

Uniwersytet Warszawski
Wydział Nauk Ekonomicznych



Studia Podyplomowe
„Metody statystyczne w biznesie.
Warsztaty z oprogramowaniem SAS”

mgr Elżbieta Szarejko

Analiza Text Mining rynku książek self-publishing
- wpływ opisu książki na jej sprzedaż

Praca dyplomowa
napisana pod kierunkiem
dr Karoliny Kuligowskiej
Katedra Informatyki Gospodarczej
i Analiz Ekonomicznych

Warszawa, wrzesień 2014

Oświadczenie kierującego pracą

Oświadczam, że niniejsza praca została przygotowana pod moim kierunkiem i stwierdzam, że spełnia ona wszystkie wymogi formalne i merytoryczne pracy studiów podyplomowych

Data

Podpis kierującego pracą

Oświadczenie autora (autorów) pracy

Świadom odpowiedzialności prawnej oświadczam, że niniejsza praca dyplomowa została napisana przeze mnie samodzielnie i nie zawiera treści uzyskanych w sposób niezgodny z obowiązującymi przepisami.

Oświadczam również, że przedstawiona praca nie była wcześniej przedmiotem procedur związanych z uzyskaniem tytułu zawodowego w wyższej uczelni.

Oświadczam ponadto, że niniejsza wersja pracy jest identyczna z załączoną wersją elektroniczną.

Data

Podpis autora (autorów) pracy

Streszczenie

Praca podejmuje problematykę opisów zamieszczanych przy e-bookach w księgarniach internetowych i ich roli w procesie decyzyjnym konsumenta. W części teoretycznej omówiono zmiany zachodzące na rynku książek oraz czynniki wpływające na ich sprzedaż. Przytoczono też wyniki badań dotyczących znaczenia opisów, zaprezentowano przegląd wybranych poradników na temat tworzenia takich tekstów oraz wskazano cechy, jakie powinien posiadać taki tekst. W części badawczej porównano za pomocą narzędzi text mining formę i treść opisów najbardziej oraz najmniej popularnych książek, a także sprawdzono zasadność porad, które można znaleźć w pozycjach kierowanych do autorów takich tekstów.

Słowa kluczowe

<opis, blurb, e-book, książka, księgarnia internetowa, sprzedaż, text mining>

Dziedzina pracy (kody wg programu Socrates-Erasmus)

Ekonomia (14300)

Spis treści

Wstęp	5
Przedmiot badań.....	5
Cel pracy	5
Charakterystyka pracy	5
Rozdział 1.....	6
Rozdział 2.....	14
Zakończenie.....	32
Bibliografia.....	34

Wstęp

Przedmiot badań

Rosnąca popularność księgarni internetowych skłoniła mnie do zbadania, jak tekstowy sposób prezentowania książek może wpływać na ich sprzedaż. Przedmiot mojej analizy w niniejszej pracy stanowią opisy zamieszczone przy e-bookach w księgarni Amazon.co.uk. Analiza obejmuje porównanie formy i treści opisów najchętniej czytanych książek oraz tych, na które klienci raczej nie zwracają uwagi.

Cel pracy

Celem analizy jest wskazanie takich cech opisu książki, które mogą wpływać na wzrost jej sprzedaży lub wręcz przeciwnie – zniechęcać do zakupu. Ustalenie, czym różnią się opisy najbardziej i najmniej popularnych książek, ma pomóc w sformułowaniu zasad tworzenia skutecznych sprzedażowo tekstów.

Charakterystyka pracy

W Rozdziale 1 omówiłam, jak wraz z rozwojem księgarni internetowych zmienił się sposób kupowania książek oraz jaką rolę w procesie decyzyjnym odgrywa obecnie blurb.

W Rozdziale 2 dokonałam analizy porównawczej opisów najpopularniejszych i najmniej popularnych książek w celu określenia, czym się różnią i jak to może wpływać na decyzję czytelników o zakupie.

W Zakończeniu podsumowałam zamierzenia niniejszej pracy i wskazałam cechy opisów, mające wpływ na popularność książek, których dotyczą.

Rozdział 1

Internet zrewolucjonizował rynek książki. Zmienił nie tylko sposób, w jaki czytelnicy kupują nowe pozycje, ale też dał im dostęp do wszelkich informacji ułatwiających podjęcie decyzji o zakupie. Oferta księgarni internetowych jest znacznie szersza niż tradycyjnych placówek, w których czytelnicy mogą wybierać jedynie spośród tego, co księgarz uznał za interesujące i co zmieściło się na półkach. W e-księgarniach czytelnicy mogą znaleźć nie tylko klasykę i nowości, ale też książki wydane kilka lat wcześniej, propozycje małych, niszowych wydawnictw, a nawet twórczość publikowaną przez autorów własnym sumptem. Nic dziwnego zatem, że udział księgarni internetowych w sprzedaży książek dynamicznie rośnie. Z danych zaprezentowanych na konferencji *Bowker's 2013 Books and the Consumer Conference* wynika, że w Wielkiej Brytanii w pierwszym kwartale 2013 r. udział podmiotów zajmujących się tylko sprzedażą książek przez Internet wyniósł 38%, a tradycyjnych księgarni – 32%.¹ Raport *VersoAdvertising.com* pokazuje, że w 2012 r. z tego kanału sprzedaży korzystało 49% Brytyjczyków. W grupie najbardziej aktywnych czytelników (kupujących powyżej 10 książek w roku) odsetek ten wynosił 65%.²

Udział e-księgarni będzie rósł wraz z popularyzacją e-booków i ich czytników, do których przekonuje się coraz więcej konsumentów. Jak wynika z badania *Bookboon*, w 2013 r. książki elektroniczne kupowało 24,4% badanych, ale 28% zapowiadało, że zacznie to robić w ciągu najbliższego roku. Aż 75% czytelników z Wielkiej Brytanii zamierzało czytać e-booki w 2015 r., a 56,1% badanych twierdziło, że książki w formie elektronicznej będą wówczas stanowiły ponad połowę ich lektur.³

Rosnące znaczenie e-booków potwierdzają badania rynku książki prowadzone przez firmę Nielsen. Wynika z nich, że w 2013 r. w Wielkiej Brytanii sprzedano 323 mln książek za kwotę 2,2 mld funtów. Sprzedaż zarówno w ujęciu ilościowym, jak i wartościowym

¹ Jo Henry, *New Trends in Publishing: BIC Seminar 2013 What are book consumers up to?*, Bowker Market Research, http://www.bic.org.uk/files/pdfs/What%20are%20book%20consumers%20up%20to_Jo%20Henry.pdf [dostęp 24.09.2014]

² Verso Advertising, *2011 Survey of Book-Buying Behavior*, January 2012, <http://www.versoAdvertising.com/Wi7survey2012/> [dostęp 24.09.2014]

³ Global eBook Survey 2013 by Bookboon, <http://bookboon.com/blog/bookboon-coms-global-ebook-survey/> [dostęp 24.09.2014]

obniżyła się o 4%, ale był to głównie efekt wysokiej bazy z 2012 r., kiedy triumfy na rynku święciła trylogia „Fifty Shades of Grey”. Co ciekawe, mimo spadku całkowitej liczby sprzedaży książek, w przypadku e-booków odnotowano wzrost aż o 20%. Nielsen szacuje, że w 2013 r. czytelnicy w Wielkiej Brytanii kupili 80 mln e-booków i zapłacili za nie łącznie 300 mln funtów.⁴ Z kolei PwC szacuje, że już w 2018 r. ilość sprzedanych e-booków w Wielkiej Brytanii będzie wyższa niż tradycyjnych książek. W tym czasie sprzedaż cyfrowych książek ma wzrosnąć z 380 milionów funtów do 1 miliarda funtów, a przychody ze sprzedaży drukowanych wydań mają się obniżyć do poziomu 912 milionów funtów. Według PwC w 2013 r. cyfrowe książki przyniosły 18% wszystkich przychodów na rynku księgarskim, ale w 2018 r. odsetek ten wzrośnie aż do 41%.⁵

Czytelnicy polubili e-księgarnie za szeroką ofertę, atrakcyjne ceny i wygodę robienia zakupów, ale także za możliwość łatwego zapoznania się z ofertą. Internauci odwiedzają witryny wirtualnych księgarni, by zobaczyć, jakie nowości pojawiły się na rynku, dowiedzieć się więcej o treści interesujących ich książek, przeczytać udostępnione przez wydawnictwo lub autora fragmenty, sprawdzić ceny, poznać opinie innych czytelników i zobaczyć, jak oceniają oni daną pozycję. Większość księgarni prezentuje listy najlepiej sprzedających się książek w poszczególnych kategoriach, dodatkowo pokazuje, co kupili inni czytelnicy danej książki, a także sugeruje, co mogłoby zainteresować potencjalnego klienta. Informacje prezentowane na stronach księgarni internetowych stanowią ważne źródło wiedzy i mają istotny wpływ na decyzje kupujących.

O tym, czy i jaką książkę konsument włoży do prawdziwego lub wirtualnego koszyka, decyduje wiele czynników, a proces decyzyjny może być bardzo różny. Czytelnicy czekający na kolejną książkę ulubionego autora raczej nie przejmą się negatywnymi opiniami lub wysoką ceną, ale do zakupu debiutu kupujących trzeba zachęcić reklamą i entuzjastycznymi recenzjami. Badania dotyczące czynników, jakie wpływają na decyzję o zakupie książek, w zależności od metodologii pokazują różne wyniki, ale zawsze na wysokich pozycjach pojawia się w nich autor, temat książki i cena. Według raportu

⁴ Nielsen, Raport Book & Consumer 2013, maj 2014

⁵ PwC, Raport Global entertainment and media outlook 2014-2018, <https://www.pwc.com/gx/en/global-entertainment-media-outlook/segment-insights/consumer-and-educational-book-publishing.jhtml> [dostęp 24.09.2014]

Versoadvertising.com wśród czynników, które są „bardzo ważne” przy wyborze książek, czytelnicy najczęściej wymieniają: opinię o autorze (52%), rekomendacje znajomych (49%), cenę (45%), recenzje (37%), wygląd okładki i zamieszczone na niej informacje (22%) oraz reklamę (14%).⁶ Z danych przedstawionych z konferencji *Books & Consumer*, udostępnionych przez *Booksellers.org.uk*, wynika natomiast, że największy wpływ na decyzje kupujących książki mają temat i autor (po 25% odpowiedzi), cena (20%) i to, czy książka jest częścią jakiejś serii (17%). Czynniki takie jak wygląd okładki i blurb (zarówno na okładce, jak i na stronie internetowej księgarni) znalazły się dalszych pozycjach z wynikiem odpowiednio 10% i 6%.⁷

Poszczególne etapy procesu decyzyjnego i czynniki, które na to wpływają, są różne w zależności od rodzaju książki i celu jej zakupu. Dane Nielsena potwierdzają, że czytelnicy decydują się wydać pieniądze, jeśli lubią autora lub serię, interesuje ich określony temat, książka ma atrakcyjną cenę lub można ją akurat kupić w promocyjnej ofercie, a także jeśli zainteresuje ich krótki opis lub pobieżne przejrzanie zawartości książki. W przypadku książek drukowanych większe znaczenie ma znajomość autora lub serii i przydatność na prezent – czynniki te były ważne zarówno dla kupujących w tradycyjnych księgarniach, jak i księgarniach online. Zakupy papierowych książek są zwykle bardziej planowane, podczas gdy e-booki, szczególnie *self-publishing*, kupuje się często pod wpływem impulsu. W przypadku książek elektronicznych większe znaczenie ma także cena i blurb.⁸

Przytoczone badania opierają się na deklaracjach konsumentów, dlatego często nie uwzględniają lub nie doceniają czynników, które mogą mieć wpływ na decyzję, choć wydają się zbyt oczywiste albo, wbrew pozorom, nieistotne. Kompletnie i odpowiednio przedstawione informacje o książce zamieszczane na stronie internetowej księgarni rzadko bywają wymieniane wśród najważniejszych czynników wpływających na decyzję kupujących, ale niewątpliwie mają duże znaczenie. Wyraźnie pokazują to wyniki badania firmy Nielsen na temat związku między metadanymi książki a wynikami sprzedażowymi. Analiza sprzedaży 100 tysięcy najpopularniejszych tytułów na rynku brytyjskim pokazała,

⁶ Verso Advertising, op.cit.

