



Relacje między kwalifikacjami – klastrowanie, reguły asocjacyjne i wizualizacja

Warszawa 12.12.2019 r.

Leopold Będkowski, Marcin Będkowski



Kontekst naszych działań: Modernizacja Zintegrowanego Rejestru Kwalifikacji

- ✓ Usprawnienie wyszukiwania i przeglądania treści
(wyszukiwanie semantyczne, opcje filtrowania, kategoryzacja i/lub tagowania treści)
- ✓ Opracowanie automatycznego raportowania i dodatkowych zapytań
(porównywanie kwalifikacji, generowanie list kwalifikacji w oparciu o wybrane kryteria, np. zawierające zadaną frazę, podobne do)
- ✓ Projektowanie aplikacji internetowych:
 - ✓ „Kompas”
 - ✓ „Ścieżki rozwoju”
 - ✓ „Wirtualny asystent”



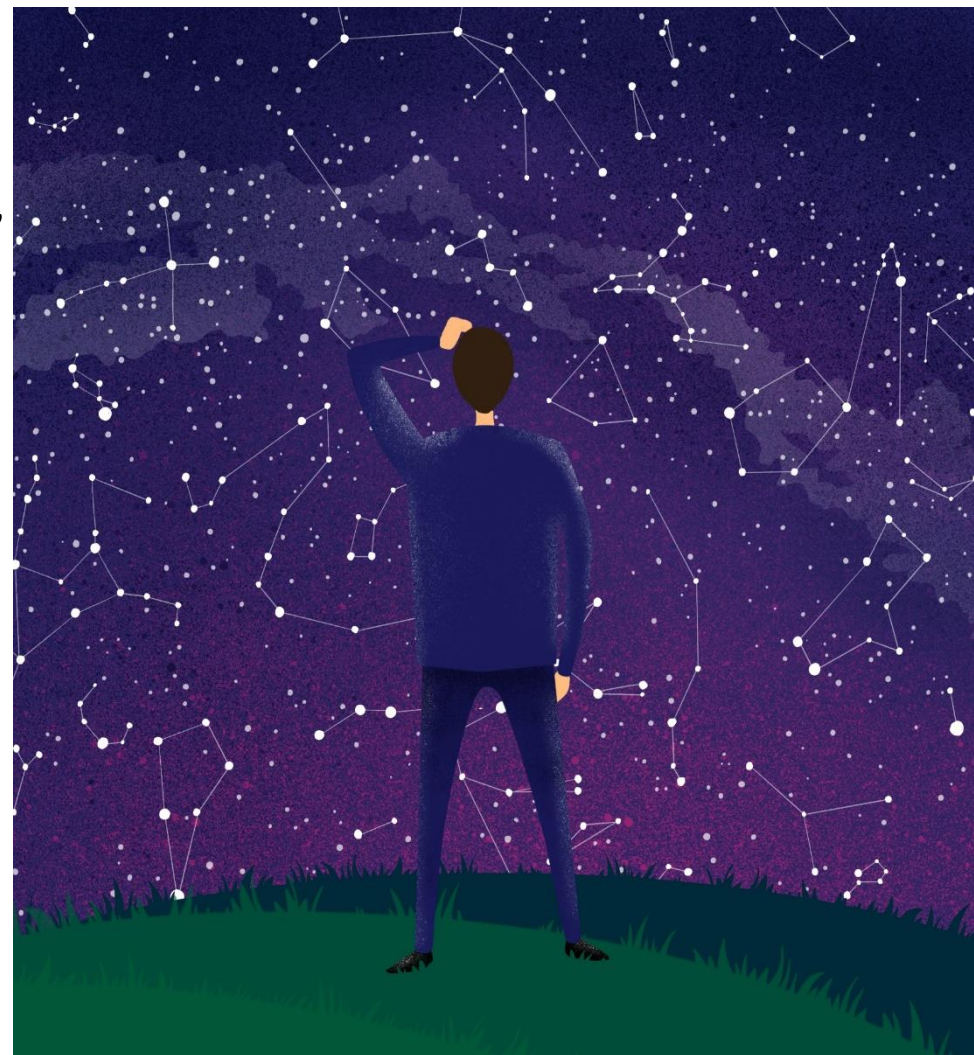
Planowane aplikacje internetowe

- ✓ **„Kompas”** – umożliwi wyszukanie kwalifikacji lub grupy kwalifikacji w ZRK poprzez kolejne przybliżenia według wybranych kryteriów lub kategorii
- ✓ **„Ścieżki rozwoju”** – wizualna reprezentacja relacji pomiędzy kwalifikacjami (poprzedzające, podobne-do, jest-częścią, itp.)
- ✓ **„Wirtualny doradca”** – konwersacyjny interfejs dla użytkownika, który umożliwi dostęp do ZRK i szeregu usług: wyszukiwania informacji, FAQ i ewentualnie usług takich jak dopasowywanie kwalifikacji do CV, testów diagnostycznych/doradczych



