

Data oddania: _____

Ocena: _____

Przemysław Lis 229940

Michał Olczak 229972

Projekt 2. Podsumowania lingwistyczne relacyjnych baz danych

1. Cel

Celem projektu jest stworzenie aplikacji okienkowej z dostępem do relacyjnej bazy danych, której główną funkcjonalnością jest lingwistyczna agregacja zbioru danych. Aplikacja umożliwi automatyczne generowanie podsumowań lingwistycznych służących do tworzenia krótkich wiadomości tekstowych.

2. Baza danych, zmienne lingwistyczne, kwantyfikatory lingwistyczne

Sekcja uzupełniona jako efekt zadania Tydzień 09 wg Harmonogramu Zajęć na WIKAMP KSR.

2.1. Charakterystyka podsumowywanej bazy danych

Jako zbiór danych wybraliśmy bazę danych Perth House Prices pochodzącą ze strony kaggle.com [1] (<https://www.kaggle.com/datasets/syuzai/perth-house-prices>). Jest to zbiór mieszkań pochodzących z miasta Perth w Australii. Znajduje się tam 33656 rekordów oraz 19 kolumn z czego 13 kolumn jest postaci liczbowej oraz 6 w postaci łańcucha znaków. Do naszego eksperymentu rozpatrzymy kolejne atrybuty:

1. PRICE - Cena z przedziału [51000; 2440000]
2. GARAGE - Liczba garaży z przedziału [1; 99]

3. LAND AREA - powierzchnia działki w metrach kwadratowych z przedziału [61; 999999]
4. FLOOR AREA - powierzchnia podłogi w metrach kwadratowych z przedziału [1; 870]
5. BUILD YEAR - rok budowy z przedziału [1868; 2017]
6. CBD DIST - odległość do centrum z przedziału [681; 59800]
7. NEAREST STN DIST - odległość do najbliższej stacji pociągów z przedziału [46; 35500]
8. LATITUDE - wysokość geograficzna z przedziału [-32.47297865; -31.45745]
9. LONGITUDE - szerokość geograficzna z przedziału [115.58273; 116.343201]
10. NEAREST SCH DIST - odległość do najbliższej szkoły z przedziału [0.0709120036079524; 23.2543724776336]

2.2. Zmienne lingwistyczne (atrybuty/własności obiektów)

Zmienne lingwistyczne dla wybranych 10 atrybutów z bazy danych, przedstawione w formie wykresów funkcji przynależności i wzorów analitycznych.

1. Cena

Przestrzeń rozważań zmiennej lingwistycznej: [51000; 2440000]

2.3. Kwantyfikatory lingwistyczne (liczności obiektów)

Jw. kwantyfikatory lingwistyczne – opisane etykietami, wykresami funkcji przynależności i wzorami analitycznymi. Uzasadnione wiedzą dziedzinową **zakresy i etykiety**. Precyzyjnie podane przestrzenie rozważań każdego kwantyfikatora lingwistycznego/rozmytego, wzory i wykresy dla każdej wartości/etykiety. Opisy własne z przypisami do literatury, tak by inżynier innej specjalności zrozumiał dalszy opis tego konkretnego ćwiczenia/eksperymentu.

3. Narzędzia obliczeniowe: wybór/implementacja.

Diagram UML pakietu obliczeń rozmytych i generatora podsumowań. Instrukcja użytkownika

Sekcja uzupełniona jako efekt zadania Tydzień 10 wg Harmonogramu Zajęć na WIKAMP KSR.

Diagram UML i zwięzły opis pakietu obliczeń rozmytych: źródło pakietu (zewnętrzny/własny/hybrydowy), przypis do literatury/źródeł. Krótka charakterystyka najważniejszych klas i podstawowych dla zadania ich metod.

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.

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 podmioty. 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.

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