

SPDB – projekty, semestr 2025L

Procedura wyboru tematu projektu do realizacji:

1. Należy przesłać e-mailem preferencje (minimum 3 tematy) do prowadzących (robert.bembenik@pw.edu.pl, grzegorz.protaziuk@pw.edu.pl) do 19.03.2025 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.
3. Osoby, które chcą realizować własny temat muszą uzgodnić cel i zakres projektu z jednym z prowadzących do dnia 19.03.2025 r.
4. Osoby, które nie prześlą preferencji w wyznaczonym terminie zostaną przypisane do tematu projektu arbitralnie wybranego przez prowadzących.

Prezentacja kluczowych elementów rozwiązania:

Kluczowe elementy rozwiązania należy przedstawić osobiście do 25.04.2025 r. Przedstawienie obejmuje:

- przedstawienie idei rozwiązania w formie prezentacji (np. w formacie .ppt);
- omówienie planowanego rozwiązania.

Przed prezentacją kluczowych elementów projektu należy uzgodnić z prowadzącym dokładny termin.

Oddanie projektu:

Projekt należy oddać osobiście do 30.05.2025 r. z zachowaniem możliwości zwolnienia z jednego ze sprawdzianów lub do 11.06.2025 r., bez możliwości zwolnienia ze sprawdzianu.

Oddanie projektu obejmuje:

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

Przed oddaniem projektu należy uzgodnić z prowadzącym dokładny termin oddania oraz przesłać odpowiednio wcześniej (absolutne minimum to jeden dzień roboczy przed oddaniem) dokumentację projektu (w tym również kody źródłowe oprogramowania, jeśli takie było napisane).

Tematy projektów

1. Analiza danych dotyczących przejazdów hulajnogami elektrycznymi w jednym z miast europejskich. Analiza powinna dotyczyć okresu min. 1 miesiąca. (Dane można znaleźć w portalu <https://data.europa.eu/pl>). W szczególności zadanie obejmuje:
 - stworzenie aplikacji do pobierania danych i pobranie potrzebnych danych,
 - analizę przemierzanych przez hulajnogi tras w dni robocze vs. dni świąteczne/weekendy,
 - znalezienie grup miejsc rozpoczynania podróży i miejsc docelowych,
 - wizualizację wyników z wykorzystaniem R, Python, PowerBI lub Tableau.

Dokumentacja projektowa powinna zawierać:

- zbiór danych i opis danych wykorzystanych w eksperymentach;
- wyniki przeprowadzonych analiz i wnioski dotyczące możliwości ich praktycznego wykorzystania;
- charakterystykę stworzonych wizualizacji.

Proponowana liczba osób w grupie: 2-3.

2. Implementacja aplikacji do określenia najkorzystniejszego czasu wyjazdu.
- Dla podanych przez użytkownika miejsc początkowych (kilku), miejsca docelowego oraz czasu (przedziału) oczekiwanego dotarcia do miejsca docelowego aplikacja powinna podać czas wyjazdu z każdego miejsca początkowego umożliwiający dotarcie do miejsca docelowego w podanym terminie w najkrótszym czasie.
 - Dla podanych przez użytkownika miejsc docelowych (kilku), miejsca początkowego oraz czasu (przedziału) możliwego wyjazdu dla każdego miejsca docelowego aplikacja powinna podać czas wyjazdu z zadanego przedziału czasu umożliwiający dotarcie do tego miejsca w najkrótszym czasie.
- W obu przypadkach należy zamodelować zmienny w czasie (np. w ciągu doby) czas przejazdu przez dane odcinki drogi.

Dokumentacja końcowa powinna zawierać:

- Opis metody wyznaczania czasu wyjazdu.
- Opis metody modelowania zmiennego czasu przejazdu przez odcinki.
- Wyniki eksperymentów pozwalające na ocenę jakości wyników.
- Informacje o implementacji (skomentowany kod źródłowy, zastosowany model danych, architektura aplikacji).

Proponowana liczba osób w grupie: 2 (realizacja jednego z wariantów)

3. Implementacja aplikacji do wyznaczania najlepszej trasy wg zadanych parametrów, opracowanie metody oraz wykonanie testów sprawdzających praktyczną użyteczność zaproponowanych rozwiązań (wyznaczanie trasy w miastach dla 5 lub więcej podanych miejsc do odwiedzenia). Podstawowymi parametrami do wyboru najlepszej trasy jest: czas i odległość. Dodatkowym wymogiem jest minimalizacja skrętów w lewo.

