz dentysta

0.181 dentysta

0.176 stomatologiczny

0.161 dentystyczny

0.152 gabinet dentystyczny

0.150 gabinet

Asystentka stomatologiczna (325101)

Najbardziej podobne dokumenty:

[1] 0.9158 item_11959 Higienistka stomatologiczna (325102)

[2] 0.7508 item_11947 Asystowanie lekarzowi dentyście i utrzymanie gabinetu w gotowości do pracy (Z.15.)

[3] 0.7347 item_12698 Ratownik medyczny (325601)

[4] 0.6841 item_11962 Ortoptystka (325906)

[5] 0.6777 item_11972 Technik dentystyczny (321402)

[6] 0.6556 item_11958 Prowadzenie działalności profilaktyczno-leczniczej pod nadzorem i na zlecenie lekarza dentysty oraz utrzymanie gabinetu w gotowości do pracy i prowadzenie promocji zdrowia(Z.14.)

[7] 0.6456 item_12271 Technik elektroradiolog (321103)

[8] 0.6439 item_12276 Opiekunka dziecięca (325905)

Najbardziej znaczące frazy:

0.421 serwer

0.386 tworzyć

0.268 klient

0.231 treść

0.216 baza dana

0.212 kop

Najbardziej podobne dokumenty:

[1] 0.4638 item_12577 Programowanie, tworzenie i administrowanie stronami internetowymi i bazami danych (EE.09.) od 1 września 2017r.

[2] 0.3346 item_12406 Tworzenie aplikacji internetowych i baz danych oraz administrowanie bazami (E.14.)

[3] 0.2705 item_12575 Technik informatyk (351203) od 1 września 2017r.

[4] 0.2223 item_12294 Technik informatyk (351203)

[5] 0.2116 item_12691 Zarządzanie projektami

Najbardziej znaczące frazy:

0.302 własne

0.287 drukować

0.249 tworzyć

0.246 wykres

0.238 za pomoc

Najbardziej podobne dokumenty:

[1] 0.3282 item_12406 Tworzenie aplikacji internetowych i baz danych oraz administrowanie bazami (E.14.)

[2] 0.2364 item_12577 Programowanie, tworzenie i administrowanie stronami internetowymi i bazami danych (EE.09.) od 1 września 2017r.

[3] 0.2261 item_12587 Obsługa oprogramowania i sprzętu informatycznego wspomagających użytkownika z niepełnosprawnością wzrokową (EE.07.) od 1 września 2017r.

[4] 0.2240 item_12294 Technik informatyk (351203)

[5] 0.2196 item_12407 Obsługa oprogramowania i sprzętu informatycznego wspomagających użytkownika z niepełnosprawnością wzrokową (E.11.)

Podsumowanie

- Pilotaże grupowania i aplikacja grupująca
 - testowanie innych miar podobieństwa, ważenia cech, ekstrakcji słów kluczowych, modeli języka
 - zbieranie danych o zawodach, ofert pracy itp.
- Uzupełnienie opisów kwalifikacji
- Wykorzystanie metod z nadzorem – przypisywanie do branż
- Refleksja nad czasownikami, reguły asocjacyjne
- Zaproszenie ekspertów do oceny grupowań