

Data oddania: _____

Ocena: _____

Julia Szymańska 224441
Przemysław Zdrzałik 224466

Projekt 2. Podsumowania lingwistyczne relacyjnych baz danych

1. Cel

Celem projektu jest stworzenie aplikacji pozwalającej na generowanie podsumowań lingwistycznych[1] w oparciu o kwantyfikatory rozmyte[2], co oznacza opisanie danych liczbowych ze zbioru danych[3] językiem quasi-naturalnym - pozornie naturalnym.

Przykładem podsumowania lingwistycznego[1] w formie

$$Q \ P \text{ jest } S \ [T] \quad (1)$$

jest: *Wiele wypadków jest przy ujemnej temperaturze $[0, 76]$* , gdzie Q jest kwantyfikatorem lingwistycznym, P podmiotem podsumowań, S sumaryzatorem, a $T[0, 1]$ stopniem prawdziwości.

Przykładem drugiego podsumowania lingwistycznego w formie:

$$Q \ P \text{ bedacych } W \text{ jest } S \ [T] \quad (2)$$

jest: *Wiele wypadków będących podczas deszczu, jest przy ujemnej temperaturze $[0.68]$* , gdzie Q jest kwantyfikatorem lingwistycznym, P podmiotem podsumowań, S sumaryzatorem, W kwantyfikatorem reprezentującym dodatkowe własności obiektów, a $T[0, 1]$ stopniem prawdziwości.

Analizowany zbiór danych zawiera liczbowe informacje o ponad 3 milionach wypadków samochodowych w 49 stanach Zjednoczonych Stanów Ameryki,

mających miejsce od lutego 2016 do grudnia 2020[3]. Zbiór danych składa się z 47 kolumn. W tym celu wykonania podsumowania lingwistycznego zostaną wykorzystane metody logiki rozmytej[4]. Logika rozmyta pozwala na opisanie wartości zapisanych językiem naturalnym za pomocą zrozumiałych określeń jak: mało, dużo, około połowy. W projekcie zostaną wykorzystane kwantyfikatory lingwistyczne względne takie jak: niewiele, około połowy oraz kwantyfikatory lingwistyczne absolutne takie jak: około jednego, około stu.

2. Charakterystyka podsumowywanej bazy danych













W programie został użyty zbiór danych[3] znajdujący się w pliku CSV, który został przekształcony w bazę danych.

Zbiór danych zawiera informacje o ponad 3 milionach wypadków samochodowych w 49 stanach Zjednoczonych Stanów Ameryki, mających miejsce od lutego 2016 do grudnia 2020. Spośród 47 kolumn znajdujących się w zbiorze danych, wybraliśmy następujące 11 kolumn:

- Czas rozpoczęcia - Start_Time - czas rozpoczęcia się wypadku w lokalnej strefie czasowej, przyjmuje wartości od 8 lutego 2016, do 31 grudnia 2020. Wartość kolumny zostanie zamieniona na wartość całkowitą oznaczającą liczbę sekund od początku 1970 roku.
- Czas zakończenia - End_Time - czas zakończenia się wypadku w lokalnej strefie czasowej, przyjmuje wartości od 8 lutego 2016, do 1 stycznia 2021. Wartość kolumny zostanie zamieniona na wartość całkowitą oznaczającą liczbę sekund od początku 1970 roku.
- Odległość - Distance - długość odcinka ulicy wyrażony w milach, na którego miał wpływ wypadek. Przyjmuje wartości zmiennoprzecinkowe od 0 do 334, gdzie zdecydowana większość danych mieści się w przedziale od 0.00 do 4.00.
- Temperatura - Temperature - temperatura powietrza wyrażona w Fahrenheit'ach, w momencie, gdy zdarzył się wypadek. Przyjmuje wartości zmiennoprzecinkowe od -16.00 do 104.00. Temperature można opisać jako bardzo zimną, zimną, umiarkowaną, ciepłą, bardzo ciepłą. Oczywiście jest to opis subiektywny.
- Temperatura odczuwalna - Wind_Chill - temperatura odczuwalna wyrażona w Fahrenheit'ach, w momencie, gdy zdarzył się wypadek. Przyjmuje wartości zmiennoprzecinkowe od -16.00 do 101.00. Temperaturę odczuwalną można opisać tak samo jak temperaturę.
- Wilgotność - Humidity - wilgotność powietrza wyrażona w procentach w momencie, gdy zdarzył się wypadek. Przyjmuje wartości zmiennoprzecinkowe od 4.00 do 100.00.
- Ciśnienie - Pressure - ciśnienie powietrza wyrażone w inches, w momencie, gdy zdarzył się wypadek. Przyjmuje wartości zmiennoprzecinkowe od 27.00 do 32.00. Ciśnienie można opisać jako wysokie, umiarkowane lub niskie.
- Widoczność - Visibility - widoczność wyrażona w milach, w momencie, gdy zdarzył się wypadek. Przyjmuje wartości zmiennoprzecinkowe od 0.00 do 12.00. Widoczność można opisać jako dobrą, ograniczoną, słabą.

- Prędkość wiatru - Wind_Speed - prędkość wiatru wyrażona w milach na godzinę, w momencie, gdy zdarzył się wypadek. Przyjmuje wartości zmiennoprzecinkowe od 0.00 do 40.00. Wiatr można opisać jako słaby, umiarkowany, silny.
- Ilość opadów - Precipitation - ilość opadów wyrażona w inches, w momencie, gdy zdarzył się wypadek. Jeśli opady nie występowały to kolumna przyjmuje wartość nan. Przyjmuje wartości zmiennoprzecinkowe od 0.00 do 0.50.

