

Obowiązki kierownika projektu – analiza *text mining*

Responsibilities of project manager – text mining analysis

Marcin Wyskwarski

Politechnika Śląska, e-mail: marcin.wyskwarski@polsl.pl, ORCID: 0000-0003-2004-330X

Streszczenie

W pracy podjęto próbę identyfikacji zakresu obowiązków kierownika projektu z wykorzystaniem analizy *text mining*. Analizie poddano treść ofert pracy pobranych z serwisu internetowego. Badania miały udzielić odpowiedzi, czy na podstawie wyników przeprowadzonej analizy *text mining* będzie można ustalić zakres obowiązków kierownika projektu. Stosownie do przyjętego celu ustalono strukturę pracy. W punkcie drugim, korzystając z badań literaturowych, omówiono zakres obowiązków kierownika projektu. Kolejny punkt obejmuje najważniejsze informacje dotyczące *text mining*. Zawarto w nim podstawowe definicje, obszary zastosowań, a także uproszony podział etapów analizy *text mining*. Punkt trzeci został podzielony na dwie części. W pierwszej omówiono sposób i miejsce zebrania danych, a także procedurę przeprowadzonej analizy. Ostatnia część punktu trzeciego przedstawia wyniki przeprowadzonej analizy. Ustalono dzięki niej najczęściej występujące w ofertach słowa oraz jaka była ich korelacja z innymi słowami w danej grupie analizowanych ogłoszeń.

Słowa kluczowe: *text mining*, zakres obowiązków, kierownik projektu, chmura słów.

Abstract

This article presents the approach of identification of responsibilities of project manager with the use of text mining analysis. This paper shows it on job advertisements gathered from internet service. The research was made to obtain an answer whether from text mining analysis the responsibilities of project manager can be created. According to that aim the structure of the paper was created. In the second part of the paper the description of project manager was presented according to the literature. Next part contains the description of text mining analysis, definitions, application areas and the simplified diversification of text mining analysis. The third one was divided into two parts. The first describes the way and place of gathering data and the procedure of analysis. The last contains the results of the analysis. It shows the most frequently used words in job advertisements and their correlation with other words in the advertisements.

Keywords: text mining, responsibilities, project manager, world cloud.

Wstęp

W dzisiejszym dynamicznie zmieniającym się i złożonym otoczeniu gospodarczym sprawne realizowanie projektów jest istotnym elementem wpływającym na przetrwanie oraz rozwój organizacji. Projekty to przede wszystkim realizujący je ludzie. Kluczową rolę wśród nich odgrywa kierownik projektu. Coraz większa liczba realizowanych projektów sprawia, że wzrasta zapotrzebowanie na pracowników na stanowisko kierownika projektu. W literaturze z obszaru zarządzania projektami oraz zarządzania ogółem można uzyskać informacje o wykonywanych zadaniach na stanowisku kierownika projektu. Analizując dostępne na rynku oferty pracy, można podjąć próbę ustalenia, jakie zadania i obowiązki zostaną przydzielone w danej organizacji.

Celem niniejszego artykułu była próba identyfikacji zakresu obowiązków na stanowisku kierownika projektu. Informacje te próbowano uzyskać poprzez analizę *text mining* ofert pracy. Miała ona udzielić odpowiedzi, jakie zadania, obowiązki czekają na przyszłego kierownika projektu. Badanie miało również odpowiedzieć na pytanie, czy stosując analizę *text mining*, która nie analizowała znaczenia poszczególnych wyrazów, można zidentyfikować zadania realizowane przez kierownika projektu.

1. Zadania kierownika projektu

Kierownik projektu napotyka wiele problemów, takich jak presja czasu, ograniczenia budżetowe, tymczasowy i nietrwały skład zespołu projektowego czy też ograniczona możliwość oddziaływania na członków zespołu, którzy często są zatrudnieni w innych komórkach i formalnie podlegają innym przełożonym [Wachowiak i in. 2004; Walczak 2014]. Do podstawowych obowiązków kierownika projektu zalicza się planowanie działań, organizowanie pracy, kierowanie zespołem, a także monitorowanie i kontrolowanie przebiegu projektu [Wachowiak i in. 2004; Pawlak 2006]. W swojej pracy odpowiada on za [Pawlak 2006]:

- określenie struktury organizacyjnej realizowanego projektu,
- ustalenie celów projektu i przedstawienie ich do zatwierdzenia,
- nadzór nad realizacją celów,
- strukturalizację projektu,
- planowanie i nadzorowanie terminów oraz kosztów,
- dobór członków grupy projektowej,
- kierowanie grupą projektową,
- zapewnienie przepływu informacji w realizowanym projekcie,
- podejmowanie istotnych dla projektu decyzji.

Wśród najczęściej wykonywanych przez kierownika czynności wymienia się między innymi [Steyn, Nicholas 2012]:

- planowanie zadań, działań i wyników projektu przez opracowanie struktury podziału prac, harmonogramów, budżetu, przydzielanie zasobów do zadań oraz sterowanie ich realizacją,
- dobór oraz organizację zespołu projektowego,

- tworzenie oraz utrzymywanie relacji z interesariuszami projektu,
- przeprowadzanie negocjacji oraz integrację związanych z projektem kierowników funkcyjnych, podwykonawców oraz kierownictwa wyższego szczebla,
- monitorowanie realizacji projektu,
- identyfikację problemów funkcjonalnych i technicznych,
- rozwiązywanie pojawiających się problemów lub szukanie odpowiednich sposobów ich rozwiązania,
- zarządzanie konfliktem oraz radzenie sobie z kryzysem,
- rekomendowanie wstrzymania, zakończenia projektu w przypadku, gdy osiągnięcie celów jest niemożliwe.

