

Data oddania: \_\_\_\_\_

Ocena: \_\_\_\_\_

Daria Rogowska 229989

Mateusz Srebnik 230004

## Projekt 2. Podsumowania lingwistyczne relacyjnych baz danych

### 1. Cel

Celem zadania jest stworzenie aplikacji, której główna funkcjonalność odpowiedzialna jest za lingwistyczną agregację zawartości, wybranego do badania, zbioru danych[4]. Program generuje opis w języku quasi-naturalnym na podstawie danych liczbowych w zbiorze. Podsumowania danych z bazy tworzone są, na podstawie elementów lingwistycznego podsumowania tj. kwantyfikatory, kwalifikatory, sumatory oraz podmiot, podanych przez użytkownika, które to stanowią interpretację informacji i wiedzy pozyskanej z dużego zbioru danych. Wygenerowane podsumowania są prezentowane w formie tekstowej i opiera się na domyślnych modelach wyrażen w języku naturalnym. Część eksperymentalna zadania stanowi określenie wpływu wyboru kwantyfikatorów, sumatorów, kwalifikatorów i ich miar jakości na wiarygodność i jakość otrzymanych podsumowań.

### 2. Baza danych, zmienne lingwistyczne, kwantyfikatory lingwistyczne

#### 2.1. Charakterystyka podsumowywanej bazy danych

W celu wykonania projektu wybrano zbiór danych SpeedDating[4], który zawiera dane zebrane na eksperymentalnych wydarzeniach 'speed dating' na przełomie lat 2002-2004, przeprowadzone przez Columbia Business School.

Speed dating to rodzaj zorganizowanego randkowania polegający na wyjątkowo krótkich spotkaniach z nieznanymi. Jeden rekord odpowiada jednemu spotkaniu. Dane opisują osobę randkującą oraz przypisanego jej partnera, oraz wrażenia i wyniki po spotkaniu. Zbiór danych może posłużyć np. w celu polepszenia jakości usług agencji matrymonialnych. Zbiór ten zawiera 8378 rekordów, tego samego typu, a każdy z nich opisany jest na 121 atrybutach, z czego wybrano 11 atrybutów do rozmowy:

1. age (oznacza wiek osoby randkującej),
2. d\_age (oznacza różnicę wieku pomiędzy osobą randkującą a jej partnerem podczas spotkania),
3. importance\_same\_race (oznacza w skali [1-10] ważność posiadania tej samej rasy dla osoby randkującej),
4. importance\_same\_religion (oznacza w skali [1-10] ważność posiadania tej samej religii dla osoby randkującej),
5. pref\_o\_intelligence (oznacza w skali [0-100] preferowaną inteligencję partnera dla osoby randkującej),
6. pref\_o\_ambitious (oznacza w skali [0-100] preferowaną ambicję partnera dla osoby randkującej),
7. tvsports (oznacza w skali [1-10] poziom zainteresowania osoby randkującej oglądaniem sportów w telewizji),
8. expected\_num\_interested\_in\_me (oznacza oczekiwaną przez osobę randkującą liczbę osób zainteresowanych nią),
9. guess\_prob\_liked (oznacza w skali [1-10] oczekiwaną szansę na to, że partner polubił osobę randkującą),
10. funny (oznacza w skali [1-10] jak bardzo zabawna jest osoba randkująca według samej siebie),
11. sincere (oznacza w skali [1-10] jak bardzo szczerą jest osoba randkująca według samej siebie)

Przykładowa wartość rekordu przedstawionego za pomocą wyżej wymienionych atrybutów:

Atrybut	wartość
age	21
d_age	6
importance_same_race	2
importance_same_religion	4
pref_o_intelligence	14
pref_o_ambitious	50
tvsports	9
expected_num_interested_in_me	3
guess_prob_liked	8
funny	7
sincere	5

Tabela 1. Tabela przedstawiająca wartości wybranych atrybutów przykładowego rekordu zbioru danych SpeedDating [4]

Wybrane atrybuty przyjmują wartości liczbowe. Ludzie jednak nie posłu-

gują się liczbami w sytuacjach gdy istnieje potrzeba opisu pewnych obiektów lub pojęć takich jak na przykład inteligencja, wygląd, szczerść (nie mówimy „on jest szczerzy w stopniu 1 na 10”, tylko „on jest fałszywy”). Dlatego też atrybutom wybranym przez nas w tym zadaniu, zostały przypisane zwyczajowe wartości lingwistyczne:

Atrybut	Zwyczajowe wartości lingwistyczne nadawane danemu atrybutowi
age	natolatek, młody, w średnim wieku, w sile wieku, stary
d_age	brak, niewielka, mała, średnia, znacząca, spora
importance_same_race	nieistotne, mało istotne, średnio ważne, znaczące, ważne
importance_same_religion	nieistotne, mało istotne, średnio ważne, znaczące, ważne
pref_o_intelligence	głupi, mało inteligentny, przeciętny, inteligentny, geniusz
pref_o_ambitious	nieambitny, średnio ambitny, ambitny
tvsports	niezainteresowany, obojętny, średnio zainteresowany, zainteresowany, pasjonat
expected_num_interested_in_me	brak, niewiele, mało, kilka, dużo, wiele
guess_prob_liked	brak, niewielka, mała, znacząca, spora
funny	nudny, trochę nudny, przeciętnie zabawny, zabawny
sincere	fałszywy, nieszczerzy, szczerzy

Tabela 2. Tabela przedstawiająca zwyczajowe wartości lingwistyczne wybranych atrybutów przykładowego rekordu zbioru danych SpeedDating [4]

Baza danych została zrealizowana w PostgreSQL 14 [5]. Poniższy zrzut ekranu przedstawia część bazy danych w programie PgAdmin 4 [6].

hat_size	name	gender	age	age_o	age_o2	age_o3	age_o4	age_o5	race	race_o	same_race	importance_same_race	importance
1	0	1 female	21	27				6 [4-6]	Asian/Pacific Islander/Asian-American	European/Caucasian-American	0 2	4	4
2	0	1 female	21	22				1 [0-1]	Asian/Pacific Islander/Asian-American	European/Caucasian-American	0 2	4	4
3	1	1 female	21	22				1 [0-1]	Asian/Pacific Islander/Asian-American	Asian/Pacific Islander/Asian-American	1 2	4	4
4	0	1 female	21	23				2 [2-3]	Asian/Pacific Islander/Asian-American	European/Caucasian-American	0 2	4	4
5	0	1 female	21	24				3 [2-3]	Asian/Pacific Islander/Asian-American	Latino/Hispanic American	0 2	4	4
6	0	1 female	21	25				4 [4-6]	Asian/Pacific Islander/Asian-American	European/Caucasian-American	0 2	4	4
7	0	1 female	21	30				9 [7-17]	Asian/Pacific Islander/Asian-American	European/Caucasian-American	0 2	4	4
8	1	1 female	21	27				6 [4-6]	Asian/Pacific Islander/Asian-American	European/Caucasian-American	0 2	4	4
9	0	1 female	21	28				7 [7-9]	Asian/Pacific Islander/Asian-American	European/Caucasian-American	0 2	4	4
10	0	1 female	21	24				3 [2-3]	Asian/Pacific Islander/Asian-American	European/Caucasian-American	0 2	4	4
11	0	1 female	24	27				3 [2-3]	European/Caucasian-American	European/Caucasian-American	1 2	5	5
12	0	1 female	24	22				2 [2-3]	European/Caucasian-American	European/Caucasian-American	1 2	5	5
13	0	1 female	24	22				2 [2-3]	European/Caucasian-American	Asian/Pacific Islander/Asian-American	0 2	5	5
14	0	1 female	24	23				1 [0-1]	European/Caucasian-American	European/Caucasian-American	1 2	5	5
15	0	1 female	24	24				0 [0-1]	European/Caucasian-American	Latino/Hispanic American	0 2	5	5
16	0	1 female	24	25				1 [0-1]	European/Caucasian-American	European/Caucasian-American	1 2	5	5
17	0	1 female	24	30				6 [4-6]	European/Caucasian-American	European/Caucasian-American	1 2	5	5
18	0	1 female	24	27				3 [2-3]	European/Caucasian-American	European/Caucasian-American	1 2	5	5
19	0	1 female	24	28				4 [4-6]	European/Caucasian-American	European/Caucasian-American	1 2	5	5
20	0	1 female	24	24				0 [0-1]	European/Caucasian-American	European/Caucasian-American	1 2	5	5
21	1	1 female	25	27				2 [2-3]	European/Caucasian-American	European/Caucasian-American	1 8	4	4
22	1	1 female	25	22				3 [2-3]	European/Caucasian-American	European/Caucasian-American	1 8	4	4