Atrybutom nadawane są opisane zwyczajowe wartości lingwistyczne ze względu na zwiększenie przystępności i ułatwienie szybkiego zrozumienia atrybutu przez człowieka, kiedy ten atrybut nie musi być dokładnie opisany. Przykładowo temperatura, mimo że zrozumiała dla człowieka w postaci liczbowej, jest łatwiejsza do szybszego zrozumienia w postaci tekstowej, a dla ludzi nie ma dużego znaczenia czy temperatura różni się o 1 czy 2 stopnie, wystarczy opisać ją słownie tak jak wcześniej podaliśmy jako bardzo zimną, zimną, umiarkowaną, ciepłą, bardzo ciepłą.

accidents	
 id	bigint
 severity	smallint
 start_time	timestamp with time zone
 end_time	timestamp with time zone
 distance	double precision
 temperature	double precision
 wind_chill	double precision
 humidity	double precision
 pressure	double precision
 visibility	double precision
 wind_speed	double precision
 precipitation	double precision

Powered by yFiles

Rysunek 1. Tabela reprezentująca omawiane dane wykonana w DBMS PostgreSQL

3. Atrybuty i liczności obiektów wyrażone zmiennymi lingwistycznymi

Poniżej zostaną przedstawione zmienne lingwistyczne[2] dla jedenastu atrybutów z bazy danych wraz z przypisanymi etykietami w formie funkcji przynależności oraz wzorów analitycznych.

3.1. Czas utrudnień w ruchu drogowym spowodowanych przez wypadek

Na podstawie znajdujących się w bazie danych pól Czas rozpoczęcia (Start.Time) oraz Czas zakończenia (End.Time) zostanie obliczony Czas utrudnień w ruchu drogowym (Duration) spowodowanych przez wypadek według wzoru:

$$Duration = End_Time - Start_Time \quad (3)$$

Przedstawienie Czasu utrudnień w ruchu drogowym (Duration) spowodowanych przez wypadek jako zmiennej lingwistycznej. Do zmiennej lingwistycznej zostały dopasowane etykiety: krótki, średni, długi.

$$\mu_{czasTrwaniaPonizejGodziny}(x) = \begin{cases} \frac{1-x}{1} & \text{dla } 0 < x \leq 1 \end{cases} \quad (4)$$

$$\mu_{czasTrwaniaOkoloDwochGodzin}(x) = \begin{cases} \frac{x}{2} & \text{dla } 0 < x \leq 2 \\ \frac{4-x}{2} & \text{dla } 2 < x \leq 4 \end{cases} \quad (5)$$

$$\mu_{czasTrwaniaOkoloCzterechGodzin}(x) = \begin{cases} \frac{x-2}{2} & \text{dla } 2 < x \leq 4 \\ \frac{6-x}{2} & \text{dla } 4 < x \leq 6 \end{cases} \quad (6)$$

$$\mu_{czasTrwaniaOkoloSzesciuGodzin}(x) = \begin{cases} \frac{x-4}{2} & \text{dla } 4 < x \leq 6 \\ \frac{8-x}{2} & \text{dla } 6 < x \leq 8 \end{cases} \quad (7)$$

$$\mu_{czasTrwaniaPonadSzescGodzin}(x) = \begin{cases} \frac{x-6}{2} & \text{dla } 6 < x \leq 8 \\ 1 & \text{dla } 8 \leq x \end{cases} \quad (8)$$

gdzie: $\mu_{czasTrwaniaPonizejGodziny}$, $\mu_{czasTrwaniaOkoloDwochGodzin}$, $\mu_{czasTrwaniaOkoloCzterechGodzin}$, $\mu_{czasTrwaniaOkoloSzesciuGodzin}$, $\mu_{czasTrwaniaPonadSzescGodzin}$ - funkcje przynależności, x - czas trwania wypadku.



Rysunek 2. Wykres funkcji przynależności zbiorów rozmytych ilustrujących wartości zmiennej lingwistycznej czas utrudnień w ruchu drogowym (Duration) spowodowanych przez wypadek.

3.2. Odległość

Przedstawienie odległości jako zmiennej lingwistycznej. Do zmiennej lingwistycznej zostały dopasowane etykiety: krótki, długi.

$$\mu_{\text{OdlegloscPonizejPolMili}}(x) = \begin{cases} 1 & \text{dla } x \leq 0.5 \end{cases} \quad (9)$$

$$\mu_{\text{OdlegloscOkoloJednejMili}}(x) = \begin{cases} x & \text{dla } 0 < x \leq 1 \\ 2 - x & \text{dla } 1 < x \leq 2 \end{cases} \quad (10)$$

$$\mu_{\text{OdlegloscOkoloTrzechMili}}(x) = \begin{cases} \frac{x-1}{2} & \text{dla } 1 < x \leq 3 \\ \frac{5-x}{2} & \text{dla } 3 < x \leq 5 \end{cases} \quad (11)$$

$$\mu_{\text{OdlegloscPonadTrzyMile}}(x) = \begin{cases} \frac{x-3}{2} & \text{dla } 3 < x \leq 5 \\ 1 & \text{dla } 5 \leq x \end{cases} \quad (12)$$

gdzie: $\mu_{\text{OdlegloscPonizejPolMili}}$, $\mu_{\text{OdlegloscOkoloJednejMili}}$, $\mu_{\text{OdlegloscOkoloTrzechMili}}$, $\mu_{\text{OdlegloscPonadTrzyMile}}$ - funkcje przynależności, x - odległość.



