

Paweł Lula, Renata Oczkowska, Sylwia Wiśniewska

Uniwersytet Ekonomiczny w Krakowie

e-mails: pawel.lula@uek.krakow.pl; renata.oczkowska@uek.krakow.pl;

sylwia.wisniewska@uek.krakow.pl

IDENTYFIKACJA OCZEKIWAŃ PRACODAWCÓW DOTYCZĄCYCH KOMPETENCJI ZATRUDNIANYCH PRACOWNIKÓW NA PODSTAWIE EKSPLORACYJNEJ ANALIZY OFERT PRACY

IDENTIFICATION OF EMPLOYERS' EXPECTATIONS ABOUT STAFF COMPETENCES BY EXPLORATORY ANALYSIS OF JOB OFFERS

DOI: 10.15611/pn.2018.507.13

JEL Classification: M51, C55, C63

Streszczenie: Za główny cel pracy przyjęto zaproponowanie metody i przeprowadzenie badań pozwalających na identyfikację kluczowych kompetencji oczekiwanych przez pracodawców u kandydatów do zatrudnienia. Proponowane podejście polega na automatycznej analizie ofert zatrudnienia publikowanych w Internecie, w trakcie której identyfikowane są fragmenty zawierające opis wymagań wobec poszukiwanych kandydatów. Następnie, za pomocą metody ukrytej alokacji Dirichleta, określone są tematy reprezentowane w rozpatrywanych tekstach. Poprzez agregację wyodrębnionych tematów konstruowane są definicje kompetencji kluczowych. Znaczenie każdej z kompetencji wyznaczone jest poprzez zsumowanie miar znaczenia tematów uznanych za składowe rozpatrywanej kompetencji. Przedstawiona metoda posłużyła do identyfikacji kompetencji oczekiwanych u pracowników zatrudnianych w charakterze administratorów systemów informatycznych.

Słowa kluczowe: kompetencje, analiza rynku pracy, eksploracyjna analiza tekstu, ukryta alokacja Dirichleta.

Summary: Identification of employers' expectations about employees competences is the main goal of the study. In order to achieve this goal a multistage procedure based on exploratory text analysis was proposed. First a web scraping process was performed to gather a set of job offers from the web site – pracuj.pl. Next Latent Dirichlet Allocation method was used for the identification of main topics represented in a corpus. Finally, key competences were defined as a composition of topics identified by LDA approach. The results show that in the case of offers concerning IT specialists technical competences play a crucial role.

Keywords: competences, labour market analysis, exploratory text analysis, Latent Dirichlet Allocation method.

1. Wstęp

Problematyka artykułu powiązana jest z zagadnieniem analizy rynku pracy, głównym zaś jego celem jest zaproponowanie metody identyfikacji kluczowych kompetencji oczekiwanych u kandydatów do zatrudnienia na podstawie wyników eksploacyjnej analizy ofert pracy.

Proponowana metoda polega na automatycznym wyodrębnieniu z ofert pracy fragmentów tekstu zawierającego opis wymagań stawianych kandydatom, a następnie analizie tak uzyskanego materiału za pomocą metody ukrytej alokacji Dirichleta w celu identyfikacji głównych tematów występujących w przetwarzanym zbiorze tekstów. W kolejnym kroku uzyskane tematy zostały ze sobą połączone w taki sposób, aby poszczególne ich grupy stanowiły opisy kluczowych kompetencji oczekiwanych u pracowników. W pracy zaproponowano również sposób określania istotności każdej z wyodrębnionych kompetencji. Zaproponowaną metodę poddano weryfikacji polegającej na analizie ofert pracy zamieszczonych w portalu *Pracuj.pl* i dotyczących stanowisk związanych z administrowaniem systemami informatycznymi.

Proponowaną metodę zaimplementowano w języku R.

2. Pojęcie i typologia kompetencji

Tematyka kompetencji należy do kluczowych zagadnień rozważanych na gruncie teorii i praktyki zarządzania zasobami ludzkimi. W najprostszym ujęciu kompetencje obejmują wiedzę, umiejętności i postawy człowieka. W pracy [Król 2006] kompetencje definiowane są jako *predyspozycje w zakresie wiedzy, umiejętności i postaw, zapewniające realizację zadań zawodowych na poziomie skutecznym i (lub) wyróżniającym, stosownie do standardów określonych przez organizację dla danego stanowiska*. Takie rozumienie kompetencji przyjęte zostało w niniejszej pracy.