Rysunek 1. Zrzut ekranu programu PgAdmin 4 przedstawiający pierwsze kilka rzędów danych [4].

## 2.2. Zmienne lingwistyczne (atrybuty/własności obiektów)

Poniżej zostały przedstawione wzory i wykresy przedstawiające poszczególne zmienne lingwistyczne.  $\mu_z(x)$  oznacza wartość funkcji przynależności zmiennej lingwistycznej  $z$  dla danego atrybutu zbioru danych zależnie od wartości  $x$ :

1. age

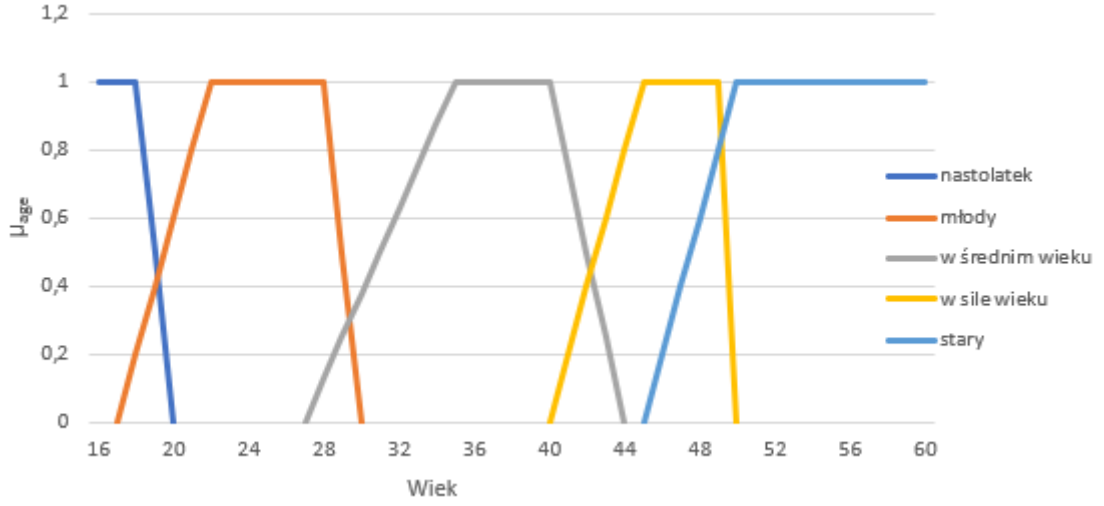
$$\mu_{nastolatek} = \begin{cases} 1 & \text{dla } x \in [16, 18] \\ -0.25x + 5 & \text{dla } x \in [18, 20] \end{cases} \quad (1)$$

$$\mu_{mody} = \begin{cases} 0.2x - 3.4 & \text{dla } x \in [17, 22] \\ 1 & \text{dla } x \in [23, 28] \\ -0.5x + 15 & \text{dla } x \in [28, 30] \end{cases} \quad (2)$$

$$\mu_{w\_rednim\_wieku} = \begin{cases} 0.125x - 3.375 & \text{dla } x \in [28, 34] \\ 1 & \text{dla } x \in [35, 40] \\ -0.25x + 11 & \text{dla } x \in [40, 44] \end{cases} \quad (3)$$

$$\mu_{w\_sile\_wieku} = \begin{cases} 0.2x - 8 & \text{dla } x \in [40, 45] \\ 1 & \text{dla } x \in [45, 49] \\ -x + 50 & \text{dla } x \in [49, 50] \end{cases} \quad (4)$$

$$\mu_{stary} = \begin{cases} 0.2x - 9 & \text{dla } x \in [45, 50] \\ 1 & \text{dla } x \in [50, 65] \end{cases} \quad (5)$$



Rysunek 2. Wykres funkcji przynależności dla atrybutu *age*.

2. d\_age

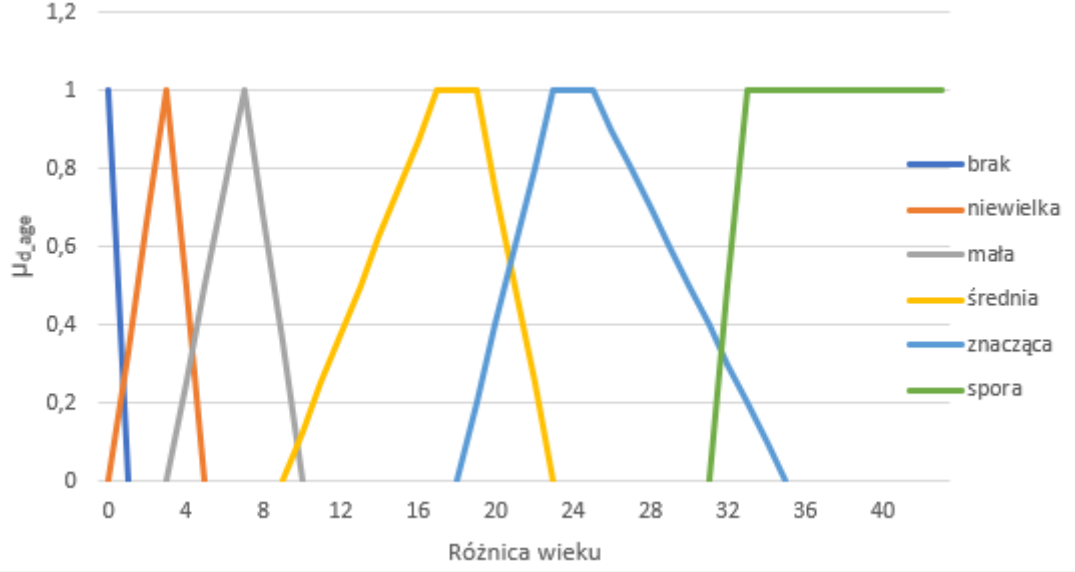
$$\mu_{brak} = \begin{cases} 1 & \text{dla } x = 0 \\ -x + 1 & \text{dla } x \in [0, 1] \end{cases} \quad (6)$$

$$\mu_{niewielka} = \begin{cases} 0.33x - 0.33 & \text{dla } x \in [1, 3] \\ 1 & \text{dla } x = 3 \\ -0.5x + 3 & \text{dla } x \in [3, 5] \end{cases} \quad (7)$$

$$\mu_{maa} = \begin{cases} 0.25x - 0.75 & \text{dla } x \in [3, 7] \\ 1 & \text{dla } x = 7 \\ -0.33x + 3.33 & \text{dla } x \in [7, 10] \end{cases} \quad (8)$$

$$\mu_{rednia} = \begin{cases} 0.125x - 1.125 & \text{dla } x \in [9, 17] \\ 1 & \text{dla } x \in [17, 19] \\ -0.25x + 5.75 & \text{dla } x \in [19, 23] \end{cases} \quad (9)$$

$$\mu_{znaczca} = \begin{cases} 0.2x - 3 & \text{dla } x \in [15, 19] \\ 1 & \text{dla } x \in [19, 22] \\ -0.1x + 3.2 & \text{dla } x \in [22, 32] \end{cases} \quad (10)$$



Rysunek 3. Wykres funkcji przynależności dla atrybutu  $d\_age$ .