Szczególnie ważną i wymagającą funkcją kierownika projektu jest ogół czynności związanych z kierowaniem zespołem projektowym. Istotne jest nie tylko utworzenie zespołu projektowego poprzez dobór właściwych członków, ale także odpowiedni podział zadań. Ważna jest także umiejętność motywowania zespołu, jego integracja, rozwiązywanie powstających konfliktów oraz usprawnianie pracy i komunikacji w zespole [Wachowiak i in. 2004; Pawlak 2006]. Kierowanie zespołem wykracza poza tradycyjnie pojmowane funkcje kierownicze. Kierownik projektu powinien dzielić się swoją wiedzą i doświadczeniem, tworzyć wśród członków zespołu poczucie zaufania i odpowiedzialności za realizowany projekt [Wachowiak i in. 2004; Walczak 2014].

Na zadania wykonywane przez kierownika projektu możemy spojrzeć także z punktu widzenia przebiegu projektu, czyli m.in. takich istotnych zagadnień, jak [Trocki 2013]:

- planowanie – zaplanowanie m.in. struktury projektu (zadania projektu i niezbędne do ich realizacji czynności), terminów projektu (ukazujących przebieg projektu w czasie), zasobów projektu (przydział zasobów do poszczególnych zadań i czynności oraz ich rozkład w czasie), kosztów projektu (budżetowanie), planu ryzyka, jakości, komunikacji i zaopatrzenia;
- organizowanie zespołu projektowego – przeprowadzenie analizy niezbędnych zasobów ludzkich (szczególnie ich kompetencji) oraz ich pozyskanie, motywowanie, rozwiązywanie powstających konfliktów, opracowanie i wdrożenie odpowiedniej strategii komunikacyjnej;
- nadzorowanie i kontrola – sprawdzanie przebiegu projektu z harmonogramem oraz zaplanowanym budżetem, delegowanie zadań członkom zespołu projektowego, kontrola jakości wykonywanych zadań;
- zamykanie projektu – przekazanie projektu zamawiającemu, rozliczenie projektu, przygotowanie dokumentacji eksploatacyjnej, opracowanie raportu z realizacji projektu, podjęcie decyzji o zakończeniu projektu, rozwiązanie zespołu projektowego.

2. Istota text miningu

Text mining jest stosunkowo młodą oraz interdyscyplinarną dziedziną, która wywodzi się m.in. z *data mining*, wyszukiwania informacji, kategoryzacji tekstu i modelowania probabilistycznego [Kao, Poteet 2007]. Na popularność *text miningu* wpływa m.in. ciągły, dynamiczny rozwój Internetu i towarzy-

szący mu wzrost znaczenia opiniotwórczej roli portali społecznościowych [Berry, Kogan 2010] oraz przechowywanie informacji w dokumentach tekstowych – Tan szacuje, że 80% informacji przedsiębiorstw jest przechowywana w dokumentach tekstowych [Tan 1999].

Według A. Gładysz *text mining* to metody, koncepcje oraz algorytmy przetwarzania zasobów tekstowych sporządzonych w językach naturalnych. Są implementowane w postaci programów komputerowych, co umożliwia zautomatyzowanie procesów przetwarzania dokumentów [Gładysz 2012]. Hearst podaje, iż *text mining* to „proces mający na celu wydobywanie z zasobów tekstowych nieznanych wcześniej informacji” [Hearst 1999]. W dużej części rozwiązań *text mining* nie analizuje się znaczenia wyrazów oraz zdań, lecz próbuje się zidentyfikować reguły i prawidłowości dotyczące występowania określonych ciągów znaków.

Najpopularniejsze obszary zastosowań *text miningu* to pozyskiwanie informacji z dokumentów, identyfikacja wiadomości zawierających określone treści, generowanie streszczeń, klasyfikacja wzorcowa, klasyfikacja bezwzorcowa (grupowanie, klasteryzacja), identyfikacja powiązań, wizualizacja oraz generowanie odpowiedzi na pytanie [Lula 2005].

Proces analizy tekstu można w dużym uproszczeniu podzielić na trzy etapy [Gładysz 2012]:

- wstępne przetwarzanie tekstu,
- budowa macierzy częstości występowania słów,
- zastosowanie klasycznych metod wywodzących się z *data miningu*.

Wstępne przetwarzanie tekstu polega na przekształceniu analizowanego dokumentu tekstowego w listę wyrazów określanych mianem „worka słów” (*bag of words*). W etapie tym pomija się informacje o kolejności słów i związkach pomiędzy nimi. Przyjmuje się, że wystąpienia słów są niezależne od siebie. Jeżeli analizowanych jest więcej dokumentów, to dla każdego tworzy się odrębny „worek słów”. Podczas tworzenia tzw. worka słów usunięte zostają wszelkie cyfry, znaki interpunkcyjne (np. kropki, przecinki, średniki, myślniki), a także słowa niewnoszące dodatkowych informacji tzw. *stop-words* (usunięte zostają m.in. spójniki, przyimki, słowa uznane w danej analizie za nieprzydatne itp.). Usunięciu mogą ulec także słowa występujące bardzo rzadko (*least frequent*) bądź bardzo często (*most frequent*) – tzw. przycinanie (*pruning*) [Gładysz 2012; Mirończuk 2012].

Istotnym procesem realizowanym w ramach wstępnego przetwarzania tekstu jest przekształcenie słów tworzących „worek słów” do postaci uznawanej za podstawową. Można w tym celu wykorzystać takie operacje, jak [Borycki, Sołdacki 2002; Mykowiecka 2007]:

- *stemming*, tzw. ekstrakcja rdzeni słów – polega na wybraniu z danego słowa części niezmiennej dla wszystkich form gramatycznych, czyli tzw. rdzenia (usunięte zostają wszelkiego rodzaju przedrostki i przyrostki);
- lematyzacja – to analiza morfologiczna zapewniająca znalezienie podstawowej formy danego wyrazu, czyli identyfikację leksemu (np. czasownik zostaje przekształcony do

bezokolicznika, a rzeczownik do mianownika liczby pojedynczej). Do przeprowadzenia lematyzacji konieczne jest utworzenie słownika lub zestawu reguł fleksyjnych dla danego języka.

