

Data oddania: _____

Ocena: _____

Mateusz Walczak 216911

Konrad Kajszczyk 216790

Zadanie 2: Lingwistyczne podsumowania baz danych*

1. Cel

Celem zadania było zaprojektowanie aplikacji desktopowej, służącej do generowania podsumowań lingwistycznych oraz obliczania ich miar jakości dla wybranej bazy danych. Dodatkowo, aplikacja powinna posiadać graficzny interfejs użytkownika, który umożliwi intuicyjne korzystanie z programu.

2. Wprowadzenie

Rozważania we wprowadzeniu rozpoczniemy od *zbioru rozmytego*, czyli najbardziej podstawowego pojęcia, bez którego analiza działania lingwistycznych podsumowań baz danych nie byłaby możliwa. Przytoczmy zatem definicję *zbioru rozmytego*:

Definicja 1. Zbiór rozmyty A w niepustej przestrzeni X [5]

$$A = \{\langle x, \mu_A(x) \rangle : x \in X\}$$

gdzie $\mu_A(x) : \mathcal{X} \rightarrow [0, 1]$ nazywamy *funkcją przynależności do zbioru rozmytego A* .

W związku z faktem, iż pojęcie *funkcji przynależności* wystąpiło w powyższej definicji, w następnym podrozdziale zajmiemy się opisem tego rodzaju funkcji, wykorzystywanych przez nas.

* SVN: <https://github.com/Walducha1908/KSR2>

2.1. Funkcje przynależności

Funkcja przynależności określa w jakim stopniu dany element przynależy do zbioru. W zbiorach rozmytych zakres wartości jakie może ona przyjmować jest rozszerzony do przedziału $[0,1]$.

W naszym programie, posłużyliśmy się dwoma rodzajami funkcji przynależności:

- funkcją trójkątną opisaną wzorem:

$$f_{troj}(x) = \begin{cases} \frac{x-a}{b-a} & \text{jeśli } a \leq x < b \\ 1 & \text{jeśli } x = b \\ \frac{c-x}{c-b} & \text{jeśli } b < x \leq c \\ 0 & \text{w przeciwnym wypadku.} \end{cases} \quad (1)$$

- oraz funkcją trapezoidalną opisaną wzorem:

$$f_{trap}(x) = \begin{cases} \frac{x-a}{b-a} & \text{jeśli } a \leq x < b \\ 1 & \text{jeśli } b \leq x \leq c \\ \frac{d-x}{d-c} & \text{jeśli } c < x \leq d \\ 0 & \text{w przeciwnym wypadku.} \end{cases} \quad (2)$$

gdzie a, b, c oraz d są parametrami funkcji przynależności - wierzchołkami trójkąta lub trapezu na wykresie.

2.2. Podsumowania lingwistyczne

W tym rozdziale skoncentrujemy się na wyjaśnieniu czym są podsumowania lingwistyczne, które stanowią główny przedmiot rozważań tego zadania.

Ogólna postać lingwistycznego podsumowania bazy danych, prezentuje się następująco [4]:

$$Q \ P \ \text{jest/są} \ S_j \ [T] \quad (3)$$

Omówmy poszczególne elementy powyższego wzoru:

- Q stanowi kwantyfikator, kwantyfikatory mogą być względne (np. *większość, prawie wszystkie, około połowa*) lub bezwzględne (np. *mniej niż 100, około 500*),
- P jest podmiotem podsumowania lingwistycznego, zestawem obiektów reprezentowanym przez krotki w bazie danych (np. *pomiary pogodowe, ludzie, samochody*),
- S_j jest sumaryzatorem, stanowi zbiór rozmyty na zbiorze wartości przyjmowanych przez j -tą kolumnę w bazie danych (np. *wysoka temperatura* w dziedzinie V_j $[-20, 40]$),
- T to stopień prawdziwości podsumowania (więcej informacji zamieszczono w rozdziale 2.3 dotyczącym miar jakości).

Zaprezentujemy przykład, dla omówionej postaci podsumowania lingwistycznego: *"Prawie wszyscy programiści zarabiają ponad 2000 złotych [0.88]"*, gdzie: *"Prawie wszyscy"* to kwantyfikator, *"programiści"* to podmiot lingwistyczny, *"zarabiają ponad 2000 złotych"* to sumaryzator, a *"[0.88]"* to stopień

prawdziwości podsumowania.

Aby zwiększyć stopień skomplikowania podsumowania lingwistycznego, można skorzystać ze złożonego sumaryzatora. Wtedy podsumowanie lingwistyczne przybiera postać:

$$Q \text{ } P \text{ jest/są } S_1 \text{ i/lub } S_2 \text{ i/lub } \dots \text{ i/lub } S_n \text{ } [T] \quad (4)$$

gdzie za pomocą słowa *lub* wyrażamy sumę sumaryzatorów zaś z wykorzystaniem słowa *i* - ich iloczyn.

Rozwińmy podany wcześniej przykład, tak aby wykorzystywał on sumaryzator złożony: *"Prawie wszyscy programiści zarabiają ponad 2000 złotych i mają bardzo drogi samochód [0.54]"*, sumaryzatorem złożonym w tym przykładzie jest *"zarabiają ponad 2000 złotych i mają bardzo drogi samochód"*.

Ostatnim etapem w naszej wędrówce po rozważaniach dotyczących podsumowań lingwistycznych, będzie podsumowanie z kwalifikatorem, które ma postać:

$$Q \text{ } P \text{ będących/mających własność } W \text{ jest/są } S_j \text{ } [T] \quad (5)$$

gdzie *W* jest kwalifikatorem - dodatkową właściwością (lub zestawem właściwości) posiadaną przez analizowane obiekty (np. *"[programiści] mający około 30 lat"*).

Kończąc nasze rozważania, zaprezentujemy jeszcze przytaczany wcześniej przykład, z wykorzystaniem kwalifikatora i sumaryzatora złożonego: *"Prawie wszyscy programiści mający około 30 lat zarabiają ponad 2000 złotych i mają bardzo drogi samochód [0.54]"*.

2.3. Miary jakości dla podsumowań lingwistycznych

Aby określić jakość naszych podsumowań zaimplementowaliśmy 11 miar jakości od T_1 do T_{11} . Poniżej zestawimy nazwy miar jakości wraz z oznaczeniami matematycznymi, którymi będziemy się posługiwać w następnych rozdziałach:

- T_1 - stopień prawdziwości,
- T_2 - stopień nieprecyzyjności,
- T_3 - stopień pokrycia,
- T_4 - stopień trafności,
- T_5 - długość podsumowania,
- T_6 - stopień nieprecyzyjności kwantyfikatora,
- T_7 - stopień licznosci kwantyfikatora,
- T_8 - stopień licznosci sumaryzator,
- T_9 - stopień nieprecyzyjności kwalifikatora,
- T_{10} - stopień licznosci kwalifikatora,
- T_{11} - długość kwalifikatora.

Dokładny opis miar i wzorów można znaleźć w rozdziałach 8.3 i 8.4 w monografii *Methods for the linguistic summarization of data - applications of fuzzy sets and their extensions*, Adam Niewiadomski [4].

3. Opis implementacji

Program został napisany w języku Java z wykorzystaniem narzędzia Maven [2], służącego do automatyzacji budowy oprogramowania. Graficzny interfejs użytkownika został zbudowany w oparciu o bibliotekę JavaFX [3].

W opisie implementacji skoncentrujemy się na jej najważniejszej części, czyli logice aplikacji.

W celu zbudowania struktury klas reprezentujących typy, parametry i własności zbiorów rozmytych oraz operacji na nich, zaimplementowaliśmy własną bibliotekę w formie pakietu w naszej aplikacji o nazwie *FuzzyLib*.

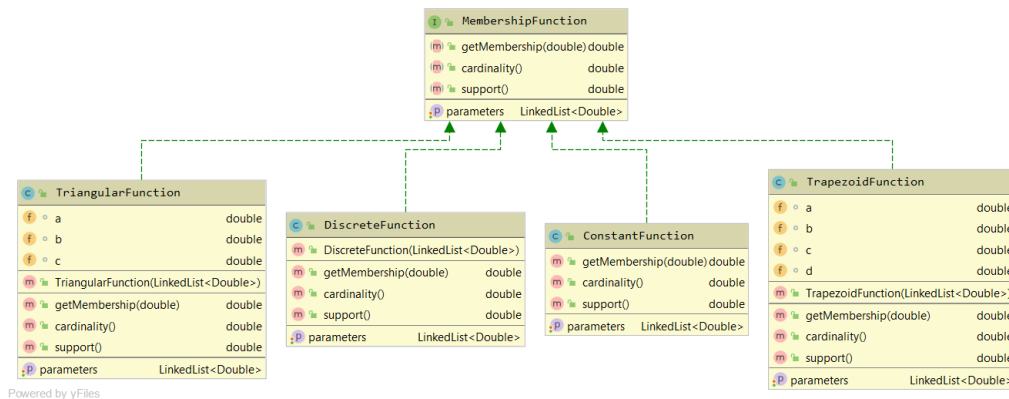
Pakiet *FuzzyLib* został podzielony na następujące podpakiety:

- *Membership*,
- *Logic*,
- *Containers*,
- *Summaries*.

W tym rozdziale omówione zostaną wszystkie wyżej wymienione podpakiety. Przedstawimy diagramy UML każdego z podpakietów a także omówimy zastosowanie poszczególnych klas.

3.1. Podpakiet Membership

Podpakiet *Membership* zawiera implementacje funkcji przynależności. Każda klasa tego podpakietu implementuje interfejs *MembershipFunction*.