### 3. importance\_same\_race

$$\mu_{\text{nieistotne}} = \begin{cases} 1 & \text{dla } x = 0 \\ -x + 1 & \text{dla } x \in [0, 1] \end{cases} \quad (11)$$

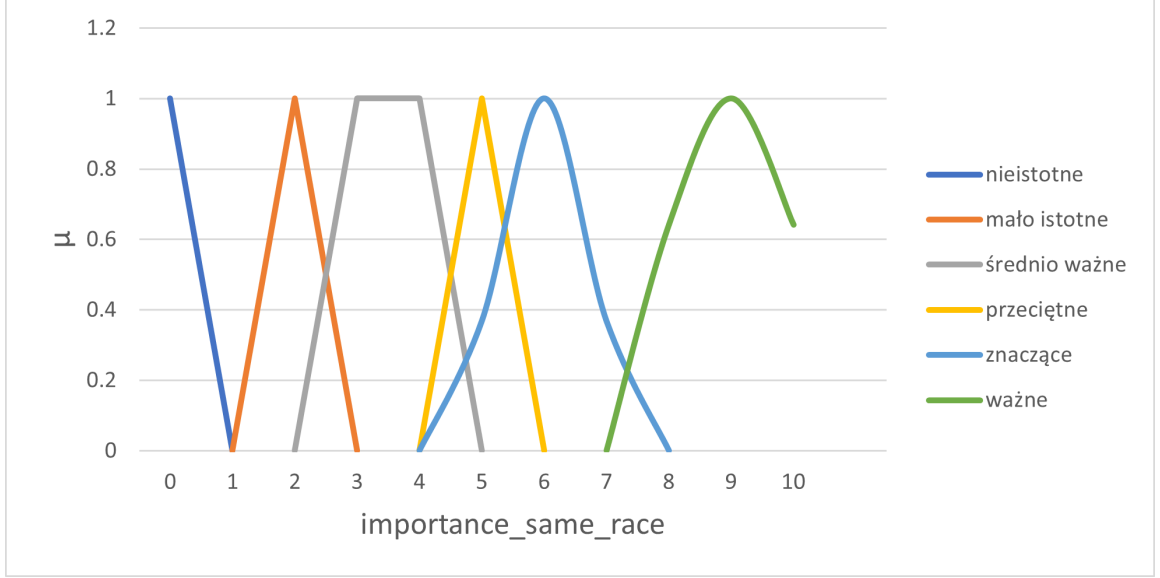
$$\mu_{\text{mało istotne}} = \begin{cases} x - 1 & \text{dla } x \in [1, 2] \\ 1 & \text{dla } x = 2 \\ -x + 3 & \text{dla } x \in [2, 3] \end{cases} \quad (12)$$

$$\mu_{\text{średnio ważne}} = \begin{cases} x - 2 & \text{dla } x \in [2, 3] \\ 1 & \text{dla } x \in [3, 4] \\ -x + 5 & \text{dla } x \in [4, 5] \end{cases} \quad (13)$$

$$\mu_{\text{przeciętne}} = \begin{cases} x - 4 & \text{dla } x \in [4, 5] \\ 1 & \text{dla } x = 5 \\ -x + 6 & \text{dla } x \in [5, 6] \end{cases} \quad (14)$$

$$\mu_{\text{znaczące}} = e^{-(x-6)^2}, \text{ dla } x \in [5, 7] \quad (15)$$

$$\mu_{\text{ważne}} = e^{-\left(\frac{x-9}{1.5}\right)^2}, \text{ dla } x \in [7, 10] \quad (16)$$



Rysunek 4. Wykres funkcji przynależności dla atrybutu *importance\_same\_race*.

#### 4. importance\_same\_religion

$$\mu_{\text{nieistotne}} = \begin{cases} 1 & \text{dla } x = 0 \\ -x + 1 & \text{dla } x \in [0, 1] \end{cases} \quad (17)$$

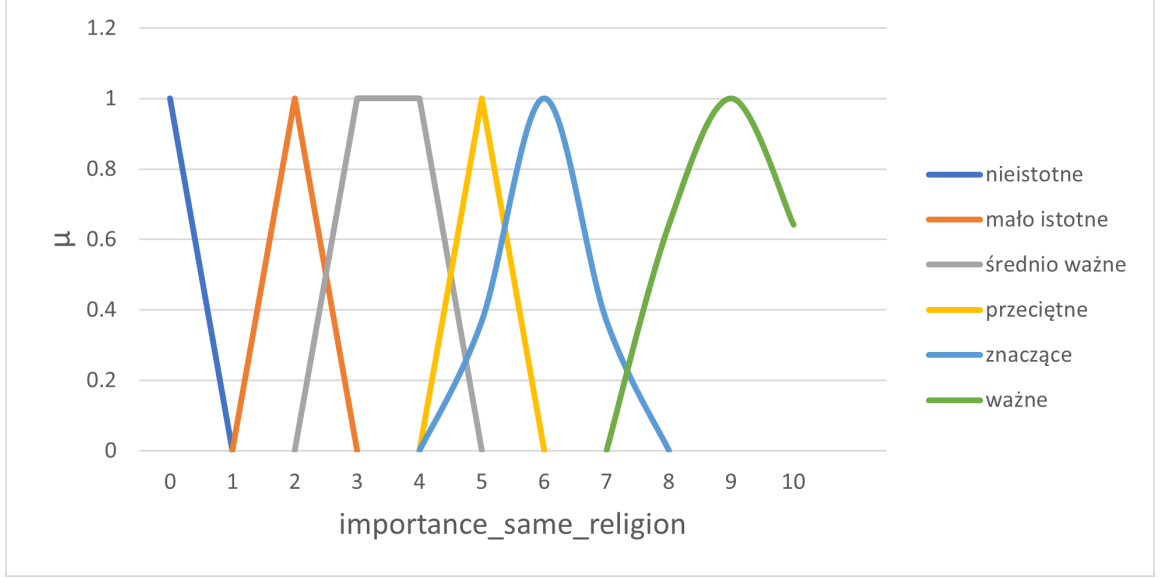
$$\mu_{\text{mało istotne}} = \begin{cases} x - 1 & \text{dla } x \in [1, 2] \\ 1 & \text{dla } x = 2 \\ -x + 3 & \text{dla } x \in [2, 3] \end{cases} \quad (18)$$

$$\mu_{\text{średnio ważne}} = \begin{cases} x - 2 & \text{dla } x \in [2, 3] \\ 1 & \text{dla } x \in [3, 4] \\ -x + 5 & \text{dla } x \in [4, 5] \end{cases} \quad (19)$$

$$\mu_{\text{przeciętne}} = \begin{cases} x - 4 & \text{dla } x \in [4, 5] \\ 1 & \text{dla } x = 5 \\ -x + 6 & \text{dla } x \in [5, 6] \end{cases} \quad (20)$$

$$\mu_{\text{znaczące}} = e^{-(x-6)^2}, \text{ dla } x \in [5, 7] \quad (21)$$

$$\mu_{\text{ważne}} = e^{-\left(\frac{x-9}{1.5}\right)^2}, \text{ dla } x \in [7, 10] \quad (22)$$



Rysunek 5. Wykres funkcji przynależności dla atrybutu *importance\_same\_religion*.