W literaturze przedmiotu istnieją różnorodne typologie kompetencji. Wśród najbardziej znanych wskazać można klasyfikację R.L. Katza [1974], który wyróżnia kompetencje techniczne, społeczne i koncepcyjne.

Złożoną typologię kompetencji proponują F. Delamare Le Deist oraz J. Winterton [2005], którzy w ramach autorskiego modelu kompetencji dokonali ich podziału na poznawcze (kognitywne), funkcjonalne, społeczne i metakompetencje.

Kolejna klasyfikacja kompetencji, opracowana przez G. Filipowicza [2014], nawiązuje do podziału kompetencji, które przedstawili R.L. Katz oraz F. Delamare Le Deist i J. Winterton. Filipowicz wyróżnia cztery zasadnicze kategorie kompetencji: społeczne, osobiste, menedżerskie i zawodowe.

Niejednoznaczność w definiowaniu pojęcia kompetencji determinuje różnorodność kategorii kompetencji. W literaturze przedmiotu funkcjonuje wiele klasyfikacji kompetencji, opracowanych na podstawie różnych kryteriów. Do najpowszechniej wskazywanych należą kompetencje: zawodowe, pozazawodowe, miękkie, twarde, personalne, społeczne, obiektywne i subiektywne.

3. Eksploracyjna analiza ofert pracy jako narzędzie identyfikacji oczekiwanych kompetencji pracowników

W trakcie badań analizie poddano oferty pracy publikowane przez portal *Pracuj.pl* (<https://www.pracuj.pl/>).

Celem badań było opracowanie metody identyfikacji, na podstawie automatycznej analizy ofert pracy, głównych kompetencji oczekiwanych przez pracodawców u zatrudnianych pracowników.

Wstępna analiza ofert pozwoliła zidentyfikować najistotniejsze czynniki utrudniające proces automatyzacji procesu przetwarzania ogłoszeń. Do najważniejszych z nich należy zaliczyć:

- zróżnicowaną strukturę ofert pracy – teksty poszczególnych ofert w serwisie *Pracuj.pl* różnią się strukturą i tytułami nagłówek poszczególnych sekcji (np. *Oczekiwania wobec Kandydatów* lub *Czego oczekujemy*),
- niejednorodny opis ofert za pomocą znaczników semantycznych języka HTML – pomimo deklarowanej w kodzie opisu oferty zgodności ze schematem *JobPosting* (<http://schema.org/JobPosting>) nie we wszystkich ofertach stosowane są właściwe znaczniki,
- dwujęzyczność ofert – oferty zamieszczane w portalu przygotowywane są w języku polskim lub angielskim.

Z uwagi na przedstawione trudności oraz potrzebę analizy dużej liczby ofert, podejmując decyzję dotyczącą typu konstruowanego modelu, wykluczono rozwiązanie oparte na regułach (przede wszystkim z uwagi na brak ujednoczonego sposobu zapisu ofert) oraz na uczeniu maszynowym przeprowadzanym w trybie z nauczycielem (gdyż tryb ten stwarzałby konieczność przygotowania dużych zbiorów uczących zawierających prawidłowy sposób interpretacji ofert). Przedstawione powody wskazują na zasadność budowy modelu uczonego w trybie bez nauczyciela, co doprowadziło do zaproponowania następującej procedury badawczej:

1. automatyczne pobranie ofert z portalu *Pracuj.pl* dotyczących zatrudnienia pracowników z określonej branży,
2. wyodrębnienie w ofertach opisu wymagań odnoszących się do kandydatów,
3. skonstruowanie modelu opisu zawartości wymagań przy wykorzystaniu metody LDA (*Latent Dirichlet Allocation*) i identyfikacja zasadniczych tematów występujących w analizowanych tekstach,
4. określenie kluczowych kompetencji oczekiwanych u zatrudnianych pracowników poprzez agregację tematów wyodrębnionych za pomocą LDA.

Szczegóły zaproponowanej procedury badawczej zostały przedstawione w dalszej części bieżącego punktu pracy.

