

Zadania projektowe z przedmiotu MED (semestr zimowy 2025/2026)

Prowadzący Grzegorz Protaziuk, email. grzegorz.protaziuk@pw.edu.pl

Konsultacje: czwartek w godz. 12.15-14.00, pok. 301, możliwość konsultacji w trybie zdalnym (teams) po uprzednim umówieniu się (np. mejlowo).

Platforma LeON kurs: <https://leon.pw.edu.pl/course/view.php?id=1976>

Wymagania

Cel: Zbadanie własności zaimplementowanego przez siebie algorytmu:

- czas wykonania dla różnych wartości parametrów wejściowych,
- wpływ zmiany wartości parametrów na uzyskiwane wyniki,
- skalowalność algorytmu względem rozmiaru danych (liczby rekordów oraz liczby atrybutów),
- oraz cele szczegółowe, jeśli są podane w opisie zadania.

Testy należy przeprowadzić dla kilku zbiorów danych

Środowisko deweloperskie

Wybór języka programowania oraz narzędzi i bibliotek należy do wykonawcy zadania, ograniczeniem jest tu legalność wykorzystania wybranych narzędzi/bibliotek do realizacji projektów studenckich.

Preferowane są popularne języki oprogramowania.

Grupy

Zadania są pomyślane jako jednoosobowe (chyba, że wprost wskazano inaczej), jednak istnieje możliwość ich realizacji w grupie maksymalnie dwuosobowej. W takim przypadku należy ustalić z prowadzącym zakres zadania projektowego.

Etapy realizacji zadania i terminy:

1. Etap 1 - do 2025.11.27 - wybór tematu.
2. Etap 2 - do 2025.12.19 - opracowanie projektu implementacji algorytmu, planu eksperymentów oraz wybór technologii do realizacji zadania – omówienie rozwiązania
3. Etap 3 - do 2026.01.23 – oddanie realizacji zadania.

Dokumentacja

Dokumentacja końcowa powinna zawierać:

- opis zadania – przyjęte założenia,
- opis projektu implementacji algorytmu (szczegóły w dalszej części dokumentu),
- opis implementacji programu – najważniejsze elementy implementacji,
- opis przeprowadzonych eksperymentów: cel, zbiory danych, sposób wykonania
- uzyskane wyniki eksperymentów oraz wnioski
- instrukcja użycia oprogramowania: specyfikacja parametrów wejściowych, sposób instalacji/uruchomienia, opis formatu danych wejściowych i wyjściowych
- kod źródłowy w osobnym pliku(ach).

Zbiory danych

<https://archive.ics.uci.edu/ml/datasets.php>,

<http://fimi.uantwerpen.be/data/>

<https://github.com/deric/clustering-benchmark>

<http://www.philippe-fournier-viger.com/spmf/index.php?link=datasets.php>

Eksperymenty można wykonać na zbiorach danych dostępnych z innych źródeł niż podane wyżej.

Procedura wyboru tematu projektu do realizacji:

1. Należy przesłać e-mailem (z frazą MED. w temacie) preferencje (3 tematy) do prowadzącego (grzegorz.protaziuk@pw.edu.pl) do 27.11.2025 r. włącznie (zapisy na temat będą realizowane wg kolejności zgłoszeń i preferencji) – z wykorzystaniem konta/adresu politechnicznego.
2. Osoby zgłaszające grupę powinny wysłać ten sam e-mail do wiadomości wszystkich pozostałych członków zespołu na adresy uczelniane.
3. Osoby, które chcą realizować własny temat muszą uzgodnić cel i zakres projektu z prowadzącym do dnia 27.11.2025 r.
4. Osoby, które nie prześlą preferencji w wyznaczonym terminie zostaną przypisane do tematu projektu arbitralnie wybranego przez prowadzącego.

Prezentacja - Etap 2

Projekt implementacji powinien zawierać informacje m.in. o strukturze implementacji (funkcje, klasy) z odniesieniem do konkretnych kroków implementowanego algorytmu, strukturach danych wykorzystywanych do przechowywania danych źródłowych oraz struktur wykorzystywanych w implementacji algorytmu oraz ocenę efektywności proponowanego rozwiązania. Projekt implementacji nie jest implementacją.

Opracowanie do etapu 2 należy przedstawić osobiście do 19 grudnia 2025r. Przedstawienie obejmuje:

- przedstawienie zakresu, wybranych technologii do realizacji zadania oraz planów testów w formie prezentacji (np. w formacie .ppt);
- omówienie projektu implementacji algorytmu (należy przygotować odpowiednią dokumentację).