5. pref\_o\_intelligence

$$\mu_{gupi} = \begin{cases} 1 & \text{dla } x \in [0, 1] \\ -0.5x + 1.5 & \text{dla } x \in [1, 3] \end{cases} \quad (23)$$

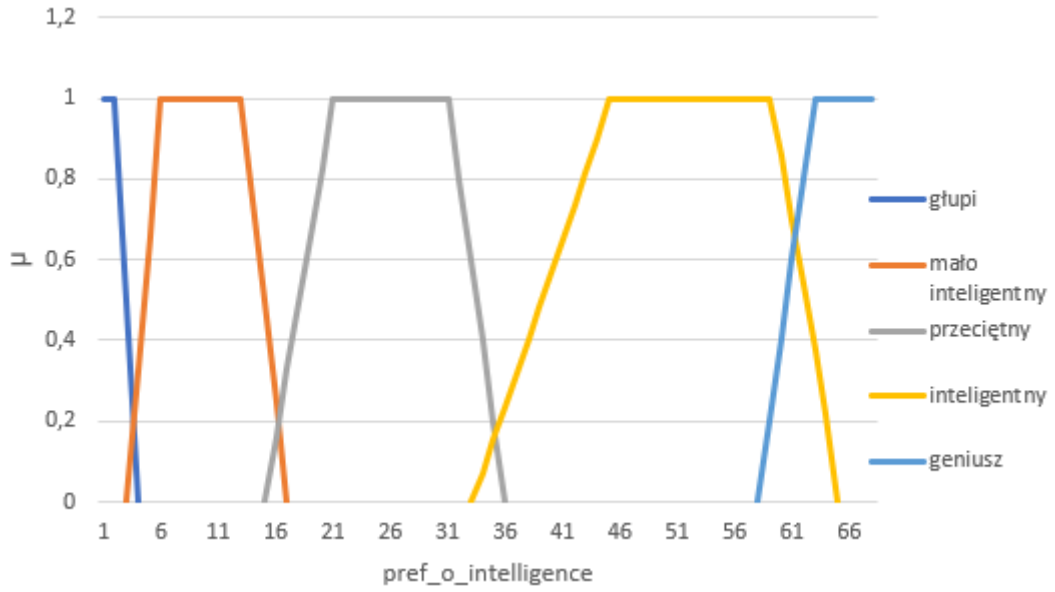
$$\mu_{mao_inteligentny} = \begin{cases} 0.33x - 0.66 & \text{dla } x \in [2, 5] \\ 1 & \text{dla } x \in [5, 12] \\ -0.25x + 4 & \text{dla } x \in [12, 16] \end{cases} \quad (24)$$

$$\mu_{przecitny} = \begin{cases} 0.166x - 2.33 & \text{dla } x \in [14, 20] \\ 1 & \text{dla } x \in [20, 30] \\ -0.2x + 7 & \text{dla } x \in [30, 35] \end{cases} \quad (25)$$

$$\mu_{inteligentny} = \begin{cases} 0.083x - 2.67 & \text{dla } x \in [32, 44] \\ 1 & \text{dla } x \in [44, 58] \\ -0.16x + 10.3 & \text{dla } x \in [58, 64] \end{cases} \quad (26)$$

$$\mu_{geniusz} = \begin{cases} 0.2x - 11.4 & \text{dla } x \in [57, 62] \\ 1 & \text{dla } x \in [62, 67] \end{cases} \quad (27)$$





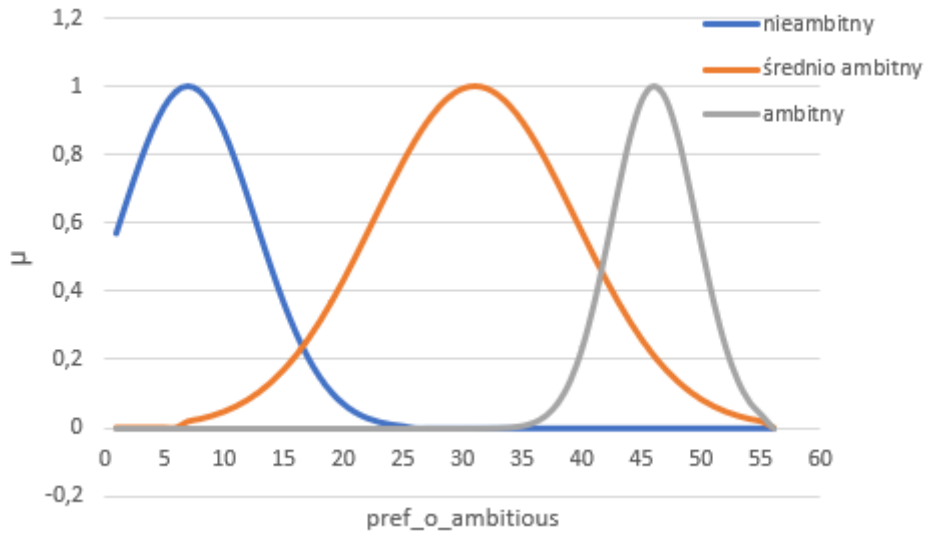
Rysunek 6. Wykres funkcji przynależności dla atrybutu *pref\_o\_intelligence*.

#### 6. pref\_o\_ambitious

$$\mu_{nieambitny} = e^{-\left(\frac{x-6}{8}\right)^2}, \text{ dla } x \in [0, 24] \quad (28)$$

$$\mu_{rednio_ambitny} = e^{-\left(\frac{x-30}{12}\right)^2}, \text{ dla } x \in [5, 54] \quad (29)$$

$$\mu_{ambitny} = e^{-\left(\frac{x-45}{5}\right)^2}, \text{ dla } x \in [30, 55] \quad (30)$$



Rysunek 7. Wykres funkcji przynależności dla atrybutu *pref\_o\_ambitious*.

#### 7. tvsports

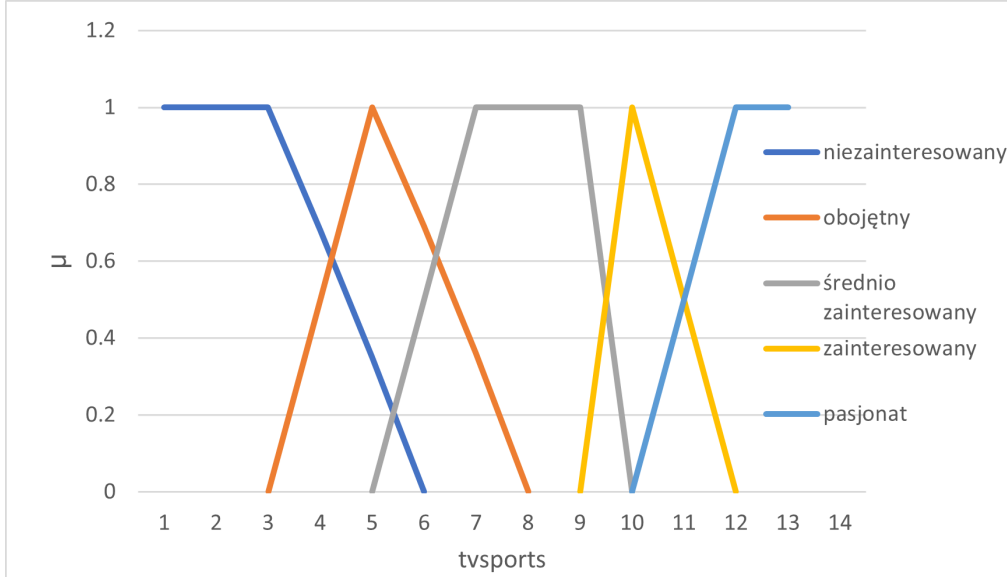
$$\mu_{niezainteresowany} = \begin{cases} 1 & \text{dla } x \in [0, 2] \\ -0.33x + 1.67 & \text{dla } x \in [2, 5] \end{cases} \quad (31)$$