Rysunek 3. Wykres funkcji przynależności zbiorów rozmytych ilustrujących wartości zmiennej lingwistycznej odległości.

3.3. Tempertura i temperatura odczuwalna

Przedstawienie temperatury oraz temperatury odczuwalnej jako zmiennej lingwistycznej. Do zmiennej lingwistycznej zostały dopasowane etykiety: bardzo zimno, zimno, umiarkowanie, ciepło, bardzo ciepło.

$$\mu_{temperaturaBardzoZimno}(x) = \begin{cases} 1 & \text{dla } x \leq 14 \\ \frac{23-x}{9} & \text{dla } 14 < x \leq 23 \end{cases} \quad (13)$$

$$\mu_{temperaturaZimno}(x) = \begin{cases} \frac{x-14}{9} & \text{dla } 14 < x \leq 23 \\ 1 & \text{dla } 23 < x < 44 \\ \frac{54-x}{10} & \text{dla } 44 < x \leq 54 \end{cases} \quad (14)$$

$$\mu_{temperaturaUmiarkowanie}(x) = \begin{cases} \frac{x-44}{10} & \text{dla } 44 < x \leq 54 \\ 1 & \text{dla } 54 < x < 63 \\ \frac{71-x}{8} & \text{dla } 63 < x \leq 71 \end{cases} \quad (15)$$

$$\mu_{temperaturaCieplo}(x) = \begin{cases} \frac{x-63}{8} & \text{dla } 63 < x \leq 71 \\ 1 & \text{dla } 71 < x < 80 \\ \frac{90-x}{10} & \text{dla } 80 < x \leq 90 \end{cases} \quad (16)$$

$$\mu_{temperaturaBardzoCieplo}(x) = \begin{cases} \frac{x-80}{10} & \text{dla } 80 < x \leq 90 \\ 1 & \text{dla } 90 \leq x \end{cases} \quad (17)$$

gdzie: $\mu_{temperaturaBardzoZimna}$, $\mu_{temperaturaZimna}$, $\mu_{temperaturaUmiarkowana}$, $\mu_{temperaturaCiepła}$, $\mu_{temperaturaBardzoCiepła}$ - funkcje przynależności, x - temperatura, temperatura odczuwalna.



Rysunek 4. Wykres funkcji przynależności zbiorów rozmytych ilustrujących wartości zmiennej lingwistycznej temperatury.

3.4. Wilgotność

Przedstawienie wilgotności jako zmiennej lingwistycznej. Do zmiennej lingwistycznej zostały dopasowane etykiety: suche, umiarkowane, wilgotne.

$$\mu_{wilgotnoscBardzoSuche}(x) = \exp\left(\frac{-(x-4)^2}{128}\right) \quad (18)$$

$$\mu_{wilgotnoscSuche}(x) = \exp\left(\frac{-(x-25)^2}{128}\right) \quad (19)$$

$$\mu_{wilgotnoscUmiarkowane}(x) = \exp\left(\frac{-(x-50)^2}{128}\right) \quad (20)$$

$$\mu_{wilgotnoscWilgotne}(x) = \exp\left(\frac{-(x-75)^2}{128}\right) \quad (21)$$

$$\mu_{wilgotnoscBardzoWilgotne}(x) = \begin{cases} \frac{x-60}{10} & \text{dla } 60 < x \leq 70 \\ 1 & \text{dla } 70 < x < 80 \\ \frac{90-x}{10} & \text{dla } 80 < x \leq 90 \end{cases} \quad (22)$$

gdzie: $\mu_{wilgotnoscBardzoSuche}$, $\mu_{wilgotnoscSuche}$, $\mu_{wilgotnoscUmiarkowane}$, $\mu_{wilgotnoscWilgotne}$, $\mu_{wilgotnoscBardzoWilgotne}$ - funkcje przynależności, x - wilgotność.



Rysunek 5. Wykres funkcji przynależności zbiorów rozmytych ilustrujących wartości zmiennej lingwistycznej wilgotności.

3.5. Ciśnienie

Przedstawienie ciśnienia jako zmiennej lingwistycznej. Do zmiennej lingwistycznej zostały dopasowane etykiety: niskie, umiarkowane, wysokie.

$$\mu_{cisnienieBardzoNiskie}(x) = \exp\left(\frac{-(x - 27)^2}{0.18}\right) \quad (23)$$

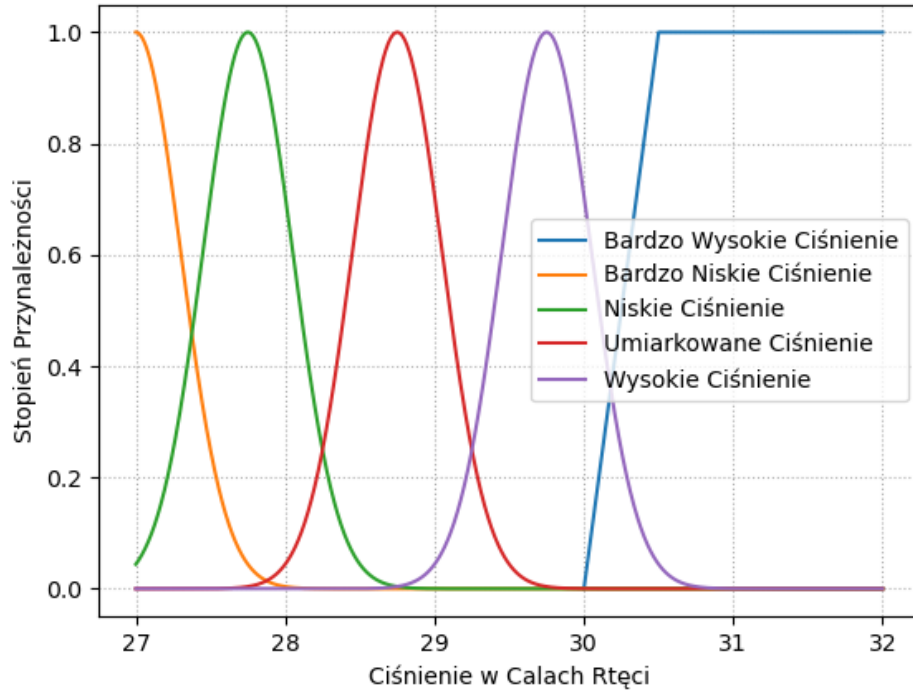
$$\mu_{cisnienieNiskie}(x) = \exp\left(\frac{-(x - 27.75)^2}{0.18}\right) \quad (24)$$