Rysunek 1. Diagram UML dla podpakietu *Membership*

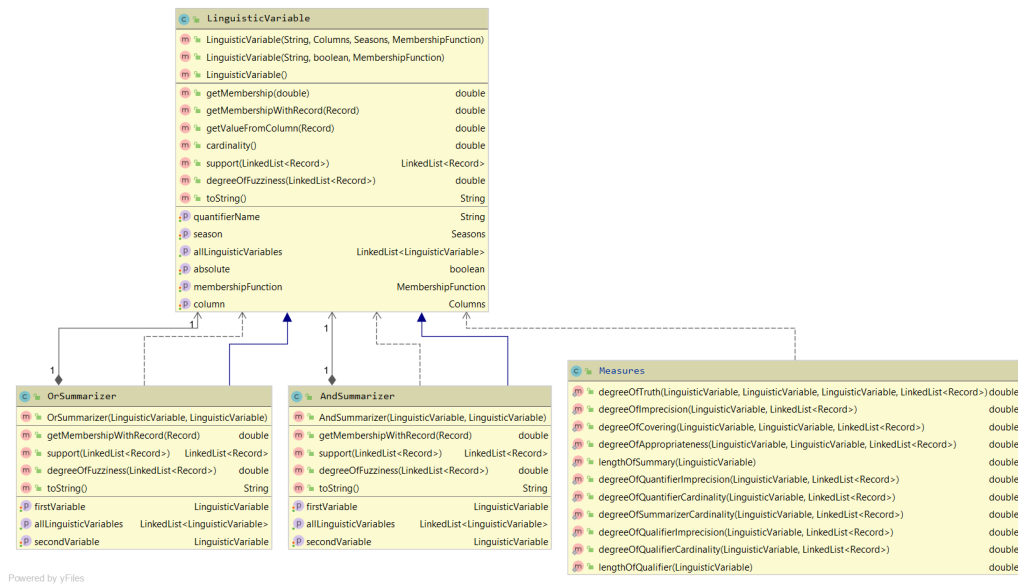
Klasy podpakietu *Membership* są następujące:

- *MembershipFunction* - interfejs implementowany przez wszystkie klasy należące do tego podpakietu,
- *TriangularFunction* - klasa implementująca trójkątną funkcję przynależności,

- *DiscreteFunction* - klasa implementująca dyskretną funkcję przynależności,
- *ConstantFunction* - klasa implementująca stałą funkcję przynależności,
- *TrapezoidFunction* - klasa implementująca trapezoidalną funkcję przynależności.

3.2. Podpakiet Logic

Podpakiet *Logic* odpowiada za implementacje zmiennych lingwistycznych, operacji sumy i iloczynu zbiorów rozmytych a także miar jakości podsumowań lingwistycznych.



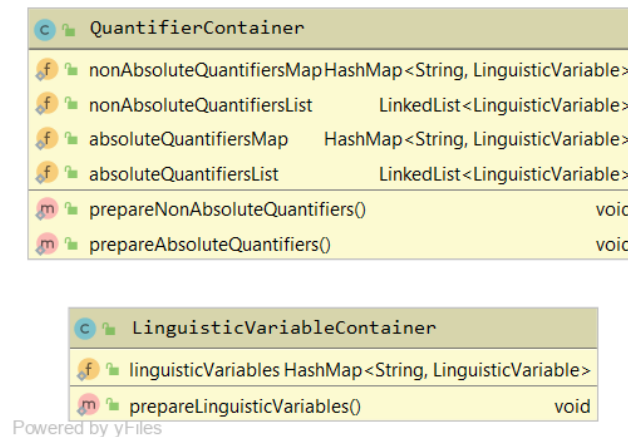
Rysunek 2. Diagram UML dla podpakietu *Logic*

Klasy podpakietu *Logic* są następujące:

- *LinguisticVariable* - klasa implementująca zmienną lingwistyczną,
- *AndSummarizer* - klasa implementująca operacje iloczynu, dziedziczy z klasy *LinguisticVariable*,
- *OrSummarizer* - klasa implementująca operacje sumy, dziedziczy z klasy *LinguisticVariable*,
- *Measures* - klasa implementująca miary jakości podsumowań lingwistycznych.

3.3. Podpakiet Containers

Podpakiet *Containers* zawiera klasy kontenerowe, tworzące i przetrzymujące zmienne lingwistyczne i kwantyfikatory.



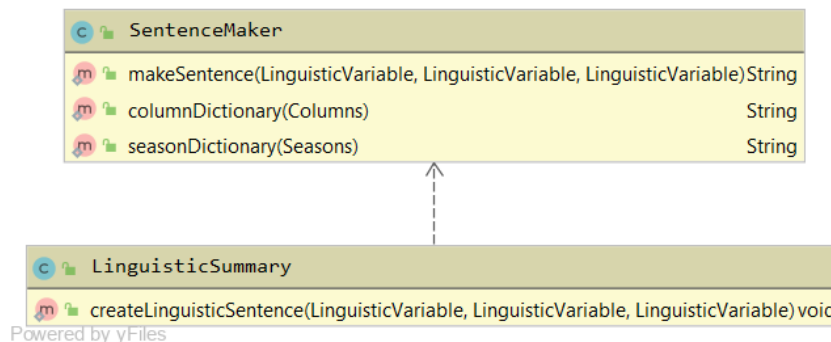
Rysunek 3. Diagram UML dla podpakietu *Containers*

Klasy podpakietu *Containers* są następujące:

- *LinguisticVariableContainer* - klasa kontenerowa dla zmiennych lingwistycznych,
- *QuantifierContainer* - klasa kontenerowa dla kwantyfikatorów.

3.4. Podpakiet Summaries

Podpakiet *Summaries* zawiera dwie klasy odpowiadające za budowę zdań podsumowania lingwistycznego w języku ludzkim (angielskim) oraz obliczenie wszystkich miar jakości dla danego podsumowania lingwistycznego - wywołanie statycznych metod klasy *Measures* podpakietu *Logic*.



Rysunek 4. Diagram UML dla podpakietu *Summaries*

Klasy podpakietu *Summaries* są następujące:

- *LinguisticSummary* - klasa odpowiadająca za obliczenie wszystkich miar jakości dla danego podsumowania,
- *SentenceMaker* - klasa odpowiadająca za budowanie zdań w języku ludzkim.

4. Materiały i metody

Wybrana przez nas baza danych zawiera historyczne pomiary pogodowe z Holandii [1]. Dane zostały zgromadzone przez KNMI (*Dutch weather institute* - Holenderski instytut pogody) na przestrzeni lat 1901-2018 i pochodzą z 50 różnych stacji pogody znajdujących się na terenie całego kraju.

Ze względu na fakt, iż oryginalna baza danych składa się z 804099 krotek, postanowiliśmy wybrać tylko niewielką część z dostępnych danych. Zdecydowaliśmy się na najnowsze dane pomiarowe - z lat 2016-2018. W ten sposób ograniczyliśmy liczbę wykorzystywanych krotek do 17000.

4.1. Wybór kolumn

W celu analizy bazy danych i tworzenia jej lingwistycznych podsumowań wybraliśmy 10 kolumn z danymi liczbowymi.

	A	B	C	D	E	F	G	H	I	J	K	L
1	STN	YYYYMMDD	FG	FHX	FHN	FXX	TG	TN	TX	T10N	Q	RH
2	380	20181231	23	30	10	60	83	74	91	70	96	19
3	370	20181231	27	40	20	70	89	73	99	72	115	5
4	350	20181231	28	40	20	80	89	75	98	72	132	5
5	375	20181231	28	50	20	90	90	73	98	71	105	1
6	290	20181231	36	60	20	90	88	74	99	73	126	2
7	275	20181231	30	50	20	90	86	72	98	70	132	1
8	279	20181231	40	60	20	100	86	78	97	77	150	-1
9	260	20181231	27	40	20	100	89	77	100	76	137	1
10	269	20181231	36	50	20	80	87	78	100	75	155	-1
11	280	20181231	40	60	20	110	88	80	99	77	174	5
12	240	20181231	45	70	30	110	90	83	109	76	205	-1
13	344	20181231	35	50	20	90	91	79	104	77	207	-1
14	215	20181231	42	60	20	90	89	82	97	80	212	0
15	235	20181231	51	80	40	110	90	83	98	77	229	-1
16	270	20181231	50	80	30	120	87	82	97	78	236	-1
17	310	20181231	48	60	30	90	88	73	102	71	335	3
18	375	20181230	32	40	20	90	79	53	92	42	137	31
19	350	20181230	29	50	20	80	82	56	93	49	139	14
20	260	20181230	26	50	20	100	85	62	95	52	112	10
21	370	20181230	33	50	20	90	80	51	91	45	156	11
22	269	20181230	33	60	20	110	82	63	92	54	119	8
23	344	20181230	36	60	20	120	86	50	97	36	141	4
24	215	20181230	43	60	20	100	87	57	99	48	127	7
25	275	20181230	35	70	20	120	77	55	90	48	138	24
26	279	20181230	39	80	20	160	76	67	84	62	148	10

Rysunek 5. Fragment widoku bazy w formacie *xlsx*

Wybrane kolumny są następujące:

- FG - średnia prędkość wiatru przez cały dzień [$0.1 \frac{m}{s}$].
- FHX - najwyższa średnia prędkość wiatru w ciągu jednej godziny [$0.1 \frac{m}{s}$].
- FHN - najniższa średnia prędkość wiatru w ciągu jednej godziny [$0.1 \frac{m}{s}$].
- FXX - najszybszy podmuch wiatru w ciągu całego dnia [$0.1 \frac{m}{s}$].
- TG - średnia dzienna temperatura [$0.1^{\circ}C$].
- TN - minimalna dzienna temperatura [$0.1^{\circ}C$].
- TX - maksymalna dzienna temperatura [$0.1^{\circ}C$].
- T10N - minimalna dzienna temperatura na wysokości 10 cm od poziomu gruntu [$0.1^{\circ}C$].

- Q - nasłonecznienie, energia słoneczna przypadająca na powierzchnię [$\frac{J}{cm^2}$].
- RH - suma opadów atmosferycznych w ciągu całego dnia [0.1mm].

Oprócz wyżej opisanych danych liczbowych, w naszej bazie znajdują się także dwie dodatkowe kolumny, służące do identyfikacji pomiaru:

- STN - numer stacji badawczej wykonującej pomiar.
- YYYYMMDD - data pomiaru w formacie opisanym przez nazwę kolumny.