⁷ Booksellers.org.uk, Raport Books & Purchasing Decisions, www.booksellers.org.uk/BookSellers/media/SiteMediaLibrary/IndustryNews/Books-Purchasing-Decisions.pdf [dostęp 24.09.2014]

⁸ Nielsen, op.cit.

że pozycje, które były opisane zgodnie ze standardami dostarczania danych bibliograficznych BIC Basic, znajdowały kilkakrotnie więcej nabywców niż książki pozbawione niektórych informacji. Normy BIC Basic obejmują 11 podstawowych elementów, takich jak m.in. ISBN, tytuł, kategoria, data publikacji, wydawca, cena i okładka. Badanie udowadnia, że braki w danych mają wyraźny wpływ na sprzedaż książek zarówno online, jak i offline. Związek między kompletem podstawowych informacji a wynikami sprzedaży był największy w kategoriach beletrystyka i literatura dla dzieci.⁹

Badanie firmy Nielsen pokazuje również, że wydawcom opłaca się dbać nie tylko o 11 podstawowych elementów opisu, ale także o 4 dodatkowe kategorie: krótki i długi opis, informacja o autorze i recenzja. Jak się okazuje, takie marketingowe informacje są szczególnie ważne dla książek sprzedawanych online – obecność jednego z tych elementów zwiększa sprzedaż o 55%, dwie takie informacje - o 71%, a w przypadku trzech wzrost wynosi 120%. Książka, która posiada wszystkie cztery elementy, znajduje średnio o 178% więcej nabywców niż bez tych danych. Zainteresowanie książką dla dzieci można zwiększyć, zamieszczając w materiałach kierowanych do dystrybutorów i księgarni krótki opis (do 350 znaków), sprzedaż beletrystyki znacząco poprawia natomiast obecność długiego opisu.¹⁰

Dane bibliograficzne i materiały promocyjne przygotowane przez autorów i wydawnictwa powinny być kompletne, rzetelne i starannie opracowane, ponieważ to na ich podstawie sprzedawcy konstruują ofertę, przypisują książkę do odpowiedniej kategorii i rekomendują określonej grupie czytelników. Wysoka jakość tych danych ma kluczowe znaczenie dla sprzedaży w księgarniach internetowych. Ich klienci nie mają możliwości fizycznego kontaktu z książką, więc decyzję o zakupie podejmują na podstawie tego, co znajdą w ofercie.

W przypadku książek kupowanych online, szczególnie e-booków, każda dodatkowa informacja, taka jak streszczenie, informacje o autorze, ocena i opinie czytelników, zwiększa szanse na przekonanie klienta do zakupu. Czytelnik (zachęcony na przykład rekomendacjami znajomych lub wzmianką o książce w mediach), który wejdzie na stronę

⁹ Nielsen, White Paper: The Link Between Metadata and Sales, styczeń 2012

¹⁰ ibidem

e-księgarni i nie dowie się nic nowego lub ciekawego na temat poszukiwanej pozycji, prawdopodobnie jej nie kupi. Szczególnie dużą rolę odgrywa również opis książki (nazywany też blurb), który zastępuje informację zamieszczaną w papierowych wydaniach na ostatniej stronie okładki, ale niekoniecznie jest jej dokładnym powtórzeniem. Ten krótki tekst, który można znaleźć przy każdej pozycji, zawiera zwykle krótkie streszczenie i zachętę do sięgnięcia po daną pozycję. Czasem opis jest uzupełniony informacjami o autorze, cytatami lub fragmentami entuzjastycznych recenzji, szczególnie jeśli są one autorstwa znanych, cenionych osobistości. Dobrze napisany blurb przyciąga uwagę, rozbudza ciekawość na temat książki i zachęca do zapoznania się z pozostałymi informacjami. Nie można powiedzieć, że od opisu zależy to, czy klient kupi książkę, ale na pewno na tym etapie procesu decyzyjnego bardzo łatwo stracić potencjalnego czytelnika. Na ostateczną decyzję o zakupie wpływają również inne elementy oferty, takie jak choćby tytuł, okładka, oceny i opinie czytelników, lecz jeśli opis nie będzie zachęcający, to nikt nie będzie czytał recenzji czy fragmentów książki.

Dane firmy Nielsen pokazują, że już samo zamieszczenie dłuższego opisu przyczynia się do wzrostu sprzedaży książki. Nie znaczy to, że można w tym miejscu wpisać cokolwiek. Blurb to nie tylko źródło informacji o treści książki, ale też jej najważniejsza i najprostsza reklama. Tylko przemyślana kompozycja i odpowiednia treść opisu pozwalają lepiej wykorzystać jego marketingowy potencjał. Potwierdza to badanie przeprowadzone przez Codex Group, w ramach którego sprawdzono reakcje i opinie internautów na temat różnych opisów ponad 7 tys. książek z kilku dziedzin. Jak się okazało, treść opisu ma znaczenie szczególnie dla beletrystyki, przy czym bardzo ważne jest, by była ona spójna z tytułem i okładką. Kluczowym elementem opisu jest nagłówek lub pierwsze zdanie – jeśli początek nie będzie wystarczająco intrygujący lub zaskakujący, znudzony klient może zrezygnować z czytania reszty tekstu.¹¹ Autorzy badania przyznają, że nie ma magicznej formuły na doskonały opis, który zapewnia wysoką sprzedaż. Inne opisy najlepiej sprawdzają się w poszczególnych kategoriach, inaczej też reagują na nie czytelnicy z różnych grup wiekowych. Analiza pozwoliła jednak z całą pewnością stwierdzić, że treść opisów ma znaczenie i wpływa na sprzedaż książek.

¹¹ Owen Laura Hazard, Jacket Copy Sells Books, So Make It Good, maj 2009, <http://www.publishingtrends.com/2009/05/jacket-copy-sells-books/> [dostęp 24.09.2014]

Nie dla wszystkich książek treść opisu ma jednakowo duże znaczenie. Książki, które są napisane przez popularnych autorów, stanowią kontynuację przygód ulubionego bohatera lub są szeroko promowane w mediach, zapewne znajdą nabywców, nawet jeśli nie będą zbyt ciekawie przedstawione na stronie księgarni. Można z całą pewnością stwierdzić, że opis nie miał wpływu na decyzję milionów klientów o zakupie ostatniej części przygód Harry’ego Pottera, ale w przypadku mniej wyczekiwanych pozycji jego znaczenie jest niezaprzeczalne. Opis jest szczególnie ważny dla książek mniej znanych autorów i nowych serii, zwłaszcza jeśli nie mogą liczyć na dodatkowe wsparcie (recenzje, rekomendacje w mediach) czy reklamę. Na blogach i w serwisach dla autorów, którzy samodzielnie publikują swoją twórczość, można znaleźć szereg porad, jak napisać świetny opis, co należy nim zawrzeć, a czego lepiej się wystrzegać.

Większość porad odnosi się do sprzedaży w księgarni Amazon, ponieważ jak wynika z badania przeprowadzonego przez Kantar Media na zlecenie Ofcom, aż 79% Brytyjczyków kupujących e-booki korzysta właśnie z oferty tej firmy.¹² Amazon ma swoje standardy dotyczące długości opisu – może on wynosić od 30 do 4000 znaków.¹³ W przypadku bardzo długich tekstów trzeba liczyć się z tym, że część opisu będzie można przeczytać dopiero po kliknięciu w „zobacz więcej”, dlatego najważniejsze treści lepiej zawrzeć w pierwszych około 200 słowach, które będą widoczne po wejściu na stronę z daną książką.

W poradniku o wymownym tytule „How To Sell More Books with Awesome Amazon Descriptions” Jim F. Kukral przekonuje, że moment, w którym zaciekawiony okładką i tytułem czytelnik przechodzi do opisu, jest punktem krytycznym w procesie decyzyjnym. Jeśli kupujący nie znajdzie w opisie nic, co go zainteresuje, zaintryguje, „zaczepi”, to po prostu wciśnie przycisk „wstecz” i poszuka innej pozycji. Jim F. Kukral przywołuje badania prowadzone przez Autor Marketing Club, z których wynika, że solidny, barwny i wciągający opis może zwiększyć sprzedaż książki o 4-15%.¹⁴ Dobry blurb powinien szybko i stanowczo przekonać czytelnika, dlaczego warto sięgnąć po tę książkę. Autor

¹² Online Copyright Infringement Tracker, Annex 1 – Individual content types, Wave 4 (Covering period March – May 2013), Prepared for Ofcom by Kantar Media, s. 56, http://stakeholders.ofcom.org.uk/binaries/research/telecoms-research/online-copyright/w4/OCI_MAIN_REPORT_W4_FINAL.pdf [dostęp 24.09.2014]

¹³ Amazon. com

¹⁴ Kukral Jim F., How To Sell More Books with Awesome Amazon Descriptions, 2013 s. 5

zwraca uwagę, że trzeba starannie dobierać słowa zachęcające do akcji, bo określenia o tym samym znaczeniu mogą różnie wpływać na decyzję konsumentów – na przykład „Dodaj do koszyka” przynosi znacznie lepsze efekty niż „Kup teraz”. Jim F. Kukral przytacza również za *Coppyblogger.com* listę 5 najbardziej przekonujących słów w języku angielskim, które szczególnie mocno wpływają na proces decyzyjny (są to słowa: You, Free, Instantly, Because i New¹⁵) i zachęca do wykorzystywania ich w opisie. Jego zdaniem trzy najważniejsze powody, które mogą zniechęcić ludzi do zakupu danej książki, to wysoka cena, obawa przed zmarnowaniem pieniędzy oraz obawa przed rozczarowaniem, wyjściem na głupka. Dobry opis powinien przekonać potencjalnych czytelników, że to im nie grozi, dlatego trzeba wyjaśnić, czego należy się spodziewać po tej pozycji i dlaczego jest warta swojej ceny. Warto uspokoić obawy czytelników, podając konkretne przykłady, że książka jest wartościowa i kupiło ją wielu ludzi - dobrze nadają się wszelkie dane liczbowe. Dobrym pomysłem może być także wyraźne określenie gatunku, na przykład „romans historyczny” czy „thriller kryminalny”. Jim F. Kukral podkreśla znaczenie nagłówka i zachęca, by w tej części opisu pochwalić się (wspomnieć o nagrodach, wysokiej sprzedaży, pozytywnych opiniach), podkreślić jakąś nietypową, wyróżniającą się cechę książki lub spróbować zaintrygować czytelnika. Nie należy przesadzać z długością opisu, ani streszczać w nim szczegółowo całej fabuły, lepiej podkreślić unikalne cechy książki i złożyć czytelnikowi konkretną obietnicę związaną z lekturą.

