

Opis wyników

1. Chmury punktów

W książkach o Mikołajku łatwo zauważyć powracające, charakterystyczne imiona bohaterów, jak Kleofas, Ananiasz czy Alcest. Co ciekawe, wyjątek stanowi tom poświęcony wakacyjnym przygodom, gdzie te postacie nie występują. Niezależnie od części, najczęściej pojawiają się słowa „powiedzieć”, „mama” i „tata”.

W chmurach wyrazów opartych na treści poszczególnych tomów innych serii widać przewagę imion typowych dla danej części, jak Paula, Leto czy Duncan. W piątym i szóstym tomie często pojawia się też słowo „Odrad” – imię jednej z postaci. We wszystkich częściach zauważalna jest również obecność terminu „Bene Gesserit”, wskazująca na wspólny motyw zakonny.

W serii chińskiej science fiction autorstwa Liu Cixina, chmury wyrazów zdominowane są przez imiona głównych bohaterów: Wang, Luo Ji i Cheng Xin. Niezmiennie powtarzają się też słowa „Ziemia” i „powiedzieć”, natomiast w tomach drugim i trzecim szczególnie wyróżnia się „statek”.

W analizie tekstów o Sherlocku Holmesie niezmiennie pojawia się jego nazwisko, a także czasowniki „wiedzieć” i „powiedzieć”. Często występuje także słowo „sprawa”, co dobrze oddaje detektywistyczny charakter opowieści. Do tego dochodzą liczne zaimki, jak „siebie” i „swój”.

2. Analiza tematów

Model LDA wykrył kilka tematów dobrze korespondujących z poszczególnymi seriami literackimi. Przykładowo, temat związany z „Mikołajkiem” był reprezentowany przez słowa takie jak „mikołaj”, „tata” i „mama”, a temat „Diuny” – przez terminy „duncan”, „bene” i „gesserit”. Sherlock Holmes został rozpoznany głównie dzięki słowu „holmes”, natomiast seria Liu – przez terminy „statek” i „kosmiczny”. Inne tematy miały bardziej ogólny charakter. W przypadku „Diuny” zauważono pewne rozproszenie między tematami, co może wskazywać na większą złożoność fabularną i obecność wielu wątków.

Model NMF oparty na odległości Frobeniusa umożliwił uzyskanie bardziej precyzyjnych i jednoznacznych tematów. Dla uzyskania wyników konieczna była redukcja wymiarowości. Słowa kluczowe silnie wiązały się z konkretnymi seriami – np. w „Mikołajku” dominowały imiona postaci, a w „Diunie” – nazwy frakcji i postaci. Przypisania książek do tematów były spójne, a dokumenty z tej samej serii trafiały zwykle do jednego klastra, co sugeruje skuteczne uchwycenie semantyki.

Największą precyzję osiągnął model NMF wykorzystujący odległość Kullbacka-Leiblera. Oprócz klarownego podziału na tematy takie jak „Mikołajek”, Holmes czy Liu, model ten potrafił skutecznie rozdzielić wielowątkową „Diunę” na kilka, powiązanych, lecz odrębnych tematów. Słowa kluczowe trafnie odzwierciedlały różne aspekty fabuły, a przypisania dokumentów były niemal jednoznaczne – co dowodzi skuteczności podejścia w klasyfikacji literatury o złożonej strukturze.

3. Analiza skupień

W analizie skupień porównano dwa podejścia do klasteryzacji tekstów: jedno z wykorzystaniem miary cosinusowej i średniego łączenia (average linkage), drugie z użyciem odległości euklidesowej i pełnego łączenia (complete linkage).

W przypadku miary cosinusowej, książki wyraźnie grupowały się według autorów i serii. Przykładowo, wszystkie tomy trylogii Liu znalazły się w jednej, spójnej grupie, co świadczy o dużym podobieństwie ich cech tekstowych. Opowiadania o Holmesie również utworzyły zwarty klaster, odzwierciedlając jednolity styl pisarski Doyle'a. Seria „Mikołajek” także zgrupowała się razem, jednocześnie odcinając się od „poważniejszej” literatury.