$$\mu_{cisnienieUmiarkowane}(x) = \exp\left(\frac{-(x - 28.75)^2}{0.18}\right) \quad (25)$$

$$\mu_{cisnienieWysokie}(x) = \exp\left(\frac{-(x - 29.75)^2}{0.18}\right) \quad (26)$$

$$\mu_{cisnienieBardzoWysokie}(x) = \begin{cases} \frac{x-30}{0.5} & \text{dla } 30 < x \leq 30.5 \\ 1 & \text{dla } 30.5 \leq x \end{cases} \quad (27)$$

gdzie: $\mu_{cisnienieBardzoNiskie}$, $\mu_{cisnienieNiskie}$, $\mu_{cisnienieUmiarkowane}$, $\mu_{cisnienieWysokie}$, $\mu_{cisnienieBardzoWysokie}$ - funkcje przynależności, x - ciśnienie.



Rysunek 6. Wykres funkcji przynależności zbiorów rozmytych ilustrujących wartości zmiennej lingwistycznej ciśnienia.

3.6. Widoczność

Przedstawienie widoczności jako zmiennej lingwistycznej. Do zmiennej lingwistycznej zostały dopasowane etykiety: brak, słaba, ograniczona, dobra.

$$\mu_{widocznoscBrak}(x) = 1 \quad \text{dla } x = 0 \quad (28)$$

$$\mu_{widocznoscSlaba}(x) = \begin{cases} 1 & \text{dla } x \leq 0.1 \\ \frac{0.1-x}{0.2} & \text{dla } 0.1 < x \leq 0.3 \end{cases} \quad (29)$$

$$\mu_{widocznoscOgraniczona}(x) = \begin{cases} \frac{x-0.1}{0.2} & \text{dla } 0.1 < x \leq 0.3 \\ 1 & \text{dla } 0.3 < x < 0.7 \\ \frac{1-x}{0.3} & \text{dla } 0.7 < x \leq 1 \end{cases} \quad (30)$$

$$\mu_{widocznoscDobra}(x) = \begin{cases} \frac{x-0.7}{0.3} & \text{dla } 0.7 < x \leq 1 \\ 1 & \text{dla } 1 < x < 2 \\ \frac{3-x}{1} & \text{dla } 2 < x \leq 3 \end{cases} \quad (31)$$

$$\mu_{widocznoscPelna}(x) = \begin{cases} \frac{x-3}{1} & \text{dla } 2 < x \leq 3 \\ 1 & \text{dla } 3 \leq x \end{cases} \quad (32)$$

gdzie: $\mu_{widocznoscBrak}$, $\mu_{widocznoscSlaba}$, $\mu_{widocznoscOgraniczona}$, $\mu_{widocznoscDobra}$, $\mu_{widocznoscPelna}$ - funkcje przynależności, x - widoczność.



Rysunek 7. Wykres funkcji przynależności zbiorów rozmytych ilustrujących wartości zmiennej lingwistycznej widoczności.

3.7. Prędkość wiatru

Przedstawienie predkości wiatru jako zmiennej lingwistycznej. Do zmiennej lingwistycznej zostały dopasowane etykiety: brak, słaby, umiarkowany, silny, wicher, huragan.

$$\mu_{wiatrBrak}(x) = 1 \quad dla \quad x = 0 \quad (33)$$

$$\mu_{wiatrSlaby}(x) = \begin{cases} 1 & dla \quad x \leq 3 \\ \frac{3-x}{0.5} & dla \quad 3 < x \leq 3.5 \end{cases} \quad (34)$$

$$\mu_{wiatrUmiarkowany}(x) = \begin{cases} \frac{x-3}{0.5} & dla \quad 3 < x \leq 3.5 \\ 1 & dla \quad 3.5 < x < 8 \\ \frac{9-x}{1} & dla \quad 8 < x \leq 9 \end{cases} \quad (35)$$

$$\mu_{wiatrSilny}(x) = \begin{cases} \frac{x-8}{1} & dla \quad 8 < x \leq 9 \\ 1 & dla \quad 9 < x < 17 \\ \frac{20-x}{3} & dla \quad 17 < x \leq 20 \end{cases} \quad (36)$$

$$\mu_{wiatrWicher}(x) = \begin{cases} \frac{x-17}{3} & dla \quad 17 < x \leq 20 \\ 1 & dla \quad 20 < x < 27 \\ \frac{30-x}{3} & dla \quad 27 < x \leq 30 \end{cases} \quad (37)$$

$$\mu_{wiatrHuragan}(x) = \begin{cases} \frac{x-40}{10} & dla \quad 30 < x \leq 40 \\ 1 & dla \quad 40 \leq x \end{cases} \quad (38)$$

gdzie: $\mu_{wiatrBrak}$, $\mu_{wiatrSlaby}$, $\mu_{wiatrUmiarkowany}$, $\mu_{wiatrSilny}$, $\mu_{wiatrWicher}$, $\mu_{wiatrHuragan}$ - funkcje przynależności, x - prędkość wiatru.