4.2. Zmienne lingwistyczne

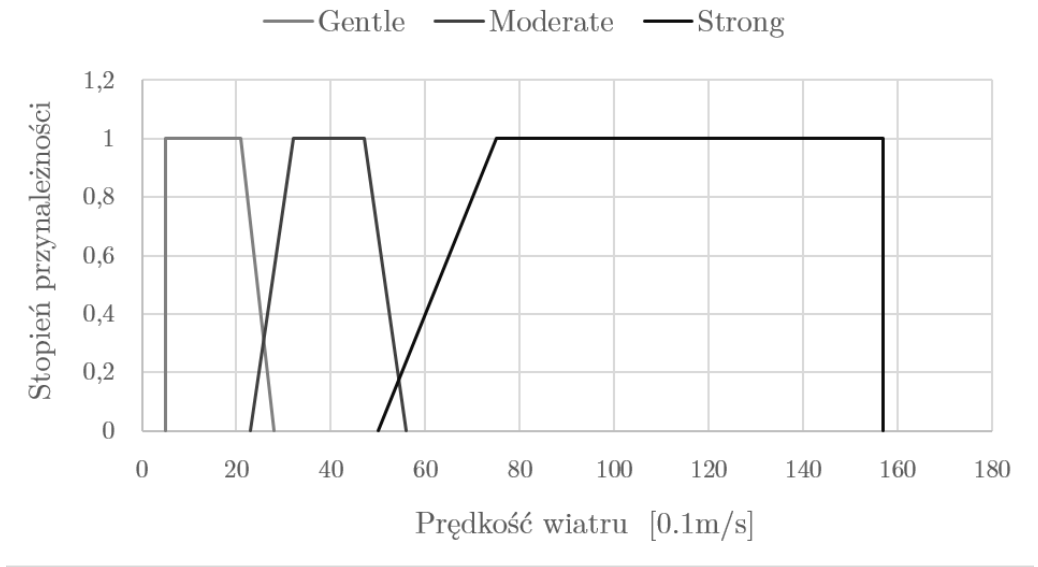
W tym rozdziale przedstawimy wzory i wykresy opisujące zaproponowane przez nas zmienne lingwistyczne. We wszystkich przypadkach, wykorzystywanymi przez nas funkcjami przynależności są funkcje trapezoidalne.¹

Aby nie duplikować treści wzorów, niepotrzebnie zwiększając w ten sposób objętość sprawozdania, zdecydowano się na zamieszczenie tabel z parametrami etykiet zmiennych lingwistycznych, odnoszącymi się do wzorów z poprzedniego podrozdziału.

¹ Wartości prezentowane w tabelach są tylko propozycjami. Autorzy sprawozdania zastrzegają sobie możliwość do ich późniejszej modyfikacji

4.2.1. Kolumna FG

Wykres opisujący zmienną lingwistyczną dla kolumny zawierającej wartości średniej prędkości wiatru przez cały dzień (FG), zamieszczono poniżej.



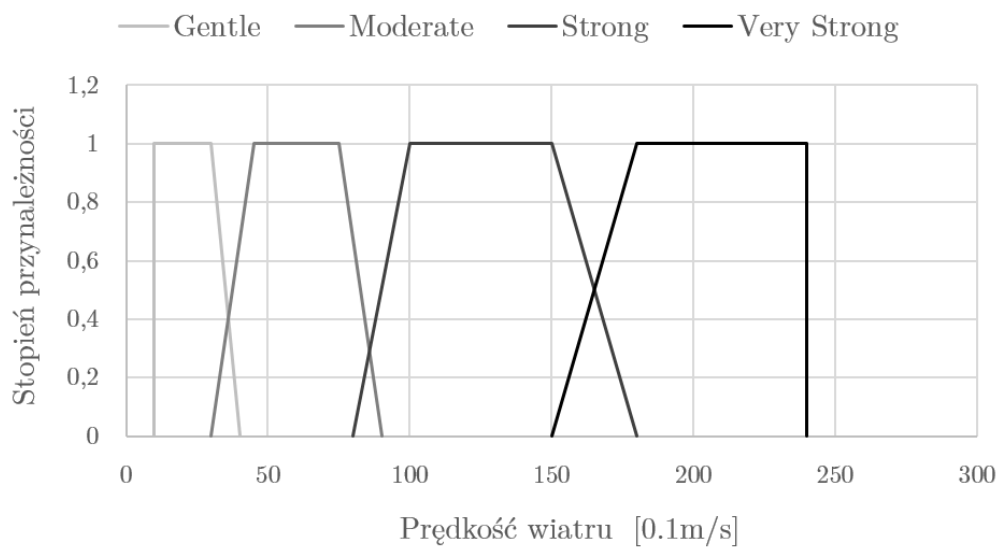
Rysunek 6. Wykres opisujący zmienną lingwistyczną dla kolumny FG.

Etykieta	a	b	c	d
Gentle	5	5	21	28
Moderate	23	32	47	56
Strong	50	75	157	157

Tabela 1. Przyporządkowane parametry funkcji trapezoidalnej dla kolumny FG.

4.2.2. Kolumna FHX

Wykres opisujący zmienną lingwistyczną dla kolumny zawierającej najwyższą średnią prędkości wiatru w przeciągu jednej godziny (FHX), zamieszczono poniżej.



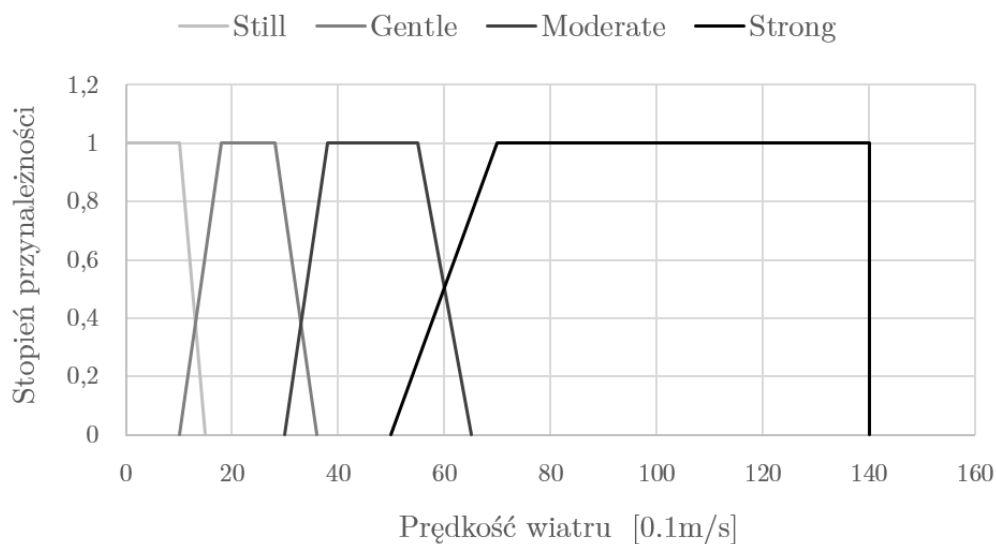
Rysunek 7. Wykres opisujący zmienną lingwistyczną dla kolumny FHX.

Etykieta	a	b	c	d
Gentle	10	10	30	40
Moderate	30	45	75	90
Strong	80	100	150	180
Very strong	150	180	240	240

Tabela 2. Przyporządkowane parametry funkcji trapezoidalnej dla kolumny FHX.

4.2.3. Kolumna FHN

Wykres opisujący zmienną lingwistyczną dla kolumny zawierającej najniższą średnią prędkości wiatru w przeciągu jednej godziny (FHN), zamieszczono poniżej.



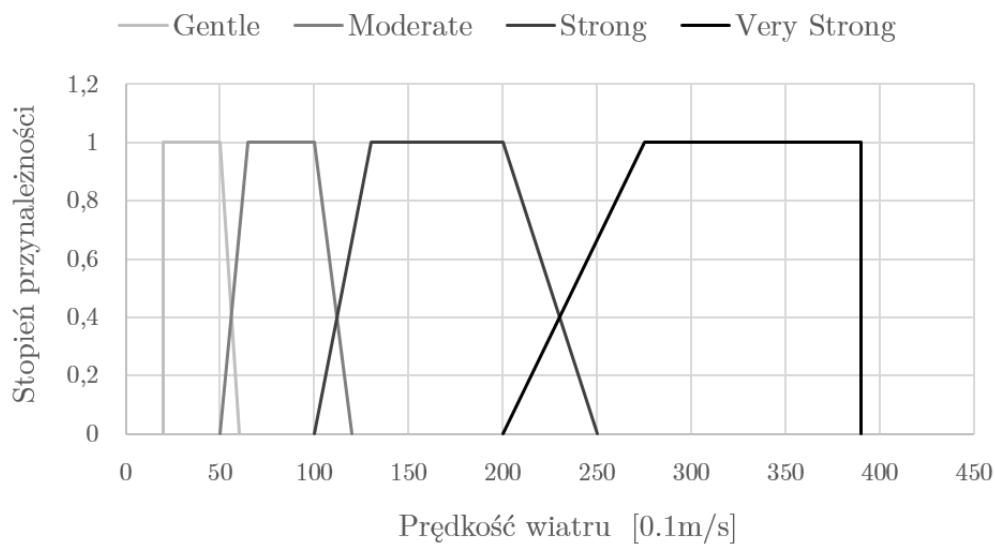
Rysunek 8. Wykres opisujący zmienną lingwistyczną dla kolumny FHN.

Etykieta	a	b	c	d
Gentle	0	0	10	15
Moderate	10	18	28	36
Strong	30	38	55	65
Very strong	50	70	140	140

Tabela 3. Przyporządkowane parametry funkcji trapezoidalnej dla kolumny FHN.

4.2.4. Kolumna FXX

Wykres opisujący zmienną lingwistyczną dla kolumny zawierającej naj-
silniejszy powiew wiatru (FXX), zamieszczono poniżej.



Rysunek 9. Wykres opisujący zmienną lingwistyczną dla kolumny FXX.