„Dlaczego?": Podobne trendy na poziomie międzynarodowym

- ✓ Dostępność i transparentność systemu kwalifikacji
- ✓ Gromadzenie i transfer kompetencji, ścieżki uczenia (instytucje certyfikujące i uczący się)
- ✓ Zapobieganie redundancji w systemie
- ✓ Porównywanie kwalifikacji między różnymi krajami





„Co?": długa lista

- ✓ **Ocena podobieństwa obiektów**
- ✓ Określanie i reprezentowanie relacji między kwalifikacjami
- ✓ **Grupowanie kwalifikacji**
- ✓ Klasyfikacja i łączenie do istniejących taksonomii i klasyfikacji
- ✓ Wsparcie procesu decyzyjnego i procesu tworzenia/opisywania kwalifikacji
- ✓ **Wsparcie procesu przypisywania poziomu ramy**



„Jak?": Podobne wyzwania?

Sądzisz, że
potrzebujemy
pomocy, żeby
to rozgryźć?



Jak wiele osób należy zaangażować,
aby porównać, pogrupować,
i otagować **10 000** kwalifikacji, spośród
których każda jest unikatowa – tzn. ma
odmienną strukturę i treść?

$$(X + Y) \times N$$

- ✓ X analityków
- ✓ Y ekspertów z dziedziny kwalifikacji
- ✓ N miesięcy



Reprezentacja kwalifikacji wg relacji poprzedzania (wizualizacja ścieżki rozwoju)

File Edit View Window Help Developer

Searching From Database

Label Relationship Tag Property

poziom

- Świadectwo ukończenia gimnazjum/szkoły podstawowej
- Matura
- Dyplom I stopnia
- Dyplom II stopnia

