

Raport projekt 2 TASS

Temat 4: Wyszukiwarka popularnych miejsc turystycznych: na podstawie danych o trasach lotów i danych demograficznych znaleźć miasta o nieproporcjonalnie dużym w stosunku do liczby mieszkańców ruchu lotniczym.

1. Fuzja danych

Na podstawie połączonych danych wyznaczono współczynnik *wsp* liczony jako stosunek liczby pasażerów obsługiwanych przez lotniska w okolicach miasta do liczby mieszkańców danego miasta. W efekcie otrzymano prawie 649 wartości *wsp* dla różnych miast z całego świata.

2. Wizualizacja

Każde z rozpatrywanych miast zostało zaznaczone na mapie Google'a za pomocą koła w konkretnym kolorze oraz etykiety przypiętej do współrzędnych geograficznych miasta. Wartość *wsp* uwzględniono w trojaki sposób:

- Koła przyjmuje jedną z następujących barw:
 - fiolet,
 - niebieski
 - zielony,
 - żółty,
 - pomarańczowy,
 - czerwony.

Gdzie kolor fioletowy przypada miastom o małych współczynnikach *wsp*, a czerwony tym o dużych. Przy czym za wartość graniczną *wsp*, po której barwa nie ulega zmianie, przyjęto $wsp_{gr} = 100$.

- Wraz ze wzrostem *wsp* rośnie promień koła. Podobnie jak w przypadku kolorów, miasta o wartości większej niż wsp_{gr} mają koła jednakowej wielkości.
- Etykieta posiada taki sam kolor, jak koło. Ponadto najeżdżając na nią wyświetla się nazwa miasta oraz dokładna wartość *wsp*.

3. Analiza wartości wskaźnika *wsk*

Podczas analizy brano pod uwagę miasta z całego świata, jednak na potrzeby tego raportu zawężono opisywany obszar do Europy.

Z początku dane naniesione na mapę wydają się spełniać założenia projektu – kolor czerwony i pomarańczowy dominują: w okolicach basenu morza Śródziemnomorskiego, w kilku skandynawskich miastach, na wyspach Kanaryjskich oraz na innych wyspach takich jak Sardynia, Kostaryka i Malta czy też w większych europejskich stolicach jak

Paryż czy Amsterdam.

Zauważono jednak, że poza miejscami, które znane są dużej liczby turystów, współczynnik *wsp* wyróżnił również mniej turystyczne miejscowości, takie jak Billund (Dania, *wsp* = 465,55), Shannon (Irlandia, *wsp* = 177,86), Faro (Portugalia, *wsp* = 100,17), Keflavik (Islandia, *wsp* = 315,76) czy Frankfurt nad Menem (Niemcy, *wsp* = 34,15). Odpowiedź na pytanie, dlaczego tak się stało, leży w specyficznych cechach tych miast, których nie uwzględnia współczynnik *wsp*:

- mała liczba ludności, stosunkowo duże lotnisko (Billund, Faro, Shannon). Billund, z powodu *Legolandu*, jest miejscowością turystyczną, ale zdecydowanie mniej niż wskazuje na to *wsp*.
- lotnisko przesiadkowe (tranzytowe), będące np. łącznikiem pomiędzy Europą a USA) – rzadko kiedy bywa portem docelowym np. Faro.
- przypadek miasta Keflavik – znajduje się tam główny międzynarodowy port lotniczy w Islandii, co w połączeniu z populacją miasta nie przekraczającą 14tyś. Daje bardzo duży współczynnik *wsp*.

4. Próba udoskonalenia współczynnika *wsp*

W celu usunięcia niedoskonałości wynikających z budowy współczynnika *wsp* zdecydowano się uwzględnić kolejny rodzaj danych – informacje o istniejących połączeniach pomiędzy lotniskami. Na ich podstawie stworzono graf skierowany *DiGraph* i przystąpiono do jego analizy.

Początkowo policzono współczynnik gronowania (ang. *clustering coefficient*) dla każdego z portów – węzły grafu. Następnie przemnożono przez niego *wsk* i wynik naniesiono na mapę. Operacja ta przyniosła jednak dwojakie efekty. Znacznie zmalał *wsp* dla miasta Frankfurt nad Menem. Jednak dla Faro, Shannon i Billund dysproporcja w stosunku do innych miast znacznie wzrosła. Ponadto na mapie znacznie „urozło” miasto Kirkenes, które znajduje się w północnej części Norwegii i zdecydowanie nie należy do miast turystycznych. Dlatego też zrezygnowano ze współczynnika gronowania.

Kolejnym krokiem było policzenie pośrednictwa (ang. *closeness*). Pomysł był następujący – miasta pełniące rolę mostów pomiędzy grupami miast zostaną uwytłuszczone a następnie usunięte. W zamyśle miały być to porty tranzytowe. Okazało się jednak, że współczynnik pośrednictwa nie tylko portów przesiadkowych, a porty centralne w poszczególnych krajach – Paryż, Amsterdam, Madryt.

Próbowano jeszcze wykluczyć z sieci wszystkie przeguby (ang. *articulation points*) oraz policzono kilka innych współczynników. Efekty były jednak bardzo niezadowolające i nie zostały w tym raporcie opisane.

5. Uruchamianie aplikacji

Aplikację napisano w języku programowania *Python*. Wynik jej działania – mapę z danymi – zapisano w pliku *mapa.html*, który znajduje się w folderze projektu *TASS2*. Plik należy

otworzyć w dowolnej przeglądarce (testowane dla *Chrome*).

Drugą opcją jest uruchomienie skryptu w środowisku *PyCharm*. W wyniku utworzony zostanie nowy plik *mapa.html*, który zostanie automatycznie otworzony w przeglądarce domyślnej użytkownika.

W celu obejrzenia map z uwzględnieniem współczynników omówionych w rozdziale 4 należy otworzyć pliki *mapa_clustering_coefficient.html* lub *mapa_betweenness centrality.html* znajdujące się w folderze projektu *TASS2*.

Do wyświetlenia mapy niezbędny jest dostęp do Internetu.