Etykieta	a	b	c	d
Gentle	20	20	50	60
Moderate	50	65	100	120
Strong	100	130	200	250
Very strong	200	275	390	390

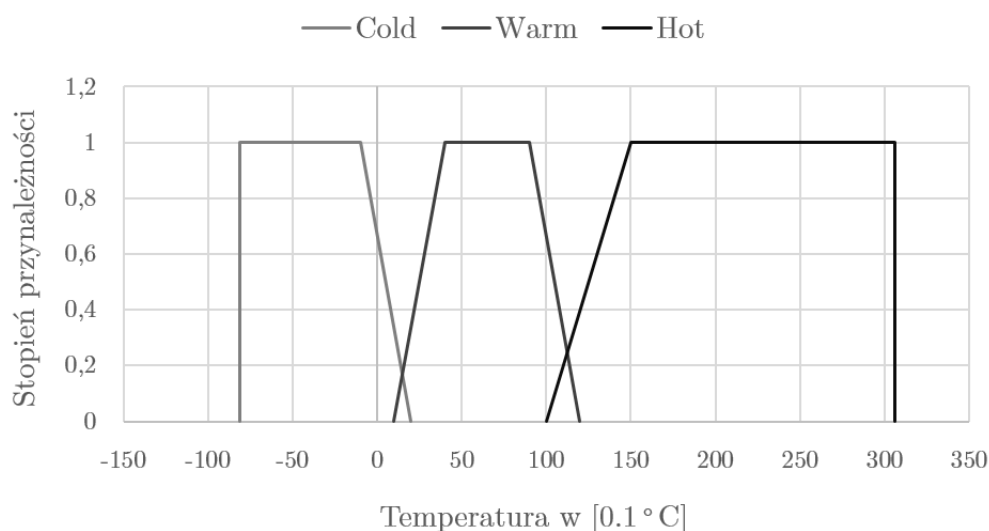
Tabela 4. Przyporządkowane parametry funkcji trapezoidalnej dla kolumny FXX.

4.2.5. Kolumna TG

W przypadku średniej dziennej temperatury (TG) oraz innych kolumn związanych z temperaturą (TN, TX oraz T10N), zdecydowaliśmy się podzielić nasze rozważania ze względu na pory roku. Dlatego też przyjęliśmy trzy różne warianty zmiennej lingwistycznej dla kolumny TG:

- TGW - dla pomiarów uzyskanych podczas astronomicznej zimy (litera *W* od *Winter*),
- TGSA - dla pomiarów uzyskanych podczas astronomicznej wiosny lub jesieni (*S* od *Spring*, *A* od *Autumn*),
- TGS - dla pomiarów uzyskanych podczas astronomicznego lata (litera *S* od *Summer*).

Rozpocznijmy od zmiennej lingwistycznej TGW.

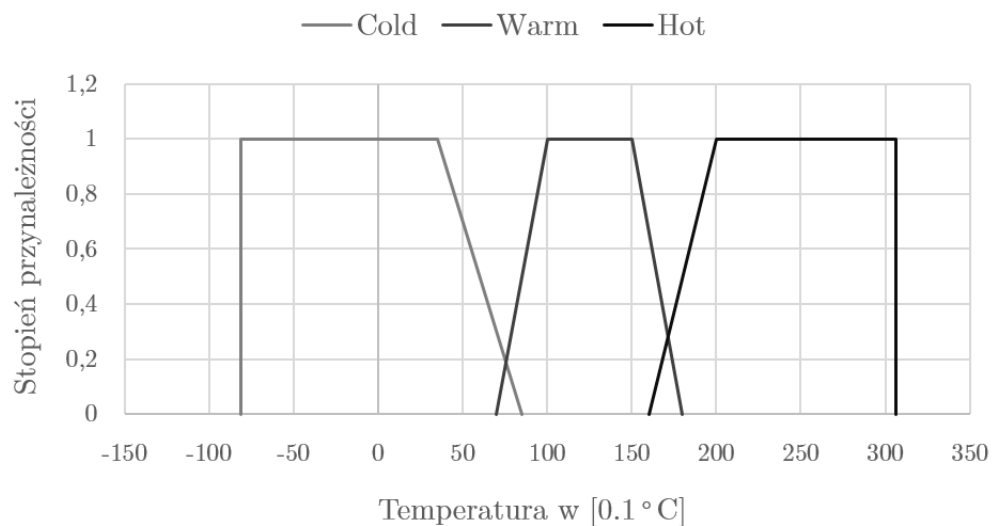


Rysunek 10. Wykres opisujący zmienną lingwistyczną dla kolumny TG dla pomiarów wykonanych astronomiczną zimą.

Etykieta	a	b	c	d
Cold	-81	-81	-10	20
Warm	10	40	90	120
Hot	100	150	306	306

Tabela 5. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TGW.

Następną prezentowaną zmienną, będzie zmienna lingwistyczna TGSA.

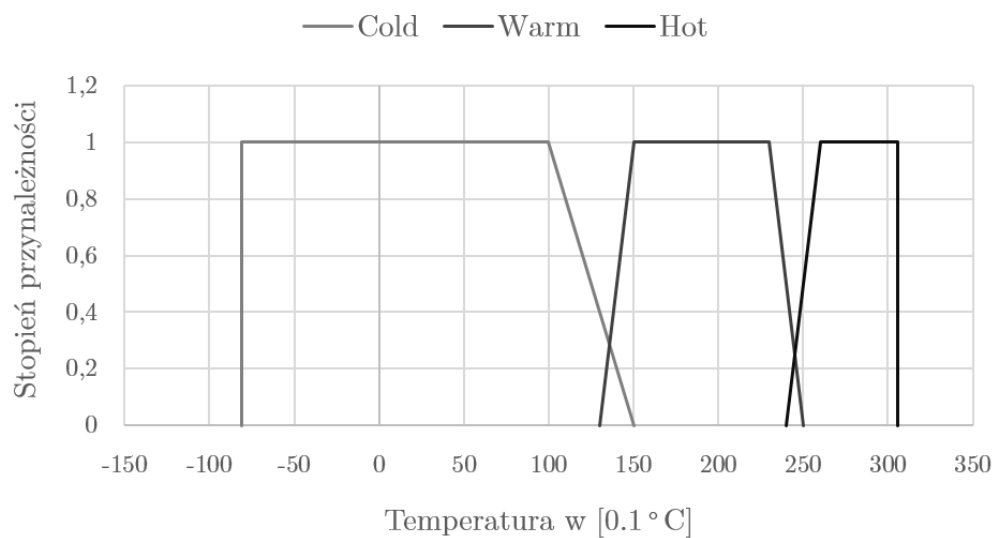


Rysunek 11. Wykres opisujący zmienną lingwistyczną dla kolumny TG dla pomiarów wykonanych astronomiczną wiosną i jesienią.

Etykieta	a	b	c	d
Cold	-81	-81	35	85
Warm	70	100	150	180
Hot	160	200	306	306

Tabela 6. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TGSA.

Ostatnią zmienną dla kolumny TG będzie zmienna dotycząca pomiarów letnich - TGS.



Rysunek 12. Wykres opisujący zmienną lingwistyczną dla kolumny TG dla pomiarów wykonanych astronomicznym latem.

Etykieta	a	b	c	d
Cold	-81	-81	100	150
Warm	130	150	230	250
Hot	240	260	306	306

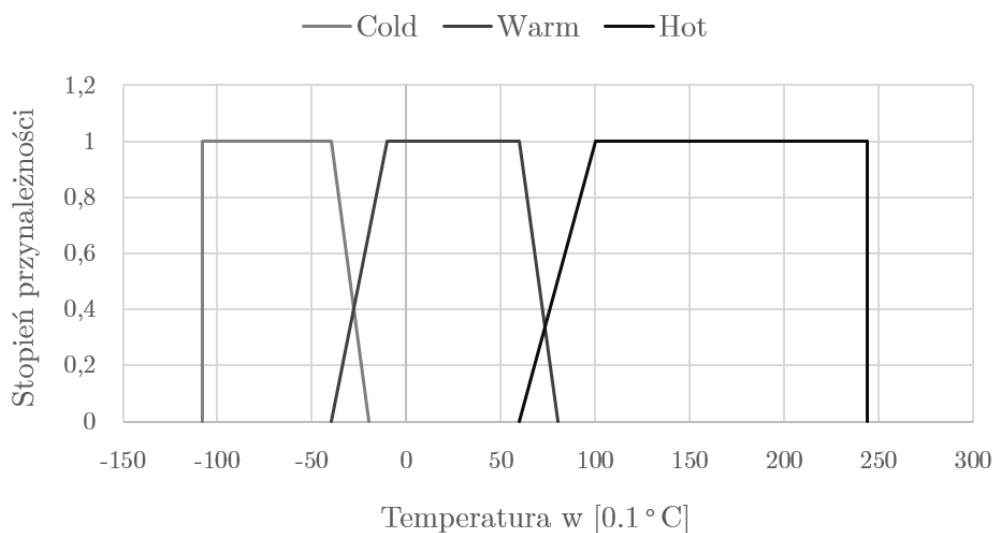
Tabela 7. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TGS.

4.2.6. Kolumna TN

Kolumna TN zawiera najniższą temperaturę powietrza w ciągu dnia. Wszystkie trzy warianty zmiennej lingwistycznej dla kolumny TN zaprezentowano poniżej:

- TNW - dla pomiarów uzyskanych podczas astronomicznej zimy,
- TNSA - dla pomiarów uzyskanych podczas astronomicznej wiosny lub jesieni,
- TNS - dla pomiarów uzyskanych podczas astronomicznego lata.

Pomiary zimowe - zmienna TNW.