Przed prezentacją należy uzgodnić z prowadzącym dokładny termin oraz przesłać/wgrać odpowiednio wcześniej (absolutne minimum to jeden dzień roboczy przed oddaniem) dokumentację projektu implementacji algorytmu.

Oddanie projektu:

Projekt należy oddać osobiście do 23.01.2026 r.

Oddanie projektu obejmuje:

1. prezentację pokazującą główne zagadnienia związane z realizowanym projektem – należy przygotować prezentację (np. w formacie .ppt);
2. pokaz działania oprogramowania
3. rozmowę dotyczącą uzyskanych wyników i wniosków.

Przed oddaniem projektu należy uzgodnić z prowadzącym dokładny termin oddania oraz przesłać/wgrać odpowiednio wcześniej (absolutne minimum to jeden dzień roboczy przed oddaniem) dokumentację końcową projektu.

Prezentacje (Etap 2 i Etap 3) mogą być zrealizowana w trybie stacjonarnym lub zdalnym. Dla trybu zdalnego wymagane jest połączenie z wizją.

Ocenianie

Projekt jest oceniany na ocenę w standardowej skali (2-5). Ocenie podlega końcowa dokumentacja projektu, jednak brak prezentacji projektu implementacji algorytmu w podanym terminie powoduje obniżenie końcowej oceny o 1 stopień.

Tematy zadań

0. Temat własny – wymaga uzgodnienia

1. Reguły asocjacyjne

Implementacja jeden z algorytmów opisany w artykułach:

- 1.1 Thanh-Long Nguyen, Bay Vo, Vaclav Snasel, Efficient algorithms for mining colossal patterns in high dimensional databases, 2017, Knowledge-Based Systems.
- 1.2] Ezeife, C.I., Su, Y. (2002). Mining Incremental Association Rules with Generalized FP-Tree. In: Cohen, R., Spencer, B. (eds) Advances in Artificial Intelligence. Canadian AI 2002. Lecture Notes in Computer.
- 1.3 Jiawei Han and Yongjian Fu, "Mining multiple-level association rules in large databases," in IEEE Transactions on Knowledge and Data Engineering, vol. 11, no. 5, pp. 798-805, Sept.-Oct. 1999, doi: 10.1109/69.806937.

2. Równoległe algorytmy wykrywania reguł asocjacyjnych

Implementacja jeden z algorytmów opisanych w artykułach:

- 2.1 R. Agrawal and J. Shafer, "Parallel Mining of Association Rules, IEEE Transactions On Knowledge And Data Engineering, Vol 8, No 6, December 1996
- 2.2 Mohammed J. Zaki, Srinivasan Parthasarathy, Wei Li, A Localized Algorithm for Parallel Association Mining, 9th Annual ACM Symposium on Parallel Algorithms and Architectures (SPAA), pp 321-330, Newport, Rhode Island, June 22-25, 1997.

3. Grupowanie

Implementacja algorytmu powinna umożliwiać grupowanie obiektów opisanych wieloma atrybutami, zarówno numerycznymi jak i nominalnymi. Należy ocenić przydatność algorytmu do grupowania obiektów opisanych atrybutami: 1) numerycznymi, 2) nominalnymi i numerycznymi.

3.1 Implementacja algorytmu CURE (Clustering Using Representative).

Literatura: Sudipto Guha, Rajeev Rastogi, Kyusok Shim, CURE: An Efficient Clustering Algorithm for Large Databases, 1998 (citeseer.ist.psu.edu)

3.2 Implementacja algorytmu NBC (Neighborhood-Based Clustering algorithm)

Listeratura: S. Zhou, Y. Zhao, J. Guan, and J. Huang, "A Neighborhood-Based Clustering Algorithm," in Advances in Knowledge Discovery and Data Mining

4. Wzorce sekwencyjne

Implementacja algorytmu do odkrywanie częsty wzorców sekwencyjnych: PrefixSpan , ERMiner

Literatura:

- 4.1 Pei, J., Han, J., Mortazavi-Asl, B., Pinto, H., Chen, Q., Dayal, U., Hsu, M., "Mining Sequential Patterns by Pattern-Growth: The PrefixSpan Approach", IEEE Trans. Knowledge and Data Engineering,
- 4.2 Fournier, P., Zida, S., Guenieche, T. and Tseng V., "ERMiner: Sequential Rule Mining using Equivalence Classes", Advanced in intelligent data Analysis,