Swoimi doświadczeniami w zakresie konstruowania opisów podzielił się również pisarz Mark Edwards, który po publikacji w *Amazon.com* swojego pierwszego thrillera dokładnie przestudiował informacje zamieszczane przy innych książkach z tej kategorii i wciąż modyfikował swój blurb, dzięki czemu jego dzieło trafiło w końcu na listę bestsellerów. W poradniku „How to Craft a Sizzling Book Description” M. Edwards przekonuje, że opis i okładka to dwa główne narzędzia sprzedaży książek w *Amazon.com* i przestrzega przed bagatelizowaniem znaczenia tych elementów. On również podkreśla znaczenie pierwszej linijki lub pierwszego zdania opisu i tłumaczy, że to największa szansa dla autora, by zwrócić uwagę kupującego i zachęcić do zakupu. M. Edwards uważa, że opisy powinny oddawać atmosferę książki – potencjalnych czytelników thrillerów trzeba postraszyć, a

¹⁵ Kukral Jim F., op.cit., s. 8

poszukujących książek o miłości wprowadzić w romantyczny nastrój. Jego zdaniem należy unikać zbyt długich opisów, lepiej skupić się za to na atrakcyjnej formie, która zaskoczy i zaintryguje czytelnika. M. Edwards uważa, że w opisach świetnie sprawdzają się pytania skierowane do pytającego, najlepiej wykorzystujące grę słów lub odwołujące się do emocji czy lęków. Ważne jest, by opis pokazał kupującemu, czego może się spodziewać, dlatego potrzebny jest jasny przekaz, że to np. kryminał z elementami horroru lub poradnik z gatunku „jak zrzucić zbędne kilogramy”. Jego zdaniem nie należy jednak zbyt szczegółowo skupiać się na samej treści książki, tylko lekko naszkicować fabułę, głównych bohaterów i ich problemy, nie wdawać się w szczegóły, a zamiast tego znaleźć coś, rozbudzi ciekawość czytelnika i zapadnie w pamięć.¹⁶

Znaczenie opisu dla sprzedaży e-booków podkreśla również Mark Coker, twórca platformy Smashwords dla autorów *self-publishing*. W swoim poradniku „The Secrets to Ebook Publishing Success” stwierdza on, że opis jest prawdopodobnie drugim (po tytule) najważniejszym elementem metadanych. Jego zdaniem opis to dla autora szansa, aby zainteresować czytelnika, zachęcić do przeczytania fragmentu, a nawet od razu do zakupu. Dobry opis jest dostosowany do grupy docelowej i stanowi próbkę talentu autora, oddaje jego styl i sposób formułowania myśli. M. Coker zwraca uwagę na jeszcze jedną ważną cechę opisów – poprawność językową. Jego zdaniem błędy ortograficzne, interpunkcyjne i gramatyczne, a nawet zwykłe literówki, mogą zepsuć nawet najlepsze wrażenie i skutecznie zniechęcić potencjalnego czytelnika.¹⁷

¹⁶ Edwards Mark, *How to Craft a Sizzling Book Description A Guide for Writers*, 2013, s. 5

¹⁷ Coker Mark, *The Secrets to Ebook Publishing Success*, 2014, s. 51

Rozdział 2

Przedmiot mojej analizy stanowią opisy zamieszczone przy książkach w księgarni internetowej *Amazon.co.uk*. Celem mojej analizy jest wskazanie i wyróżnienie takich cech opisu książki, które powodują jej dużą (lub małą) popularność, a co za tym idzie – wpływają na jej zakup i pobranie przez użytkowników księgarni. Cel ten zrealizowałam poprzez porównanie tekstów, które można znaleźć przy najpopularniejszych i najmniej popularnych książkach z kategorii *Literature & Fiction*.

Żeby ograniczyć wpływ innych czynników (marketingowych, reklamowych, promocyjnych), mogących przekładać się na liczbę sprzedanych egzemplarzy, postanowiłam w analizie uwzględnić tylko opisy e-booków, które można bezpłatnie pobrać na czytnik Kindle. Zabieg ten nie tylko pozwolił pominąć znaczenie ceny, ale również czasu i kosztów wysyłki, które mogą mieć wpływ na decyzję potencjalnego czytelnika. Większość książek, których opisy analizowałam, to *self-publishing* (tj. książki wydane samodzielnie przez ich autorów), co oznacza, że prawdopodobnie nie były zbyt szeroko promowane, a nazwiska autorów raczej nie zaistniały jeszcze w świadomości czytelników.

Ponadto w swojej analizie uwzględniłam książki opublikowane w ciągu ostatnich 90 dni, czyli takie, które stosunkowo niedawno pojawiły się w ofercie i jeszcze nie zdążyły wyrobić sobie marki; można się spodziewać, że mają one mniej ocen i komentarzy niż pozycje, które są dostępne od dawna. Uważam, że w przypadku nowo pojawiających się książek to właśnie opis ma największe znaczenie i najbardziej przekłada się na decyzję o zakupie książki (w tym przypadku o pobraniu).

W celu wyszukania pozycji spełniających opisane wyżej założenia, wybrałam następujące opcje: Kindle Store\Kindle eBooks\Literature & Fiction\Free: Last 90 days¹⁸. Otrzymałam listę 1349 książek, które posortowałam według kryterium popularności „popularity”. Następnie stworzyłam dwa zbiory danych. W pierwszym znalazły się opisy 100 najbardziej popularnych książek. W drugim zbiorze miały się znaleźć opisy 100 najrzadziej pobieranych pozycji. Okazało się jednak, że większość najmniej popularnych

¹⁸ http://bit.do/Amazon_free_kindle_books

książek była napisana w języku innym niż angielski (najczęściej po niemiecku). Ponieważ zależało mi na tym, żeby wszystkie książki z moich zbiorów miały jednakowe szanse u czytelników, a nawet najciekawszy opis nie zachęci do czytania książki w języku, którego się nie zna, dlatego pominęłam wszystkie pozycje obcojęzyczne, nawet jeśli sam opis był napisany po angielsku. W efekcie końcowym zbior WORST zawiera opisy 100 najrzadziej pobieranych pozycji anglojęzycznych.

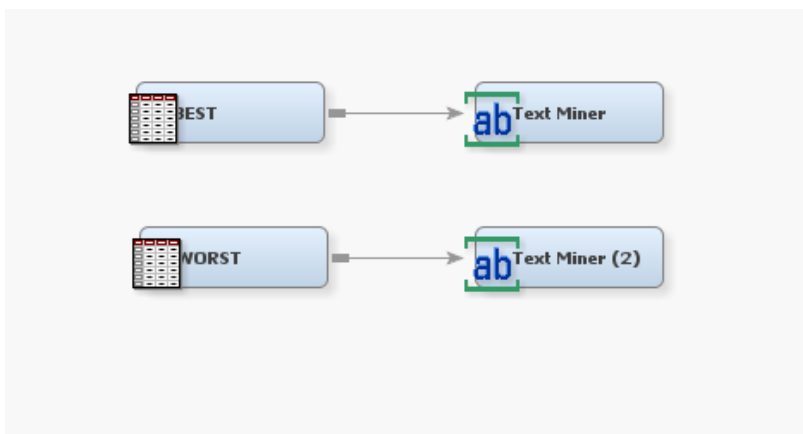
2.1 Analiza zbiorów danych

Tabele z danymi mają po pięć kolumn – tytuł, opis i zdanie (nagłówek opisu lub jego pierwsze zdanie) oraz liczba wyrazów i znaków w każdym opisie. Po zaimportowaniu danych otrzymałam dwa zbiory BEST i WORST. Analizy zrealizowano przy użyciu narzędzia SAS Text Miner 4.2, oprogramowania działającego w środowisku SAS Enterprise Miner 6.2.

Opisy z obu zbiorów wyraźnie różnią się długością. W zbiorze BEST opisy mają średnio 198 słów i 1120 znaków, natomiast w zbiorze WORST średnia liczba wyrazów w opisie wynosi 107, a średnia liczba znaków 617, czyli prawie dwukrotnie mniej. Widać więc wyraźnie, że klienci e-księgarni chętniej pobierają te książki, o których mają więcej informacji. Opisy na pewno nie mogą być przesadnie długie, ale prawdopodobnie 100 słów to za mało, by przekonać klienta, że warto poświęcić czas na daną pozycję.

W dalszej kolejności analizę tekstów przy użyciu węzła Text Miner przeprowadziłam równoległe dla obu zbiorów. Rozpoczęłam od utworzenia źródła danych (w tym przypadku BEST oraz WORST), które połączyłam z węzłem Text Miner, jak ilustruje Rysunek 1.

Rys. 1. Źródła danych obu zbiorów połączone z węzłem Text Miner

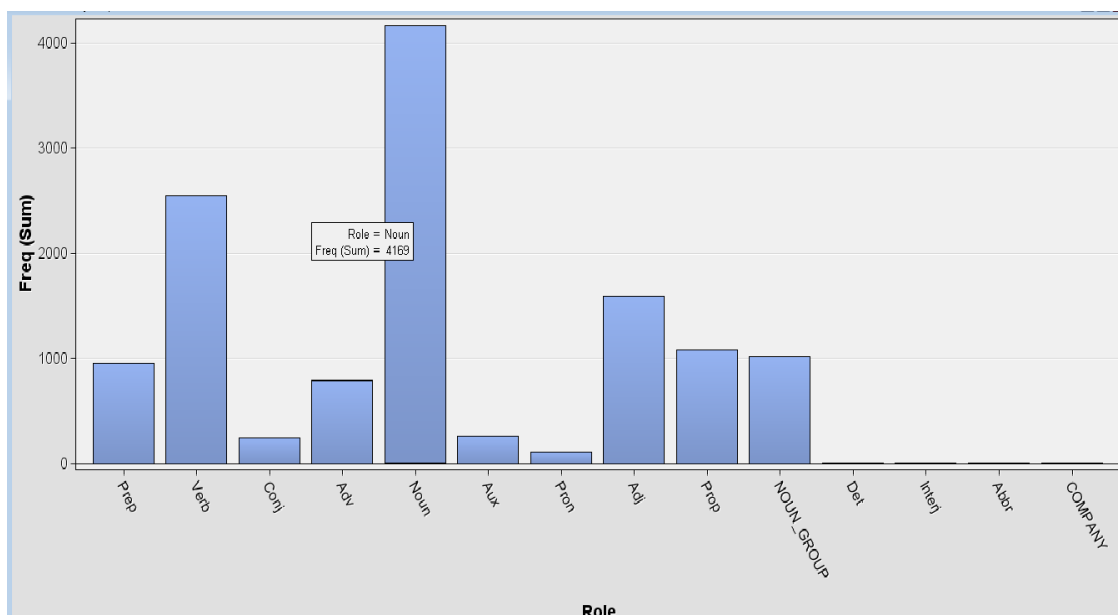


Źródło: opracowanie własne.