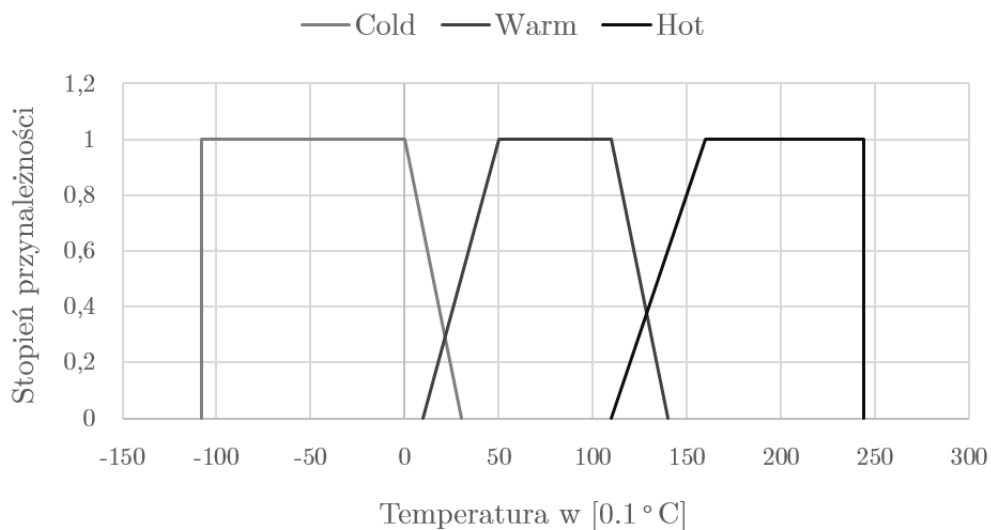


Rysunek 13. Wykres opisujący zmienną lingwistyczną dla kolumny TN dla pomiarów wykonanych astronomiczną zimą.

Etykieta	a	b	c	d
Cold	-108	-108	-40	-20
Warm	-40	-10	60	80
Hot	60	100	244	244

Tabela 8. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TNW.

Pomiary wiosenne i jesienne - zmienna TNSA.

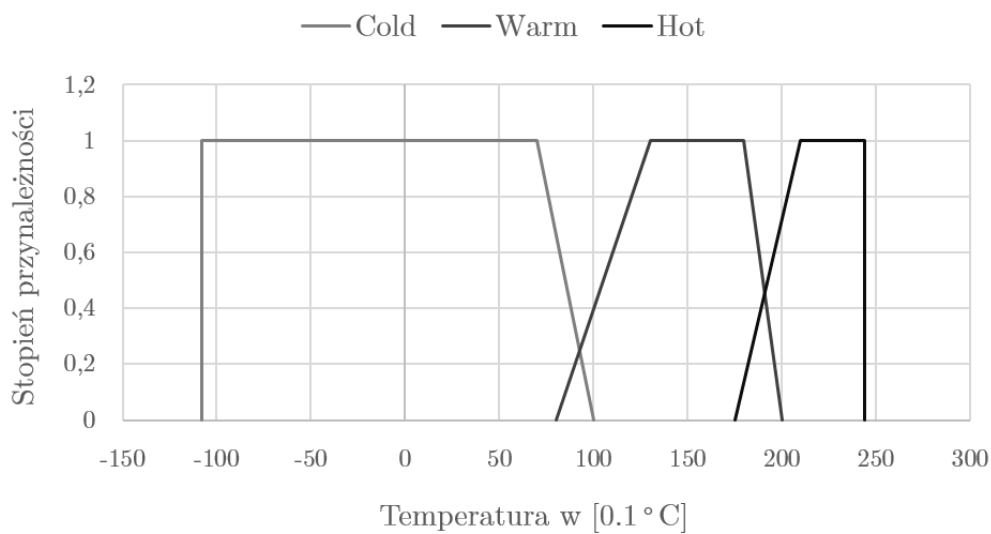


Rysunek 14. Wykres opisujący zmienną lingwistyczną dla kolumny TN dla pomiarów wykonanych astronomiczną wiosną i jesienią.

Etykieta	a	b	c	d
Cold	-108	-108	0	30
Warm	10	50	110	140
Hot	110	160	244	244

Tabela 9. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TNSA.

Pomiary letnie - zmienna TNS.



Rysunek 15. Wykres opisujący zmienną lingwistyczną dla kolumny TN dla pomiarów wykonanych astronomicznym latem.

Etykieta	a	b	c	d
Cold	-108	-108	70	100
Warm	80	130	180	200
Hot	175	210	244	244

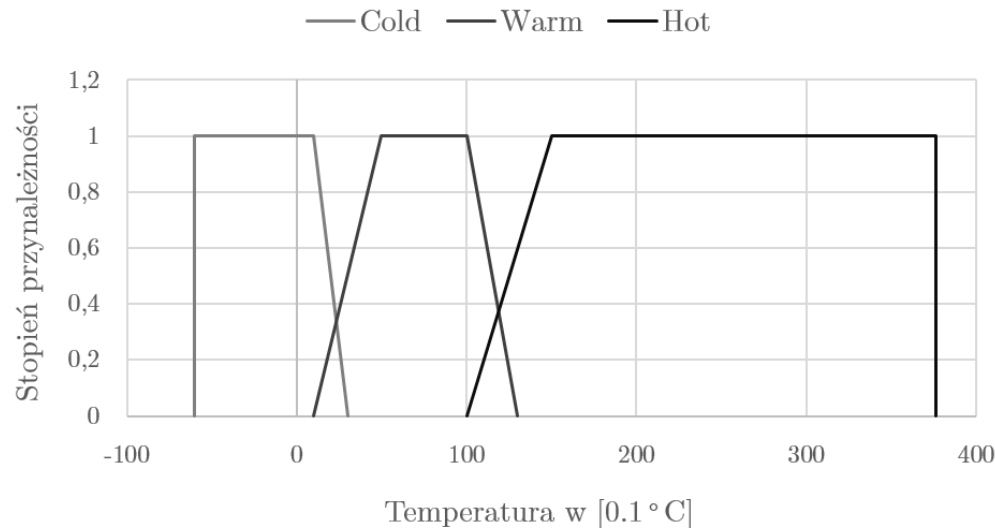
Tabela 10. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TNS.

4.2.7. Kolumna TX

Kolumna TX zawiera najwyższą temperaturę powietrza w ciągu dnia. Wszystkie trzy warianty zmiennej lingwistycznej dla kolumny TX zaprezentowano poniżej:

- TXW - dla pomiarów uzyskanych podczas astronomicznej zimy,
- TXSA - dla pomiarów uzyskanych podczas astronomicznej wiosny lub jesieni,
- TXS - dla pomiarów uzyskanych podczas astronomicznego lata.

Pomiary zimowe - zmienna TXW.

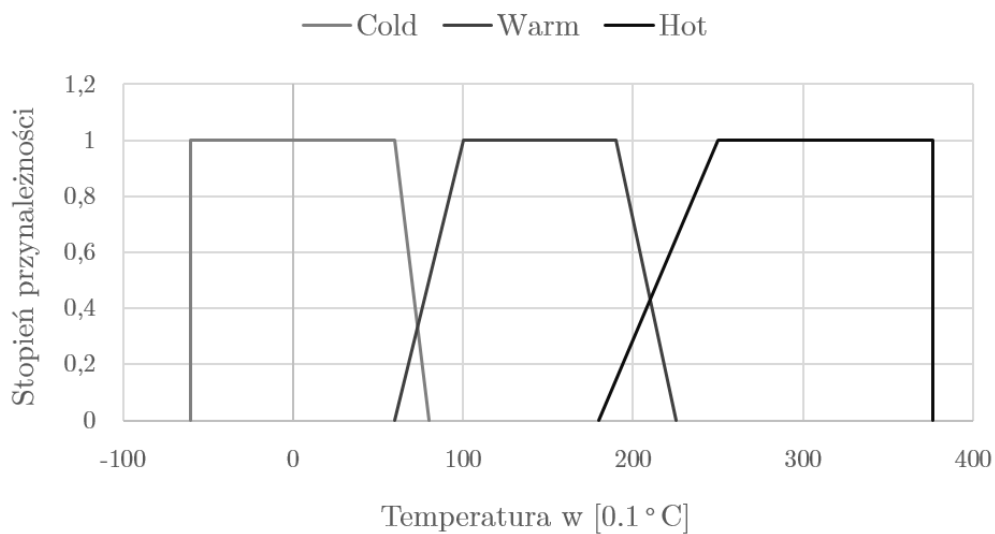


Rysunek 16. Wykres opisujący zmienną lingwistyczną dla kolumny TX dla pomiarów wykonanych astronomiczną zimą.

Etykieta	a	b	c	d
Cold	-60	-60	10	30
Warm	10	50	100	130
Hot	100	150	376	376

Tabela 11. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TXW.

Pomiary wiosenne i jesienne - zmienna TXSA.

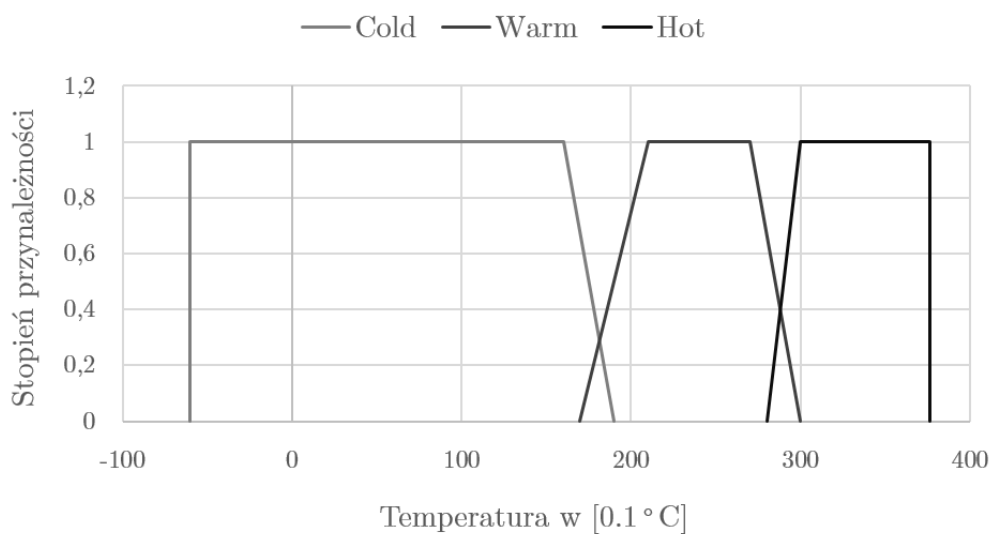


Rysunek 17. Wykres opisujący zmienną lingwistyczną dla kolumny TX dla pomiarów wykonanych astronomiczną wiosną i jesienią.