Do budowy macierzy częstości występowania słów stosuje się tzw. model przestrzeni wektorowej (*Vector Space Model*). W modelu tym dokumenty oraz znajdujące się w nich słowa są prezentowane w postaci macierzy. Każdy z analizowanych dokumentów jest reprezentowany przez osobny wektor, który przedstawia liczbę wystąpień poszczególnych słów [Gładysz 2012; Mirończuk 2012]. Macierzowa postać reprezentacji dokumentów oraz związanych z nimi wyrażen (wyrazów) nazywana jest macierzą dokumentów-wyrażen (*document – term matrix*). Wiersze tej macierzy reprezentują analizowane dokumenty, a kolumny znajdujące się w nich słowa. W zależności od zastosowanego sposobu kodowania informacji w macierzy uzyskuje się różne odmiany reprezentacji przestrzenno-wektorowej tekstu. Wśród często wykorzystywanych wymienia się reprezentację boolowską (binarną), częstościową występowania wyrażen (*Term Frequency – TF*), odwrotnej częstości dokumentu (*Inverse-Documentfrequency – IDF*), mieszaną TF-IDF, logarytmiczną, ważoną logarytmiczną, okapi BM25 [Mirończuk 2012]. Dokumenty prezentowane w macierzy częstości występowania określa się w lingwistyce informatycznej mianem korpusu dokumentów, czyli kolekcji dokumentów, które będą analizowane.

Po przeprowadzeniu wstępnego przetwarzania tekstu oraz utworzeniu macierzy częstości występowania słów można wykonać kolejne działania związane z analizowanym tekstem, np. grupowanie dokumentów tekstowych (tzw. klasteryzacja), obliczenie częstości występowania słów, obliczenie korelacji między słowami, wizualizację przeprowadzonych obliczeń itd.

3. Identyfikacja obowiązków kierownika projektu

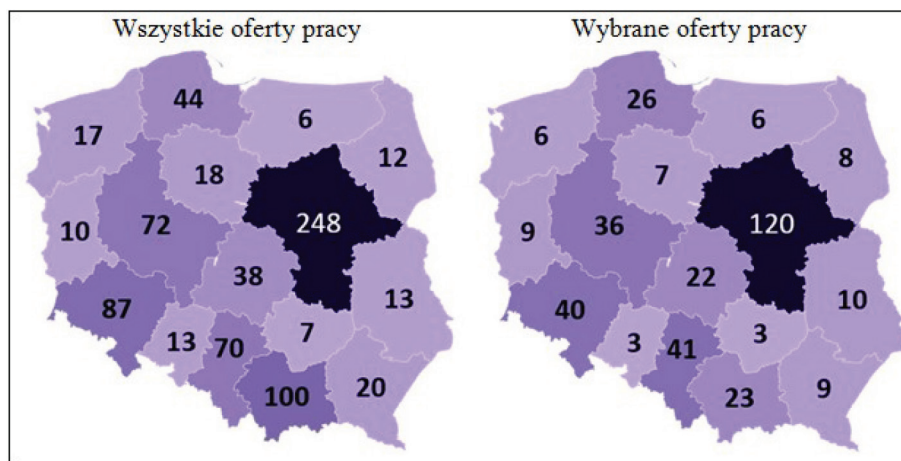
3.1. Źródło danych oraz przebieg analizy *text mining*

Dane wykorzystane w analizie to oferty pracy pobrane z serwisu www.pracuj.pl. Zebrano je w dniach od 9 do 11 kwietnia 2018 r. W celu wyszukania ofert (ogłoszeń) w serwisie www.pracuj.pl wpisano wyłącznie frazę „kierownik projektu”. Po wstępnej wyrywkowej analizie treści przedstawionych ofert okazało się, że wśród 775 wyszukanych ogłoszeń znalazły się również takie, w których użyto innej nazwy niż „kierownik projektu”, oraz takie, których treść była napisana w języku angielskim.

Do dalszej analizy zostały wybrane tylko ogłoszenia, które spełniały poniższe dwa warunki:

- ogłoszenie dotyczyło stanowiska „kierownik projektu” (stanowisko to mogło być również nazwane inaczej, np. za pomocą synonimów, lub z użyciem słów napisanych w języku angielskim),
- ogłoszenie było napisane w języku polskim.

W przypadku gdy znajdująca się w ogłoszeniu nazwa stanowiska budziła wątpliwość, czy oferta na pewno dotyczy stanowiska



Rysunek 1. Liczba ofert pracy w poszczególnych województwach

Źródło: opracowanie własne.

ska „kierownika projektu” (np. „koordynator projektu”, *project coordinator*, *program manager*), analizowano treść ogłoszenia (szczególnie zakres obowiązków) i podejmowano decyzję o odrzuceniu lub akceptacji ogłoszenia. Z 775 wyszukanych początkowo ofert do dalszej analizy wybrano 369. Ich liczbę w poszczególnych województwach prezentuje rys. 1.

W analizowanych 369 ofertach zostały użyte następujące określenia dla stanowiska „kierownik projektu”: *project manager* (142 ogłoszenia), kierownik projektu (128 ogłoszeń), koordynator projektu (54 ogłoszenia), lider projektu (11 ogłoszeń), menedżer projektu (11 ogłoszeń), młodszy kierownik projektu (11 ogłoszeń), *junior project manager* (5 ogłoszeń), *senior project manager* (3 ogłoszenia), *program manager* (2 ogłoszenia), *project coordinator* (1 ogłoszenie) oraz szef projektu (1 ogłoszenie).

Dla każdego wybranego ogłoszenia utworzono osobny plik z rozszerzeniem txt, w którym zapisano fragment opisujący zakres obowiązków. Ta część ogłoszenia była często określana za pomocą następujących słów: „główne obowiązki”, „kluczowe zadania”, „podstawowe obowiązki”, „główne zadania”, „zakres obowiązków”, „obowiązki”, „opis stanowiska”, „twoje zadania”, „zadania na stanowisku”, „będziesz odpowiedzialny za”, „wyzwania, na jakie możesz liczyć”, „w naszej firmie będziesz mógł/będziesz odpowiedzialny”, „co będziesz robił w naszym zespole”. Pliki zostały umieszczone w 16 folderach, po jednym folderze dla każdego województwa.

