

Eksploracja tekstów
ćwiczenia 5
Zajęcia 14.06.2022 i 22.06.2022

Zadanie 1. Podobnie jak na poprzedniej liście zdefiniujemy *drzewo rozbioru* zdania jako sposób postawienia w tym zdaniu (zagnieżdżonych) nawiasów, w taki sposób, by każdy nawias zawierał dokładnie dwa elementy (wyrazy, albo frazy w nawiasie), a ponadto, by te nawiasy oddawały jakoś strukturę gramatyczną zdania. Przykładowo, dla zdania *Młody krokodyl szybko pożerał kąpiące się różowiotkie prosięta* drzewem rozbioru mogłoby być:

((Młody krokodyl) ((szybko pożerał) ((kąpiące się) (różowiotkie prosięta))))

Zaproponuj algorytm, który wyznacza drzewo rozbioru, za pomocą algorytmu PageRank (możesz też korzystać z zanurzeń słów).

Zadanie 2. Znajdź w Internecie 2 różne metody inicjalizacji algorytmu K-średnich (inne niż losowanie K wektorów z rozkładem jednostajnym). Przedstaw je, starając się, by opis był na tyle dokładny, by umożliwiał napisanie programu realizującego te metody.

Zadanie 3. Przedstaw, jak w wybranym przez Ciebie języku efektywnie realizuje się obliczenia na wektorach (macierzach) rzadkich i w jaki sposób może to przyspieszyć obliczenia dokumentów podobnych do danego.

Zadanie 4. Zapoznaj się ze stroną polskiej Słowosieci (<http://plwordnet.pwr.wroc.pl/wordnet/>). Odpowiedz na następujące pytania:

- a) Ile znaczeń (i jakie) ma słowo *skóra*?
- b) Jakie są hiperonimy tego słowa (czy jest ich mniej niż znaczeń)?
- c) Co to jest fuzyzymia? (podaj przykłady)
- d) Co to są derywatywy i deminutywy?

Zadanie 5. Czy Słowosiec (z relacją *is_a*) jest drzewem? Przedstaw (skompresowane) ścieżki dla najwyższego hiperonimu dla wybranego znaczenia słowa *zamek* (niezwiązanego z hokejem).

Zadanie 6. To kontynuacja poprzedniego zadania. Odpowiedz na następujące pytania:

- a) Jak się nazywa przeciwieństwo meronimii?
- b) Jakie są rodzaje meronimii (relacji *is_a_part_of*)? Spróbuj sobie wyobrazić różne aspekty tej relacji i sprawdź otoczenie wybranych słów. Są co najmniej 3 rodzaje tej relacji.

Zadanie 7. Zaproponuj sposób wykorzystania Słowosieci przy wyznaczaniu dokumentów podobnych i przy odpowiadaniu na pytania.