- 2 1/1
- 3 96/96
- 4 119/119
- 5 23/24
- 6 5039/5039
- 7 3001/3001
- 8 0/981

Szkolnictwo zawodowe

Dyplomy II stopnia

Dyplomy III stopnia

Dyplomy I stopnia

0 2

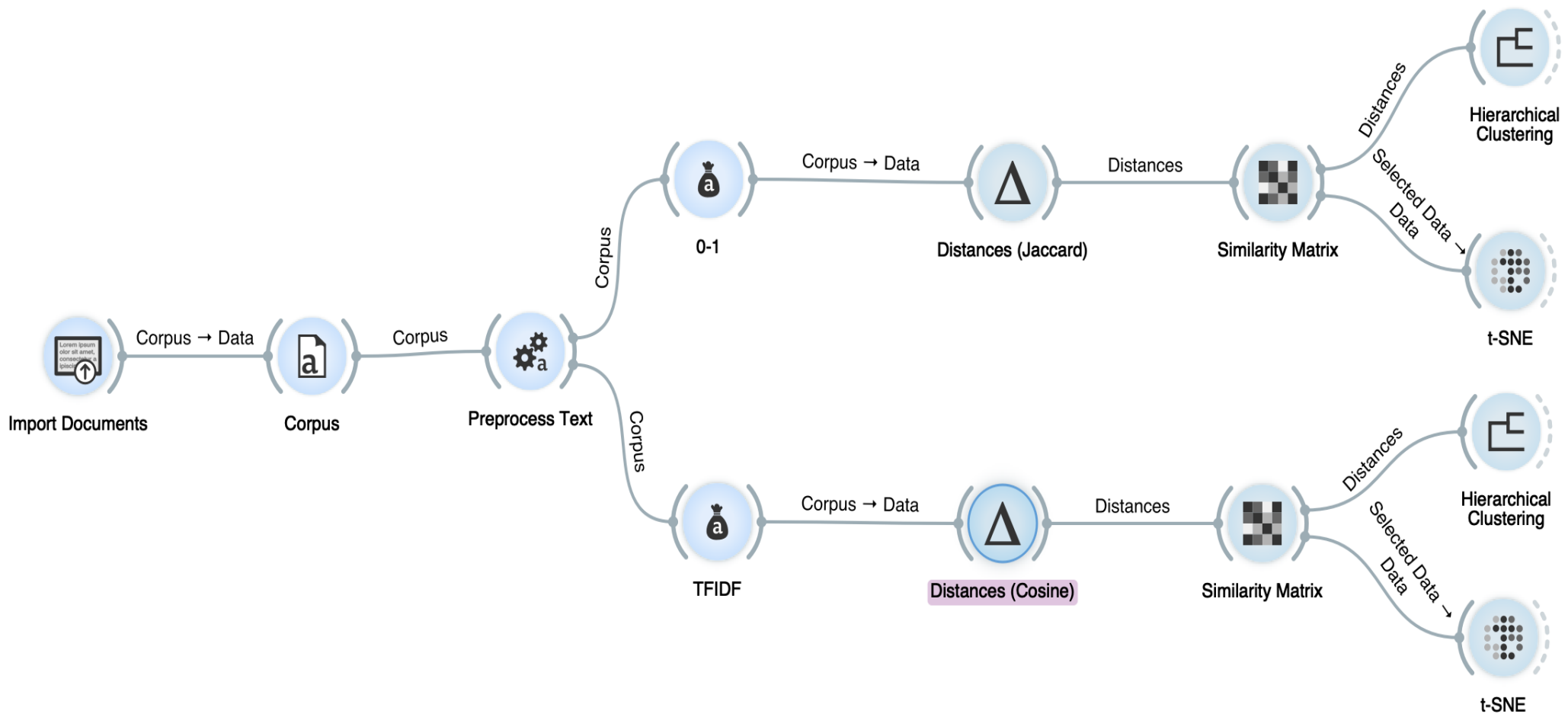




Podobieństwo i grupowanie



Przykładowy potok przetwarzania w Orange





Opis podejścia

Podstawa porównania	Nazwa + charakterystyka + efekty uczenia się + [kwalifikacje składowe]
Cechy	zlematyzowane n-gramy
Liczba cech	ok. 44000 [!]
Metoda ważenia	TFIDF
Miara podobieństwa	cosinus



Przykład przetwarzania danych: „Atomiazacja” efektów uczenia się (EU) – nietrywialne dla dla j. polskiego

Przy pomocy narzędzi NLP, zatomizowaliśmy EU i wydobyliśmy i zlematyzowaliśmy odpowiednie słowa:

- (Uczeń) **opisuje** i **wyjaśnia** budowę **młotka** i **gwoździ**



- **Opisuje** budowę **młotka** +
Wyjaśnia budowę **młotka** +
Opisuje budowę **gwoździ** +
Wyjaśnia budowę **gwoździ**



- (**opisywać**, budowa, **wyjaśniać**, **młotek**, **gwóźdź**)



Jaccard

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$$



TFIDF

$$tfidf_{i,j} = \frac{n_{i,j}}{\sum_k n_{k,j}} \cdot \log \frac{|D|}{|d : t_i \in d|}$$