Analizę *text mining* przeprowadzono w następujących etapach¹:

- wstępne przetwarzanie tekstu,
- utworzenie korpusów analizowanych dokumentów, budowa macierzy częstości występowania słów,
- wykorzystanie klasycznych metod pochodzących z *data mining*.

W ramach wstępnego przetwarzania tekstu każdy utworzony plik tekstowy został zamieniony w tzw. worek słów. Czynności te wykonano z wykorzystaniem aplikacji Notepad++ v.7.3.3 oraz RStudio v.1.0.136. W etapie tym wykonano następujące działania:

- usunięte zostały wszystkie znaki z wyjątkiem liter,
- duże litery zostały zamienione na małe,
- usunięte zostały słowa uznawane za nieprzydatne (np. spójniki, przymyki itp.) – w tym celu wykorzystano samodzielnie utworzoną tzw. *stop listę* (*stop-words*),
- przekształcono słowa do ich wersji uznawanej za podstawową,
- usunięto słowa niebędące rzeczownikami oraz przymiotnikami,
- każdy wyraz został umieszczony w osobnej linii.

Dostępny w RStudio pakiet do analizy *text mining* nie wspiera języka polskiego. Dlatego proces przekształcenia słów do postaci podstawowej, a także wybór przymiotników i rzeczowników przeprowadzono, korzystając ze słownika morfologicznego „polimorfologik 2.1”. Słownik ten ma formę pliku tekstowego w kodowaniu UTF. W chwili korzystania z niego zawierał on 4 811 854 linii tekstu i był dostępny na portalu Github². Po zaimportowaniu słownika do programu RStudio przyjął on postać tabeli składającej się z trzech kolumn: forma podstawowa, forma odmieniona, znaczniki gramatyczne. Sprowadzenie słowa do postaci podstawowej wiązało się z jego wyszukaniem w kolumnie „forma odmieniona” i wstawieniem w jego miejsce słowa z kolumny „forma podstawowa”. Jeżeli danego słowa nie udało się odnaleźć w słowniku, to pozostawało ono w dokumencie bez zmian. W takiej sytuacji nie uwzględniano również formy gramatycznej takiego słowa.

W kolejnym etapie analizy *text mining* utworzono cztery korpusy dokumentów. Po jednym dla województw, w których minimalna liczba ogłoszeń wynosiła 40 (województwo dolnoślą-

¹ Uproszczonego podziału zaproponowany przez A. Gładysz [2012].

² <https://github.com/morfologik/polimorfologik/releases/tag/2.1>.

Docs	Terms						
	budżet	dokumentacja	dostawca	działanie	etap	finansowy	
Z_16.txt	0	0	1	0	0	1	
Z_17.txt	2	2	0	1	0	1	
Z_18.txt	1	0	0	0	0	0	
Z_19.txt	0	0	0	0	1	0	

Rysunek 2. Fragment macierzy dokumentów-wyrażeń dla województwa śląskiego

Źródło: opracowanie własne.

Docs	Terms					Docs	Terms				
	harmonogram	klient	projekt	realizacja			harmonogram	klient	projekt	realizacja	
A.txt	0	0	0	0	0	A.txt	0	0	0	0	
B.txt	1	0	0	0	0	B.txt	1	0	3	1	
C.txt	1	1	0	0	0	C.txt	1	1	0	0	
D.txt	1	1	1	0	0	D.txt	4	4	2	0	
E.txt	1	1	1	1	1	E.txt	3	3	1	1	
<code>> findAssocs(dtm, "harmonogram", 0)</code>						<code>> findAssocs(dtm, "harmonogram", 0)</code>					
\$harmonogram						\$harmonogram					
klient projekt realizacja						klient projekt realizacja					
0.61 0.41 0.25						0.97 0.37 0.11					

Rysunek 3. Przykładowe macierze dokumentów-wyrażeń oraz wartość korelacji dla słowa „harmonogram”

Źródło: opracowanie własne.

skie – korpus nr 1, mazowieckie – korpus nr 2, śląskie – korpus nr 3). Czwarty korpus objął ogłoszenia pochodzące ze wszystkich województw. Następnie dla każdego korpusu utworzono macierz dokumentów-wyrażeń (*document – term matrix*) z częstotliwościową reprezentacją występowania wyrażenia (*Term Frequency – TF*). Przy tworzeniu macierzy korzystano z tzw. modelu przestrzeni wektorowej (*Vector Space Model*). Fragment macierzy³ dla województwa śląskiego przedstawia rys. 2.

W ostatnim etapie analizy dla wszystkich korpusów wyszukano najczęściej występujące słowa. Obliczono również korelację ośmiu najczęściej występujących słów z innymi słowami. Korelację obliczono z zastosowaniem funkcji `findAssocs()`, która bazuje na standardowej funkcji `cor()` dostępnej w pakiecie statystycznym języka R. Na rysunku 3 przedstawione zostały dwie przykładowe macierze dokumentów-wyrażeń wraz z obliczoną wartością korelacji dla słowa „harmonogram”. W dokumentach wchodzących w skład macierzy po lewej stronie rys. 3 słowa (harmonogram, klient, projekt, realizacja) wystąpiły co najwyżej raz w każdym z dokumentów (A, B, C, D, E). W dokumentach tworzących macierz po prawej stronie rys. 3 słowa „harmonogram”, „klient”, „projekt” wystąpiły częściej niż jeden raz (np. słowa „harmonogram” oraz „klient” wystąpiły po 3 razy w dokumencie E). Inna liczba słów w przykładowych macierzach sprawiła, że wartość korelacji dla słowa „harmonogram” ma inną wartość. Korelacja o wartości 1 oznacza, że dane dwa słowa zawsze występują razem (w takiej samej ilości) w dokumentach. Wartość 0 oznacza, że słowa nigdy nie wystąpiły razem.