W moich zbiorach aż 3 zmienne mają przypisaną rolę *Text*, ale ponieważ analiza dotyczy opisów, dlatego w zakładce *Variable* ustawiłam „yes” w kolumnie *Use* tylko przy zmiennej opis, a przy pozostałych tę opcję zmieniłam na „no”. Z kolei zmieniłam w zakładce *Parse* opcję przy *Terms on single documents* na „yes”, by w uwzględnić w analizie wszystkie występujące w opisach słowa, nawet te, które pojawiły się tylko raz w danym dokumencie (opisie). Pozostałe opcje zostawiam bez zmian.

Pierwszy krok mojej analizy polegał na porównaniu częstotliwości występowania poszczególnych części mowy w obu zbiorach, co można zaobserwować na Rysunku 2 i Rysunku 3 zamieszczonych dalej.

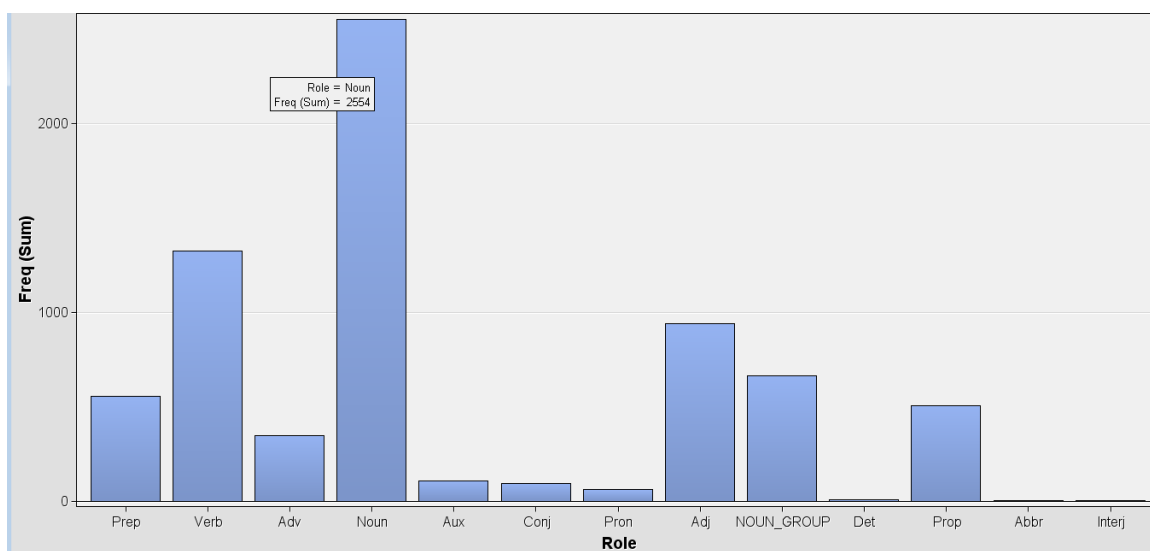
Rys 2. Zbiór BEST - Wykres przedstawiający części mowy zidentyfikowane w opisach



Źródło: opracowanie własne.

Jak można zauważyć na Rysunku 2, w 100 najskuteczniejszych opisach najwięcej wyrazów zostało rozpoznanych jako rzeczowniki (4169). W tekstach występuje też sporo czasowników (2545), przymiotników (1592), nazw własnych (1075), grup rzeczowników (1015) i przysłówków (788).

Rys. 3. Zbiór WORST - Wykres przedstawiający części mowy zidentyfikowane w opisach



Źródło: opracowanie własne.

Rysunek 3. pokazuje, że w opisach 100 najmniej popularnych książek również najczęściej pojawiały się rzeczowniki (2554), czasowniki (1323) i przymiotniki (932). Znały się w nich również 663 grupy rzeczowników, 535 nazwy własne, a także 348 przysłówki.

Częstotliwość występowania poszczególnych części mowy w obu zbiorach jest bardzo podobna, co świetnie widać na wykresach - mają one prawie jednakowy kształt. Można zauważyć, że w tekstach w pierwszym zbiorze nieco częściej pojawiają się czasowniki i przymiotniki, co mogłoby świadczyć o tym, że lepiej do zakupu książki przekonują opisy dynamiczne i barwne, ale różnice nie są zbyt duże, by można było stanowczo formułować takie wnioski. Istotne znaczenie wydaje się mieć natomiast większa częstotliwość nazw własnych w opisach najlepiej sprzedających się książek. Może to oznaczać, że klienci lubią konkret – wolą wiedzieć, kto jest bohaterem lub gdzie dzieje się akcja. Problem ten zbadałam dokładniej w dalszej części analizy.

Z obu wykresów wynika, że pozostałe części mowy mają niewielki udział w tekstach. Łatwiej jest analizować dane, jeśli jest ich mniej, dlatego w kolejnych krokach pominęłam mniej ważne (zarówno pod względem ilości, jak i znaczenia) wyrazy. Przed wykonaniem tego kroku, sprawdziłam jednak, czy i jak często w tekstach pojawiają się zaimki „you” lub „I”, które mogą wskazywać, w której osobie napisany jest opis. W jednym z poradników dla początkujących autorów zalecano zwracanie się bezpośrednio do czytelnika¹⁹, co wymagałoby użycia słowa „you”. Częstotliwość występowania wymienionych zaimków osobowych sprawdziłam w tabeli *Terms* w oknie *Interactive*, co przedstawiono na Rysunku 4.

¹⁹ Kukral Jim F., op.cit., s. 10

Rys. 4. Tabela przedstawiająca częstotliwość występowania zaimków

Terms						
TERM	FREQ	# DOCS	KEEP	WEIGHT	ROLE ▲	ATTRIBUTE
how much	1	1	<input checked="" type="checkbox"/>	1.0	Pron	Alpha
i	151	23	<input type="checkbox"/>	0.421	Pron	Alpha
it	161	59	<input type="checkbox"/>	0.207	Pron	Alpha
itself	3	3	<input type="checkbox"/>	0.762	Pron	Alpha
me	31	12	<input type="checkbox"/>	0.5	Pron	Alpha
myself	2	2	<input type="checkbox"/>	0.85	Pron	Alpha
no one	6	5	<input checked="" type="checkbox"/>	0.662	Pron	Alpha
nobody	1	1	<input type="checkbox"/>	1.0	Pron	Alpha
one	27	23	<input checked="" type="checkbox"/>	0.335	Pron	Alpha
she	285	59	<input type="checkbox"/>	0.183	Pron	Alpha
so	1	1	<input type="checkbox"/>	1.0	Pron	Alpha
some	1	1	<input type="checkbox"/>	1.0	Pron	Alpha
someone	7	7	<input type="checkbox"/>	0.578	Pron	Alpha
that	134	62	<input type="checkbox"/>	0.148	Pron	Alpha
theirs	1	1	<input type="checkbox"/>	1.0	Pron	Alpha
themselves	3	3	<input type="checkbox"/>	0.762	Pron	Alpha
there	13	11	<input type="checkbox"/>	0.49	Pron	Alpha
they	115	50	<input type="checkbox"/>	0.22	Pron	Alpha
this	29	23	<input type="checkbox"/>	0.343	Pron	Alpha
two	2	2	<input checked="" type="checkbox"/>	0.85	Pron	Alpha
us	2	2	<input type="checkbox"/>	0.85	Pron	Alpha
we	9	7	<input type="checkbox"/>	0.591	Pron	Alpha
what	28	19	<input type="checkbox"/>	0.383	Pron	Alpha
whatever	4	4	<input type="checkbox"/>	0.7	Pron	Alpha
where	14	13	<input type="checkbox"/>	0.45	Pron	Alpha
which	9	7	<input type="checkbox"/>	0.591	Pron	Alpha
who	69	43	<input type="checkbox"/>	0.222	Pron	Alpha
you	81	26	<input type="checkbox"/>	0.38	Pron	Alpha
yours	1	1	<input type="checkbox"/>	1.0	Pron	Alpha
yourself	4	4	<input type="checkbox"/>	0.7	Pron	Alpha

Źródło: opracowanie własne.

Jak widać na przedstawionym Rysunku 4, w zbiorze BEST „I” pojawia się 151 razy w 23 dokumentach, „me” użyto 31 razy w 12 opisach, a „myself” występuje 2 razy w 2 tekstach. W opisach pojawiły się również zaimki „we” (9 razy w 7 dokumentach) i „us” (2 razy w 2 dokumentach). Zaimek „you”, którego użyto 81 razy, można znaleźć w 26 opisach, w 4 dokumentach występuje słowo „yourself”, a w jednym pojawia się forma „yours”.

Z kolei w zbiorze WORST „I” pojawia się jedynie 11 razy w 4 dokumentach, a „me” tylko 4 razy w 3 dokumentach. Zaimka „we” użyto 13 razy w 8 dokumentach, „us” występuje 3 razy w 2 dokumentach, raz pojawia się też słowo „ours”. Zaimek „you” występuje 49 razy w 20 opisach, jednokrotnie użyto też słowa „yourself”.

Jak się okazuje, zaimek „you” można znaleźć w co czwartym opisie najpopularniejszych książek. W drugim zbiorze tekstów forma wskazująca na drugą osobę pojawia się wprawdzie rzadziej, ale występuje co najmniej raz w co piątym dokumencie. Można zatem stwierdzić, że choć zwracanie się bezpośrednio do czytelnika może być dobrym pomysłem na wyróżnienie się, to raczej nie przesądza jednoznacznie o popularności książki.

Bardziej interesująca wydała mi się ogromna różnica w występowaniu form wskazujących na pierwszą osobę. Użyłam opcji *Filter Terms*, by przyjrzeć się dokładniej dokumentom ze zbioru BEST, w których występuje „I”, co widnieje na Rysunku 5.