3.1. Pobranie tekstów ofert

Na potrzeby analizy pobrano oferty z portalu *Pracuj.pl* z kategorii *IT – Administracja*. Do realizacji tego zadania wykorzystano pakiet programu R o nazwie *rvest*

[Wickham 2016], który dostępny jest pod adresem: <https://CRAN.R-project.org/package=rvest>. Pobieranie ofert zostało zrealizowane dwuetapowo. Wpierw przeanalizowano strony zawierające katalog ofert z wybranej kategorii i pobrano odnośniki do poszczególnych ofert, a następnie pobrano kod HTML opisujący poszczególne oferty.

3.2. Wyodrębnienie opisu wymagań

Przyjęto, że opis wymagań znajduje się w sekcji *description*, której wystąpienie przewiduje schemat *JobPosting* (opis tego typu sekcji ma postać `<div id="description">`). Oczekiwania formułowane w odniesieniu do kandydatów przyjmują w ofertach postać listy, w której jeden element odnosi się do jednego wymagania. W trakcie pobierania opisów wymagań dokonano podziału zamieszczonej w ofercie listy oczekiwań na elementy składowe, które w trakcie dalszej analizy traktowane były jako oddzielne dokumenty.

3.3. Skonstruowanie modelu LDA i identyfikacja tematów

Metoda ukrytej alokacji Dirichleta (*Latent Dirichlet Allocation*, LDA) została zaproponowana w [Blei, Ng, Jordan 2003]. Pozwala ona na identyfikację zasadniczych tematów występujących w korpusie oraz przedstawienie każdego dokumentu jako mieszanki wyodrębnionych tematów.

Zakłada ona, że dostępny jest słownik V złożony z LV terminów:

$$\mathbf{V} = [v_1 \quad \dots \quad v_{LV}]^T. \quad (1)$$

oraz korpus D składający się z LD dokumentów:

$$\mathbf{D} = [D_1 \quad \dots \quad D_{LD}]^T. \quad (2)$$

W trakcie analizy dokumentów tworzony jest zbiór T składający się z LT tematów:

$$\mathbf{T} = [t_1 \quad \dots \quad t_{LT}]^T. \quad (3)$$

Każdy z tematów definiowany jest poprzez informację o rozkładzie występujących w nim słów. Definicje tematów ujęte są w postaci macierzy Φ :

$$\Phi = \begin{bmatrix} \phi_{1,1} & \dots & \phi_{1,LV} \\ \dots & \dots & \dots \\ \phi_{LT,1} & \dots & \phi_{LT,LV} \end{bmatrix}. \quad (4)$$

Występujące w macierzy (4) elementy $\phi_{i,j}$ mogą być interpretowane jako prawdopodobieństwa wystąpienia j -tego słowa w i -tym temacie.

Udział poszczególnych tematów w rozpatrywanych dokumentach opisany jest za pomocą macierzy Θ :

$$\Theta = \begin{bmatrix} \theta_{1,1} & \cdots & \theta_{1,LT} \\ \cdots & \cdots & \cdots \\ \theta_{LD,1} & \cdots & \theta_{LD,LT} \end{bmatrix}, \quad (5)$$

której elementy θ_{ij} mogą być interpretowane jako prawdopodobieństwa wystąpienia j -tego tematu w i -tym dokumencie.

Szacowanie modelu LDA polega na wyznaczeniu macierzy Φ i Θ na podstawie dostępnego zbioru dokumentów. Zadanie to wykonywane jest zwykle poprzez zastosowanie procedury symulacyjnej opartej na próbkowaniu Gibbsa [Heinrich 2004]. Na potrzeby niniejszej pracy wykorzystano implementację algorytmu dostępnego w pakiecie języka R o nazwie *topicmodels* [Grün, Hornik 2011].

3.4. Określenie kluczowych kompetencji oczekiwanych u zatrudnianych pracowników

Proponowana metoda identyfikacji kluczowych kompetencji oczekiwanych u zatrudnianych pracowników składała się z trzech etapów:

- a) oszacowanie ważności tematów zidentyfikowanych za pomocą metody LDA w poszczególnych ofertach pracy,
- b) oszacowanie ważności tematów w analizowanym zbiorze ofert,
- c) zdefiniowanie składowych kluczowych kompetencji i określenie ich znaczenia.