Dokumentacja końcowa powinna zawierać:

- Opis metody wyznaczania najlepszej drogi. Opis modelu danych.
- Opis architektury aplikacji.
- Informacje o implementacji (wykorzystane algorytmy + skomentowany kod źródłowy).
- Opis i wyniki przeprowadzonych testów.

Proponowana liczba osób w grupie: 2

4. Implementacja aplikacji pozwalającej na wyznaczenie optymalnej trasy turystycznej (możliwej do przebycia pieszo, samochodem, komunikacją miejską, lub kombinacją tych sposobów) w celu odwiedzenia zadanych miejsc (miejscami są podane przez użytkownika obiekty turystyczne i/lub restauracje na trasie wycieczki). Dla każdego miejsca użytkownik określa zakładany czas pobytu (przedział np. do 2 do 3 godzin). Każde miejsce ma podane godziny otwarcia. Dodatkowo użytkownik podaje: miejsce i godzinę rozpoczęcia wycieczki, może podać: godzinę zakończenia, oraz że chce odwiedzić dane miejsce (ew. typ miejsca) w określonym przedziale czasu. Wynikiem działania, oprócz wyznaczenia i wizualizacji trasy na mapie, powinna być informacja o czasie potrzebnym na realizację wycieczki.

Dokumentacja końcowa powinna zawierać:

- Wyniki eksperymentów pozwalające na ocenę jakości wyników.
- Informacje o implementacji (skomentowany kod źródłowy, zastosowany model danych, architektura aplikacji).

Proponowana liczba osób w grupie: 2-3

5. Porównanie wydajności modułów przestrzennych w najnowszych wersjach systemów zarządzania bazami danych (MySQL, SQL Server, PostgreSQL, Oracle, MongoDB, Casandra,...). W ramach realizacji projektu należy przetestować czas realizacji zapytań typu:

- punktowe: podać wszystkie obiekty, które zawierają dany punkt;
- przecięcie: znaleźć obiekty, które mają część wspólną z danym odcinkiem/innym obiektem;
- zawieranie: podać wszystkie obiekty zawarte w danym obszarze (obszar: prostokąt lub większy obiekt);
- zapytania dotyczące sąsiadów: wyznaczanie k najbliższych sąsiadów, podać wszystkie obiekty (określonego typu) znajdujące się w zadanej odległości;
- obliczanie pola obiektów;
- obliczanie odległości,
- zapytań skorelowanych.

oraz skalowalność systemu względem ilości danych.

Do testowania należy wykorzystać dane udostępnione przez serwisy oferujące dane przestrzenne (np. serwis OpenStreetMap <http://www.openstreetmap.org/>).

Dokumentacja końcowa powinna zawierać:

- Schemat danych wraz z opisem.
- Informacje o zastosowanych indeksach przestrzennych i nieprzestrzennych.
- Informacje o danych: krótka charakterystyka (rozmiar danych, ile obiektów, jakich typów, etc.).
- Zapytania wykorzystane w testach – skrypt do wykonania w SZBD
- Plan wykonania zapytań – przynajmniej jeden na dany typ zapyłania.
- Wyniki testów oraz komentarz do testów.

Proponowana liczba osób w grupie: 2 (jeden SZBD na osobę).

6. Zbadania możliwości przechowywania danych o grafach i wykonywania operacji na tych danych oferowanych wybranej najnowszej wersji SZBD (MySQL, SQL Server, PostgreSQL, Oracle, MongoDB, Casandra,...). W ramach realizacji projektu należy sprawdzić sposoby zapisu danych o grafach oraz możliwości realizacji typowych operacji na grafach: m.in. edycja danych, znajdowanie wierzchołków sąsiadujących, pobieranie ścieżki, znajdowanie wierzchołków o zadanych parametrach, sprawdzenie występowania cykli itp.

Dokumentacja końcowa powinna zawierać:

- Schemat danych wraz z opisem.
- Informacje o zastosowanych indeksach przestrzennych i nieprzestrzennych.
- Informacje o danych: krótka charakterystyka (rozmiar danych, ile obiektów, jakich typów, etc.).
- Zapytania wykorzystane w testach – skrypt do wykonania w SZBD.
- Wyniki testów oraz komentarz do testów.

Proponowana liczba osób w grupie: 2 (jeden SZBD na osobę).