$$\mu_{\text{obożtny}} = \begin{cases} 0.5x - 1 & \text{dla } x \in [2, 4] \\ 1 & \text{dla } x = 4 \\ -0.33x + 2.34 & \text{dla } x \in [4, 7] \end{cases} \quad (32)$$

$$\mu_{\text{rednio\_zainteresowany}} = \begin{cases} 0.5x - 2 & \text{dla } x \in [4, 6] \\ 1 & \text{dla } x \in [6, 8] \\ -x + 9 & \text{dla } x \in [8, 9] \end{cases} \quad (33)$$

$$\mu_{\text{zainteresowany}} = \begin{cases} x - 8 & \text{dla } x \in [8, 9] \\ 1 & \text{dla } x = 9 \\ -0.5x + 5.5 & \text{dla } x \in [9, 11] \end{cases} \quad (34)$$

$$\mu_{\text{pasjonat}} = \begin{cases} 0.5x - 4.5 & \text{dla } x \in [9, 11] \\ 1 & \text{dla } x \in [11, 12] \end{cases} \quad (35)$$



Rysunek 8. Wykres funkcji przynależności dla atrybutu *tvsports*.

8. expected\_num\_interested\_in\_me

$$\mu_{\text{brak}} = \begin{cases} 1 & \text{dla } x \in [0, 1] \\ 0.5x + 1.5 & \text{dla } x \in [1, 3] \end{cases} \quad (36)$$

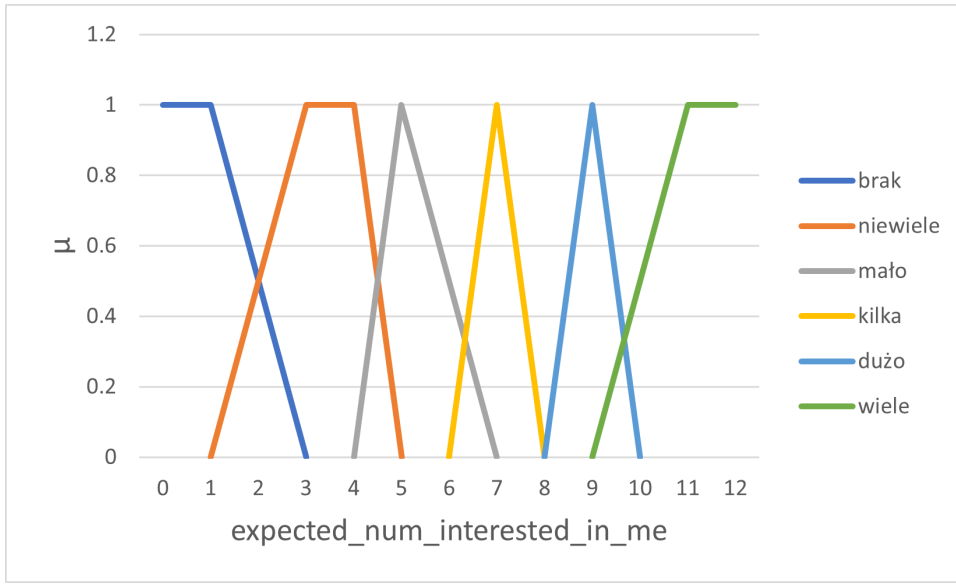
$$\mu_{\text{niewiele}} = \begin{cases} 0.5x - 0.5 & \text{dla } x \in [1, 3] \\ 1 & \text{dla } x \in [3, 4] \\ -x + 5 & \text{dla } x \in [4, 6] \end{cases} \quad (37)$$

$$\mu_{\text{mao}} = \begin{cases} x - 4 & \text{dla } x \in [4, 5] \\ 1 & \text{dla } x = 5 \\ 0.5x + 3.5 & \text{dla } x \in [5, 7] \end{cases} \quad (38)$$

$$\mu_{kilka} = \begin{cases} x - 6 & \text{dla } x \in [6, 7] \\ 1 & \text{dla } x = 7 \\ -x + 4 & \text{dla } x \in [7, 8] \end{cases} \quad (39)$$

$$\mu_{duo} = \begin{cases} x + 8 & \text{dla } x \in [7, 9] \\ 1 & \text{dla } x = 9 \\ -x + 10 & \text{dla } x \in [9, 11] \end{cases} \quad (40)$$

$$\mu_{wiele} = \begin{cases} 0.5x - 4.5 & \text{dla } x \in [8, 11] \\ 1 & \text{dla } x \in [11, 12] \end{cases} \quad (41)$$



Rysunek 9. Wykres funkcji przynależności dla atrybutu *expected\_num\_interested\_in\_me*.

9. guess\_prob\_liked

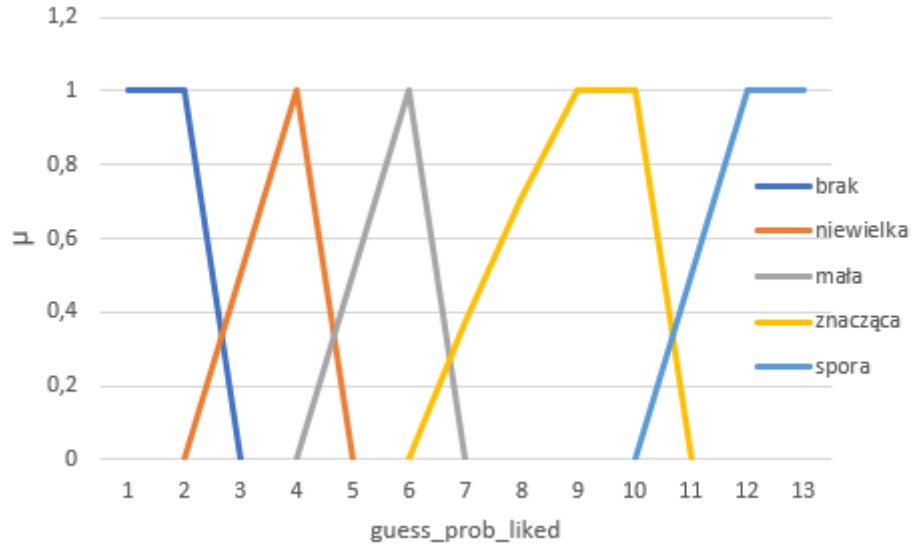
$$\mu_{brak} = \begin{cases} 1 & \text{dla } x \in [0, 1] \\ -x + 2 & \text{dla } x \in [1, 2] \end{cases} \quad (42)$$

$$\mu_{niewielka} = \begin{cases} 0.5x - 0.5 & \text{dla } x \in [1, 3] \\ 1 & \text{dla } x = 3 \\ -x + 4 & \text{dla } x \in [3, 4] \end{cases} \quad (43)$$

$$\mu_{maa} = \begin{cases} 0.5x - 1.5 & \text{dla } x \in [3, 5] \\ 1 & \text{dla } x = 5 \\ -x + 6 & \text{dla } x \in [5, 6] \end{cases} \quad (44)$$

$$\mu_{znaczaca} = \begin{cases} 0.34x - 1.67 & \text{dla } x \in [5, 8] \\ 1 & \text{dla } x \in [8, 9] \\ -x + 9 & \text{dla } x \in [9, 10] \end{cases} \quad (45)$$

$$\mu_{spora} = \begin{cases} 0.5x - 4.5 & \text{dla } x \in [9, 11] \\ 1 & \text{dla } x \in [11, 12] \end{cases} \quad (46)$$



Rysunek 10. Wykres funkcji przynależności dla atrybutu *guess\_prob\_liked*.