Etykieta	a	b	c	d
Cold	-60	-60	60	80
Warm	60	100	190	225
Hot	180	250	376	376

Tabela 12. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TXSA.

Pomiary letnie - zmienna TXS.



Rysunek 18. Wykres opisujący zmienną lingwistyczną dla kolumny TX dla pomiarów wykonanych astronomicznym latem.

Etykieta	a	b	c	d
Cold	-60	-60	160	190
Warm	170	210	270	300
Hot	280	300	376	376

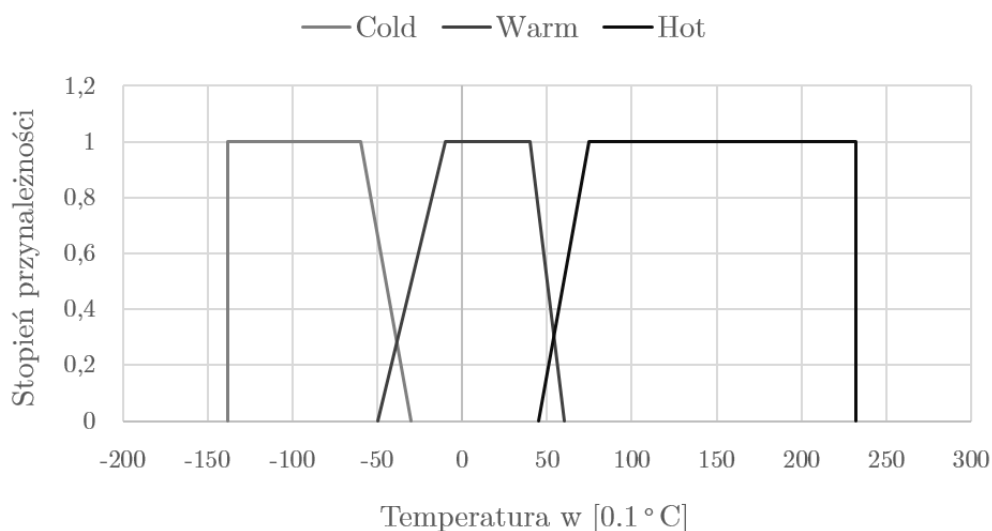
Tabela 13. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej TXS.

4.2.8. Kolumna T10N

Kolumna T10N zawiera najmniejszą temperaturę w ciągu dnia zmierzoną na wysokości 10cm od poziomu gruntu. Wszystkie trzy warianty zmiennej lingwistycznej dla kolumny T10N zaprezentowano poniżej:

- T10NW - dla pomiarów uzyskanych podczas astronomicznej zimy,
- T10NSA - dla pomiarów uzyskanych podczas astronomicznej wiosny lub jesieni,
- T10NS - dla pomiarów uzyskanych podczas astronomicznego lata.

Pomiary zimowe - zmienna T10NW.

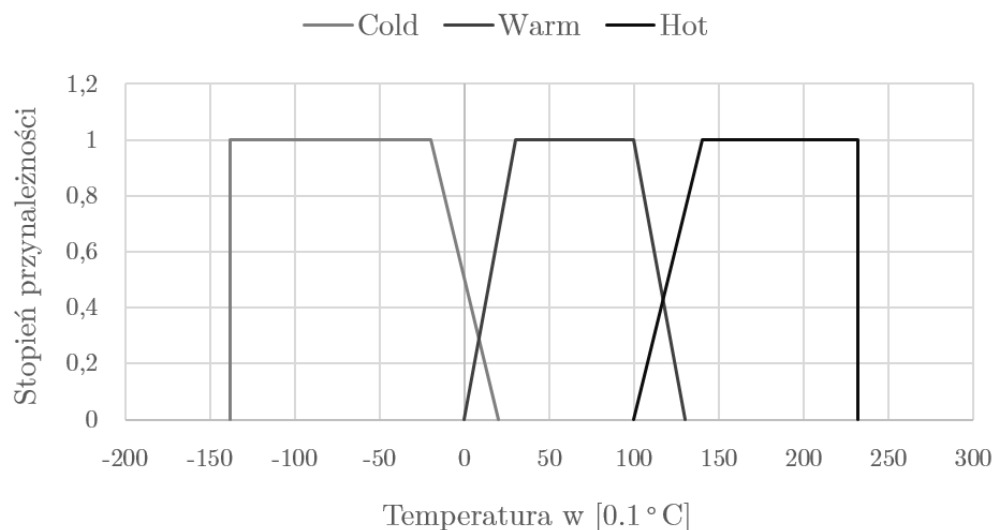


Rysunek 19. Wykres opisujący zmienną lingwistyczną dla kolumny T10N dla pomiarów wykonanych astronomiczną zimą.

Etykieta	a	b	c	d
Cold	-138	-138	-60	-30
Warm	-50	-10	40	60
Hot	45	75	232	232

Tabela 14. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej T10NW.

Pomiary wiosenne i jesienne - zmienna T10NSA.

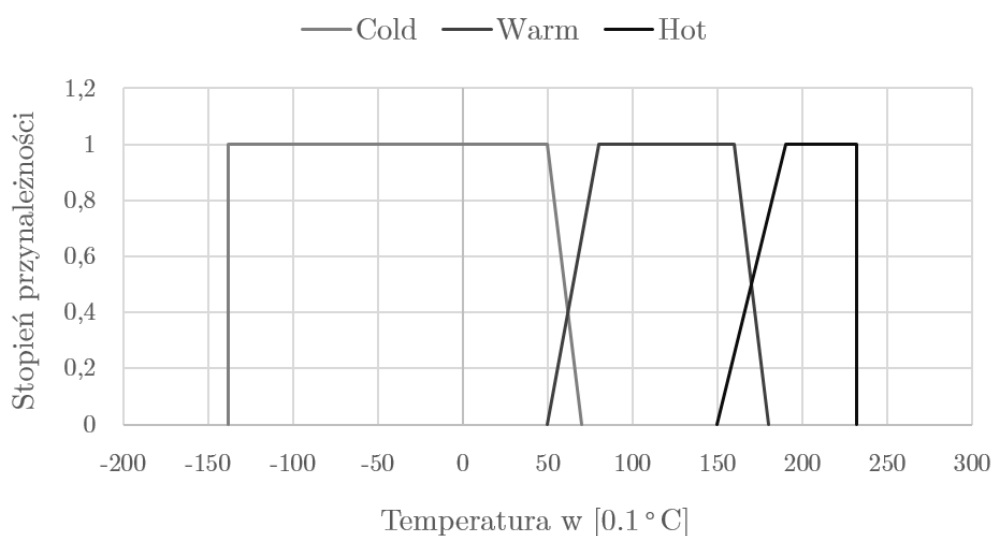


Rysunek 20. Wykres opisujący zmienną lingwistyczną dla kolumny T10N dla pomiarów wykonanych astronomiczną wiosną i jesienią.

Etykieta	a	b	c	d
Cold	-138	-138	-20	20
Warm	0	30	100	130
Hot	100	140	232	232

Tabela 15. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej T10NSA.

Pomiary letnie - zmienna T10NS.



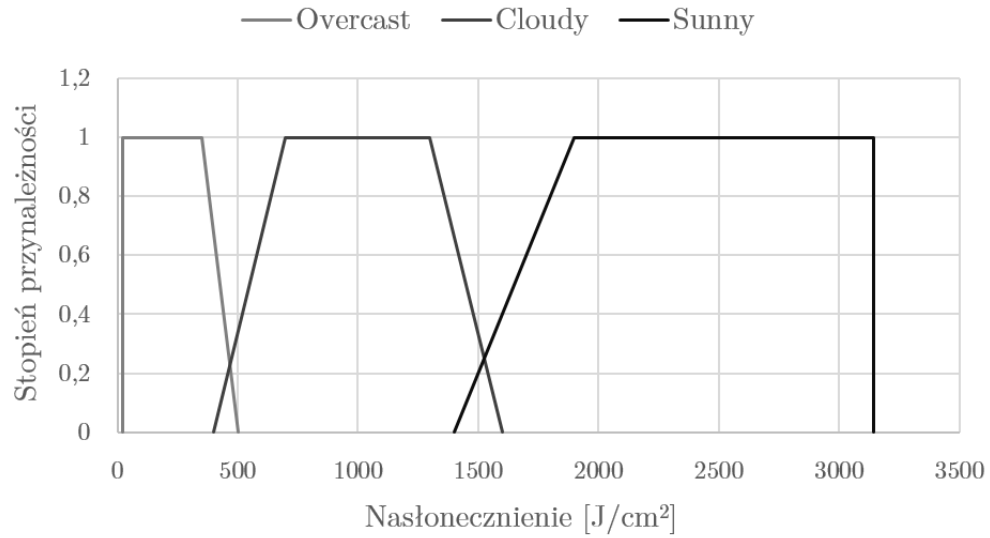
Rysunek 21. Wykres opisujący zmienną lingwistyczną dla kolumny T10N dla pomiarów wykonanych astronomicznym latem.

Etykieta	a	b	c	d
Cold	-138	-138	50	70
Warm	50	80	160	180
Hot	150	190	232	232

Tabela 16. Przyporządkowane parametry funkcji trapezoidalnej dla zmiennej T10NS.

4.2.9. Kolumna Q

Wykres opisujący zmienną lingwistyczną dla kolumny zawierającej wartości nasłonecznienia (Q), zamieszczono poniżej.



