

MED opis projektu

Semestr 2022L

Dr inż. Jarosław Chudziak

jaroslaw.chudziak@pw.edu.pl

Konsultacja: wtorki 17.00 - 18.30

Meeting: online MS Teams i/lub pokój 316

Zakres projektu

1. W projekcie wdrażamy modele z obszaru zastosowania w zakresie eksploracji tekstu.
2. W ramach projektu realizujemy dwa zadania (modele, podprojekty)
3. Pierwszy model jest wybierany ze wstępnie zdefiniowanej listy. Realizacja tego etapu może być inspirowana rozwiązaniami przedstawionymi w punkcie [1]
4. Drugi wybierany jest z listy artykułów prezentujących rozszerzenia i/lub modyfikacje algorytmu bazowego (bazy Google Scholar, ResearchGate, etc)
5. Preferowane zestawy danych wejściowych: Twitter lub fora internetowe

Metoda realizacji

Realizujemy projekty w zespołach 2 osobowych.

Projekt składa się z następujących etapów:

I. Etap 0 (**12 kwietnia**)

1. Inicjowanie spotkania
2. Odbieranie tematów
 - a. Podstawowe algorytmy
3. Szczegółowe informacje na ten temat
 - a. Zidentyfikuj problem
 - b. Określanie zestawu danych
 - c. Przygotowanie środowisk
4. Weryfikacja zrozumienia
5. Rozpoczęcie prac

II. Etap 1 (**10 maja**):

6. Pierwsza wersja raportu
 - a. Zarys
 - b. Opis algorytmu w postaci pseudokodu
 - c. Definicja danych badawczych i doświadczalnych
 - d. Planowanie eksperymentów
 - e. Zaplanuj zakres i sposób wizualizacji
7. Identyfikacja niezbędnych źródeł informacji umożliwiających skuteczną i sprawną realizację projektu (serwisy tematyczne czy Git'cie)
 - a. Uzupełnianie bibliografii i netografii
8. Implementacja algorytmu
9. Testowanie
10. Przeprowadzanie eksperymentów
11. Wizualizacja wyników
12. Analiza wyników
13. Omówienie wyników i rozszerzeń
14. Wyszukiwanie artykułów z algorytmami kandydującymi do drugiego etapu
15. Wybór drugiego algorytmu

- III. Etap od 2 do (**7 czerwca**):
1. Kolejne wersje raportu tworzone równolegle z realizacją projektu
 2. Konstruowanie algorytmu
 3. Testowanie
 4. Przeprowadzanie eksperymentów
 5. Wizualizacja wyników
 6. Analiza wyników (porównanie obu algorytmów)
 7. Omówienie wyników i rozszerzeń
- IV. Etap 3 (**14 czerwca**)
1. Prezentacja projektu do oceny
 2. Dyskusja i ocena końcowa

Następujące elementy są zaliczane i oceniane:

1. Zdefiniowanie problemu w kontekście możliwości algorytmu
2. Zrealizowany model w postaci programu w języku Python
3. Sposób przygotowania danych
4. Sposób przygotowania i przeprowadzenia eksperymentów
5. Sposób zwizualizowania i omówienia rezultatów
6. Sprawozdanie końcowe

Tematy zadań dla etapu 1:

1. DBSCAN
2. Charm
3. SPAM
4. TNS
5. GSP
6. dEclat
7. NBC
8. Grain

Narzędzia/środowiska do projektowania i współpracy:

1. Środowiska do realizacji projektów w Pythonie (z odpowiednimi bibliotekami)
2. MS Teams
3. Git (np. Gitlab)
4. Word lub Overleaf
5. Trello lub Planner
6. E-mail

Źródła wstępne bibliografia i netografia:

[1] <https://www.philippe-fournier-viger.com/spmf/> - otwarta biblioteka uczenia maszynowego i eksploracji danych

[2] Jean-Marc Adamo, Data Mining for Association Rules and Sequential Patterns, 2001

[3] Jiawei Han, Micheline Kamber, Jian Pei, Data Mining Concepts and Techniques, wydanie 3, 2012

[4] Charu C. Aggarwal • Jiawei Han, Frequent Pattern Mining, 2014

[5] Chengqi Zhang Shichao Zhang, Association Rule Mining, 2002