

Zadanie dotyczy metryki Marczewskiego-Steinhaus, wykorzystywanej w naukach przyrodniczych. Wykonany program ma polegać na obliczaniu macierzy odległości między kolekcjami, odszukanie najbardziej oddalonych par kolekcji i mających tę samą zawartość. Do obliczenia odległości wykorzystujemy ułamek, w którego mianowniku jest łączna suma zbiorów, a w liczniku różnica tych zbiorów. Danymi są między innymi : gatunki zwierząt oraz roślin danych ekosystemów.

a) Analiza dotycząca treści i postaci dopuszczalnych zestawów danych i

oczekiwanych wyników

Przygotowanie danych wejściowych dla programu:

- Na początek konieczne jest utworzenie pliku wejściowego, gdzie każdy wiersz będzie zawierał zbiór elementów, stanowiących jedną kolekcję, elementy powinny być oddzielone od siebie przecinkami w wierszach.

W jaki sposób odczytywać wyniki i interpretacja:

- Naszym wynikiem końcowym powinna być macierz odległości, według procedury objaśnianej u góry. Wewnątrz macierzy, ukazane liczby powinny przedstawiać kolejno w pierwszym wierszu porównanie zbioru pierwszego z pierwszym, pierwszego z drugim, następnie z trzecim, w taki sposób do ostatniego zbioru. W przypadku drugiej linijki porówna drugi zbiór z pierwszym, drugi z drugim, itd. Analogicznie będą ukazane wyniki w następnych wierszach. Trzeba nadmienić, iż porównanie zbiorów w odwrotnej kolejności np. pierwszego z drugim zbiorem będzie miało taki sam wynik, jak porównanie drugiego z pierwszym.
- Wyniki w macierzy ukazane będą w postaci liczb, mieszczących się w zakresie od 0 do 1. Wyniki równe zero będą oznaczały zbiory o tej samej zawartości. Im bliżej zera tym większa część elementów zbioru jest taka sama. Najbardziej oddalone od siebie odległości znajdują się blisko 1, są to również najbardziej zróżnicowane pary kolekcji.
- Identyczne kolekcje będą miały wartość 0, a najbardziej oddalonym (zróżnicowanym) zostanie przypisana wartość 1.

b) Algorytm:

Program ma za zadanie obliczanie odległości między poszczególnymi klasami obiektów, w taki sposób abyśmy w danych wyjściowych uzyskali odpowiedź na pytanie: Czy zwierzęta

danego gatunku pokrywają się w podanych ekosystemach naturalnych?

- Tworzona jest pusta lista, w której będą się później znajdowały dane z kolekcjami
- Program będzie teraz pracował w pętli. Po otrzymaniu od użytkownika danych w pliku .txt. Odczyta je linijka po linijce. Linia dzielona przez przecinek.
- Dla każdego wiersza(linijki) powinna zostać utworzona nowa, pusta lista w pętli, do której będą dodawane, w następnych liniach kodu elementy kolekcji.
- Przecinek w linijce pliku wejściowego rozdziela tekst i tworzy poszczególne elementy. Po każdym kolejnym przecinku będzie znajdował się inny element.
- Elementy wszystkie teraz powinny być dodawane do wcześniejszej pustej listy kolekcji, a kolekcje do listy z danymi. (strip - służy tutaj do usunięcia białych znaków np. spacji na końcu linii)
- Tworzymy: zmienną z rezultatami i pustą listą, zmienna o nazwie największe różnice z przypisaną wartością zero, pustą listę przypisaną do zmiennej największe różnice oraz zmienną te same kolekcje z pustą listą.
- Następnie program tworzy macierz, w której tworzeniu pomocna nam operacja set działająca podobnie jak lista, ale zawierająca nie powtarzające się (unikalne) elementy.
- W macierzy tej będą znajdowały się elementy list z kolekcjami. Zadaniem macierzy jest obliczenie odległości na podstawie danego wzoru w podpunkcie a). Zależy nam na tym, aby nasz program wyświetlił nam zaokrąglone liczby, które następnie dzięki funkcji .append zostaną dodane do rzędu macierzy.
- Chcemy, aby wyniki macierzy zostały przypisane do odpowiedniej listy:
 - 1) Lista 1 - ukaże nam pomiędzy którymi zbiorami występuje największe zróżnicowanie
 - 2) Lista 2 - w niej mają znaleźć się te same zbiory
- Przypisanie wyników macierzy do tych list, umożliwi nam wydrukowanie komunikatów zarówno dla największego zróżnicowania między zbiorami i w przypadku obecnych takich samych zbiorów elementów, abyśmy nie potrzebowali odnajdywać tej wartości w macierzy np. gdyby była ona dużej wielkości.