Podobieństwo cosinusowe

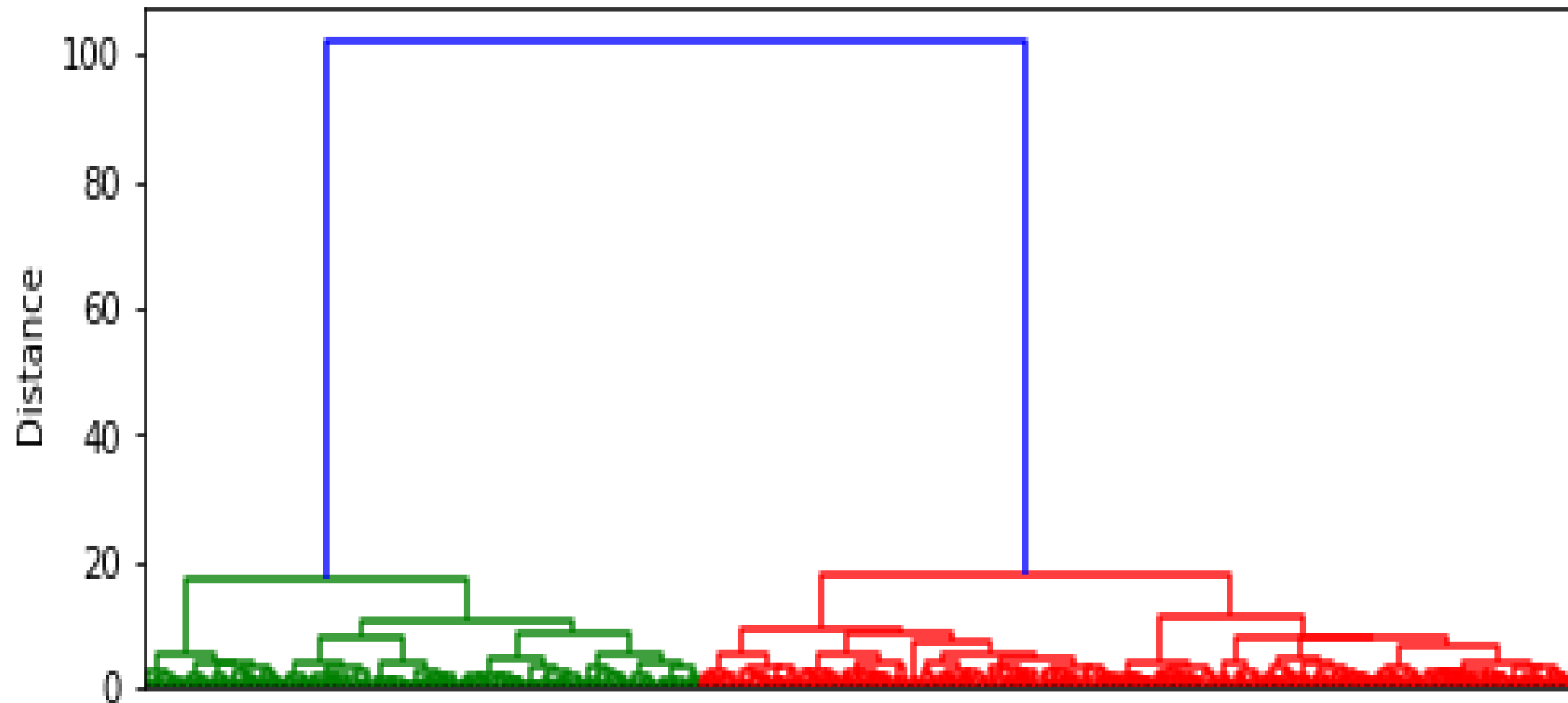
$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\| \|\mathbf{B}\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$



... ale spójrzymy na rezultaty



Przykład 1. Dendrogram grupowania hierarchicznego (aglomeracyjnego)





Przykład 2. Dendrogram grupowania hierarchicznego (aglomeracyjnego) – klaster 4 poziomowy