Rys. 5. Tabela przedstawiająca dokumenty, w których występuje zaimek „I”

OPIS	TYTUŁ	ZDANIE	ZNAKI	WYRAZY
My name is Hope Sinclair, and I'm a liar. I've deceived everyone, including my own brothers, in an effort to hide what I do and who I really am. It's not that I don't want to be honest, but I really have no choice. I thought I was strong, that I was doing fine, until I saw Jason Sutherland last New Year's Eve, and he stripped me bare in more ways than one. Jason is everything I want, but a man I can never have. We've known each other most of our lives, but I've avoided him since I turned eighteen because I know he's dangerous to me, to my façade. Turns out I was right. After our night together last New Year's Eve, I haven't been able to forget him, or the way he set my body on fire. I crave Jason like an addictive drug now, that fateful night haunting me still. Is my One Night With a Billionaire going to be my downfall or my salvation? **This is a prequel novella to Billionaire Unmasked, Jason and Hope's book in The Billionaire's Obsession series. ** 18+ only.	One Night ...	My name i...	983.0	186.0
This, my children, is the Fairy Chronicles. The stories I tell are ancient truths from a time when the world I created was filled with joy and love and peace and happiness. Then one day, something very, very horrible happened. The truth begins with a very, very bad human king and a very, very horrible day. Who that horrible king was, what happened on that very, very horrible day, and what becomes of my fairy and human children, well, you must find that out for yourself.	The Fairy ...	This, my c...	475.0	86.0
Ebola, Terrorism, and Hope In 1989 the Ebola virus mutated to into an airborne strain that infected humans for the first time on American soil in Reston, Virginia. Through belated containment efforts and luck, nobody died. Now, in the remote East African village of Kapchorwa, the Ebola virus has mutated into another airborne strain without losing any of its deadly potency. In this thriller, terrorists stumble across this new, fully lethal strain and while the world fearfully watches the growing epidemic in West Africa as Sierra Leone goes into country-wide lockdown, only a few Americans are aware of Ebola K and the danger it poses to be the deadliest pandemic in the history of mankind. Can they do anything to protect				

TERM ▲	FREQ	# DOCS	KEEP	WEIGHT	ROLE	ATTRIBUTE
i	151	23	<input type="checkbox"/>	0.421	Pron	Alpha

Źródło: opracowanie własne.

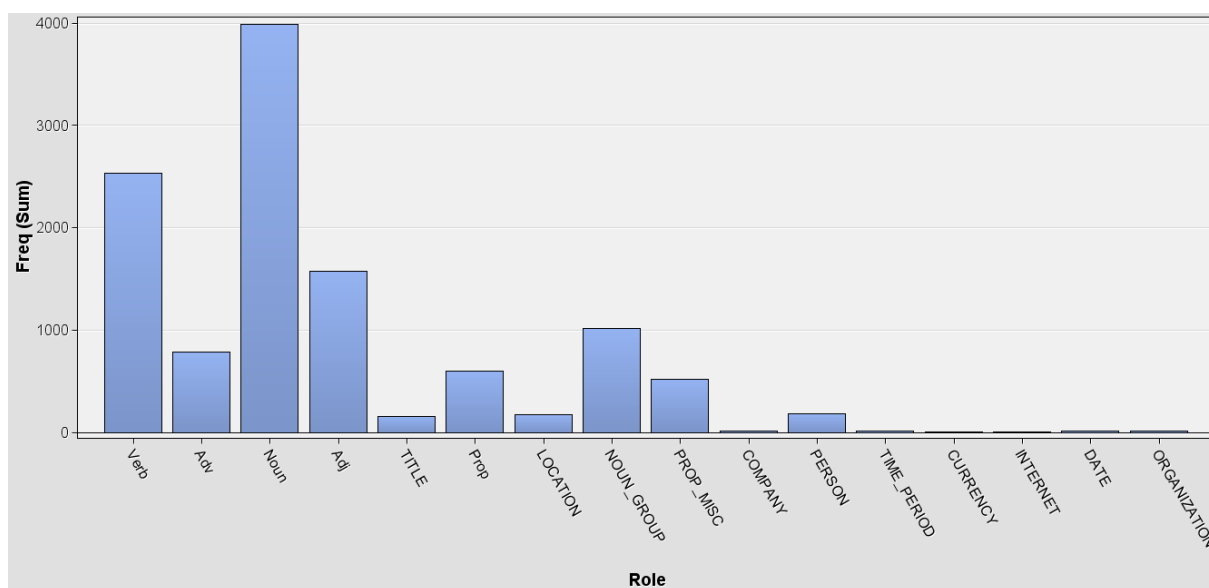
Jak się okazało, większość z nich jest rzeczywiście w całości pisana w pierwszej osobie, tylko w kilku „I” pojawia się jedynie w przytoczonych entuzjastycznych recenzjach książki (np. „One of the Best romance book I’ve read this year”) lub cytatach. Może to oznaczać, że klienci e-księgarni lubią taką formę opisu - być może, prowadząc opowieść w pierwszej osobie, łatwiej zwrócić uwagę czytelnika i zaintrygować go, a jednocześnie pokazać, czego może się spodziewać po lekturze.

Po wyjaśnieniu tej kwestii pozbyłam się części mowy, które nie miały żadnego znaczenia w dalszej analizie. Zrobiłam to za pomocą opcji *Ignore Parts of Speech* - odznaczyłam *Aux, Conj, Det, Interj, Part, Prep* i *Pron*. W kolejnych etapach analizy pod uwagę brałam zatem tylko: przymiotniki, przysłówki, rzeczowniki, liczby, czasowniki, nazwy własne i przymiotniki odczasownikowe. Po tej operacji różnica w liczbie terminów w tabeli *Term*

nie zmniejszyła się jednak znacząco - w zbiorze BEST spadła z 5150 do 4940, a w zbiorze WORST z 3686 do 3507.

W kolejnej części analizy przyjrzałam się nazwom własnym, które na wykresach przedstawiających częstotliwość występowania wydawały się być dość liczną grupą wyrazów. Intuicja podpowiadała mi, że może chodzić o imiona i nazwiska bohaterów opisywanych książek. Aby to sprawdzić, zazaczyłam „yes” przy opcji *Find Entities*. W *Type od Entities* nie ograniczyłam się jednak do kategorii *People's name*, by nie pozbawić się szansy na zaobserwowanie innych zależności i zostawiłam zaznaczone wszystkie dostępne opcje. W wynikach otrzymałam m.in. następujące wykresy częstotliwości wyrazów, którym przypisano poszczególne role, jak można zauważyć na Rysunku 6.

Rys. 6. Zbiór BEST - Wykres przedstawiający jednostki specjalne zidentyfikowane w opisach

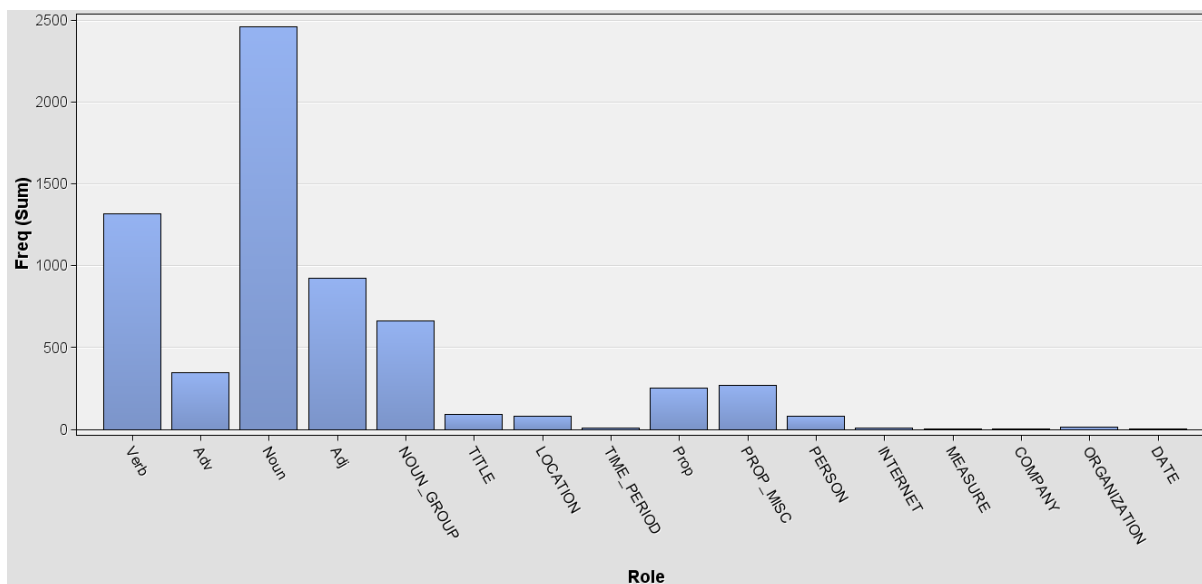


Źródło: opracowanie własne.

W zbiorze BEST spośród nowych grup najliczniejsze są PROP_MISC (520 wyrazów), PERSON (184) i LOCATION (172) oraz TITLE (156).

Z kolei w zbiorze WORST rozkład terminów według ich roli wygląda podobnie. Najliczniejsze są grupy PROP_MISC (271 słów), TITLE (94), LOCATION (83) i PERSON (79), co pokazuje Rysunek 7.

Rys. 7. Zbiór WORST - Wykres przedstawiający jednostki specjalne zidentyfikowane w opisach



Źródło: opracowanie własne.

Przedstawione wykresy częstotliwości poszczególnych części mowy wyraźnie pokazują, że najczęściej występują w opisach nazwy własne określone jako PROP_MISC (*proper noun with an ambiguous classification*). Po pobieżnym przejrzaniu w tabeli *Term* w oknie *Interactive* terminów, którym przypisaną taką rolę, stwierdziłam jednak, że jest to zbyt zróżnicowana grupa, by dało się wyciągnąć z tego jakiegokolwiek wniosku.

Z wykresów wynika również, że w opisach najpopularniejszych książek role osoby pojawiają się częściej niż w tych, które znalazły mniej czytelników. Po połączeniu wszystkich terminów z rolą PERSON w jeden zbiór (za pomocą opcji *Treat as a equivalent term*) okazało się, że imiona i nazwiska (przypuszczalnie bohaterów książek) pojawiają się aż w 58 tekstach ze zbioru BEST i 37 ze zbioru WORST.

Rys. 8. Imiona i nazwiska pojawiające się w tekstach opisów

TERM	FREQ	# DOCS	KEEP	WEIGHT	ROLE ▲	ATTRIBUTE
abby raines	184	58	<input checked="" type="checkbox"/>	1.0	PERSON	Entity
dorian marshall	1	1				
dixon black	1	1				
dirk rockwell	1	1				
diana dwayne	1	1				
debbie mccrae	1	1				
daphne du mauri...	1	1				
dan bradley	3	1				
damien andrews	1	1				
lee child	3	2				
collin morgan	1	1				
coe	1	1				

Źródło: opracowanie własne.

Tej samej metody użyłam, żeby dowiedzieć się, że wyrazy z rolą TITLE, czyli inne niż imię i nazwisko określenie osoby (np. *sister*, *student* czy *farmer*), można znaleźć w 58 dokumentach ze zbioru BEST i w 38 tekstach ze zbioru WORST. Na podstawie tych danych można przypuszczać, że opisy, z których czytelnik dowiaduje się, jak się nazywa i kim jest bohater, skuteczniej zachęcają do zakupu książki. Na pewno nie zaszkodzi też zaznaczenie w opisie, gdzie toczy się akcja - nazwy geograficzne pojawiły się bowiem w 57 dokumentach ze zbioru BEST i w 36 tekstach ze zbioru WORST.