c) Dyskusja wyników:

Poniżej zostały zamieszczone zawartości zestawu plików tekstowych, przekazanych

do sprawdzenia działania programu w tym projekcie.

- Plik "puszcze" zawierał kolejno gatunki z trzech różnych puszczy na terenie Polski:

Karpackiej, Białowieskiej, Augustowskiej.

Puszcza Karpacka:

- Ryś euroazjatycki, Bocian czarny, Niedźwiedź brunatny, Żubr, Orzeł Przedni

Puszcza Białowieska:

- Żuraw, Bocian czarny, Ryś eurazjatycki, Wilk szary,

Puszcza Augustowska:

- Ryś euroazjatycki, Zając Szarak, Wilk szary, Jeleń szlachetny, Orlik krzykliwy

- Dla zestawu danych z pliku "puszcze":

Plik zawiera dane na temat gatunków zwierząt, zamieszkujących kolejno Puszcze

Karkonoska, Białowieską i Augustowska. Nasza macierz składa się w tym przypadku z 3 rzędów, które odczytujemy zaczynając od lewego górnego rogu.

Zauważamy, że po przekątnej znajdują wyniki równe 0, pomimo, że w zestawie danych wejściowych, brak jest identycznych kolekcji, którym taki wynik jest przypisany zgodnie z interpretacją. Wynika to z tego, iż doszło do porównania zbiorów Puszczy Karkonoskiej z Karkonoską, kolejno Białowieskiej z Białowieską oraz Augustowskiej z Augustowską.

W pierwszym rzędzie, drugiej kolumnie otrzymaliśmy wynik 0,714, sugerujący wysoką różnorodność między kolekcjami Puszczy Karkonoskiej i Białowieskiej. Inaczej mówiąc, niewiele jest tych samych gatunków, które zamieszkują oba te tereny jednocześnie. Taki sam wynik wypadnie, gdy dojdzie do porównania Białowieskiej z Karkonoską, w drugim rzędzie, pierwszej kolumnie. Jednak to wynik porównania gatunków na terenie Puszczy Karkonoskiej z tymi rezydującymi na obszarze Puszczy Augustowskiej jest największy, o wartości 0,889.(3 kolumna, 1 rząd i odwrotnie). Oznacza to, że największe zróżnicowanie gatunków występuje pomiędzy tymi zbiorami. Posiadamy jeszcze w macierzy jeden wynik również o wartości 0,714, który interpretujemy tak samo jak w przypadku Puszczy Karkonoskiej i Białowieskiej. Jednakże odnosi się on tym razem do gatunków z Puszczy Białowieskiej oraz Puszczy Augustowskiej.(2 rząd, 3 kolumna i na odwrót) Taka sama wartość tych dwóch porównań jest zupełnie losowa, zależy jedynie od pokrywających się gatunków w kolekcjach.

1. Dane zostały oparte na gatunkach występujących w poszczególnych Parkach

Narodowych Polski:

Babiogórski Park Narodowy:

- Jodła pospolita, Buk zwyczajny, Rogownica alpejska, Jarzęb pospolity

Świętokrzyski Park Narodowy:

- Dąb szypułkowy, Modrzew europejski, Jodła pospolita, Buk zwyczajny, Buczyzna karpacka, Jarzęb pospolity

Tatrzański Park Narodowy:

- Jodła pospolita, Modrzew europejski, Rogownica alpejska, Sasanka alpejska, Jarzęb pospolity, Buk Zwyczajny

Bieszczadzki Park Narodowy:

- Kwaśne buczyny, Jodła Pospolita, Buczyzna karpacka

Tym razem naszym podmiotem rozważania są gatunki roślin, głównie drzew w Parkach Narodowych Polski, które zostały wypisane u góry. Otrzymaliśmy w macierzy następujące wyniki:

- 0,000 - wynik takie występują ponownie po przekątnej, tak jak w przypadku danych dla Puszczy. Zostały porównane poszczególne zbiory z samymi sobą np. Park Bieszczadzki z Parkiem Bieszczadzkim (4 kolumna, 4 rząd)

- 0,571 - dla porównania drzew występujących na terenie Babiogórskiego i Świętokrzyskiego. Możemy zauważyć, że powtarzającymi się gatunkami są Jodła Pospolita, Buk Zwyczajny, Jarzęb Pospolity. Ten wynik oznacza, iż trochę więcej jest gatunków różnych w tych dwóch parkach, niż gatunków występujących na obu terenach.

- 0,333 - Park Babiogórski oraz Tatrzański. Gatunki pokrywające się to Jodła pospolita, Rogownica alpejska, Jarzęb pospolity, Buk Zwyczajny. Informuje nas ten wynik o przewyższającej liczbie gatunków występujących w obydwu parkach ponad te, które występują tylko w jednym z nich.

- 0,833 - Park Babiogórski i Bieszczadzki. Wysoki wynik, bliski 1, odpowiada niewielkiej liczbie takich samych gatunków drzew w obu miejscach.

Spoglądając na dane wejściowe widzimy, że jedynie jest to Jodła Pospolita.

Pozostałe gatunki są charakterystyczne tylko dla jednego z dwóch obszarów.

- 0,500 - wynik odpowiadający porównaniu drzew z terenów Parku

Świętokrzyskiego i Tatrzańskiego. Występuje zarówno tyle samo gatunków powtarzających się, ile jest gatunków różniących dwie kolekcje.

- 0,714 - Narodowy Park Świętokrzyski w zestawieniu z Parkiem Bieszczadzkim. Gatunków wspólnych dla każdej tych kolekcji jest niewielka, są to: Jodła Pospolita, Buczyzna karpacka.

- 0,875 - największy z wszystkich możliwych, otrzymanych wyników w tej macierzy, tutaj występuje największe zróżnicowanie między kolekcjami. W Tatrzańskim Parku Narodowym gatunkami różniącymi są Modrzew europejski, Rogownica alpejska, Sasanka alpejska, Jarzęb pospolity, Buk Zwyczajny. Natomiast w Bieszczadzkim Kwaśne buczyny, Buczyzna karpacka.

Jedynym wspólnym gatunkiem dla obydwu terytoriów jest Jodła Pospolita.

Na podstawie objaśnionych wyników sprawdziliśmy, które z wypisanych Parków Narodowych Polski posiadają największą ilość wspólnych gatunków, a które najmniej.

Problem może pojawić się w momencie kiedy byśmy chcieli wiedzieć jakie są to konkretnie gatunki. Dla niewielkiej liczby kolekcji nie jest to bardzo trudne zadanie, tak w przypadku wybranych przez nas danych, zamieszczonych w plikach. Jednak problem stanowiły by pliki o parunastu, parudziesięciu kolekcjach. Aczkolwiek jest to problem, którego mamy możliwość obejść. Jednym ostatnich zadań naszego programu jest wydrukowanie komunikatów, pomiędzy, którymi zbiorami występuje największe zróżnicowanie. To zawęży znacznie nasz obszar poszukiwań najczęściej do dwóch kolekcji, które na spokojnie możemy ze sobą porównać odrębnie.

→ Dane zostały oparte na gatunkach występujących w poszczególnych morzach świata:

Morze Śródziemne:

- Sardynka europejska, Makrela, Wstęgór królewski, Wargacz zielony

Morze Północne:

- Sardynka europejska, Śledź oceaniczny, Makrela, Dorsz atlantycki

Morze Czarne:

- Sardynka europejska, Wstęgór królewski

Morze Bałtyckie:

- Sardynka europejska, Śledź oceaniczny, Makrela, Dorsz atlantycki