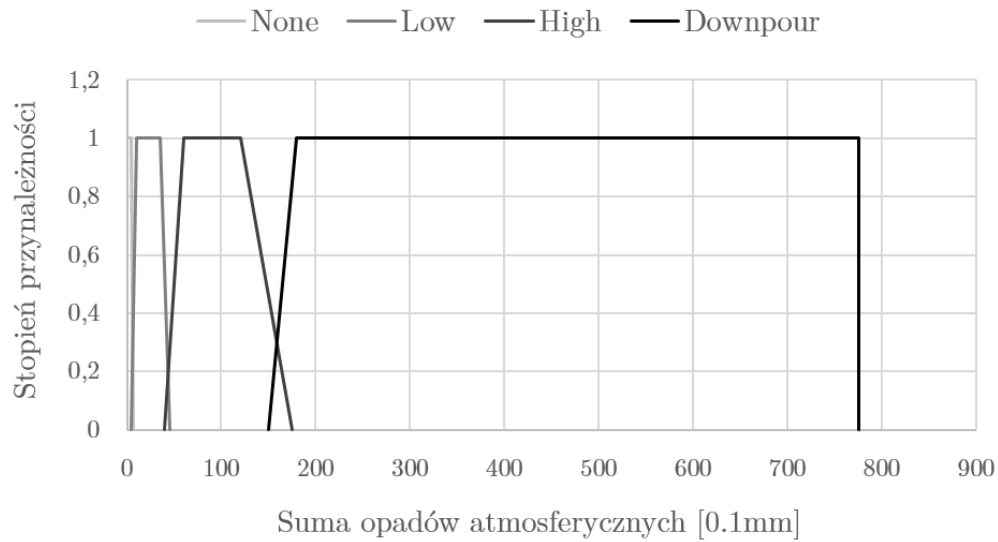
Rysunek 22. Wykres opisujący zmienną lingwistyczną dla kolumny Q

Etykieta	a	b	c	d
Overcast	24	24	350	500
Cloudy	400	700	1300	1600
Sunny	1400	1900	3145	3145

Tabela 17. Przyporządkowane parametry funkcji trapezoidalnej dla kolumny Q.

4.2.10. Kolumna RH

Wykres opisujący zmienną lingwistyczną dla kolumny zawierającej sumę opadów atmosferycznych w ciągu całego dnia (RH), zamieszczono poniżej.



Rysunek 23. Wykres opisujący zmienną lingwistyczną dla kolumny RH

Etykieta	a	b	c	d
None	-1	-1	5	7
Low	5	10	35	45
High	40	60	120	175
Downpour	150	180	776	776

Tabela 18. Przyporządkowane parametry funkcji trapezoidalnej dla kolumny RH.

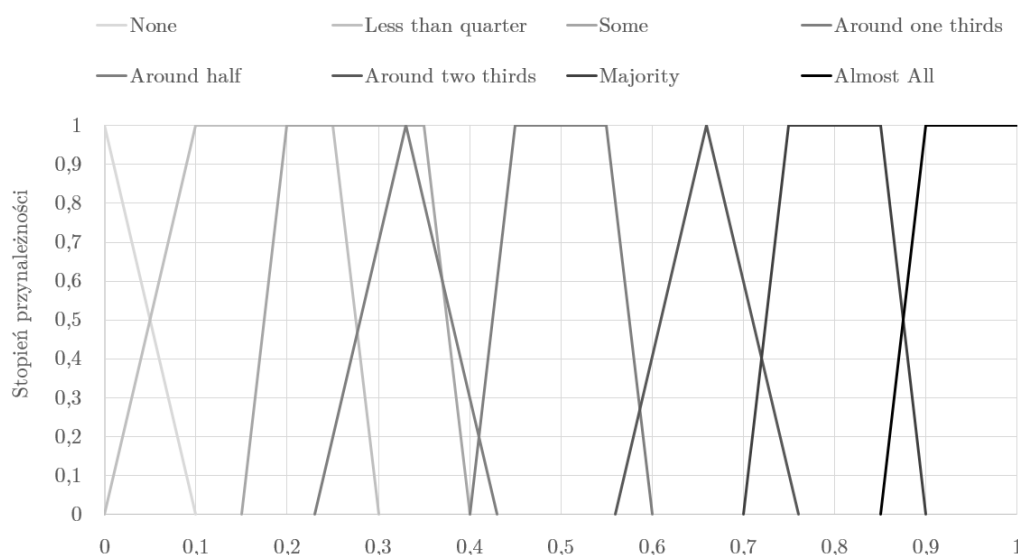
4.3. Kwantyfikatory

W tym rozdziale skoncentrujemy się na zaproponowanych przez nas kwantyfikatorach. Prezentację rozpoczniemy od kwantyfikatorów względnych, aby następnie omówić kwantyfikatory bezwzględne.

W przypadku kwantyfikatorów, wykorzystywanymi przez nas funkcjami przynależności są zarówno funkcje trapezoidalne jak i funkcje trójkątne. Za każdym razem w tabelach podano, z jakiej funkcji skorzystano przy definicji danej etykiety, co ma swoje odzwierciedlenie na prezentowanych wykresach.

4.3.1. Kwantyfikatory względne

Wykres ilustrujący wszystkie kwantyfikatory względne, zamieszczono poniżej.



Rysunek 24. Kwantyfikatory względne

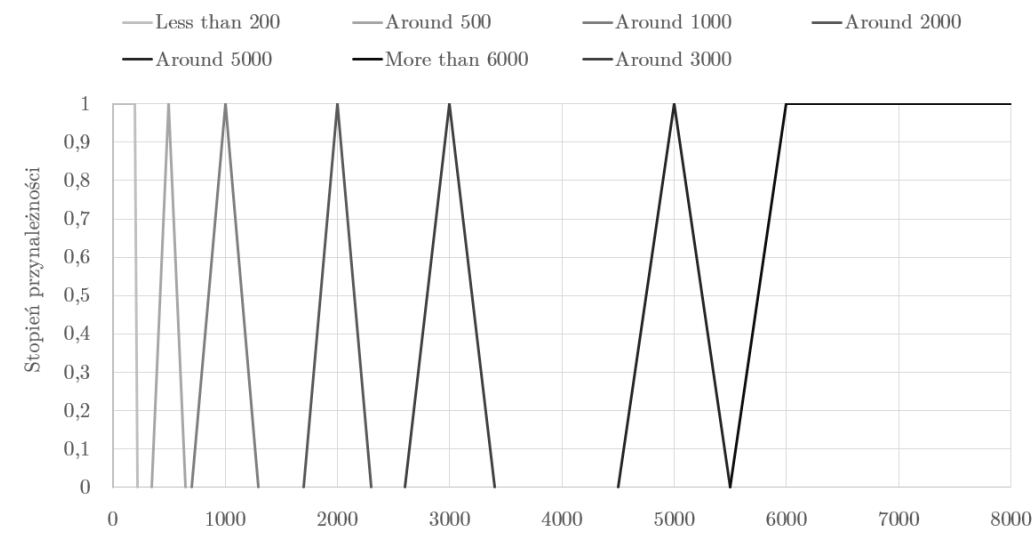
Tabela zawierająca parametry i nazwy funkcji przynależności, dla poszczególnych kwantyfikatorów, prezentuje się następująco.

Kwantyfikator	Funkcja przynależności	a	b	c	d
None	Trójkątna	0	0	0.1	-
Less than quarter	Trapezoidalna	0	0.1	0.25	0.3
Some	Trapezoidalna	0.15	0.2	0.35	0.4
Around one thirds	Trójkątna	0.23	0.33	0.43	-
Around half	Trapezoidalna	0.4	0.45	0.55	0.6
Around two thirds	Trójkątna	0.56	0.66	0.67	-
Majority	Trapezoidalna	0.7	0.75	0.85	0.9
Almost all	Trapezoidalna	0.85	0.9	1	1

Tabela 19. Funkcje przynależności kwantyfikatorów względnych - nazwy wraz z parametrami.

4.3.2. Kwantyfikatory bezwzględne

Wykres ilustrujący wszystkie kwantyfikatory bezwzględne, zamieszczono poniżej.



Rysunek 25. Kwantyfikatory bezwzględne

Tabela zawierająca parametry i nazwy funkcji przynależności, dla poszczególnych kwantyfikatorów, prezentuje się następująco.

Kwantyfikator	Funkcja przynależności	a	b	c	d
Less than 200	Trapezoidalna	0	0	200	220
Around 500	Trójkątna	350	500	650	-
Around 1000	Trójkątna	700	1000	1300	-
Around 2000	Trójkątna	1700	2000	2300	-
Around 3000	Trójkątna	2600	3000	3400	-
Around 5000	Trójkątna	4500	5000	5500	-
More than 6000	Trapezoidalna	5500	6000	8000	17000

Tabela 20. Funkcje przynależności kwantyfikatorów bezwzględnych - nazwy wraz z parametrami.

5. Wyniki

Przeprowadzone eksperymenty zostały podzielone na trzy etapy, w których porównywać będziemy różne zależności:

- W pierwszym etapie sprawdzimy jak zmieniać się będą miary podsumowań dla poszczególnych kwantyfikatorów.
- W drugim etapie przeanalizujemy zdania bardziej złożone - sumaryzatory wykorzystujące spójniki *AND* oraz *OR*.
- W trzecim etapie skoncentrujemy się na różnicy w rezultatach uzyskiwanych w wyniku wykorzystania kwalifikatora w opozycji do zdań, w których kwalifikator nie jest używany.