3.2. Wyniki przeprowadzonej analizy

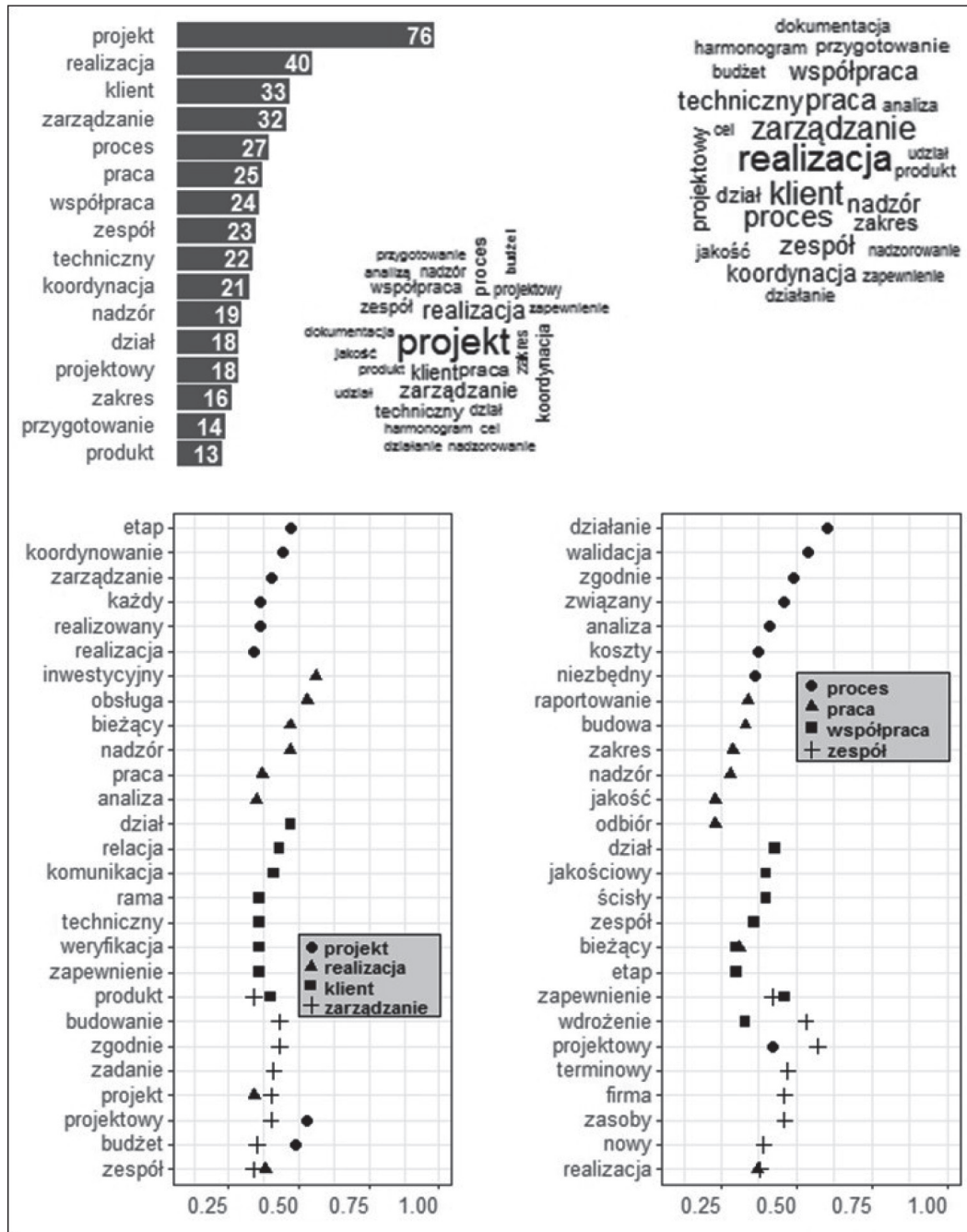
W lewej górnej części rys. 4, 5, 6, i 7 zaprezentowano szesnaście słów, których najczęściej używano w danym korpusie. Posortowano je malejąco według liczby ich wystąpień. Najczęściej używane słowa przedstawiono również w postaci tzw. chmury słów (górną część rys. 4, 5, 6, i 7). Im słowo częściej występowało w korpusie, tym jest większe. Opracowano po dwie chmury słów dla każdego z korpusu. Chmura znajdująca się po lewej stronie zawiera wszystkie najczęściej używane słowa. Z chmury po prawej stronie celowo usunięto słowo „projekt”, co sprawiło, że pozostałe słowa stały się bardziej widoczne. Na przykład w chmurze słów dla województwa dolnośląskiego (rys. 4) na pierwszy plan wysuwa się słowo „realizacja”, które wystąpiło w korpusie 40 razy.

Przedstawione na rys. 4, 5, 6, i 7 chmury mają różne rozmiary. Wynika to z przyjęcia kryteriów liczbowych wyboru słów do chmury, tj. maksymalnej liczby słów w chmurze oraz minimalnej wymaganej liczby wystąpień danego słowa w korpusie. Chmury słów zostały wygenerowane przy tych samych ustawieniach.

We wszystkich czterech korpusach najczęściej występowały odpowiednio słowa „projekt” oraz „realizacja”. Analizując kolejne osiem najczęściej występujących wyrazów, można zauważyć, że takie słowa jak „realizacja”, „zarządzanie”, „klient”, „zespół”, „współpraca”, „praca”, były użyte w każdym z czterech korpusów (rys. 4, 5, 6, 7). Oczywiście kolejność tych słów była różna w poszczególnych korpusach.

Dolny fragment rys. 4, 5, 6, i 7 prezentuje korelację ośmiu najczęściej występujących słów z innymi słowami występującymi w analizowanym korpusie. Lewa strona prezentuje korelację dla pierwszych czterech najczęściej używanych słów, a prawa

³ Macierz ma rozmiary 41 x 62 (41 – dokumentów, 62 – wyrażenia (słów)).