Rysunek 8. Wykres funkcji przynależności zbiorów rozmytych ilustrujących wartości zmiennej lingwistycznej prędkości wiatru.

3.8. Opady

Przedstawienie opadów jako zmiennej lingwistycznej. Do zmiennej lingwistycznej zostały dopasowane etykiety: brak, niewielkie, umiarkowane, duże.

$$\mu_{opadyBrak}(x) = 1 \quad \text{dla } x = 0 \quad (39)$$

$$\mu_{opadyNiewielkie}(x) = \begin{cases} 1 & \text{dla } x \leq 0.1 \\ \frac{0.1-x}{0.1} & \text{dla } 0.1 < x \leq 0.2 \end{cases} \quad (40)$$

$$\mu_{opadyUmiarkowane}(x) = \begin{cases} \frac{x-0.1}{0.1} & \text{dla } 0.1 < x \leq 0.2 \\ 1 & \text{dla } 0.2 < x < 0.3 \\ \frac{0.35-x}{0.05} & \text{dla } 0.3 < x \leq 0.35 \end{cases} \quad (41)$$

$$\mu_{opadyDuze}(x) = \begin{cases} \frac{x-0.35}{0.05} & \text{dla } 0.3 < x \leq 0.35 \\ 1 & \text{dla } 0.35 < x < 0.4 \\ \frac{0.45-x}{0.05} & \text{dla } 0.4 < x \leq 0.45 \end{cases} \quad (42)$$

$$\mu_{opadyBardzoDuze}(x) = \begin{cases} \frac{x-0.4}{0.05} & \text{dla } 0.4 < x \leq 0.45 \\ 1 & \text{dla } 0.45 \leq x \end{cases} \quad (43)$$

gdzie: $\mu_{opadyBrak}$, $\mu_{opadyNiewielkie}$, $\mu_{opadyUmiarkowane}$, $\mu_{opadyDuze}$, $\mu_{opadyBardzoDuze}$ - funkcje przynależności, x - opady.



Rysunek 9. Wykres funkcji przynależności zbiorów rozmytych ilustrujących wartości zmiennej lingwistycznej opadów.

3.9. Kwantyfikator lingwistyczny względny

Do kwantyfikatora lingwistycznego względnego[2] zostały dopasowane etykiety: niewiele, około 1/4, około połowy, większość, prawie wszystkie.

$$\mu_{niewiele}(x) = \exp\left(\frac{-x^2}{0.02}\right) \quad (44)$$

$$\mu_{okoloJednejCzwartej}(x) = \exp\left(\frac{-(x - 0.25)^2}{0.02}\right) \quad (45)$$

$$\mu_{okoloPolowy}(x) = \exp\left(\frac{-(x - 0.5)^2}{0.02}\right) \quad (46)$$

$$\mu_{wiekszosc}(x) = \exp\left(\frac{-(x - 0.75)^2}{0.02}\right) \quad (47)$$

$$\mu_{prawieWszystkie}(x) = \exp\left(\frac{-(x - 1)^2}{0.02}\right) \quad (48)$$

gdzie: $\mu_{niewiele}$, $\mu_{okoloJednejCzwartej}$, $\mu_{okoloPolowy}$, $\mu_{wiekszosc}$, $\mu_{prawieWszystkie}$ - kwantyfikatory, x - stosunek liczby obiektów posiadających cechę do wszystkich rozważanych obiektów.



Rysunek 10. Wykres funkcji przynależności kwantyfikatorów lingwistycznych względnych.

3.10. Kwantyfikator lingwistyczny absolutny

Do kwantyfikatora lingwistycznego absolutnego[2] zostały dopasowane etykiety: poniżej 10, około 50, około 100, między 100 a 200, około 200, ponad 200.

$$\mu_{ponizej10}(x) = \begin{cases} \frac{10-x}{10} & \text{dla } 0 < x \leq 10 \end{cases} \quad (49)$$

$$\mu_{okolo10}(x) = \begin{cases} \frac{x}{10} & \text{dla } 0 < x \leq 10 \\ \frac{20-x}{10} & \text{dla } 10 < x \leq 20 \end{cases} \quad (50)$$

$$\mu_{okolo50}(x) = \begin{cases} \frac{x-10}{25} & \text{dla } 10 < x \leq 35 \\ 1 & \text{dla } 35 < x < 65 \\ \frac{90-x}{25} & \text{dla } 65 < x \leq 90 \end{cases} \quad (51)$$