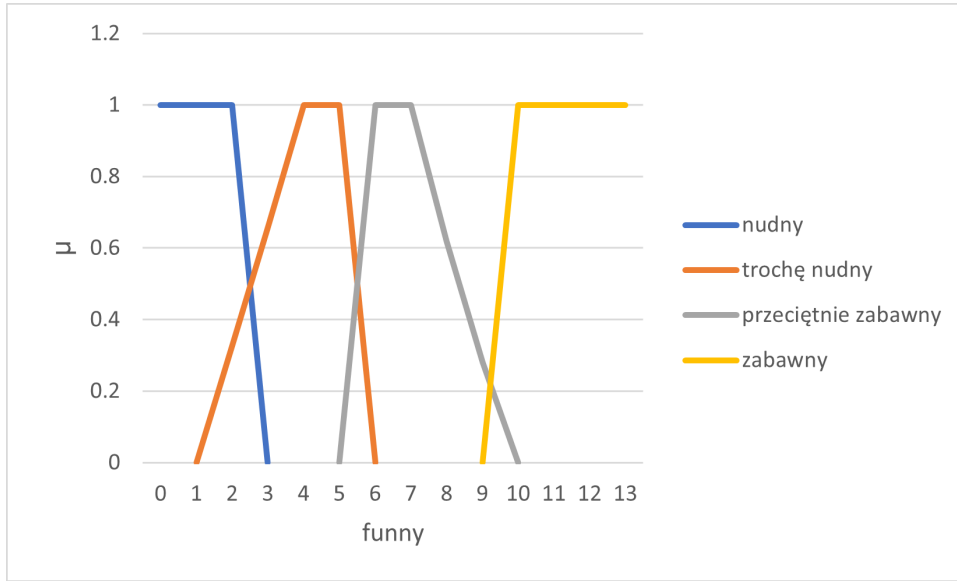
10. funny

$$\mu_{nudny} = \begin{cases} 1 & \text{dla } x \in [0, 2] \\ -0.5x - 1.5 & \text{dla } x \in [2, 3] \end{cases} \quad (47)$$

$$\mu_{troch_nudny} = \begin{cases} 0.33x - 0.34 & \text{dla } x \in [1, 4] \\ 1 & \text{dla } x \in [4, 5] \\ -x + 6 & \text{dla } x \in [5, 6] \end{cases} \quad (48)$$

$$\mu_{przecitnie_zabawny} = \begin{cases} x - 5 & \text{dla } x \in [5, 6] \\ 1 & \text{dla } x \in [6, 7] \\ -0.34x - 3.34 & \text{dla } x \in [7, 10] \end{cases} \quad (49)$$

$$\mu_{zabawny} = \begin{cases} x - 9 & \text{dla } x \in [9, 10] \\ 1 & \text{dla } x \in [10, 13] \end{cases} \quad (50)$$



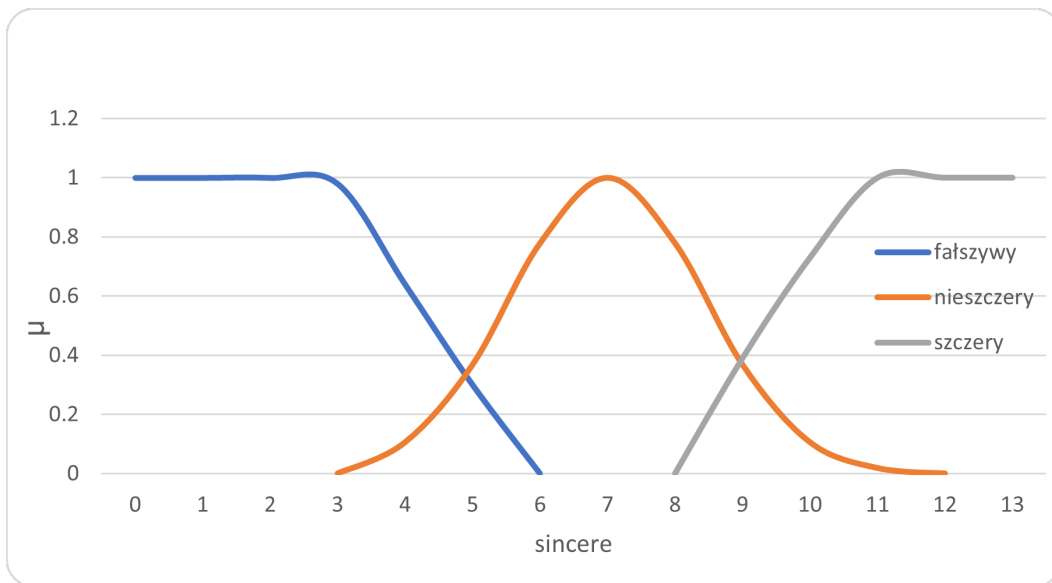
Rysunek 11. Wykres funkcji przynależności dla atrybutu *funny*.

11. sincere

$$\mu_{faszywy} = \begin{cases} 1 & \text{dla } x \in [0, 2] \\ -0.34x + 2 & \text{dla } x \in [2, 6] \end{cases} \quad (51)$$

$$\mu_{nieszczery} = e^{-\left(\frac{x-7}{2}\right)^2}, \text{ dla } x \in [4, 11] \quad (52)$$

$$\mu_{szczerzy} = \begin{cases} 0.34x - 2.67 & \text{dla } x \in [8, 11] \\ 1 & \text{dla } x \in [11, 13] \end{cases} \quad (53)$$



Rysunek 12. Wykres funkcji przynależności dla atrybutu *sincere*.