7. Zadane jest miejsce początkowe i miejsce docelowe podróży. Napisać aplikację, która wskaże miejsca do odwiedzenia (dowolnie wybrane tzw. punkty zainteresowanie - POI) oraz wyznaczy trasę przejazdu przy spełnieniu ograniczeń dotyczących trasy: wydłużenia czasu przejazdu, długości trasy oraz okresu w czasie podróży, w którym dany punkt ma być odwiedzony (np. po godzinie od rozpoczęcia podróży). Dodatkowo użytkownik określa liczbę POI do odwiedzenia oraz czas podróży (przedział) pomiędzy kolejnymi POI.

Aplikacja powinna zawierać moduł do pokazania wyznaczonej trasy na mapie. Do napisania aplikacji należy wykorzystać dostępne serwisy oferujące dane przestrzenne (np. serwis OpenStreetMap <http://www.openstreetmap.org/>).

Dokumentacja końcowa powinna zawierać:

- Opis metody wyszukiwania.
- Wyniki eksperymentów pozwalające na ocenę jakości wyników wyszukiwania.
- Informacje o implementacji (skomentowany kod źródłowy, zastosowany model danych, architektura aplikacji).

Proponowana liczba osób w grupie: 1

8. Implementacja algorytmu odkrywania kolokacji przestrzennych przedstawionego w artykule Yan Huang, Shashi Shekhar, and Hui Xiong. "Discovering colocation patterns from spatial data sets: a general approach". In: IEEE Transactions on Knowledge and data engineering 16.12 (2004), pp. 1472–1485

Dokumentacja końcowa powinna zawierać:

- Opis (pseudokod) algorytmu, jeżeli były dokonywane zmiany w porównaniu z algorytmem oryginalnym.
- Wyniki testów (wykresy + komentarze).
- Skomentowany kod.

Proponowana liczba osób w grupie: 2

9. Implementacja wybranego algorytmu eksploracji danych.

Algorytmy, które można badać/implementować są opisane w artykułach, których listę zamieszczono na końcu bieżącego dokumentu. Są to przykładowe algorytmy. Możliwe jest wybranie innego algorytmu (po wcześniejszym uzgodnieniu).

Dokumentacja końcowa powinna zawierać:

- Opis zastosowanego modelu
- Opis (pseudokod) algorytmu, jeżeli były dokonywane zmiany w porównaniu z algorytmem oryginalnym.
- Wyniki testów (wykresy + komentarze).
- Skomentowany kod.

Proponowana liczba osób w grupie: 1

10. Wizualizacja analitycznych zapytań SQL.

Należy utworzyć przykładową bazę danych w systemie PostGIS, a następnie zaimportować do niej przykładowe dane (np. dotyczące przestępstw). Po załadowaniu danych do odpowiednich tabel i wypełnieniu przestrzennej b.d., należy dokonać wizualizacji wyników zapytań SQL. Przykładowe wizualizacje obejmują: analizę zdarzeń w zadanym poligonie, prezentację zdarzeń w określonym przedziale czasowym, zastosowanie bufora dla wybranych lokalizacji (np. zdarzenia w określonym buforze) oraz identyfikację najbliższych sąsiadów (nearest neighbors). Należy założyć przykładowe scenariusze uszczegóławiające co jest celem analiz.

Dokumentacja końcowa powinna zawierać:

- Schemat danych wraz z opisem.
- Informacje o zastosowanych indeksach przestrzennych i nieprzestrzennych.
- Informacje o danych: krótka charakterystyka (rozmiar danych, ile obiektów, jakich typów, etc.).
- Zapytania wykorzystane w testach – skrypt do wykonania w SZBD oraz ich wizualizacje
- Wyniki testów oraz komentarz do testów.

Proponowana liczba osób w grupie: 1-2

11. Porównanie możliwości realizacji zapytań o dane przestrzenne w środowiskach R/Python oraz PostGIS lub innym wybranym SZPBD.

Celem projektu jest analiza porównawcza dostępnych w wybranych środowiskach funkcji do wyszukiwania danych przestrzennych w zbiorach. W zadaniu należy wskazać dostępne możliwości wyszukiwania: na podstawie relacji topologicznych, relacji odległościowych, możliwości złączeń zbiorów danych przestrzennych, funkcji agregujących.

Dokumentacja projektowa powinna zawierać:

- opis środowiska, w którym zostały wykonane testy;
- zbiory danych i opis danych wykorzystanych w eksperymentach;
- informacje o sposobie wczytania/zapisania danych, w szczególności do systemu bazodanowego;
- wyniki przeprowadzonych testów, w tym wnioski i ocena (co się udało pokazać daną metodą, gdzie to widać, jaka jest interpretacja);
- skrypt R/Python, który umożliwi ponowne wykonanie prezentowanych w raporcie badań.