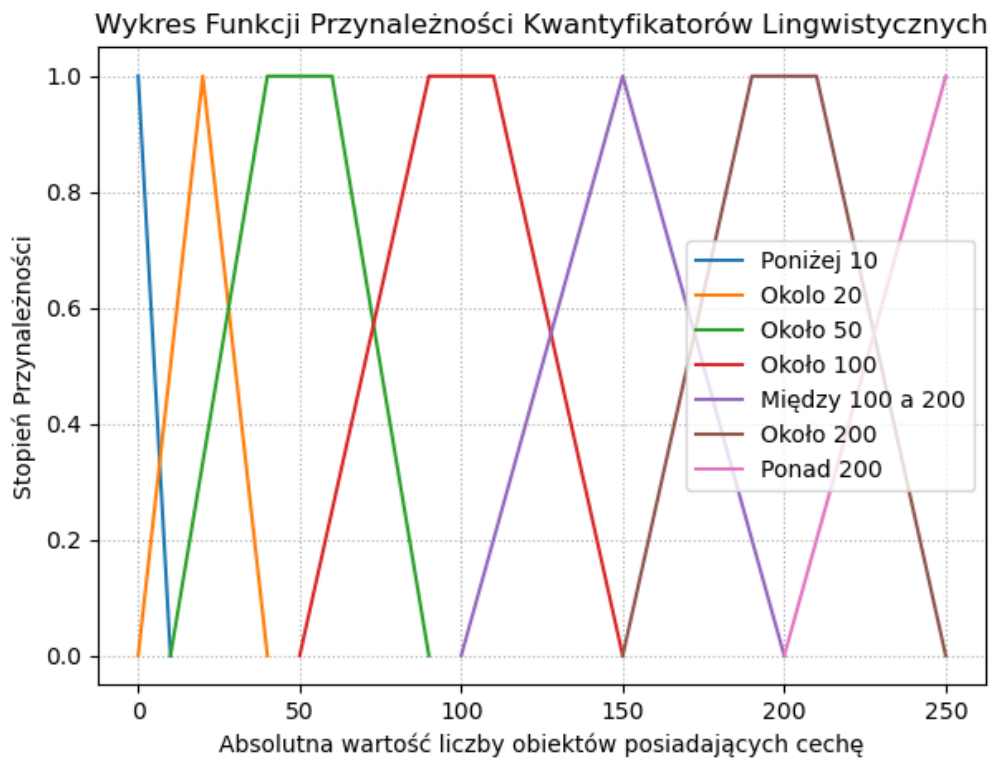
$$\mu_{okolo100}(x) = \begin{cases} \frac{x-50}{40} & \text{dla } 50 < x \leq 90 \\ 1 & \text{dla } 90 < x < 110 \\ \frac{150-x}{50} & \text{dla } 110 < x \leq 150 \end{cases} \quad (52)$$

$$\mu_{miedzy100A200}(x) = \begin{cases} \frac{x}{50} & \text{dla } 100 < x \leq 150 \\ \frac{200-x}{50} & \text{dla } 150 < x \leq 200 \end{cases} \quad (53)$$

$$\mu_{okolo200}(x) = \begin{cases} \frac{x-150}{40} & \text{dla } 150 < x \leq 190 \\ 1 & \text{dla } 190 < x < 210 \\ \frac{250-x}{40} & \text{dla } 210 < x \leq 250 \end{cases} \quad (54)$$

$$\mu_{Ponad200}(x) = \begin{cases} \frac{x}{50} & \text{dla } 200 < x \leq 250 \\ 1 & \text{dla } 250 \leq x \end{cases} \quad (55)$$

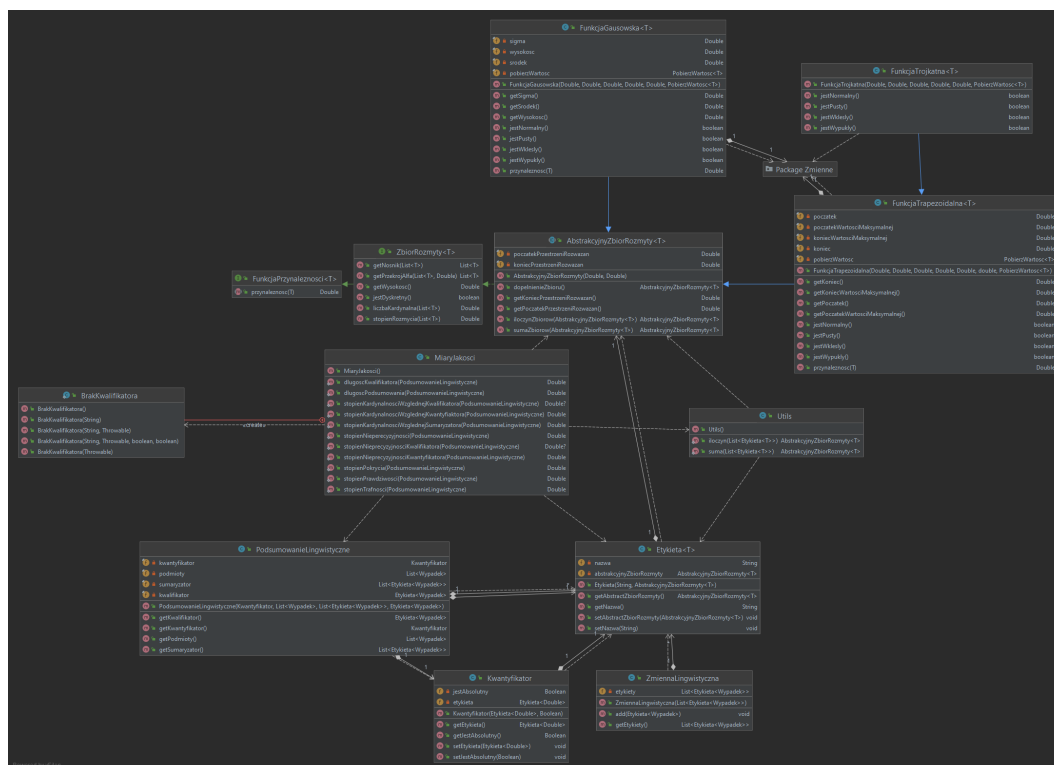
gdzie: $\mu_{ponizej10}$, $\mu_{okolo10}$, $\mu_{okolo50}$, $\mu_{okolo100}$, $\mu_{miedzy100A200}$, $\mu_{okolo200}$, $\mu_{Ponad200}$ - kwantyfikatory absolutne, x - absolutna wartość liczby obiektów posiadających cechę.