Rysunek 4. Analiza korpusu dla województwa dolnośląskiego

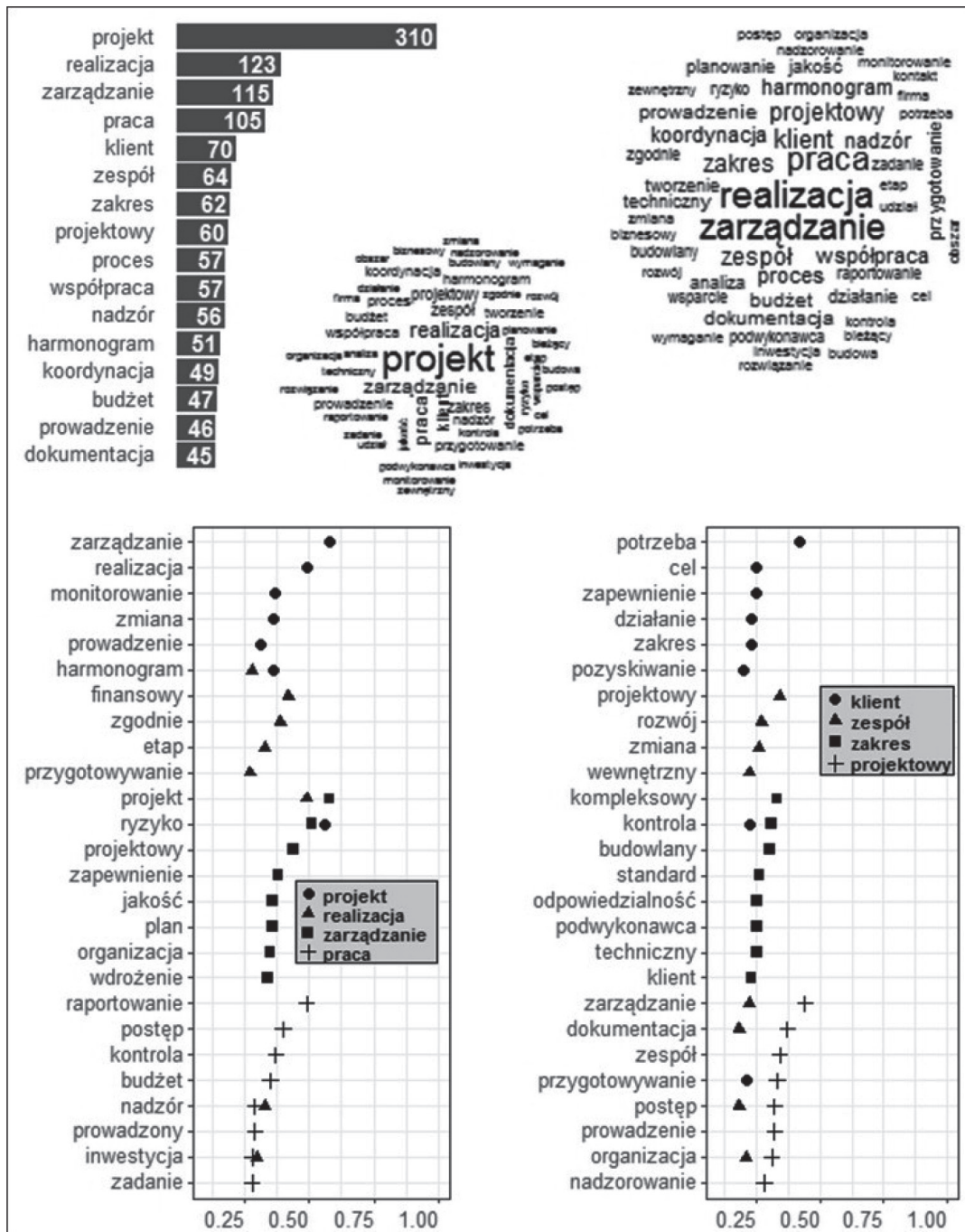
Źródło: opracowanie własne.

dla kolejnych czterech. Na przykład słowo „projekt” (76 użyc), czyli najczęściej używane słowo w ofertach pracy m.in. dla województwa dolnośląskiego (rys. 4), miało najwyższą korelację z następującymi słowami „etap”, „koordynowanie”, „zarządzanie”, „każdy”, „realizowany” itd. Wartość korelacji dla słowa „projekt” przedstawia znak „●”, słowa „realizacja” znak „▲”, słowa „klient” znak „■”, a słowa „zarządzanie” znak „+”. Dla lepszej przejrzystości dane posortowano w porządku malejącym

według liczby wystąpień danego słowa w korpusie, a następnie według wartości korelacji tego słowa z innymi słowami.