- ✓ Diagnostowanie i naprawa podzespołów i zespołów pojazdów samochodowych (M.18.)
- ✓ Diagnostowanie i naprawa podzespołów i zespołów pojazdów samochodowych (MG.18.) od 1 września 2017r.
- ✓ Organizacja i prowadzenie procesu obsługi pojazdów samochodowych (M.42.)
- ✓ Organizacja i prowadzenie procesu obsługi pojazdów samochodowych (MG.43.) od 1 września 2017r.
- ✓ **Użytkowanie pojazdów, maszyn, urządzeń i narzędzi stosowanych w rolnictwie (M.1.)**
- ✓ **Obsługa techniczna oraz naprawa pojazdów, maszyn i urządzeń stosowanych w rolnictwie (M.2.)**
- ✓ **Organizacja prac związanych z eksploatacją środków technicznych stosowanych w rolnictwie (M.43.)**
- ✓ **Eksploatacja systemów mechatronicznych w rolnictwie (M.46.)**
- ✓ **Eksploatacja systemów mechatronicznych w rolnictwie (MG.42.) od 1 września 2017r.**
- ✓ **Zakładanie i prowadzenie upraw ogrodnich (R.5.)**
- ✓ **Planowanie i organizacja prac ogrodnich (R.18.)**
- ✓ **Technik ogrodnik (314205)**
- ✓ **Organizacja i nadzorowanie produkcji rolniczej i pszczelarskiej (R.17.)**
- ✓ **Organizacja i nadzorowanie produkcji rolniczej (R.16.)**
- ✓ **Organizacja chowu i hodowli koni (R.19.)**
- ✓ **Prowadzenie chowu, hodowli i inseminacji zwierząt (R.9.)**
- ✓ **Prowadzenie chowu i inseminacji zwierząt (RL.10.) od 1 września 2017r.**
- ✓ **Technik weterynarii (324002)**
- ✓ **Technik weterynarii (324002) od 1 września 2017 r.**
- ✓ **Wykonywanie czynności pomocniczych z zakresu usług weterynaryjnych (R.10.)**
- ✓ **Wykonywanie czynności pomocniczych z zakresu realizacji zadań inspekcji weterynaryjnej (R.11.)**



Przykład 2. Dendrogram grupowania hierarchicznego (aglomeracyjnego) – klaster 4 poziomowy

- ✓ Diagnostowanie i naprawa podzespołów i zespołów pojazdów samochodowych (M.18.)
- ✓ Diagnostowanie i naprawa podzespołów i zespołów pojazdów samochodowych (MG.42.)
- ✓ Organizacja i prowadzenie procesu obsługi pojazdów samochodowych (M.42.)
- ✓ Organizacja i prowadzenie procesu obsługi pojazdów samochodowych (MG.43.) od 1 września 2017r.
- ✓ Użytkowanie pojazdów, maszyn, urządzeń i narzędzi stosowanych w rolnictwie (M.1.)
- ✓ Obsługa techniczna oraz naprawa pojazdów, maszyn i urządzeń stosowanych w rolnictwie (M.45.)
- ✓ Organizacja prac związanych z eksploatacją środków technicznych stosowanych w rolnictwie (M.46.)
- ✓ Eksploatacja systemów mechatronicznych w rolnictwie (M.46.)
- ✓ Eksploatacja systemów mechatronicznych w rolnictwie (MG.42.) od 1 września 2017r.
- ✓ Zakładanie i prowadzenie upraw ogrodnich (R.5.)
- ✓ Planowanie i organizacja prac ogrodnich (R.18.)
- ✓ Technik ogrodnik (314205)
- ✓ Organizacja i nadzorowanie produkcji rolniczej i pszczelarskiej (R.17.)
- ✓ Organizacja i nadzorowanie produkcji rolniczej (R.16.)
- ✓ Organizacja chowu i hodowli koni (R.19.)
- ✓ Prowadzenie chowu, hodowli i inseminacji zwierząt (R.9.)
- ✓ Prowadzenie chowu i inseminacji zwierząt (RL.10.) od 1 września 2017r.
- ✓ Technik weterynarii (324002)
- ✓ Technik weterynarii (324002) od 1 września 2017 r.
- ✓ Wykonywanie czynności pomocniczych z zakresu usług weterynaryjnych (R.10.)
- ✓ Wykonywanie czynności pomocniczych z zakresu realizacji zadań inspekcji weterynaryjnej (R.11.)

