STATYSTYKA NA TROPIE BRUDNYCH PIENIĘDZY

PRAWO ZIPFA

PRAWO ZIPFA

SFORMUŁOWANE W 1935 ROKU PRZEZ GEORGE'A ZIPFA PRAWO MÓWI, ŻE W DOWOLNYM TEKŚCIE CZĘSTOTLIWOŚĆ WYSTĘPOWANIA DANEGO SŁOWA $f(r)$ JEST ODWROTNIE PROPORCJONALNA DO JEGO RANGI r (POZYCJI W RANKINGU NAJPOPULARNIEJSZYCH SŁÓW).

$$f(r) \propto \frac{1}{r}$$

OZNACZA TO, ŻE n-TE NAJCZĘSTSZE SŁOWO WYSTĘPUJE n RAZY RZADZIEJ, NIŻ TO NAJCZĘSTSZE

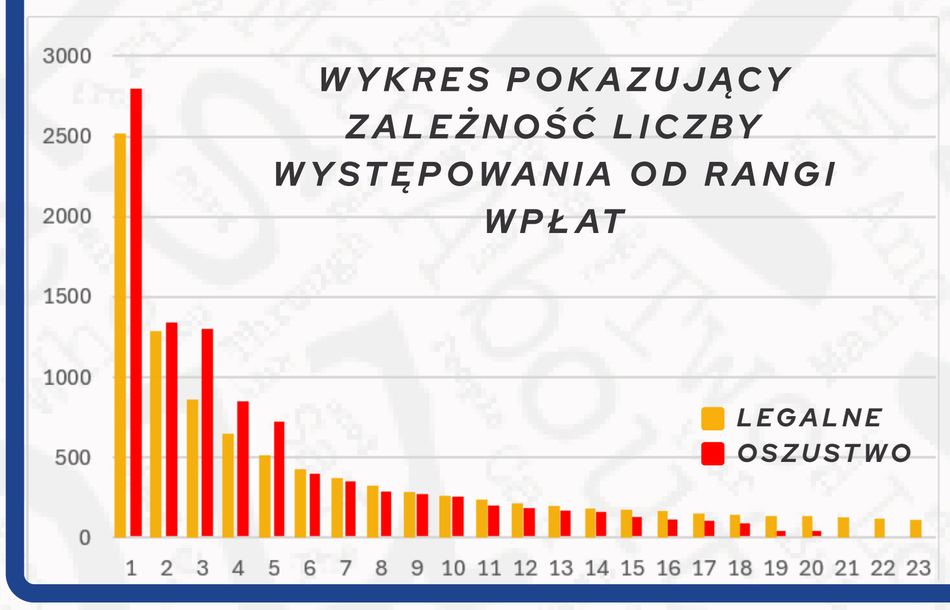
...A JEDNAK TO SAMO

CO ZASKAKUJĄCE CHARAKTERYSTYCZNY ROZKŁAD ZIPFA NIE DOTYCZY TYLKO TEKSTU, PODOBNE KRZYWE MOŻEMY ZAUWAŻYĆ M.IN. W ROZKŁADZIE:

- LICZEBNOŚCI POPULACJI MIAST
- SIŁ TRZĘSIEŃ ZIEMI
- RUCHU W WITRYNACH INTERNETOWYCH
- NUT W UTWORACH KLASYCZNYCH
- SEKWENCJI NUKLEOTYDÓW W DNA
- WYKORZYSTANIA PAMIĘCI CACHE
- LICZBY PRACOWNIKÓW W FIRMACH
- LICZBY OFIAR W WOJNACH
- WIELKOŚCI TRANSAKCJI BANKOWYCH

WYKRYWANIE ANOMALII FINANSOWYCH

GDZIE PRAWO ZIPFA PRZYDAJE SIĘ W PRAKTYCE? NA PRZYKŁAD PRZY PRANIU BRUDNYCH PIENIĘDZY. W ZDROWYM SYSTEMIE FINANSOWYM WIELKOŚCI I CZĘSTOTLIWOŚCI TRANSAKCJI PODLEGAJĄ ŚCISŁYM ROZKŁADOM. WYSTĘPUJĄ W NIM MILIONY MIKROPŁATNOŚCI I ZALEDWIE GARSTKA GIGANTYCZNYCH PRZELEWÓW KORPORACYJNYCH. TAKI SYSTEM JEST WYSOCE PRZEWIDYWALNY.