Przystępując do realizacji pierwszego etapu zaproponowanej metody, należy pamiętać, że w trakcie budowy modelu LDA pojęcie dokumentu nie odnosiło się do pełnego tekstu oferty zatrudnienia, ale do pojedynczego punktu zamieszczonego w ogłoszeniu (który zwykle dotyczył jednego wymagania). Można więc przyjąć, że jednej ofercie odpowiadało n dokumentów wchodzących w skład korpusu. Ważność każdego z tematów w rozpatrywanej ofercie wyznaczono jako maksimum z prawdopodobieństw wystąpienia rozpatrywanego tematu w dokumentach składających się na ofertę. Schemat obliczeń dla k -tej oferty pracy przedstawia tabela 1. Oszacowanie ważności poszczególnych tematów w rozpatrywanej k -tej ofercie znajduje się w ostatnim wierszu tabeli.

Chcąc obliczyć ważność poszczególnych tematów w rozpatrywanym zbiorze ofert, zsumowano wyznaczone oszacowania ważności tematów dokonane dla poszczególnych ofert. Uzyskany w ten sposób wektor poddawany jest normalizacji (dokonywanej poprzez podzielenie poszczególnych jego elementów przez sumę wszystkich przechowywanych w nim wartości). Schemat obliczeń przedstawia tabela 2 (oszacowanie ważności tematów dla całego zbioru ofert przedstawione jest w ostatnim wierszu tabeli).

Ostatnim etapem analizy było zdefiniowanie kompetencji kluczowych poprzez wskazanie tematów mogących stanowić ich składowe. Zatem kompetencja kluczo-

Tabela 1. Schemat szacowania ważności tematów w rozpatrywanej ofercie pracy

Oferta	Dokument (punkt z opisu wymagań)	Temat 1	...	Temat LT
k	d_1	$\theta_{1,1}$...	$\theta_{1,LT}$
	d_2	$\theta_{2,1}$...	$\theta_{2,LT}$
	d_3	$\theta_{3,1}$...	$\theta_{3,LT}$

	d_n	$\theta_{n,1}$...	$\theta_{n,LT}$
	O_k	$\pi_{k,1} = \max_i(\theta_{i,1})$...	$\pi_{k,LT} = \max_i(\theta_{i,LT})$

Źródło: opracowanie własne.

Tabela 2. Schemat szacowania ważności tematów w zbiorze ofert

Oferta	Temat 1	...	Temat LT
1	$\pi_{1,1}$...	$\pi_{1,LT}$
2	$\pi_{2,1}$...	$\pi_{2,LT}$
...
K	$\pi_{K,1}$...	$\pi_{K,LT}$
Zbiór ofert	$\omega_1 = \frac{\sum_{k=1}^K \pi_{k,1}}{\sum_{t=1}^{LT} \omega_t}$...	$\omega_{LT} = \frac{\sum_{k=1}^K \pi_{k,LT}}{\sum_{t=1}^{LT} \omega_t}$

Źródło: opracowanie własne.

wa KK_q może zostać zdefiniowana jako podzbiór zbioru tematów wyodrębnionych za pomocą metody LDA:

$$KK_q \subseteq \mathbf{T}. \quad (6)$$

Przy czym dla każdych dwóch kompetencji kluczowych KK_q i KK_r zachodzi

$$KK_q \cap KK_r = \emptyset. \quad (7)$$

Znaczenie kluczowej kompetencji KK_q wyznaczyć można poprzez zsumowanie mierników ważności tematów pełniących funkcję składowych rozpatrywanej kompetencji:

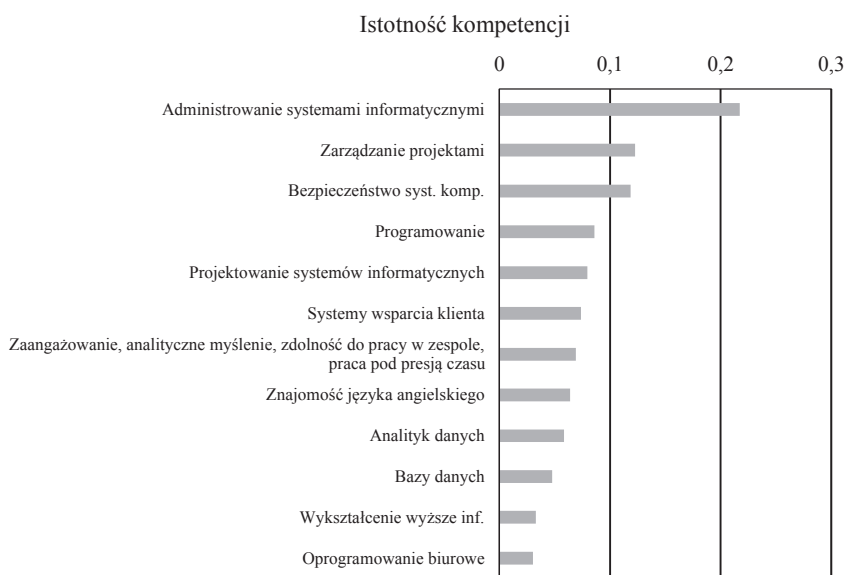
$$I(KK_q) = \sum_{s \in KK_q} \omega_s. \quad (8)$$