Proponowana liczba osób w grupie: 2.

12. Porównanie możliwości przekształceń danych przestrzennych w środowiskach R/Python oraz PostGIS lub innym wybranym SZPBD.

Celem projektu jest analiza porównawcza dostępnych w wybranych środowiskach funkcji do przekształcania danych przestrzennych. W zadaniu należy zbadać dostępne funkcje dotyczące: upraszczania danych, wyznaczania środków oraz buforów, transformacji afiniczne, przycinania obiektów, tworzenia unii geometrycznych oraz metod transformacji typów.

Dokumentacja projektowa powinna zawierać:

- opis środowiska, w którym zostały wykonane testy;
- zbiory danych i opis danych wykorzystanych w eksperymentach;
- informacje o sposobie wczytania/zapisania danych, w szczególności do systemu bazodanowego;
- wyniki przeprowadzonych testów, w tym wnioski i ocena (co się udało pokazać daną metodą, gdzie to widać, jaka jest interpretacja);
- skrypty, który umożliwią ponowne wykonanie prezentowanych w raporcie badań.

Proponowana liczba osób w grupie: 2.

13. Zbadanie funkcjonalności dostępnej w środowisku R/Python dotyczącej predykcji przestrzennej metodą kriging (uniwersalny, zwykły, prosty), predykcji IDW (IDW - inverse distance weighted interpolation) oraz regresji liniowej. Celem zadania jest ocena dostępnych metod pod kątem ich przydatności do analizy danych przestrzennych: poprawność działania, czas wykonania w zależności od wielkości danych, łatwość użycia, dostępność materiałów pomocniczych.

Dokumentacja projektowa powinna zawierać:

- opis metod reprezentacji i wizualizacji;
- opis środowiska, w którym zostały wykonane testy;
- zbiory danych i opis danych wykorzystanych w eksperymentach;
- wyniki przeprowadzonych testów, w tym wnioski i ocena (co się udało pokazać daną metodą, gdzie to widać, jaka jest interpretacja);
- skrypt R/Python, który umożliwi ponowne wykonanie prezentowanych w raporcie badań.

Proponowana liczba osób w grupie: 1.

14. Zbadanie funkcjonalności dostępnej w środowisku R/Python dotyczącej regresji przestrzennej, w tym autoregresji. Celem zadania jest ocena dostępnych metod pod kątem ich przydatności do analizy danych przestrzennych: poprawność działania, czas wykonania w zależności od wielkości danych, łatwość użycia, dostępność materiałów pomocniczych.

Dokumentacja projektowa powinna zawierać:

- opis metod reprezentacji i wizualizacji;
- opis środowiska, w którym zostały wykonane testy;
- zbiory danych i opis danych wykorzystanych w eksperymentach;
- wyniki przeprowadzonych testów, w tym wnioski i ocena (co się udało pokazać daną metodą, gdzie to widać, jaka jest interpretacja);
- skrypt R/Python, który umożliwi ponowne wykonanie prezentowanych w raporcie badań.

Proponowana liczba osób w grupie: 1.

15. Zbadanie funkcjonalności dostępnej w środowisku R/Python dotyczącej tworzenia map: w tym map statycznych, animowanych, interaktywnych. Badanie należy przeprowadzić na różnej wielkości danych oraz różnego typu danych (w tym, jeżeli jest to możliwe, na danych dostępnych w środowisku R/Python).

Dokumentacja projektowa powinna zawierać:

- opis metod reprezentacji i wizualizacji;
- opis środowiska, w którym zostały wykonane testy;
- zbiory danych i opis danych wykorzystanych w eksperymentach;
- wyniki przeprowadzonych testów, w tym wnioski i ocena (co się udało pokazać daną metodą, gdzie to widać, jaka jest interpretacja);
- skrypt R/Python, który umożliwi ponowne wykonanie prezentowanych w raporcie badań.

Proponowana liczba osób w grupie: 1.

Literatura

Analiza danych przestrzennych w R

- Lovelace, Robin, Jakub Nowosad, and Jannes Muenchow. *Geocomputation with R*. Chapman and Hall/CRC Press, 2019 (<https://geocompr.robinlovelace.net/>)
- Bivand, Roger S., et al. *Applied spatial data analysis with R*. New York: Springer, 2013 (książka jest dostępna online w bibliotece PW)
- Brunsdon, Chris, and Lex Comber. *An introduction to R for spatial analysis and mapping*. Sage, 2015.