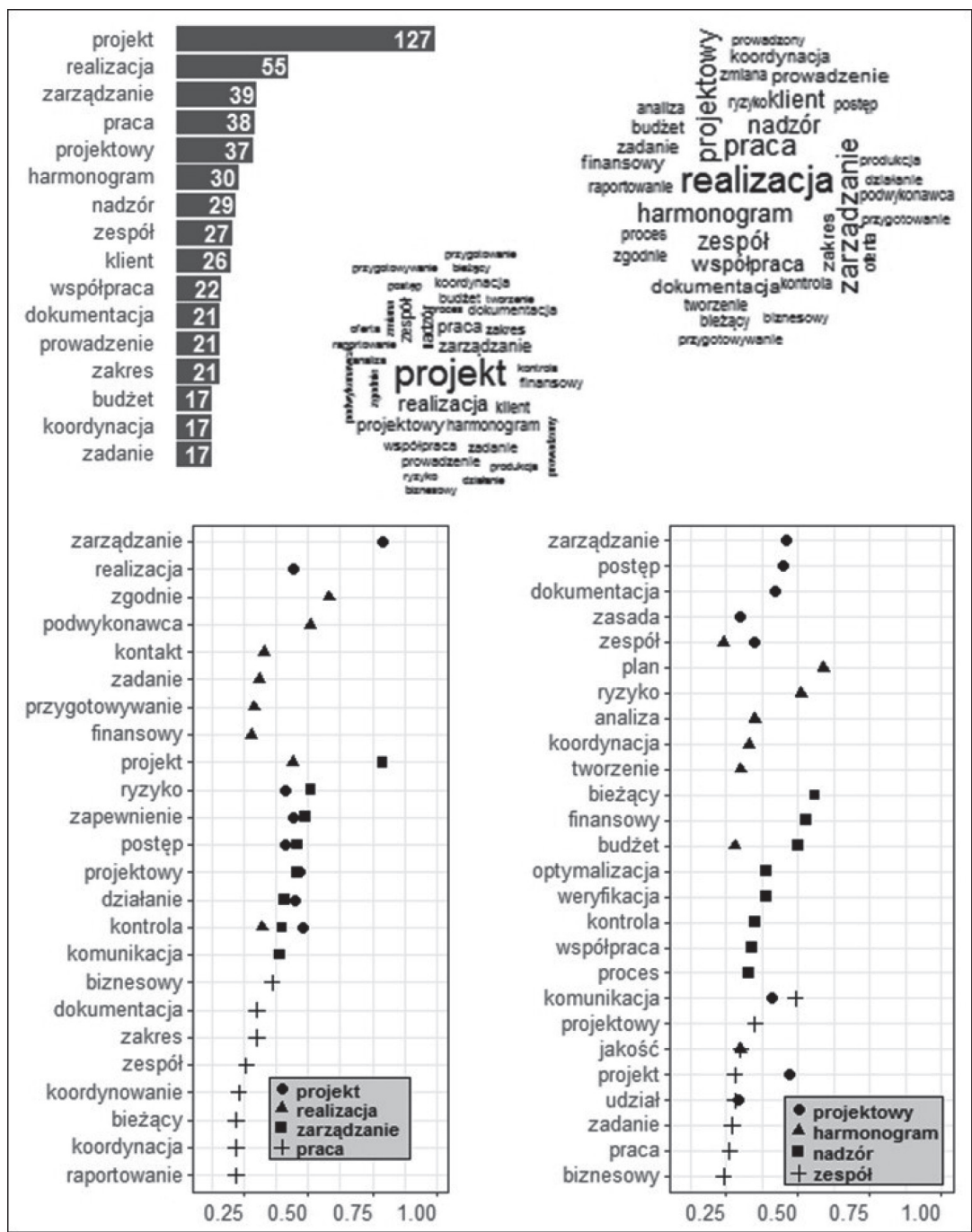
Dodatkowych spostrzeżeń na temat analizowanych ogłoszeń może dostarczyć analiza wykresów przedstawiających korelację pomiędzy słowami. Na przykład w korpusie nr 1 słowo „klient” miało największą korelację ze słowami „dział”, „relacja”, „kommunikacja”, „rama”⁴, „techniczny”, „weryfikacja”, „zapewnienie”, „produkt” (rys. 4), w korpusie nr 2 ze słowa-

⁴ Słowo to powstało z przekształcenia słowa „ramach” pochodzącego np. z frazy „(...) w ramach swoich obowiązków...”.



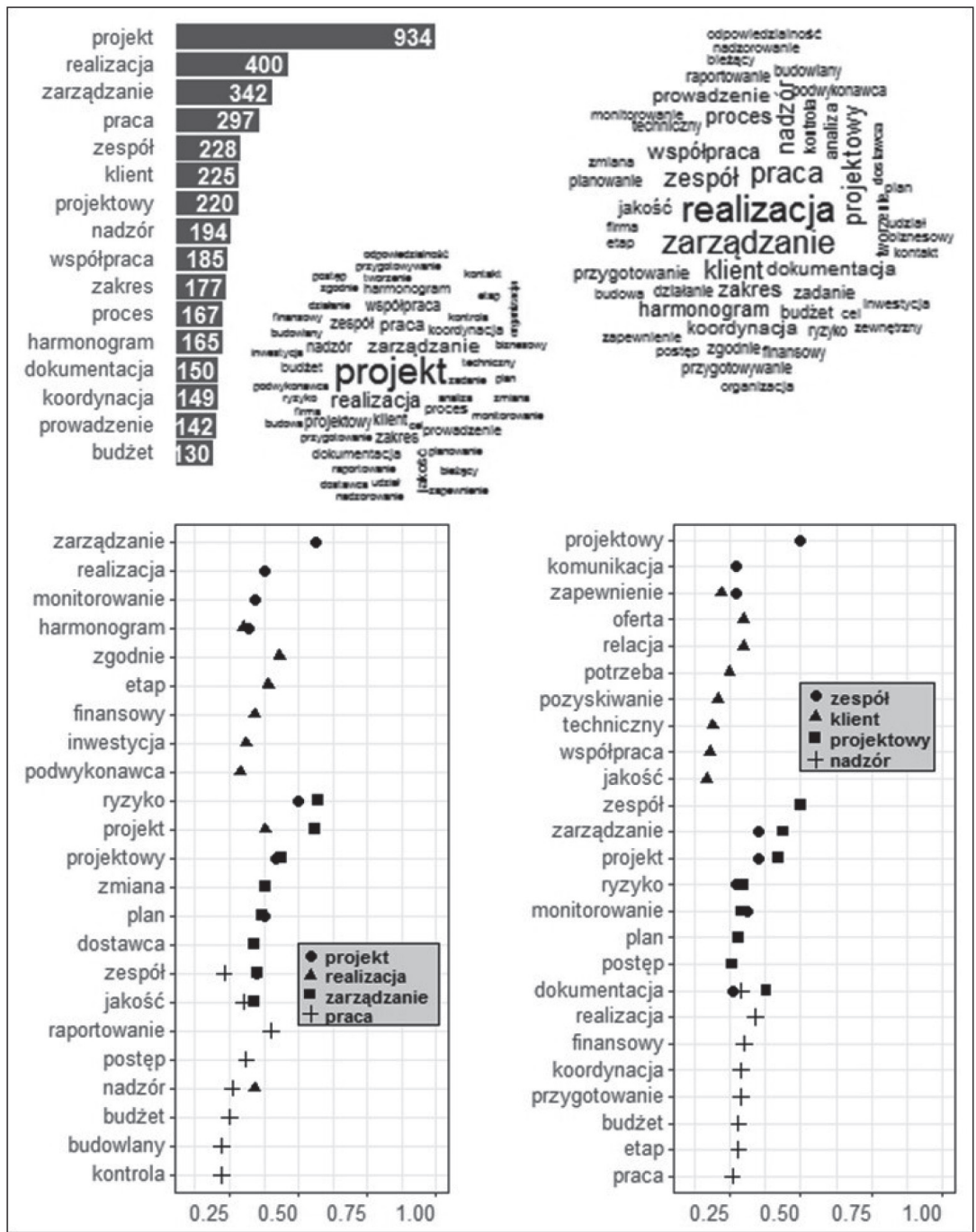
Rysunek 5. Analiza korpusu dla województwa mazowieckiego

Źródło: opracowanie własne.