Naprawa i obsługa pojazdów

Obsługa pojazdów i systemów mechatronicznych w rolnictwie

Uprawy ogrodnicze i produkcja rolnicza

Hodowla zwierząt

Usługi weterynaryjne



Przykład 3. Problematyczne łączenie kwalifikacji rynkowych. Najniższy poziom hierarchii

Wynik wcześniejszy:

- ✓ Praca z dzieckiem metodą Marii Montessori
- ✓ Montowanie i serwisowanie przyłączy oraz instalacji wewnętrzzbudynkowych w technologii światłowodowej

Wynik obecny:

- ✓ Praca z dzieckiem metodą Marii Montessori
- ✓ Prowadzenie terapii środowiskowej dzieci i młodzieży

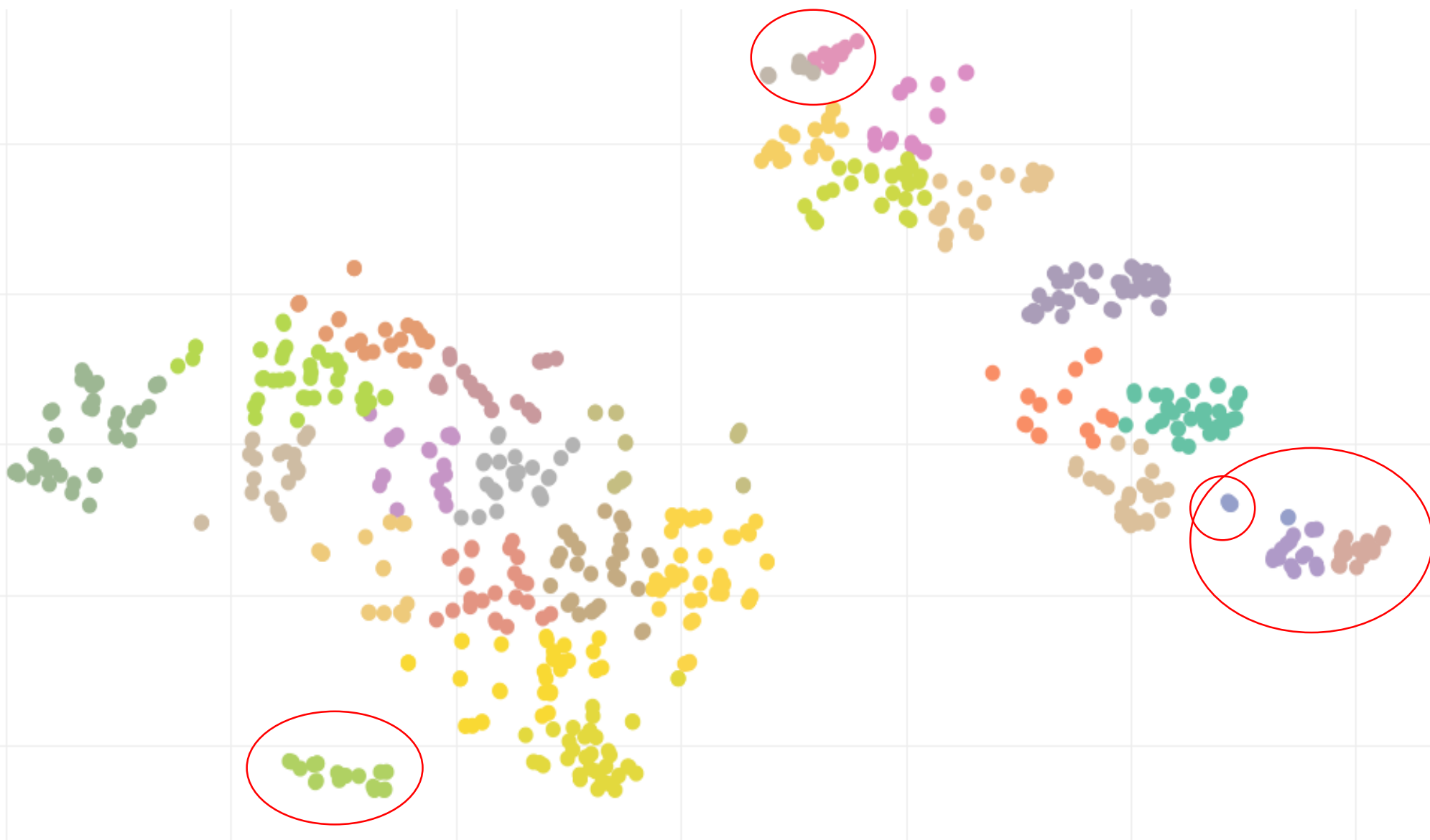


Przykład 3. Problematiczne łączenie kwalifikacji rynkowych (ciąg dalszy)

- ✓ Odzyskiwanie danych z dysków twardych HDD
- ✓ Certyfikat umiejętności komputerowych – poziom podstawowy
- ✓ Tworzenie witryn internetowych
- ✓ Programowanie i obsługiwane procesu druku 3D
- ✓ **Praca z dzieckiem metodą Marii Montessori**
- ✓ **Prowadzenie terapii środowiskowej dzieci i młodzieży**
- ✓ Prowadzenie szkoleń metodami aktywizującymi
- ✓ Zarządzanie obsługą spraw kadrowych
- ✓ Montowanie i serwisowanie przyłączy oraz instalacji wewnętrznych w technologii światłowodowej
- ✓ Serwis napojów mieszanych i alkoholi
- ✓ Pilotowanie imprez turystycznych
- ✓ [... +5]



Przykład 4. Wizualizacja kwalifikacji w przestrzeni 2-wymiarowej przy pomocy t-SNE z grupowaniem k-średnich (kolory) ([demo](#))





Wnioski

- ✓ Dotyczące danych:
 - ✓ Połączenie kwalifikacji poprawia wyniki (nasylenie słowami kluczowymi)