Rysunek 11. Wykres funkcji przynależności kwantyfikatorów lingwistycznych absolutnych.

4. Narzędzia obliczeniowe: projekt (wybór, implementacja) i diagram UML pakietu obliczeń rozmytych. Diagram UML generatora podsumowań

4.1. Diagram pakietu obliczeń rozmytych



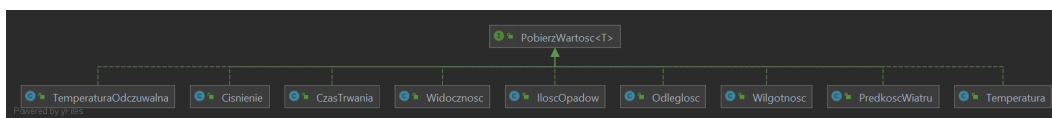
Rysunek 12. Diagram uml pakietu obliczeń rozmytych.

Program został wykonany w języku Java 11 LTS z użyciem narzędzia Maven. W celu wykonania projektu stworzyliśmy własny pakiet obliczeń rozmytych - obliczeniaRozmyte. Pakiet ten jest przystosowany do pracy przy użyciu naszego dyskretnego zbioru danych o wypadkach samochodowych w Stanach Zjednoczonych Ameryki. Pakiet posiada klasy oraz interfejsy niezbędne do wykonywania obliczeń rozmytych i generowania podsumowań lingwistycznych. Interfejs FunkcjaPrzynależności posiada jedną metodę pozwalającą na obliczenie przynależności obiektu do zbioru[9]. Interfejs ten rozszerzany jest przez interfejs ZbiórRozmyty który pozwala na obliczenie różnych własności zbioru rozmytego: wysokość[4], liczbę kardynalna[4], stopień rozmycia[5] oraz pozwala na pobranie zbiorów klasycznych będących odpowiednio nośnikami[9] i przekrojem alfa zbioru rozmytego[9]. Interfejs ten jest implementowany przez abstrakcyjną klasę AbstrakcyjnyZbiórRozmyty posiadającą informacje o początku i końcu przestrzeni rozważań[4], i pozwalającą na wyznaczenie dopełnienia zbioru oraz iloczynu i sumy zbioru z innym zbiorem rozmytym[4]. Klasa ta implementowana jest przez klasy odpowiadające odpowiednio

funkcji gaussowskiej[9] i trapezoidalnej[9], a po funkcji trapezoidalnej następnie dziedziczy funkcja trójkątna[9]. W celu pobrania odpowiedniego pola z obiektu w celu obliczenia przynależności klasy implementujące AbstrakcyjnyZbiórRozmyty posiadają jako swoje pole interfejs PobierzWartość, który wybiera z klasy odpowiednie pole podlegające podsumowaniu. Odpowiednie implementacje klasy AbstrakcyjnyZbiórRozmyty posiadają też informacje o tym czy zbiór przez nie reprezentowany jest normalny[9], pusty[9], wklęsły[9] lub wypukły[9]. Zbiory rozmyte są natępnie używane przy tworzeniu etykiety[5], gdzie każda posiada swoją nazwę i odpowiedni zbiór rozmyty. Używając Etykiet reprezentujemy Kwantyfikatory[5], kwalifikatory[5] oraz sumaryzatory[5]. Do obliczania iloczynu lub sumy wielu zbiorów rozmytych używamy klasy Utils posiadającą odpowiednie statyczne metody. Kwantyfikator posiada etykietę oraz informacje czy jest on absolutny[5]. Klasa zmienna lingwistyczna reprezentuje zmienną lingwistyczną poprzez listę Etykiet. Klasa podsumowanie lingwistyczne reprezentuje podsumowanie lingwistyczne poprzez kwantyfikator, klasyfikator, sumaryzator oraz podmioty. Dla każdego podsumowania możemy policzyć miary T1 - T11 korzystając z klasy MiaryJakości.

Klasa AbstrakcyjnyZbiórRozmyty jest abstrakcyjną klasą, z której dziedziczą klasy FunkcjaGaussowska oraz FunkcjaTrapezoidalna. Ta abstrakcyjna klasa zawiera pola odpowiadające początkowi oraz końcowi przestrzeni rozważań[4]. Posiada metody zwracające dopełnienie zbioru, iloczyn oraz sumę zbiorów, a także informację czy zbiór jest gęsty czy dyskretny. Klasa FunkcjaTrapezoidalna posiada pola dla wartości potrzebnych do wygenerowania funkcji przynależności takie jak: początek, koniec oraz początek i koniec przedziału, w którym funkcja przynależności przyjmuje wartość 1. Klasa FunkcjaGaussowska posiada pola dla wartości potrzebnych do wygenerowania funkcji przynależności takie jak: sigma, wysokość oraz środek. Klasa FunkcjaTrojkatna dziedziczy po klasie FunkcjaTrapezoidalna, gdzie początek i koniec przedziału, w którym funkcja przynależności przyjmuje wartość 1,