W poradnikach dla początkujących pisarzy pojawiały się zalecenia, by w opisie kierować do potencjalnego czytelnika pytania, próbować go zaskoczyć, zaintrygować lub wzbudzić w nim emocje²⁰. Aby stwierdzić, czy tego typu porady są uzasadnione, sprawdziłam, jak często w analizowanych tekstach pojawiają się znaki zapytania i wykrzykniki. Po zaznaczeniu „yes” przy opcji *Punctuation* w zakładce *Parse*, w tabeli częstotliwości obok wyrazów pojawiły się znaki interpunkcyjne. Jak się okazało, w zbiorze BEST znak zapytania pojawia się 102 razy i występuje w 47 dokumentach, a wykrzyknik występuje 62 razy w 29 dokumentach. Autorzy skutecznych opisów dość często używali też cudzysłowu (78 razy w 15 dokumentach) i dwukropka (75 razy w 33 dokumentach).

²⁰ Kukral Jim F., op.cit., s. 6

W opisach należących do zbioru WORST znaki interpunkcyjne inne niż przecinek czy kropka pojawiają się znacznie rzadziej. Znak zapytania postawiono w nich 49 razy w 24 dokumentach, wykrzyknika użyto 29 razy w 13 dokumentach, cudzysłów pojawił się 48 razy w 13 dokumentach, a dwukropek 14 razy w 11 tekstach.

Odpowiednio zastosowane znaki interpunkcyjne mogą sprawić, że tekst będzie nie tylko lepiej zrozumiały, ale też bardziej zróżnicowany i dynamiczny, co w przypadku opisów wydaje się mieć znaczenie. Przytoczone dane zdają się potwierdzać, że opisy, które wyrażają emocje (o czym świadczą wykrzykniki), budzą w czytelniku wątpliwości lub skłaniają do przemyśleń (znaki zapytania), skuteczniej zachęcają do przeczytania książki.

Jim F. Kukral w swoim poradniku zalecał, by rozwiewać wątpliwości klientów danymi liczbowymi, ponieważ mają one dużą moc przekonywania. Aby sprawdzić, czy obecność liczb w opisach poprawia ich skuteczność, porównałam, jak często pojawiają się one w tekstach ze zbiorów BEST i WORST. Po zaznaczeniu „yes” przy opcji *Numbers* w zakładce *Parse*, w tabeli częstotliwości pojawiły się występujące w analizowanych tekstach liczby. Po połączeniu wszystkich wyrażen z rolą *Num* w jeden zbiór (za pomocą *Treat as a equivalent term*) okazało się, że liczby w różnej postaci (zapisane słownie i za pomocą cyfr) pojawiły się w zbiorze BEST aż 175 razy w 66 dokumentach.

Rys. 9. Liczby pojawiające się w tekstach opisów

TERM	FREQ	# DOCS	KEEP	WEIGHT	ROLE	ATTRIBUTE
000	175	66	<input checked="" type="checkbox"/>	1.0	Num	Num
2015	1	1				
7,400	1	1				
4.5	1	1				
100,000	1	1				
2114	1	1				
one	39	29				
2014	3	3				
thousand	1	1				
000	1	1				
24,000	1	1				
three	9	7				
1915	1	1				
25	1	1				
5	5	5				
80	1	1				
2012	1	1				
12	2	2				
100	3	3				
hundred	2	1				
ten	1	1				
3	6	6				
1918	1	1				
2013	1	1				
24	1	1				
20,000	1	1				

Źródło: opracowanie własne.

Najczęściej używano słów „one” (39 razy w 29 dokumentach) i „two” (24 razy w 19 dokumentach), często też te wartości występowały jako cyfry: 1 (17 razy w 12 dokumentach) i 2 (8 razy w 7 dokumentach).

W zbiorze WORST liczby pojawiały się trochę rzadziej – 115 razy w 57 dokumentach. Tu również najpopularniejsze były „one” (28 razy w 21 dokumentach) i „two” (17 razy w 16 tekstach), ale tylko w postaci słownej.

Liczby, które najliczniej występują w analizowanych zbiorach, nie są chyba tymi, o które chodziło ekspertowi – zalecane dane liczbowe miałyby zapewne wskazywać np. na liczbę sprzedanych egzemplarzy lub zachęcać do przeczytania 35 ciekawych opowiadań. Z przeprowadzonej analizy wynika, że liczby są ważne i często można spotkać je w opisach, ale ich obecność nie tłumaczy różnic w skuteczności tych testów.

Ostatnim etapem analizy opisów była próba znalezienia kluczowych słów i wiodących tematów poruszanych w tych tekstach. Trudno to jednak zrobić na podstawie tabel zawierających po kilka tysięcy wyrażen, z których żadne nie jest na pierwszy rzut oka wyjątkowo ważne i wyróżniające. W opisach ze zbioru BEST w czołówce najczęściej występujących słów można znaleźć m.in.: „be”, „have”, „not”, „life”, „book”, „find” czy „man”. Terminy, które powtarzają się w największej liczbie dokumentów, to m.in.: „never”, „come”, „now”, „love” czy „time”. W opisach ze zbioru WORST wyniki są podobne. Najwięcej razy w analizowanych tekstach pojawiają się m.in. słowa: „be”, „have”, „life”, „not”, „story” i „world”. Sortowanie według liczby dokumentów, w których pojawia się przynajmniej raz dane słowo, daje praktycznie takie same wyniki - w czołówce są wszystkie wymienione wcześniej wyrażenia. Terminy te są dość powszechne i nie da się na ich podstawie wyciągnąć jakichś trafnych wniosków na temat zawartości opisów lub wskazać różnic między tekstami z obu zbiorów.

Aby ułatwić, a właściwie umożliwić, ocenę treści i tematyki opisów, przeprowadziłam redukcję wymiarów macierzy i klasteryzację. Analizowane zbiory nie są liczne i zawierają raczej krótkie teksty, dlatego zdecydowałam się na redukcję wymiarów metodą *Roll Up terms*. W zakładce *Transform* zmieniłam następujące ustawienia: *Roll up Terms* - “yes”,

No. of Rolled-up Terms - 100 i *Drop Other Terms* – “yes”. W zakładce *Cluster* zazaczyłam “yes” przy *Autmatically Cluster* i *Ignore Outliers*, a przy *What to Cluster* wybrałam opcję *Roll Up Term*. Pozostałe ustawienia zostawiłam domyślne. Wyniki takiej redukcji wymiarów, a tym bardziej klasteryzacji, nie były jednak zadowalające. Metodą prób i błędów próbowałam stworzyć w miarę jasny i czytelny podział dokumentów na grupy, ale nie udało mi się tego osiągnąć. Ustaliłam, że w miarę dobre efekty osiąga się przy zwijaniu do 120 wyrażzeń (*No. of Rolled-up Terms* - 120) i przy 8 słowach opisujących klastery (*Desriptive Terms* – 8), więc takie ustawienia przyjąłam do dalszej analizy. Otrzymałam dwie tabele zawierające po 120 terminów, które niosą najwięcej informacji o dokumentach z obu zbiorów, co ilustruje Rysunek 10.

Rys. 10. Terminy niosące najwięcej informacji – odpowiednio o zbiorze BEST i WORST

Term	Role	Attribute	Freq	# Documents	Keep	Weight
+ do	Verb	Alpha	85	43Y		0.24629
+ know	Verb	Alpha	53	34Y		0.2751
+ make	Verb	Alpha	39	29Y		0.29672
never	Adv	Alpha	35	25Y		0.32193
no	Adj	Alpha	38	24Y		0.33635
now	Adv	Alpha	31	24Y		0.32741
+ story	Noun	Alpha	34	24Y		0.33309
love	Noun	Alpha	38	23Y		0.35062
more	Noun	Alpha	28	23Y		0.33567
+ time	Noun	Alpha	31	23Y		0.34679
+ want	Verb	Alpha	31	23Y		0.3371
ever	Adv	Alpha	24	21Y		0.35365
+ book	Noun	Alpha	70	21Y		0.47783
+ thing	Noun	Alpha	30	21Y		0.36826
+ world	Noun	Alpha	31	21Y		0.37397
best	Adj	Alpha	25	19Y		0.3837
out	Verb	Alpha	20	19Y		0.36591
series	Noun	Alpha	24	19Y		0.37869
up	Adv	Alpha	25	19Y		0.37916
+ night	Noun	Alpha	21	19Y		0.36892
only	Adv	Alpha	21	18Y		0.38323
+ day	Noun	Alpha	27	18Y		0.40551
+ discover	Verb	Alpha	24	17Y		0.40372
+ heart	Noun	Alpha	19	17Y		0.39362
+ read	Verb	Alpha	31	17Y		0.43942
only	Adj	Alpha	17	16Y		0.40377
own	Adj	Alpha	22	16Y		0.41731
real	Adj	Alpha	18	16Y		0.41339
will	Noun	Alpha	19	16Y		0.40943
+ body	Noun	Alpha	18	16Y		0.40709
+ friend	Noun	Alpha	22	16Y		0.41731
+ keep	Verb	Alpha	20	16Y		0.42598
+ leave	Verb	Alpha	22	16Y		0.41731
+ meet	Verb	Alpha	16	16Y		0.39924
out	Adv	Alpha	16	15Y		0.43008
people	Noun	Alpha	20	15Y		0.43165

Term	Role	Attribute	Freq	# Documents	Keep	Weight
+ life	Noun	Alpha	51	35Y		0.26167
+ story	Noun	Alpha	37	29Y		0.28406
+ do	Verb	Alpha	34	27Y		0.30843
+ find	Verb	Alpha	29	25Y		0.31032
+ world	Noun	Alpha	30	22Y		0.34929
when	Adv	Alpha	23	21Y		0.34531
+ new	Adj	Alpha	28	19Y		0.38799
+ year	Noun	Alpha	27	18Y		0.40002
+ book	Noun	Alpha	21	17Y		0.40164
+ time	Noun	Alpha	18	16Y		0.40581
+ come	Verb	Alpha	18	15Y		0.42254
+ young	Adj	Alpha	17	15Y		0.42019
now	Adv	Alpha	16	14Y		0.43557
+ day	Noun	Alpha	18	14Y		0.44557
+ man	Noun	Alpha	17	14Y		0.45561
+ know	Verb	Alpha	16	13Y		0.45438
+ live	Verb	Alpha	14	13Y		0.44844
+ make	Verb	Alpha	14	13Y		0.44844
all	Noun	Alpha	16	12Y		0.4803
first	Adj	Alpha	13	12Y		0.46618
only	Adv	Alpha	14	12Y		0.47806
short	Adj	Alpha	14	12Y		0.46994
up	Adv	Alpha	13	12Y		0.46618
+ begin	Verb	Alpha	15	12Y		0.47974
+ friend	Noun	Alpha	15	12Y		0.47974
+ see	Verb	Alpha	14	12Y		0.46994
+ thing	Noun	Alpha	12	12Y		0.46041
+ woman	Noun	Alpha	14	12Y		0.46994
+ word	Noun	Alpha	12	12Y		0.46041
no	Adj	Alpha	16	11Y		0.51083
+ short stor...	NOUN_GR...	Alpha	11	11Y		0.4793
best	Adj	Alpha	11	10Y		0.50667
love	Noun	Alpha	16	10Y		0.51793
on	Adv	Alpha	10	10Y		0.5
+ lead	Verb	Alpha	10	10Y		0.5
girl	Noun	Alpha	11	9Y		0.53404

Źródło: opracowanie własne.