 - ✓ Przykład kwalifikacji rynkowych świadczy o potrzebie:
 - ✓ eliminacji wspólnych efektów uczenia się z kwalifikacji zawodowych (BHP, języki, praca w grupie)?
 - ✓ pominięcia kwalifikacji cząstkowych ze szkolnictwa zawodowego?
- ✓ Dotyczące cech:
 - ✓ 44000 cech to wręcz niedorzecznie dużo
 - ✓ ale wyniki są relatywnie dobre
 - ✓ dobre podejście bazowe, które można ulepszyć w kolejnych etapach
- ✓ Dotyczące grupowania hierarchicznego:
 - ✓ potrzeba zbadania stabilności grupowań
 - ✓ Ward nie jest optymalną metodą?



Nasze wyniki postrzegamy jako „proof of concept”

- ✓ zautomatyzowanego porównywania w kwalifikacji i określania interpretowalnego stopnia podobieństwa
- ✓ zautomatyzowanej ekstrakcji fraz kluczowych w kwalifikacjach
- ✓ grupowanie kwalifikacji niezależnie od istniejących taksonomii i klasyfikacji



Dalsze prace

- ✓ Pilotaż metod grupowania i aplikacja – testowanie innych podejść
 - ✓ wykorzystanie WordNetu
 - ✓ modele językowe oparte na wektorach dystrybucyjnych (word2vec, fasttext, ELMo, USE...)
 - ✓ ARTM (modelowanie tematyczne)
 - ✓ łączenie modeli (ensembling)
- ✓ Gromadzenie danych dotyczących zawodów, ofert pracy itp. dla potrzeb trenowania modeli
- ✓ Konsultacje z ekspertami i ocena rezultatów
- ✓ Studium wykonalności dotyczące chatbota
- ✓ Trzy aplikacje wspierające użytkowników ZRK



Reguły asocjacyjne



Czasowniki a efekty uczenia się

Efekty uczenia się powinny być możliwie **jednoznaczne** i **precyzyjne**.