Analiza danych przestrzennych w Python-ie

- Toms, S., Rees, E. V., & Crickard, P. (2018). *Mastering Geospatial Analysis with Python*. Packt Publishing Ltd.
- Spatial data programming with Python, <https://geobgu.xyz/py/>
- Bonny McClain, *Python for Geospatial Data Analysis: Theory, Tools, and Practice for Location Intelligence*, O'Reilly Media, 2022

Grupowanie

- Guha S., Rastogi R., Shim K., *ROCK: A robust clustering algorithm for categorical attributes*, Proceedings of the International Conference on Data Engineering, Sydney 1999, pp. 512–521.
- Karypis G., Han E., Kumar V., *CHAMELEON: A hierarchical clustering algorithm using dynamic modeling*, IEEE Computer: Special Issue on Data Analysis and Mining, vol. 32, no. 8, 1999, pp. 68–75.
- Ng R. T., Han J., *Efficient and effective clustering methods for spatial data mining*, Proc. 20th Int. Conf. on Very Large Data Bases, Morgan Kaufmann, Santiago 1994, pp. 144–155.
- Ester M., Kriegel H.-P., Sander J.: *Algorithms and applications for spatial data mining*, Geographic Data Mining and Knowledge Discovery, Research Monographs in GIS, Taylor and Francis, 2001
- Estivill-Castro V., Lee I.: *AMOEB: hierarchical clustering based on spatial proximity using Delaunay diagram*, In Proceedings of the 9th International Symposium on Spatial Data Handling, 2000
- Estivill-Castro V., Lee I.: *AUTOCLUST: automatic clustering via boundary extraction for mining massive point-data sets*, Proceedings of the 5th International Conference on Geocomputation, 2000
- Zhang T., Ramakrishnan R., Linvy M., *BIRCH: an efficient data clustering method for very large databases*, Proc. ACM SIGMOD Int. Conf. on Management of Data, ACM Press, 1996, pp. 103–114.

Klasyfikacja

- Koperski K., Han J., Stefanovic N.: *An efficient two-step method for classification of spatial data*, Proc. Int. Symp. on Spatial Data Handling (SDH '98), 1998
- Frank, R., Ester, M. and Knobbe, A. *A multi-relational approach to spatial classification*. Proceedings KDD. 2009. Dostępne także wideo autora (15.02.2011): http://videlectures.net/kdd09_frank_mrasc/.

Asocjacje/Kolokacje

- Xiong H., Shekhar S., Huang Y., Kumar V., Ma X., Yoo J. S., *A framework for discovering co-location patterns in data sets with extended spatial objects*, In Proc. 2004 SIAM International Conference on Data Mining (SDM), 2004.
- Zhang X., Mamoulis N., Cheung D. W., Shou Y., *Fast mining of spatial collocations*, KDD, 2004.
- Koperski K., Han J.: *Discovery of spatial association rules in geographic information databases*, Proceedings of 4th International Symposium on Large Spatial Databases, August 1995
- Morimoto Y.: *Mining Frequent Neighboring Class Sets In Spatial Databases*, KDD'01, San Francisco 2001
- Shekhar S, Huang Y.: *Discovering Spatial Co-Location Patterns: A summary of results*, In Proc of SSTO, Redondo Beach, 2001
- Wan, Y., Zhou, J. and Bian, F. *CODEM: A novel spatial co-location and de-location patterns mining*. Shandong : International School of Software, Wuhan University, 2008, s. 576 - 580.

Wykorzystanie danych z sieci społecznościowych LBSN

- Enrico Palumbo, Giuseppe Rizzo, Raphaël Troncy, and Elena Baralis. *Predicting your next stop-over from location-based social network data with recurrent neural networks*. In RECSYS 2017, 2nd ACM International Workshop on Recommenders in Tourism (RecTour'17), CEUR Proceedings Vol. 1906, August 27-31, 2017, Como, Italy, Como, ITALIE, 08 2017.
- Lian, Jianxun, et al. "Restaurant survival analysis with heterogeneous information." *Proceedings of the 26th International Conference on World Wide Web Companion*. 2017.
- Otaki, Keisuke, and Yukino Baba. *Maximal Likelihood Itinerary Planning with User Interaction Data*. (2022).