Przeglądając zawartość 120 terminów dla obu zbiorów, można spostrzec, że wiele terminów się powtarza, choć występują one z inną częstotliwością. W obu zbiorach można znaleźć wiele popularnych czasowników (np. „meet”, „look”), przymiotników (np. „best”, „young”), a także rzeczowników (m.in. „love”, „friend”, „child”, „death”, „mystery” czy „heart” - wszystkie w zbiorze BEST pojawiają się nieco częściej). Tylko w zbiorze BEST występują m.in. słowa: „die”, „killer”, „lover”, „romance”, „secret” czy „sexy”, co może świadczyć o tym, że dużą popularnością cieszą się książki o miłości i zbrodniach. W tabeli

reprezentującej teksty ze zbioru WORST znalazły się natomiast terminy „dark”, „dream”, „experience”, „fantasy”, „alien”, „strange”, „journey” czy „school”.

Podział na klastry wg najważniejszych terminów ze zbioru BEST wskazuje na istnienie dwóch dużych grup, ale dość trudno je jednoznacznie opisać - w jednej tematy zdają się oscylować wokół życia rodzinnego i przyjaciół, w drugiej mogłyby się natomiast znaleźć opisy thrillerów i książek o miłości. Do klastra 1. zostały przyporządkowane 3 teksty – opisujące go słowa sugerują, że być może są to opisy kryminałów, jak widać na Rys. 11.

Rys. 11 Zbiór BEST – wyniki klastrowania

Clusters				
#	DESCRIPTIVE TERMS	FREQ	PERCENTAGE	RMS STD.
1	+ enjoy, + page, + contain, + character, + keep, + mystery, + read, + die	3	0.0297029702...	0.2592724...
2	+ bring, love, + thriller, will, + book, + leave, out, + girl	23	0.2277227722...	0.1793966...
3	husband, never, + body, family, + thing, + other, + discover, + friend	35	0.3465346534...	0.1575977...
4	+ learn, + start, + hot, + look, + want, + know, + do	1	0.0099009900...	0.0

Źródło: opracowanie własne.

W wyniku klasteryzacji terminów ze zbioru WORST otrzymałam 6 grup, w tym 2 dość liczne, ale tak jak poprzednio zauważyłam problem z opisaniem ich. Klaster 3. zdaje się skupiać opisy książek o przyjaciółkach i spotkaniach. Dość oczywista jest też tematyka 4 opisów, które znalazły się w klastrze 4., opisywanym m.in. terminami „alien”, „strange” czy „unknown”. Trudno natomiast określić zawartość klastra 5., opisywanego m.in. słowami „world”, „eye” i „young”, do którego trafiło aż 48 tekstów. Omawiane klastry przedstawiono na Rysunku 12.

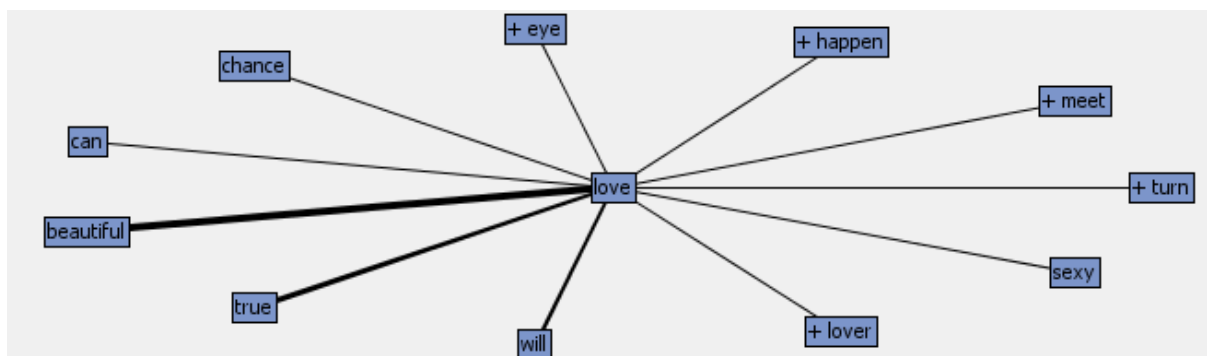
Rys. 12. Zbiór WORST – wyniki klastrowania

Clusters				
#	DESCRIPTIVE TERMS	FREQ	PERCENTAGE	RMS STD. ▲
6	+ land, + begin	1	0.01	0.0
3	will, + side, again, + collection, + friend, + meet, + story, never	22	0.22	0.1492404...
5	+ make, + live, + world, + young, now, + eye, + come, + do	48	0.48	0.1607229...
2	+ look, + know, + tale, + start, ever, up, + child, + thing	3	0.03	0.1919131...
4	+ novel, + time, alien, + determine, + explore, strange, unknown, + power	4	0.04	0.2131565...
1	+ heart, + year, old, high, journey, can, love, school	2	0.02	0.2179896...

Źródło: opracowanie własne

Duża liczba terminów powtarzających się w obu zbiorach skłoniła mnie do sprawdzenia za pomocą widoku *Concept View*, z jakimi wyrazami się one łączą. Słowo „love” z zbioru BEST najczęściej (co najmniej 7 razy) występuje w dokumentach razem z „beautiful”, „turn”, „meet” i „will”, jak to ilustruje Rysunek 13.

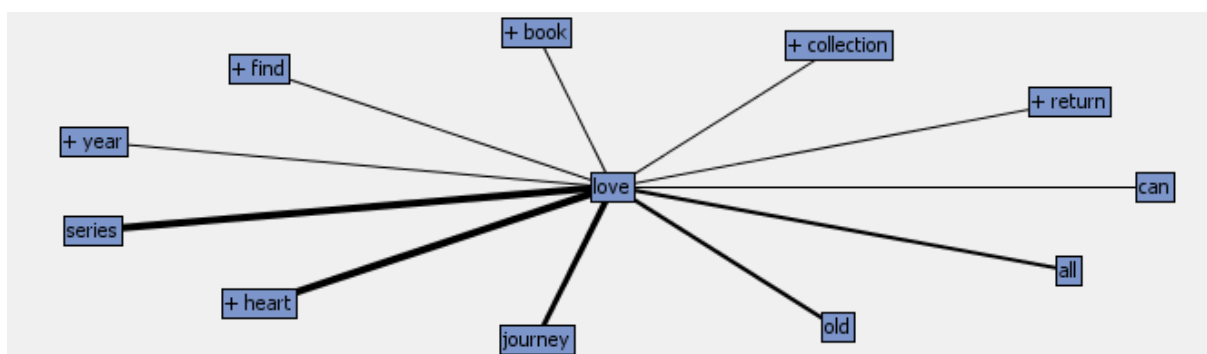
Rys. 13. Zbiór BEST - wyniki *concept linking* dla wyrażenia *love*



Źródło: opracowanie własne

W zbiorze WORST termin ten najczęściej (co najmniej 4 razy) pojawia się w towarzystwie „find”, „year”, „series”, „heart” i „journey”, co można zobaczyć na Rysunku 14.

Rys. 14. Zbiór WORST - wyniki *concept linking* dla wyrażenia *love*



Źródło: opracowanie własne

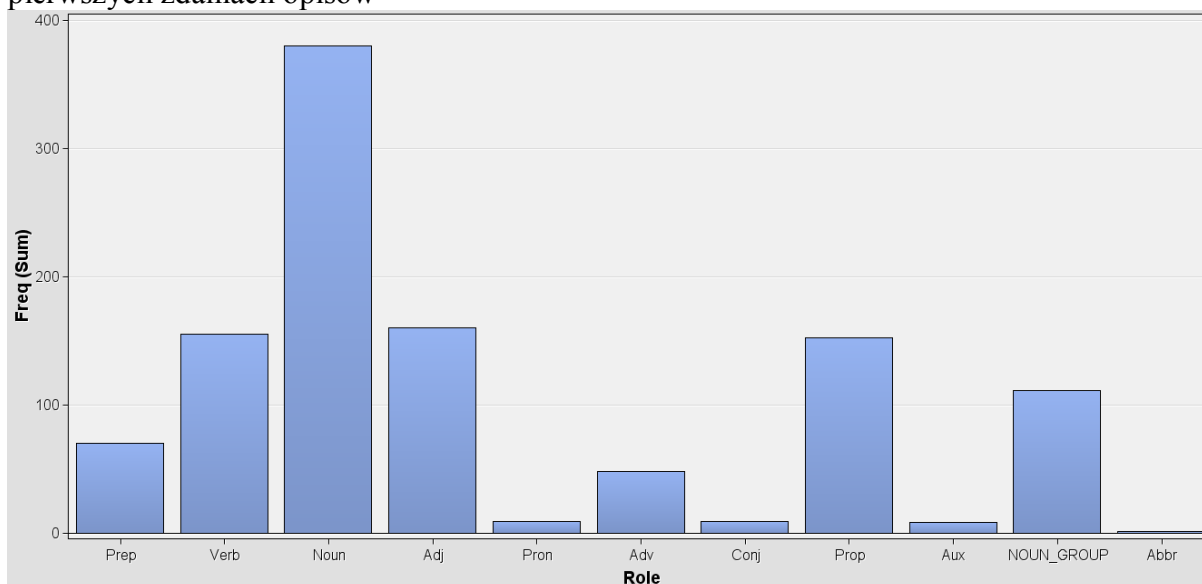
Podobne analizy przeprowadzone dla innych terminów potwierdziły, że te same słowa w obu zbiorach łączą się zupełnie innymi określeniami, jednak trudno znaleźć w tym jakąkolwiek regułę lub zależność.

Redukcja wymiarów tabeli *Term* do 120 najważniejszych określeń i podział na klastry ułatwiły na pewno analizę, ale mimo to nie udało mi się określić jednoznacznie najważniejszych tematów i trendów w każdym ze zbiorów, a tym bardziej wskazać różnic między nimi. Wydaje się oczywiste, że treść opisów książek ma wpływ na ich popularność, ale nie potrafię za pomocą dostępnych narzędzi potwierdzić tego na przykładzie analizowanych tekstów.

Autorzy poradników na temat pisania dobrych opisów książek podkreślają znaczenie nagłówka lub pierwszej linijki tekstu²¹, dlatego przeanalizowałam dodatkowo treść i budowę zdań rozpoczynających opisy ze zbiorów BEST i WORST. Tym razem w oknie *Variables* „yes” zostawiłam tylko przy zmiennej zdanie i już na początku wyrzuciłam z obserwacji mniej ważne dla treści części mowy.

Jak się okazało, wykresy przedstawiające częstotliwość występowania poszczególnych części mowy w pierwszych zdaniach opisów nie są już takie podobne do siebie, co można zaobserwować na zamieszczonych kolejno Rysunkach 15 i 16.