W odniesieniu do nich projektowane są formy sprawdzenia czy osoby ubiegające się o kwalifikacje, osiągnęły te efekty (np. egzaminy)

Przykład: „(...) **definiuje** pojęcie prawdopodobieństwa”

Taksonomie celów kształcenia (np. taksonomia Blooma).

Czasowniki operacyjne, np. „rozdziela”, „uzasadnia”, „montuje” > potocznie używane czasowniki takie jak „wie”, „rozumie” czy „potrafi”.



Polska Rama Kwalifikacji a czasowniki używane do opisu efektów uczenia się:

Charakterystyki poziomów PRK odnoszą się do:

- ✓ rodzajów czynności (np. wykonywać, planować, rozwiązywać),
- ✓ obiektów czynności (np. podstawowe zjawiska, uwarunkowania zjawisk, teorie dotyczące zjawisk) oraz
- ✓ kontekstu ich wykonywania (np. w przewidywalnych warunkach, w zmiennych i nieprzewidywalnych warunkach, pod kierunkiem bardziej doświadczonej osoby)



REGUŁY ASOCJACYJNE

Problem: jaka jest zależność między poziomem PRK a czasownikami występującymi w kwalifikacji?



CO MÓWI INTUICJA?

	PRK 2	PRK 3	PRK 4	PRK 5	PRK 6	PRK 7
administrować			2			
akceptować				1	3	
akompaniować				1		
aktualizować	6	103	151	29	2	
aktywizować		1	2			
analizować	2	236	331	119	9	2
antycypować				1		
archiwizować		2	18	3		
...						



Analizy asocjacyjne - geneza

- ✓ Analiza koszyka zakupów
- ✓ Regułami asocjacyjnymi są proste implikacje w postaci:

„Jeśli ..., to ...”,

np.: **Jeśli** klient kupił chleb, **to** kupił również masło



PRZYKŁAD 1: ARCHIWIZOWAĆ

Reguła	Wsparcie	Ufność	Lift
{udzielać, archiwizować} ⇒ {PRK=4}	0,0247	1,00	2,028
{identyfikować, archiwizować} ⇒ {PRK=4}	0,0247	0,94	1,909
{archiwizować} ⇒ {PRK=4}	0,0278	0,85	1,738



PRZYKŁAD 2: WYBIERAĆ

Reguła	Wsparcie	Ufność	Lift
{wybierać} ⇒ {PRK=5}	0,0124	0,800	7,611
{dobierać, wybierać} ⇒ {PRK=5}	0,0108	0,875	8,325
{stosować, wybierać} ⇒ {PRK=5}	0,0124	0,888	8,4575



PRZYKŁAD 3: OKREŚLAĆ

Reguła	Wsparcie	Ufność	Lift
{określać, posługiwać, stosować, rozpoznawać, usuwać} ⇒ {PRK=4}	0,0123	1,00	2,028
{określać, sporządzać, charakteryzować, udzielać, identyfikować} ⇒ {PRK=4}	0,0231	1,00	2,028
{określać, wyjaśniać, obsługiwać, dokonywać, <i>charakteryzować</i> } ⇒ {PRK=4}	0,0247	1,00	2,028



Dalsze prace

- ✓ Ulepszenie metody:
 - ✓ dokonanie dezambiguacji czasowników (wyznaczenie reguł dla czasowników w określonych znaczeniach)
opisywać1
opisywać2
 - ✓ odniesienie do taksonomii Blooma i innych
 - ✓ ważenie cech – TFIDF, cracovian?
- ✓ Możliwość wykorzystania:
 - ✓ zbadanie spójności opisów i błędów się w nich pojawiających
 - ✓ model deskrypcyjny / predykcyjny – system rekomendacyjny dla decydentów



Dziękuję za uwagę!

l.bedkowski@ibe.edu.pl
m.bedkowski@ibe.edu.pl

Educational Research Institute

IQS Project Office

Górczewska 8, 01-180 Warsaw, Poland

phone: +48 22 24 17 100, +48 22 24 17 111

e-mail: rejestr@ibe.edu.pl

<http://rejestr.kwalifikacje.gov.pl> | <http://www.ibe.edu.pl>