Rysunek 6. Analiza korpusu dla województwa śląskiego

Źródło: opracowanie własne.



Rysunek 7. Analiza korpusu dla wszystkich województw

Źródło: opracowanie własne.

mi „potrzeba”, „cel”, „zapewnienie”, „działanie”, „zakres”, „pozyskiwanie”, „kontrola” (rys. 5), a w korpusie nr 4 ze słowami „zapewnienie”, „oferta”, „relacja”, „potrzeba”, „pozyskiwanie”, „techniczny”, „współpraca”, „jakość”. Z tych słów można ułożyć następujące frazy, np. zapewnienie komunikacji, pozyskiwanie klientów, zapewnienie komunikacji z klientem, utrzymywanie relacji z klientem, pozyskiwanie klientów, współpraca z klientem, kontrola jakości.

4. Zakończenie

Kierownik projektu to jeden z kluczowych elementów w procesie realizacji projektu. Z przeprowadzonych przez S. Spałek badań wynika, że trzy najważniejsze czynniki sukcesu projektu są związane właśnie z tym stanowiskiem [Spałek 2004]. W związku z tym pracodawcy dążą do zatrudnienia osób o określonych kompetencjach, skłonnych do wykonywania powierzonych im zadań i obowiązków.

Literatura przedmiotu dostarcza informacji o zakresie obowiązków i najczęściej wykonywanych przez kierownika projektu zadaniach. Przeprowadzone z wykorzystaniem analizy *text mining* badania miały sprawdzić, jaki jest obecnie ustalony przez pracodawców zakres obowiązków kierownika projektu.

Należy zaznaczyć, że wykorzystana analiza text miningowa nie brała pod uwagę znaczenia wyrazów oraz zdań. Pozwoliła jedynie zidentyfikować pewne reguły i prawidłowości związane z występowaniem słów oraz korelację pomiędzy nimi. Przeglądając zaprezentowane wyniki, trudno jest dokładnie ustalić cały zakres obowiązków. Natomiast doskonale widać, na co pracodawcy zwracają uwagę, co najczęściej akcentują, ustalając zakres obowiązków dla kierownika projektu.

Literatura

- Berry M.W., Kogan J., 2010, *Text Mining: Applications and Theory*, Text, Wiley.
- Borycki Ł., Sołdacki P., 2002, *Automatyczna klasyfikacja tekstów*, Mat. III Krajowej Konferencji: Multimedialne i Sieciowe Systemy Informacyjne, Wrocław, s. 473-481.
- Gładysz A., 2012, *Zastosowanie metod eksploracyjnej analizy tekstu w logistyce*, Logistyka, nr 3, s. 643-651. Available at: <http://yadda.icm.edu.pl/baztech/element/bwmeta1.element.baztech-article-BUS6-0042-0021> (accessed: 28 June 2018).
- Hearst M.A., 1999, *Untangling text data mining*, ACL '99 Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics. Morristown, NJ, USA: Association for Computational Linguistics, s. 3-10.
- Kao A., Poteet S.R., 2007, *Natural Language Processing and Text Mining*, *Natural Language Processing and Text Mining*, red. A. Kao, S.R. Poteet, Springer London, London.
- Lula P., 2005, *Text mining jako narzędzie pozyskiwania informacji. Czym jest text mining?*, s. 67-84, https://media.statsoft.pl/_old_dnn/downloads/text_mining_jako_narzedzie_pozyskiwania.pdf.
- Mirończuk M., 2012, *Przegląd metod i technik eksploracji danych tekstowych*, *Studia i Materiały Informatyki Stosowanej*, 4(6), s. 25-42, [https://repozytorium.ukw.edu.pl/bitstream/handle/item/3527/Przegląd metod i technik eksploracji danych tekstowych.pdf?sequence=1&isAllowed=y](https://repozytorium.ukw.edu.pl/bitstream/handle/item/3527/Przegl%C4%85d%20metod%20i%20technik%20eksploracji%20danych%20tekstowych.pdf?sequence=1&isAllowed=y).
- Mykowiecka A., 2007, *Inżynieria lingwistyczna. Komputerowe przetwarzanie tekstów w języku naturalnym*, Polsko-Japońska Wyższa Szkoła Technik Komputerowych, Warszawa.
- Pawlak M., 2006, *Zarządzanie projektami*, PWN, Warszawa.
- Spalek S., 2004, *Krytyczne czynniki sukcesu w zarządzaniu projektami*, Wydawnictwo Politechniki Śląskiej, Gliwice.
- Steyn H., Nicholas J.M., 2012, *Zarządzanie projektami. Zastosowanie w biznesie, inżynierii i nowoczesnych technologiach*, Wolters Kluwer, Warszawa.
- Tan A.-H., 1999, *Text mining: The state of the art and the challenges*, Proceedings of the PAKDD 1999, Workshop on Knowledge Discovery from Advanced Databases, 8, s. 65-70.
- Trocki M., 2013, *Nowoczesne zarządzanie projektami*, PWE, Warszawa.
- Wachowiak P. i in., 2004, *Kierowanie zespołem projektowym*, Difin, Warszawa.
- Walczak R., 2014, *Podstawy zarządzania projektami. Metody i przykłady*, Difin, Warszawa.