PRZESTĘPCY PIORĄCY BRUDNE PIENIĄDZE NIE MOGĄ WPŁACIĆ DO BANKU MILIONA DOLARÓW NARAZ, GDYŻ URUCHOMI TO SYSTEMY ALARMOWE. ZAMIAST TEGO ROZKŁADAJĄ KWOTĘ NA WIELE MNIEJSZYCH TRANSAKCJI. TAKI SZTUCZNY ZASTRZYK GOTÓWKI NISZCZY KRZYWĄ ZIPFA. TAKIE ZJAWISKO ROZBIJANIA KWOT NOSI NAZWĘ *SMURFINGU*. POWTARZAJĄCE SIĘ TRANSAKCJE TUŻ PONIŻEJ PROGU ALARMOWEGO TWORZĄ NA WYKRESIE NIENATURALNY GARB.

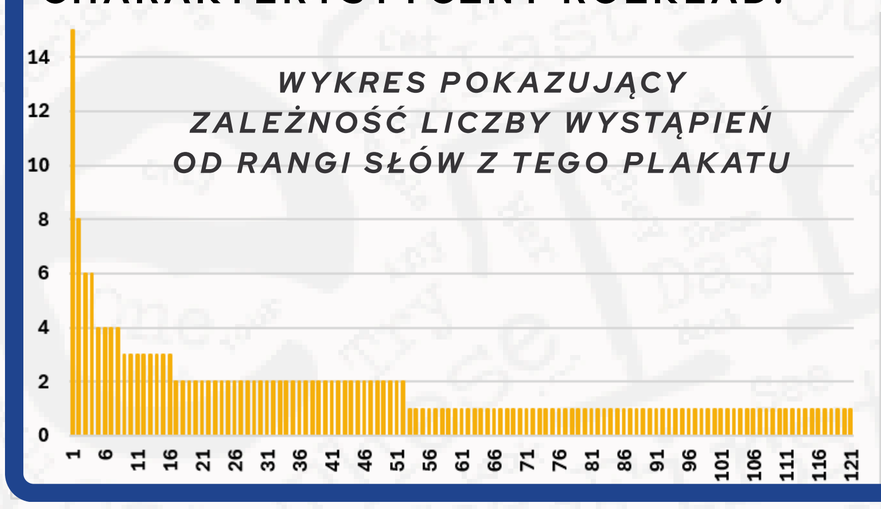
HAPAX LEGOMENA

HAPAX LEGOMENA TO ELEMENTY WYSTĘPUJĄCE W DANYM ZBIORZE TYLKO RAZ. ZGODNIE Z PRAWEM ZIPFA, TWORZĄ ONE NIESKOŃCZENIE DŁUGI, PŁASKI OGON ROZKŁADU. TO ONE NADAJĄ TEKSTOWI SENS.



ZIPFOWSKI PLAKAT

CZY TO DZIAŁA W PRAKTYCE? ZOBACZMY! NAJCZĘŚCIEJ WYSTĘPUJĄCYM SŁOWEM NA CAŁYM PLAKACIE JEST *„W"* Z PIĘTNASTOMA WYSTĄPIENIAMI, DRUGIM NAJCZĘŚCIEJ WYSTĘPUJĄCYM SŁOWEM JEST *„ZIPFA"* WYSTĘPUJĄCE 8 RAZY! NAWET NA TAK MAŁYM ZBIORZE DANYCH MOŻEMY ZAUWAŻYĆ CHARAKTERYSTYCZNY ROZKŁAD!



ZASADA PARETO

SUMUJĄC WARTOŚCI POCZĄTKOWYCH 20% ROZKŁADU MOŻEMY ZAUWAŻYĆ, ŻE ODPOWIADAJĄ ZA 80% SUMY WSZYSTKICH WARTOŚCI. PRZYKŁADOWO 20% POPULACJI ODPOWIADA ZA 80% PRZESTĘPSTW, A 20% PRODUKTÓW ODPOWIADA ZA 80% PRZYCHODÓW PRZEDSIĘBIORSTWA.

ADAM DRAŻYK