5.1. Wpływ wyboru kwantyfikatora na jakość podsumowania

W pierwszym etapie, wygenerowano następujące zdania:

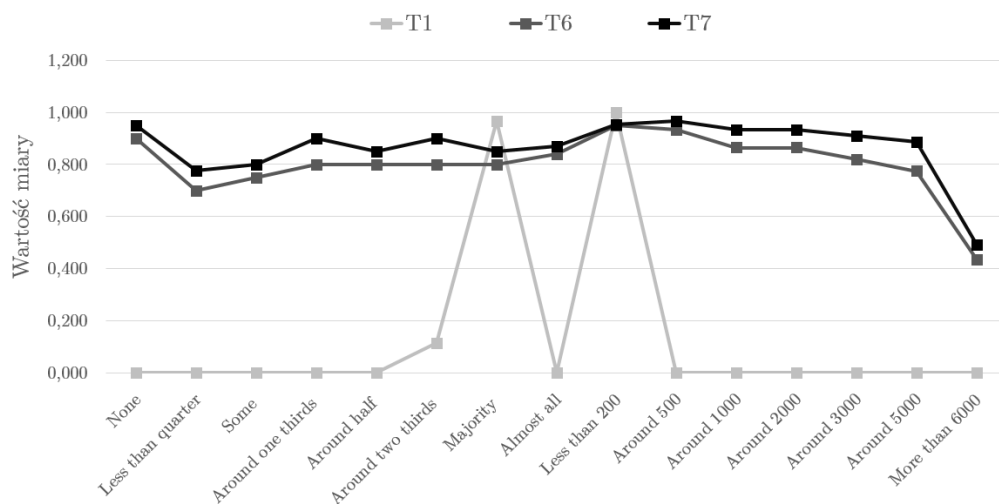
- Zdanie pierwsze: *X of summer measures with cold daily average temperature have cloudy insolation* - wykorzystujące zmienną lingwistyczną TGS jako kwalifikator (etykieta *Cold*) oraz zmienną Q jako sumaryzator (etykieta *Cloudy*).
- Zdanie drugie: *X of measures with very strong strongest wind blow have overcast insolation* - wykorzystujące zmienną lingwistyczną FXX jako kwalifikator (etykieta *Very strong*) oraz zmienną Q jako sumaryzator (etykieta *Overcast*).
- Zdanie trzecie: *X of measures with none precipitation have gentle daily wind speed average.* - wykorzystujące zmienną lingwistyczną RH jako kwalifikator (etykieta *None*) oraz zmienną FG jako sumaryzator (etykieta *Gentle*).

gdzie *X* jest jednym ze zdefiniowanych przez nas kwantyfikatorów.

Miary $T_2 - T_5$ oraz $T_8 - T_{11}$ osiągały stałe wartości w obrębie każdego ze zdań, dlatego też ich wartości zostały zamieszczone poza tabelami, w których zestawiono tylko te miary, których wartości były zmienne wewnątrz jednego eksperymentu.

5.1.1. Zdanie pierwsze

Wartości miar, które były jednakowe dla każdego kwantyfikatora, prezentują się następująco: $T_2 = 0,537$, $T_3 = 0,726$, $T_4 = 0,374$, $T_5 = 1$, $T_8 = 0,797$, $T_9 = 0,844$, $T_{10} = 0,954$, $T_{11} = 1$.



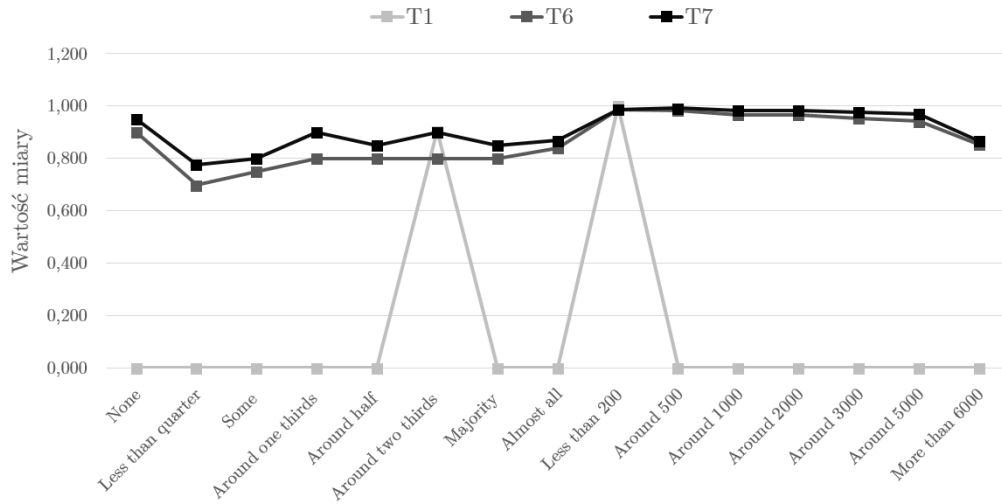
Rysunek 26. Wykres przedstawiający wyniki eksperymentu dla zdania pierwszego

Kwentyfikator	T_1	T_6	T_7
None	0,000	0,900	0,950
Less than quarter	0,000	0,700	0,775
Some	0,000	0,750	0,800
Around one thirds	0,000	0,800	0,900
Around half	0,000	0,800	0,850
Around two thirds	0,116	0,800	0,900
Majority	0,968	0,800	0,850
Almost all	0,000	0,840	0,870
Less than 200	1,000	0,950	0,953
Around 500	0,000	0,932	0,966
Around 1000	0,000	0,865	0,932
Around 2000	0,000	0,865	0,932
Around 3000	0,000	0,820	0,910
Around 5000	0,000	0,775	0,887
More than 6000	0,000	0,437	0,493

Tabela 21. Tabela przedstawiająca wyniki eksperymentu dla zdania pierwszego

5.1.2. Zdanie drugie

Wartości miar, które były jednakowe dla każdego kwantyfikatora, prezentują się następująco: $T_2 = 0,690$, $T_3 = 0,633$, $T_4 = 0,358$, $T_5 = 1$, $T_8 = 0,976$, $T_9 = 0,979$, $T_{10} = 0,991$, $T_{11} = 1$.



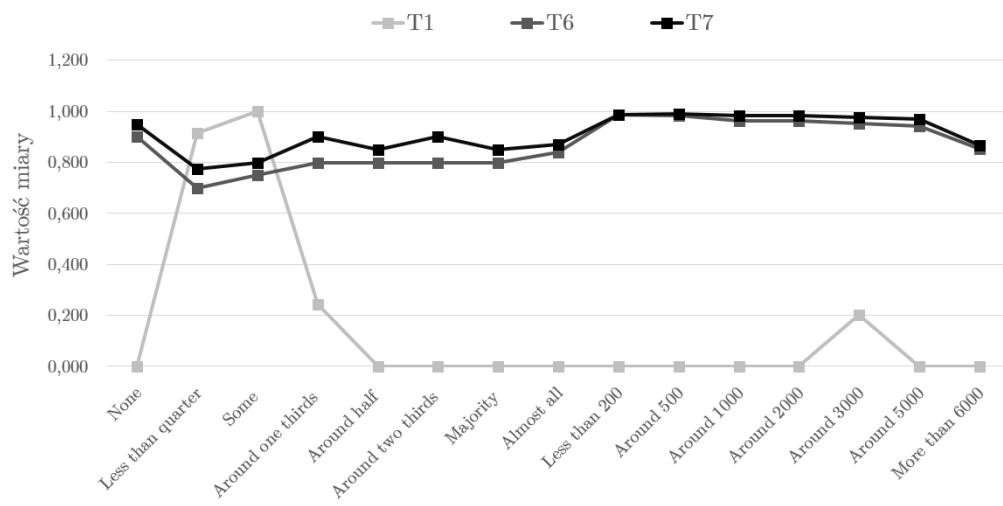
Rysunek 27. Wykres przedstawiający wyniki eksperymentu dla zdania drugiego

Kwantyfikator	T_1	T_6	T_7
None	0,000	0,900	0,950
Less than quarter	0,000	0,700	0,775
Some	0,000	0,750	0,800
Around one thirds	0,000	0,800	0,900
Around half	0,000	0,800	0,850
Around two thirds	0,905	0,800	0,900
Majority	0,000	0,800	0,850
Almost all	0,000	0,840	0,870
Less than 200	1,000	0,987	0,988
Around 500	0,000	0,982	0,991
Around 1000	0,000	0,965	0,982
Around 2000	0,000	0,965	0,982
Around 3000	0,000	0,953	0,976
Around 5000	0,000	0,941	0,971
More than 6000	0,000	0,853	0,868

Tabela 22. Tabela przedstawiająca wyniki eksperymentu dla zdania drugiego

5.1.3. Zdanie trzecie

Wartości miar, które były jednakowe dla każdego kwantyfikatora, prezentują się następująco: $T_2 = 0,742$, $T_3 = 0,332$, $T_4 = 0,141$, $T_5 = 1$, $T_8 = 0,999$, $T_9 = 0,373$, $T_{10} = 1$, $T_{11} = 1$.



Rysunek 28. Wykres przedstawiający wyniki eksperymentu dla zdania trzeciego

Kwantyfikator	T_1	T_6	T_7
None	0,000	0,900	0,950
Less than quarter	0,915	0,700	0,775
Some	1,000	0,750	0,800
Around one thirds	0,242	0,800	0,900
Around half	0,000	0,800	0,850
Around two thirds	0,000	0,800	0,900
Majority	0,000	0,800	0,850
Almost all	0,000	0,840	0,870
Less than 200	0,000	0,987	0,988
Around 500	0,000	0,982	0,991
Around 1000	0,000	0,965	0,982
Around 2000	0,000	0,965	0,982
Around 3000	0,201	0,953	0,976
Around 5000	0,000	0,941	0,971
More than 6000	0,000	0,853	0,868

Tabela 23. Tabela przedstawiająca wyniki eksperymentu dla zdania trzeciego

5.2. Wpływ wykorzystania operacji sumy i iloczynu na jakość podsumowania

Praca w toku

5.3. Wpływ wykorzystania kwalifikatora na jakość podsumowania

Praca w toku

6. Dyskusja

Praca w toku

7. Wnioski

Praca w toku

Literatura

- [1] Baza danych - "*Historical weather in the Netherlands 1901-2018*"
- [2] Narzędzie Maven
<https://maven.apache.org/>.
- [3] Biblioteka JavaFX
<https://openjfx.io/>
- [4] Methods for the linguistic summarization of data - applications of fuzzy sets and their extensions, Adam Niewiadomski, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2008
- [5] Zadeh, L. A.: 1965, 'Fuzzy sets'. *Inf. and Control* 8, 338–353.