Koherencja i Prawo Zipfa

Jarosław Kołodziej
Przemysław Kożuch
Radosław Brzostek

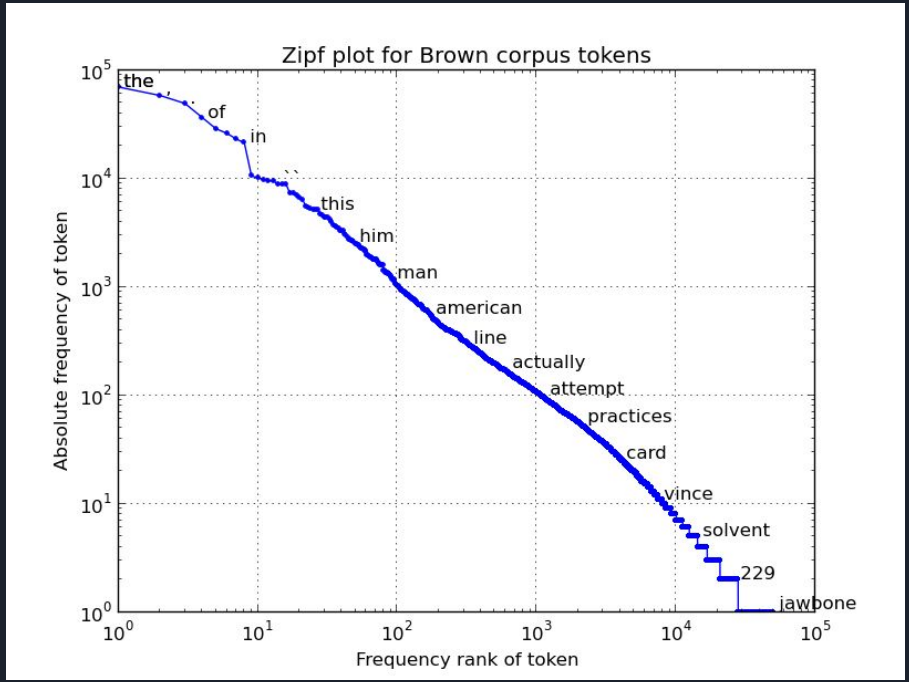
Czym jest prawo Zipfa

Wikipedia głosi:
“Prawo empiryczne głoszące, że wiele rodzajów danych tworzonych przez ludzi lub odnoszących się do ich zachowań cechuje charakterystyczny rozkład wartości, w którym dystrybucja częstotliwości występowania poszczególnych wartości jest odwrotnie proporcjonalna do ich rangi statystycznej.”

Co oznacza iż drugi najczęściej występujący element zbioru występuje o ½ rzadziej od pierwszego a trzeci występuje ⅓ raza rzadziej i tak dalej..

Czym jest prawo Zipfa

Mówiąc o języku angielskim, biorąc za przykład korpus “**Brown Corpus of American English text**” można zaobserwować, że słowo **the**stanowi 7% wszystkich słów w całym korpusie, na drugim miejscu mamy **of**, które składa się na 3.5%, następnie **and** z około 1.75% . . .



Czym jest prawo Zipfa

Prawo to można opisać wzorem:

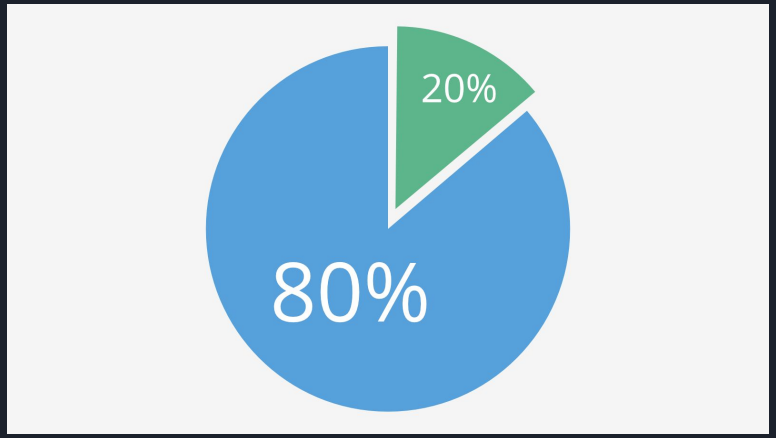
wielkość_zbioru ÷ ranga_elementu = ilość_wystąpień_elementu

Gdzie **rangą elementu** nazywamy miejsce w liście sortowanej od najczęściej występujących elementów

Zasada Pareta

Zasada ta mówi, że możemy założyć z pewnym poziomem pewności, iż w danym zdarzeniu 20% akcji jest odpowiedzialne za 80% wyników.

W przenoszeniu to na corpus słów możemy powiedzieć iż 20% unikalnych słów jest odpowiedzialne za 80% wszystkich słów użytych w danych zbiorze



Zasada Pareta

To samo zjawisko możemy zauważyć w innych dziedzinach, jak na przykład:

- 20% ludzi posiada 80% nieruchomości
- 20% pacjentów wykorzystuje 80% zasobów szpitalnych
- 20% klientów jest odpowiedzialnych za 80% dochodów firmy
- 20% ludzi posiada 80% dochodów świata
    - A 20% z tych ludzi posiada 80% z tych 20% i tak dalej
- itd.

Koherencja

W języku, do komunikacji z innymi, w każdej rozmowie używamy implikacji do przedstawienia jakiejś akcji oraz powodu dla tej akcji, co najczęściej dzieje się w postaci np.:

Bartek założył dzisiaj kurtkę. Jest dzisiaj bardzo zimno.

Z logicznego punktu widzenia możemy założyć, iż **z powodu**, że jest dzisiaj zimno **Bartek założył kurtkę.** Koherencja oznacza właśnie, iż spójność logiczna tych dwóch zdań została w całej informacji zachowana i jest ona zrozumiała.

Koherencja

Jeśli zamiast poprzedniego zdania powiedzielibyśmy:

Bartek założył dzisiaj kurtkę. Zbliża się koniec studiów.

Możemy tutaj stwierdzić, iż koherencja pomiędzy tymi dwoma zdaniami nie została zachowana. Moglibyśmy założyć, iż założenie kurtki ma jakiś związek z końcem studiów Bartka, jednak nie jest to związek jakiego moglibyśmy się spodziewać i domyślać z podanych informacji z tych dwóch zdań.

Stwierdzić można, iż zbiór informacji z tych dwóch zdań jest niekoherentny.

Koherencja

Koherencja jest podzielona na kategorie:

- Narracja
  - czasowość
- Przyczynowość
  - związek przyczynowo skutkowy
- Dopasowanie
- Podobieństwo