Dendrogram oparty na odległości euklidesowej również ujawniał pewne zależności, jednak grupowania były mniej wyraziste. Klaster „Diuny” nie był już tak spójny – niektóre tomy łączyły się dopiero na dalszych etapach hierarchii. Książki Liu pozostały w pobliżu, choć pojawiło się więcej odgałęzień. Teksty „Mikołajka” tworzyły zwartą grupę, co może wynikać z prostoty języka i powtarzalności słownictwa. W przypadku opowiadań Doyle'a pojawiło się kilka podgrup, co może świadczyć o większej różnorodności strukturalnej opowiadań.

Podsumowując, miara cosinusowa pozwoliła na uzyskanie bardziej intuicyjnych i spójnych tematycznie klastrów. Z kolei podejście euklidesowe bardziej uwypukliło różnice w długości i liczbie słów, co w analizie literatury może prowadzić do mniej trafnych grupowań.

4. Ngramy

W pierwszym tomie trylogii „Problem trzech ciał” najczęściej pojawiającym się 3-gramem było „(problem, trzy, ciało)” – bezpośrednie nawiązanie do tytułu. Pojawiały się też frazy jak „(baza, Czerwony, Brzeg)”, odnoszące się do kluczowej lokalizacji. W „Ciemnym lesie” dominowały frazy typu „(program, wpatrywać, Ściana)” i „(rada, Obrona, planeta)”, akcentujące militarno-polityczny wymiar fabuły. Fraza „(bitwa, dzień, sąd)” przywołuje kulminacyjne momenty walki. W „Końcu śmierci” n-gramy „(Cheng, Xin, powiedzieć)” i „(fragment, przeszłość, czas)” wskazują na przesunięcie uwagi na główną bohaterkę i refleksję nad historią.

W serii o Mikołajku analiza n-gramów ujawnia powtarzalną, dziecięcą strukturę narracji. W tomie „Mikołajek” dominuje zwrot „(a, potem, powiedział)”, dobrze oddający prosty, przyczynowo-skutkowy sposób opowiadania. Często pojawiają się też wypowiedzi dorosłych, np. „(Mikołaj, powiedział, tata)” czy „(pani, powiedzieć, Mikołaj)”. W „Nieznanych przygodach” częsty jest 3-gram „(powiedzieć, tata, Kleofas)”, akcentujący relacje między postaciami. Z kolei w „Nowych przygodach” powtarzają się frazy „(Mikołaj, powiedział, mama)” i „(powiedział, Alcest, potem)”, które ukazują schematyczną strukturę narracyjną całej serii.

W „Psie Baskervillów” i „Przygodach Sherlocka Holmesa” dominuje wątek śmierci Sir Karola Baskervilla, natomiast w „Znaku czterech” kluczowa jest postać z „drewnianą nogą”. „Dolina Trwogi” wskazuje na znaczenie postaci takich jak profesor Moriarty, zaś „Studium w Szkarlacie” silnie akcentuje rolę Lake City.

W pierwszym tomie dominują zwroty takie jak „(piór, książniczki, Irulan)” oraz „(zdawać, siebie, sprawa)”, sugerujące refleksyjny charakter narracji. W *Mesjaszu Diuny* wyróżnia się „(Gaius, Helena, Mohiam)”, co może wskazywać na istotną rolę tej postaci. W *Dzieciach Diuny* powracają introspektywne frazy oraz pojawia się „(Bene, Gesserit, plan)”, sugerując działania zakonu. *Bóg Imperator Diuny* przynosi „(ekologiczny, transformacja, Diuna)”, prawdopodobnie odnoszący się do przemian planety. W późniejszych tomach, takich jak *Heretycy Diuny*, *Kapitał Diuny* i kontynuacje, powracają zarówno znane motywy („zdawać, siebie, sprawa”), jak i nowe („(Murbella, powiedzieć, Duncan)”), co wskazuje na rozwój uniwersum i jego postaci.