### 2.3. Kwantyfikatory lingwistyczne (liczności obiektów)

Poniżej zostały przedstawione wzory i wykresy przedstawiające poszczególne kwantyfikatory lingwistyczne, zarówno względne jak i absolutne.  $\mu_z(x)$  oznacza wartość funkcji przynależności kwantyfikatora lingwistycznego  $z$  dla danego atrybutu zbioru danych zależnie od wartości  $x$ :

$$\mu_{\text{Prawie żadne}} = \begin{cases} 1 & \text{dla } x \in [0, 50] \\ 2 - 0.02x & \text{dla } x \in [50, 100] \end{cases} \quad (54)$$

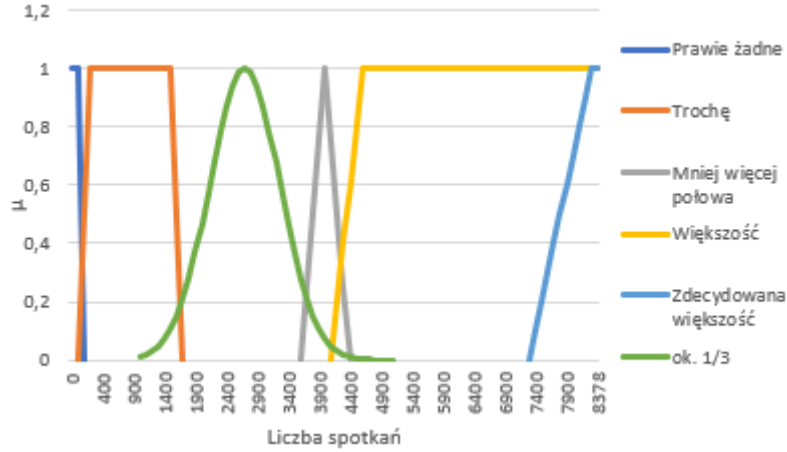
$$\mu_{\text{Trochę}} = \begin{cases} 0.005x & \text{dla } x \in [0, 200] \\ 1 & \text{dla } x \in [200, 1500] \\ -0.005x & \text{dla } x \in [1500, 1700] \end{cases} \quad (55)$$

$$\mu_{\text{ok. } 1/3} = e^{-\left(\frac{x-2700}{800}\right)^2}, \text{ dla } x \in [1000, 5000] \quad (56)$$

$$\mu_{\text{Mniej więcej połowa}} = \begin{cases} 0.0025x - 9.4725 & \text{dla } x \in [3789, 4189] \\ -0.0025x + 11.4725 & \text{dla } x \in [4189, 4589] \end{cases} \quad (57)$$

$$\mu_{\text{Większość}} = \begin{cases} 0.002x - 8.2 & \text{dla } x \in [4100, 4600] \\ 1 & \text{dla } x \in [4600, 8378] \end{cases} \quad (58)$$

$$\mu_{\text{Zdecydowana większość}} = \begin{cases} 0.001x - 7.3 & \text{dla } x \in [7300, 8300] \\ 1 & \text{dla } x \in [8300, 8378] \end{cases} \quad (59)$$



Rysunek 13. Wykres funkcji przynależności dla kwantyfikatorów względnych użytych w zadaniu

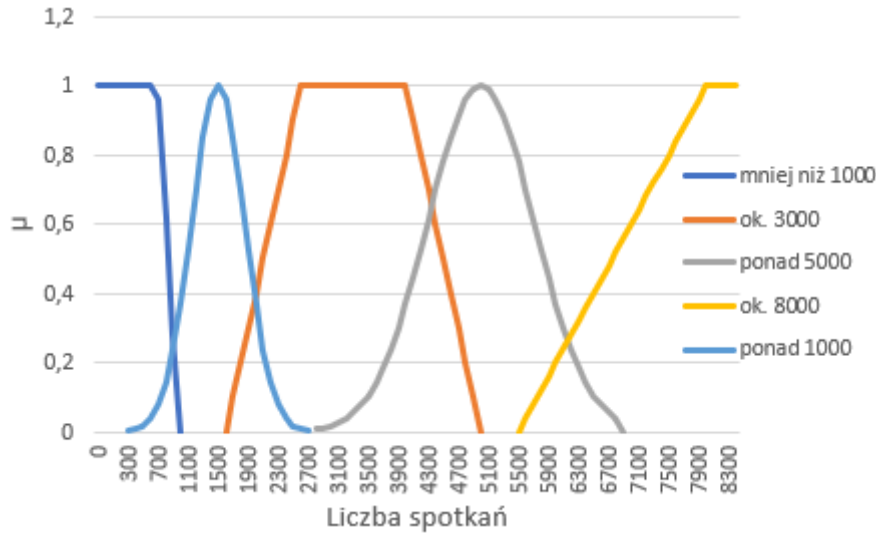
$$\mu_{\text{mniej niż 1000}} = \begin{cases} 1 & \text{dla } x \in [0, 600] \\ 0.0034x - 3.34 & \text{dla } x \in [600, 1000] \end{cases} \quad (60)$$

$$\mu_{\text{ponad 1000}} = e^{-\left(\frac{x-1500}{500}\right)^2}, \text{ dla } x \in [300, 2700] \quad (61)$$

$$\mu_{\text{ok. 3000}} = \begin{cases} 0.001x - 1.6 & \text{dla } x \in [1600, 2600] \\ 1 & \text{dla } x \in [2600, 4000] \\ 0.001x + 5 & \text{dla } x \in [4000, 5000] \end{cases} \quad (62)$$

$$\mu_{\text{ponad 5000}} = e^{-\left(\frac{x-5000}{1000}\right)^2}, \text{ dla } x \in [2800, 6900] \quad (63)$$

$$\mu_{\text{ok. 8000}} = \begin{cases} 0.0004x - 2.2 & \text{dla } x \in [5500, 8000] \\ 1 & \text{dla } x \in [8000, 8400] \end{cases} \quad (64)$$



Rysunek 14. Wykres funkcji przynależności dla kwantyfikatorów absolutnych użytych w zadaniu

### 3. Narzędzia obliczeniowe: wybór/implementacja. Diagram UML pakietu obliczeń rozmytych i generatora podsumowań. Instrukcja użytkownika

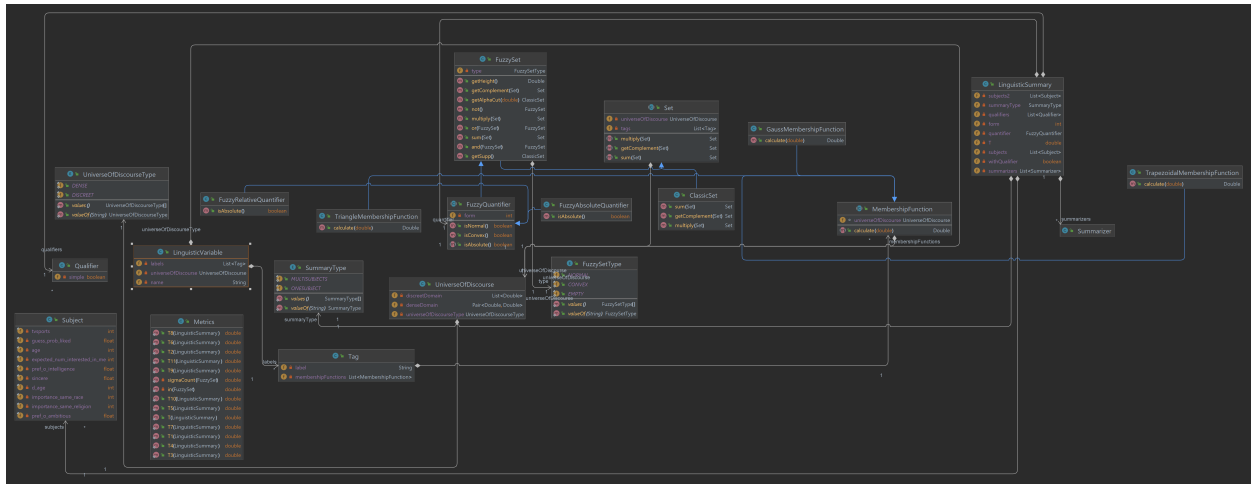
Aplikacja została napisana w technologii Java w wersji 17, z wykorzystaniem biblioteki JavaFX [7] w celu stworzenia prostego okienkowego interfejsu użytkownika.