Rys. 15. Zbiór BEST - Wykres przedstawiający części mowy zidentyfikowane w pierwszych zdaniach opisów

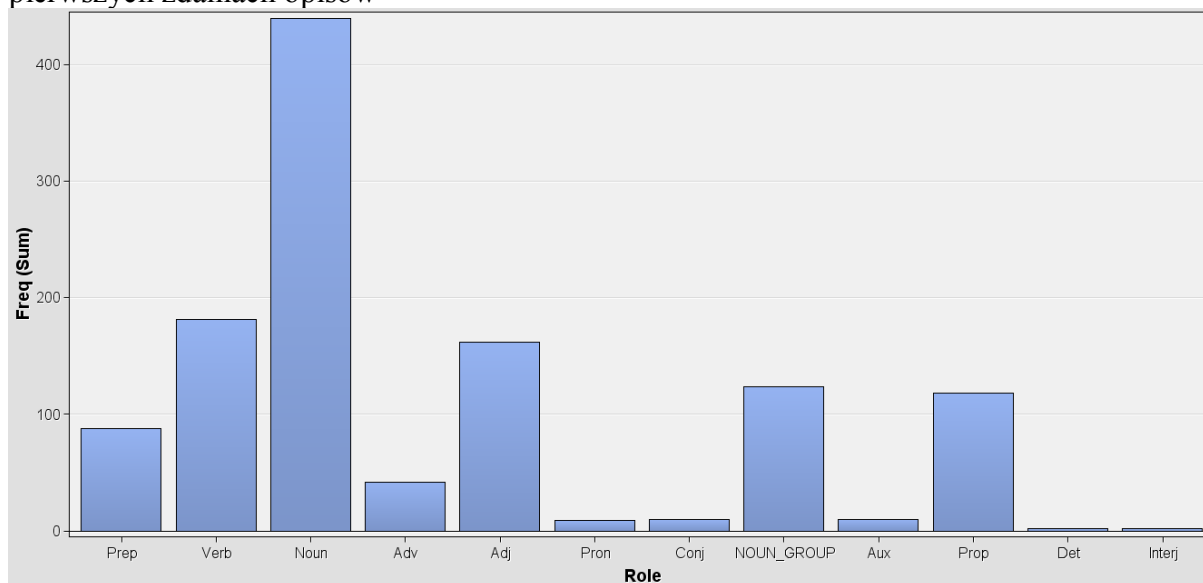


Źródło: opracowania własne.

²¹ Kukral Jim F., op.cit., s. 9; Edwards Mark, op.cit., s. 5

W zbiorze BEST najwięcej jest rzeczowników (380), przymiotników (160), czasowników (155) i nazw własnych (152).

Rys. 16. Zbiór WORST - Wykres przedstawiający części mowy zidentyfikowane w pierwszych zdaniach opisów



Źródło: opracowanie własne.

W zdaniach ze zbioru WORST również przeważają rzeczowniki (440), ale drugą pod względem częstotliwości występowania grupą są czasowniki (181), za nimi plasują się przymiotniki (162). Wyraźnie mniej jest natomiast nazw własnych (118).

W pierwszych zdaniach opisów najpopularniejszych książek można zauważyć nieco większy udział czasowników i przymiotników oraz przysłówków, co może świadczyć o tym, że barwny, niosący więcej treści początek lepiej się sprawdza. Najbardziej widoczną różnicą jest jednak większa liczebność słów zidentyfikowanych jako nazwy własne.

Tak jak w analizie opisów użyłam opcji *Find Entities*, by zidentyfikować w tekstach jednostki specjalne. Jak się okazało, w zbiorze BEST pojawia 30 wyrażeń z rolą PERSON (w 28 dokumentach) i 18 terminów z rolą TITLE (w 17 dokumentach). W zbiorze WORST wyróżniono odpowiednio 21 (w 19 dokumentach) i 23 takie słowa (w 17 dokumentach). Różnice w częstotliwości występowania innych wyrażeń zidentyfikowanych jako *Entities* w obu zbiorach są również niewielkie, zatem trudno powiedzieć, z czego wynika zauważalny na wykresie większy udział nazw własnych w tekstach ze zbioru BEST.

Kolejnym krokiem było sprawdzenie, jak często pojawiają się w pierwszych zdaniach opisu znaki interpunkcyjne. Jak się okazało, w tekstach ze zbioru BEST wykrzyknika użyto 13 razy w 11 dokumentach, a znaku zapytania tylko raz w jednym dokumencie. W zbiorze WORST zarówno wykrzyknik, jak i znak zapytania występują tylko 4 razy w 4 dokumentach. Zastosowanie wykrzyknika w zaledwie co dziesiątym tekście ze zbioru BEST to za mało, by z pewną stanowczością stwierdzić, że taki zabieg zwiększa skuteczność opisu. Wydaje się jednak, że może to być całkiem dobry sposób na zwrócenie uwagi czytelnika.

Analiza treści pierwszych zdań opisu pokazała, że w obu zbiorach teksty te są dość podobne. W tabeli *Term* znalazło się 878 terminów występujących w zdaniach ze zbioru BEST. Słowa o najwyższej częstotliwości, które mogą mieć wartość wyróżniającą, to „love” (10 razy w 10 dokumentach), „new” (8 razy w 8 dokumentach) i „know” (7 razy w 6 dokumentach). Połączyłam w jedną grupę określenia gatunków literackich (m.in. „novel”, „story”, romance), aby sprawdzić, czy często taka informacja o książce pojawia się już na początku opisu. Jak się okazało, słowa z tej grupy wystąpiły łącznie 19 razy w 17 dokumentach.

Tabela *Term* z terminami ze zdań ze zbioru WORST zawiera 881 wyrazów. Słowa o najwyższej częstotliwości, które mogą wносить jakąś wartość znaczeniową, to „life” (11 razy w 10 dokumentach), „book” i „young” (po 7 razy w 7 dokumentach). Na gatunek literacki wskazano 19 razy w 15 dokumentach.

Trudno więc znaleźć istotne różnice między pierwszymi zdaniami w dwóch grupach analizowanych opisów, a tym samym wskazać, co można zrobić, by przyciągnąć uwagę czytelników i zachęcić do przeczytania całego tekstu.

Zakończenie

Celem mojej pracy było porównanie opisów najbardziej i najmniej popularnych książek oraz wskazanie najlepszych zasad tworzenia tego typu tekstów. Nie udało się stworzyć przepisu na blurb, który sprzeda każdą książkę, ale wyniki moich analiz wyraźnie pokazują, że nie warto ograniczać się streszczenia w kilku zdaniach fabuły. Dobrze sprawdzają się wszelkie próby wyróżnienia się i urozmaicenia struktury tekstu, zarówno za pomocą interpunkcji, jak też poprzez zwracanie się bezpośrednio do czytelnika, zadawanie mu pytań czy snucie opowieści w pierwszej osobie. Na pewno dzięki takim zabiegom łatwiej jest zwrócić uwagę kupującego i skłonić do zapoznania się z innymi informacjami o książce (np. opiniami czytelników czy fragmentami), a może nawet do zakupu.

Porównanie treści opisów ze zbioru BEST i WORST nie dało jasnych wyników, które słowa mają szczególną moc przekonywania czytelników do zakupu, a których lepiej się wystrzegać. Warto jednak zauważyć, że w zbiorze skutecznych sprzedażowo opisów jednym z najczęściej pojawiających się słów było „love” (zarówno jako rzeczownik, jak i czasownik), co może oznaczać, że nie zaszkodzi zaznaczyć w opisie, że w książce pojawia się wątek miłosny, nawet jeśli nie jest to jej główny temat.

Rozczarowujące wydają się wyniki analizy pierwszych zdań. Nie udało mi się wskazać żadnych istotnych różnic między początkiem opisów najpopularniejszych i najmniej popularnych książek. Stało się tak być może dlatego, że w większości przypadków były to typowe pierwsze zdania opisu, a nie zręcznie sformułowane nagłówki, które z założenia mają zwracać uwagę i przekonywać do zakupu.

Brak znaczących różnic między opisami może wynikać ze specyfiki badanych zbiorów. Część autorów mogła uznać, że możliwość pobrania książki bezpłatnie powinna być wystarczającą zachętą dla czytelników, dlatego opisy pełniły bardziej rolę informacyjną niż marketingową.

Interpretację niewielkich różnic między opisami utrudnia brak informacji o liczbie pobranych książek z obu zbiorów. Nie wiadomo, czy najpopularniejsze e-booki znalazły na przykład kilka lub kilkanaście razy więcej nabywców niż te najrzadziej wybierane, czy też raczej o ich miejscu w czołówce zdecydowali pojedynczy klienci.

Trzeba również zauważyć, że kategoria *Literature & Fiction* jest dość pojemna - obejmuje kilka gatunków literackich i dość szeroką tematykę. Być może trendy byłyby bardziej wyraźne, a wspólne cechy złych i dobrych opisów lepiej widoczne, gdyby analiza obejmowała bardziej jednorodny zbiór książek (np. tylko romanse albo tylko thrillery).

Bibliografia

- Booksellers.org.uk, Raport Books & Purchasing Decisions,
www.booksellers.org.uk/BookSellers/media/SiteMediaLibrary/IndustryNews/Books-Purchasing-Decisions.pdf [dostęp 24.09.2014]
- Coker Mark, The Secrets to Ebook Publishing Success, 2014
- Edwards Mark, How to Craft a Sizzling Book Description A Guide for Writers, 2013
- Global eBook Survey 2013 by Bookboon,
<http://bookboon.com/blog/bookboon-coms-global-ebook-survey/> [dostęp 24.09.2014]
- Jo Henry, New Trends in Publishing: BIC Seminar 2013 What are book consumers up to?,
Bowker Market Research,
<http://www.bic.org.uk/files/pdfs/What%20are%20book%20consumers%20up%20to%20Jo%20Henry.pdf> [dostęp 24.09.2014]
- Kukral Jim F., How To Sell More Books with Awesome Amazon Descriptions, 2013
- Nielsen, Raport Book & Consumer 2013, maj 2014
- Nielsen, White Paper: The Link Between Metadata and Sales, styczeń 2012
- Online Copyright Infringement Tracker, Annex 1 – Individual content types, Wave 4
(Covering period March – May 2013), Prepared for Ofcom by Kantar Media,
http://stakeholders.ofcom.org.uk/binaries/research/telecoms-research/online-copyright/w4/OCI_MAIN_REPORT_W4_FINAL.pdf [dostęp 24.09.2014]
- Owen Laura Hazard, Jacket Copy Sells Books, So Make It Good, maj 2009,
<http://www.publishingtrends.com/2009/05/jacket-copy-sells-books/> [dostęp 24.09.2014]
- PwC, Raport Global entertainment and media outlook 2014-2018,
<https://www.pwc.com/gx/en/global-entertainment-media-outlook/segment-insights/consumer-and-educational-book-publishing.jhtml> [dostęp 24.09.2014]
- Verso Advertising, 2011 Survey of Book-Buying Behavior, January 2012,
<http://www.versoadvertising.com/Wi7survey2012/> [dostęp 24.09.2014]