są tą samą wartością. Klasa `AbstractZbiorRozmyty` implementuje interfejs `ZbiorRozmyty` zawierający metody pozwalające uzyskać wysokość zbioru[4], przeroj alfa, nośnik, liczba kardynalna[4] oraz stopień rozmycia [?]. Gdzie funkcja zwracająca nośnik zwraca listę elementów, dla których wartość funkcji przynależności jest większa od zera. Funkcja zwracająca przkerój alfa zwraca listę elementów, dla których wartość funkcji przynależności jest większa niż zadana w parametrze wartość. Interfejs `ZbiorRozmyty` dziedziczy interfejs `FunkcjaPrzynaleznosci` zawierający metodę `przynaleznosc` pozwalającą obliczyć wartość funkcji przynależności. Implementacje tej metody korzystają z pakietu `Zmienne`. Pakiet posiada klasę `ZbiorKlasyczny` reprezentującą klasyczne zbiory, klasa dziedziczy po klasie `FunkcjaTrapezoidalna`. W naszym programie zbiór klasyczny reprezentowany jest jako zbiór trapezoidalny, w którym funkcja przynależności przyjmuje wartości tylko 0 lub 1 (prostokąt). Klasa `Etykieta` zawiera w sobie nazwę etykiety oraz implementację abstrakcyjnej klasy `AbstrakcyjnyZbiorRozmyty`. W pakeicie posiadamy również klasę `Klasyfikator`, który posiada swoją nazwę, obiekt klasy `Etykieta` oraz pole boolean czy jest klasyfikatorem absolutnym. Klasa `PodsumowanieLingwistyczne` posiada obiekt klasy `Klasyfikator`, listę `Wypadków` jako podmioty, listę etykiet jako sumaryzatory oraz obiekt klasy `Etykieta` jako kwalifikator. Klasa `Utils` pozwala na policzenie iloczynu oraz sumy `AbstrakcyjnychZbiorowRozmytych` oraz zwraca `AbstrakcyjnyZbiorRozmyty`. Klasa `MiaryJakosci` zawiera 11 metod pozwalających na obliczenie danej miary jakości podsumowania lingwistycznego, niektóre z metod mogą rzucić wyjątek będący obiektem klasy `BrakKwalifikatora`.



Rysunek 13. Diagram uml pakietu Zmienne.

W programie został zaimplementowany pakiet Zmienne posiadający interfejs PobierzWartosc, który implementuje 9 klas reprezentujących analizowane pola z obiektu Wypadek.

4.2. Diagram UML generatora podsumowań. Krótka instrukcja użytkownika

Diagram UML generatora podsumowań (warstwy obliczeniowej oraz interfejsu użytkownika). Krótki ilustrowany opis jak użytkownik może korzystać z aplikacji, w szczególności wprowadzać parametry podsumowań, odczytywać wyniki oraz definiować własne etykiety i kwantyfikatory. Wersja JRE i inne wymagania niezbędne do uruchomienia aplikacji przez użytkownika na własnym komputerze.

Sekcja uzupełniona jako efekt zadania Tydzień 11 wg Harmonogramu Zajęć na WIKAMP KSR.

5. Jednopodmiotowe podsumowania lingwistyczne. Miary jakości, podsumowanie optymalne

Wyniki kolejnych eksperymentów wg punktów 2.-4. opisu projektu 2. Listy podsumowań jednopodmiotowych 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.

Sekcja uzupełniona jako efekt zadania Tydzień 11 wg Harmonogramu Zajęć na WIKAMP KSR.

6. Wielopodmiotowe podsumowania lingwistyczne i ich miary jakości

Wyniki kolejnych eksperymentów wg punktów 2.-4. opisu projektu 2. Uzasadnienie i metoda podziału zbioru danych na rozłączne podmioty. Listy podsumowań wielopodmiotowych i tabele/rankingi podsumowań dla danych atrybutów obowiązkowe i dokładnie opisane w „captions” (tytułach), konieczny opis kolumn i wierszy tabel. Konieczne uwzględnienie wszystkich 4-ch form podsumowań wielopodmiotowych.

****** Możliwe sformułowanie zagadnienia wielopodmiotowego 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.

7. 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 prezentuje 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.

8. 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] I. Superson, A. Niewiadomski, POZYSKIWANIE WIEDZY Z RELACYJNYCH BAZ DANYCH: WIELOPODMIOTOWE PODSUMOWANIA LINGWISTYCZNE, Politechnika Łódzka
- [2] A. Niewiadomski, Rozmyte metody inteligentnej interpretacji danych, tom 10, 2006, 546-547
- [3] 2021 Kaggle Inc [internetowa społeczność związana z analizą danych], US Accidents (3 million records – updated) A Countrywide Traffic Accident Dataset (2016 - 2020) [przeglądany 24 kwietnia 2021], Dostępny w: <https://www.kaggle.com/sobhanmoosavi/us-accidents> Oficyna Wydawnicza EXIT, Warszawa, 2008.
- [4] Zadeh L.A., A computational approach to fuzzy quantifiers in natural languages. Computers and Maths with Applications, nr 9, 1983, 149-183

- [5] A. Niewiadomski, Zbiory rozmyte typu 2. Zastosowania w reprezentowaniu informacji. Seria „Problemy współczesnej informatyki” pod redakcją L. Rutkowskiego. Akademicka Oficyna Wydawnicza EXIT, Warszawa, 2019.
- [6] S. Zadrozny, Zapytania nieprecyzyjne i lingwistyczne podsumowania baz danych, EXIT, 2006, Warszawa
- [7] A. Niewiadomski, Methods for the Linguistic Summarization of Data: Applications of Fuzzy Sets and Their Extensions, Akademicka
- [8] Zadeh, L. A.: 1965, ‘Fuzzy sets’. Inf. and Control 8, 338–353.
- [9] dr inż. Anna Bryniarska, Podstawy teorii systemów rozmytych z zadaniami, Dostępny w: https://a.bryniarska.po.opole.pl/images/nsi/skrypt_r2_28listopada.pdf,

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