Implementacja została podzielona na dwa moduły warstwy logiki i widoku, wykorzystany został wzorzec projektowy Model View Controller. Z kolei część odpowiedzialna za logikę aplikacji składa się z poszczególnych pakietów:

1. **functions** - pakiet reprezentujący funkcje przynależności (klasa abstrakcyjna `MembershipFunction` i rozszerzające ją klasy) oraz przestrzeń rozważań (klasa `UniverseOfDiscourse`).
2. **fuzzy** - pakiet reprezentujący poszczególne elementy podsumowania, jak i samo podsumowanie lingwistyczne oraz metryki użyte to obliczeń jakości wykonywanych operacji. Zawiera między innymi klasy: `Metrics` -

reprezentującą metryki oraz LinguisticVariable reprezentującą zmienną lingwistyczną

3. set - pakiet reprezentujący obiekt zbioru, zarówno rozumianego w sensie klasycznym jak i rozmytym. Zawiera klasę abstrakcyjną Set oraz rozszerzające ją klasy ClassicSet oraz FuzzySet.
4. quantifier - pakiet odpowiedzialny za reprezentacje kwantyfikatorów absolutnych (klasa FuzzyAbsoluteQuantifier) jak i względnych (klasa FuzzyRelativeQuantifier) oraz etykiet (klasa Tag).



Rysunek 15. Diagram UML modułu logiki

Poniżej zamieszczono poszczególne widoki GUI aplikacji:

Rysunek 16. Widok GUI aplikacji dla użytkownika niezaawansowanego pozwalający na generowanie podsumowań poprzez wybranie między innymi kwantyfikatora, podmiotu oraz określenie wag dla poszczególnych metryk



The screenshot shows a software window with three tabs at the top: 'Niezaawansowany', 'Zaawansowany' (selected), and 'Wyniki'. In the 'Zaawansowany' tab, there are three sub-tabs: 'Sumatory', 'Kwantyfikatory' (selected), and 'Kwalifikatory'. The main area contains the following controls:

- 'Wybierz typ obiektu:' followed by a dropdown menu.
- 'Wybierz zmienną:' followed by a dropdown menu.
- 'Podaj etykiety:' followed by a text input field and a 'Dodaj' button.
- 'Funkcja przynależności:' followed by a dropdown menu and a 'Usuń' button.
- 'Podaj współrzędne punktów:' followed by two columns of input fields. The first column has fields for X1, X2, and X3. The second column has fields for Y1 and Y2. A 'Dodaj' button is located below these fields.

Rysunek 17. Widok GUI aplikacji dla użytkownika zaawansowanego pozwalający na definiowanie własnych etykiet i funkcji przynależności dla nowych kwantyfikatorów, sumaryzatorów i kwalifikatorów

The screenshot shows the 'Wyniki' tab selected. The table header includes 'Podsumowanie', 'T2', 'T3', 'T4', 'T5', 'T6', 'T7', 'T8', 'T9', 'T10', 'T11', 'T', and 'Zapis'. The table body is empty, displaying the message 'No content in table'. At the bottom left, there is a 'Zapisz do pliku' button.

Podsumowanie	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T	Zapis
No content in table												

Rysunek 18. Widok GUI aplikacji przedstawiający wyniki podsumowań

## 4. Jednopođmiotowe podsumowania lingwistyczne.

### Miary jakości, podsumowanie optymalne

Wyniki kolejnych eksperymentów wg punktów 2.-4. opisu projektu 2. Listy podsumowań jednopođmiotowych i tabele/rankingi podsumowań dla danych atrybutów obowiązkowe i dokładnie opisane w „captions” (tytułach), konieczny opis kolumn i wierszy tabel. Dla każdego podsumowania podane miary jakości oraz miara jakości podsumowania optymalnego. **Wzorów podsumowań ani miar nie należy przepisywać ani cytować, wystarczy podać literaturę, ale należy skomentować co oznaczają i jaką informacje niosą wybrane miary w wybranych przypadkach.** Sekcja uzupełniona jako efekt zadania Tydzień 11 wg Harmonogramu Zajęć na WIKAMP KSR.

## 5. Wielopođmiotowe podsumowania lingwistyczne i ich miary jakości

Wyniki kolejnych eksperymentów wg punktów 2.-4. opisu projektu 2. Uzasadnienie i metoda podziału zbioru danych na rozłączne pođmioty. Listy podsumowań wielopođmiotowych i tabele/rankingi podsumowań dla danych atrybutów obowiązkowe i dokładnie opisane w „captions” (tytułach), konieczny opis kolumn i wierszy tabel. **Wzorów podsumowań ani miar nie należy przepisywać ani cytować, wystarczy podać literaturę, ale należy skomentować co oznaczają i jaką informacje niosą wybrane miary w wybranych przypadkach.** Konieczne uwzględnienie wszystkich 4-ch form podsumowań wielopođmiotowych.

\*\* Możliwe sformułowanie zagadnienia wielopođmiotowego podsumowania optymalnego \*\*.

\*\* Ewentualne wyniki realizacji punktu „na ocenę 5.0” wg opisu Projektu 2. i ich porównanie do wyników z części obowiązkowej \*\*.

Sekcja uzupełniona jako efekt zadania Tydzień 12 wg Harmonogramu Zajęć na WIKAMP KSR.

## 6. Dyskusja, wnioski

Dokładne interpretacje uzyskanych wyników w zależności od parametrów klasyfikacji opisanych w punktach 3.-4 opisu Projektu 2. Szczególnie istotne są wnioski o charakterze uniwersalnym, istotne dla podobnych zadań. Omówić i wyjaśnić napotkane problemy (jeśli były). Każdy wniosek/problem powinien mieć poparcie w przeprowadzonych eksperymentach (odwołania do konkretnych wyników: tabel i miar jakości). Ocena które wybrane kwantyfikatory, sumaryzatory, kwalifikatory i/lub ich miary jakości mają małe albo duże znaczenie dla wiarygodności i jakości otrzymanych agregacji/podsumowań. Dla końcowej oceny jest to najważniejsza sekcja sprawozdania, gdyż prezen-

tuje poziom zrozumienia rozwiązywanego problemu.

\*\* Możliwości kontynuacji prac w obszarze logiki rozmytej i wnioskowania rozmytego, zwłaszcza w kontekście pracy inżynierskiej, magisterskiej, naukowej, itp. \*\*

**Sekcja uzupełniona jako efekt zadań Tydzień 11 i Tydzień 12 wg Harmonogramu Zajęć na WIKAMP KSR.**

## **7. Braki w realizacji projektu 2.**

Wymienić wg opisu Projektu 2. wszystkie niezrealizowane obowiązkowe elementy projektu, ewentualnie podać merytoryczne (ale nie czasowe) przyczyny tych braków.

## **Literatura**

- [1] A. Niewiadomski, Zbiory rozmyte typu 2. Zastosowania w reprezentowaniu informacji. Seria „Problemy współczesnej informatyki” pod redakcją L. Rutkowskiego. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2019.
- [2] S. Zadrozny, Zapytania nieprecyzyjne i lingwistyczne podsumowania baz danych, EXIT, 2006, Warszawa
- [3] A. Niewiadomski, Methods for the Linguistic Summarization of Data: Applications of Fuzzy Sets and Their Extensions, Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2008.
- [4] SpeedDating Dataset, <https://www.openml.org/search?type=data&sort=runs&status=active&id=40536>
- [5] strona internetowa PostgreSQL, <https://www.postgresql.org>
- [6] strona internetowa PgAdmin 4, <https://www.pgadmin.org>
- [7] JavaFX, <https://openjfx.io>

Literatura zawiera wyłącznie źródła recenzowane i/lub o potwierdzonej wiarygodności, możliwe do weryfikacji i cytowane w sprawozdaniu.