Przeprowadzona wcześniej normalizacja gwarantuje, że wskaźniki ważności kompetencji sumują się do jedności.

4. Rezultaty analizy ofert pracy dotyczących administratorów systemów informatycznych

W trakcie badań przeanalizowano 1618 ofert pracy dotyczących administratorów w branży IT. W dokumentach tych zidentyfikowano 39 426 zapisów dotyczących wymagań formułowanych wobec kandydatów. Każdy z tych zapisów potraktowano jako oddzielny dokument wchodzący w skład przetwarzanego korpusu.

W wyniku zastosowania metody LDA zidentyfikowano 35 tematów. Decyzję dotyczącą liczby wyodrębnionych tematów podjęto w wyniku analizy wzrokowej sposobu kształtowania się miary nieokreśloności (*perplexity*) wyznaczonej dla różnej liczby klas (zauważono, że począwszy od wspomnianej powyżej liczby klas, tempo spadku wskaźnika nieokreśloności dla modeli z większą liczbą klas wyraźnie się zmniejsza).



Rys. 1. Istotność kompetencji kluczowych wyznaczona na podstawie analizy ofert zatrudnienia w grupie administratorów systemów informatycznych

Źródło: opracowanie własne.

Następnie, analizując prawdopodobieństwa wystąpienia poszczególnych słów w wyodrębnionych tematach, podjęto próbę ich interpretacji i nadania im nazw. Z uwagi na dwujęzyczność ofert większość zagadnień reprezentowana była przez tematy opisywane za pomocą słów w języku polskim i w języku angielskim. W kolejnym kroku określono kompetencje kluczowe poprzez wskazanie ich składowych. Wyodrębniono 12 kompetencji kluczowych:

- Bezpieczeństwo systemów komputerowych,

- Znajomość języka angielskiego,
- Projektowanie systemów informatycznych,
- Zarządzanie projektami,
- Systemy wsparcia użytkownika/klienta,
- Administracja systemami komputerowymi,
- Analityk danych,
- Wykształcenie wyższe informatyczne,
- Bazy danych,
- Programowanie,
- Znajomość oprogramowania biurowego,
- Zaangażowanie, analityczne myślenie, praca pod presją czasu, zdolność do pracy w zespole.

Istotność wyodrębnionych kompetencji kluczowych przedstawiona została na rys. 1.

5. Wnioski końcowe

Wydaje się, że zaproponowana w pracy metoda pozwoliła zrealizować główny cel pracy, jakim była identyfikacja oraz określenie istotności kluczowych kompetencji pracowników na podstawie tekstów ofert pracy.

Analiza kompetencji oczekiwanych u pracowników zatrudnianych na stanowiskach administratorów systemów informatycznych wskazuje, że za szczególnie ważne należy uznać kompetencje twarde. Warto również zauważyć, że pracodawcy stosunkowo niewielką wagę przywiązują do wykształcenia formalnego, przejawiającego się ukończeniem studiów wyższych z zakresu informatyki.

Porównanie oczekiwań pracodawców z ocenami kompetencji osób poszukujących zatrudnienia może stanowić podstawę do oszacowania luki kompetencyjnej istniejącej na współczesnym rynku pracy.

Przeprowadzone badania pozwalają również na sformułowanie wniosków dotyczących zastosowanej procedury badawczej. Zastosowanie metody LDA pozwoliło w sposób prawidłowy zidentyfikować główne treści zawarte w ogłoszeniach. Warto podkreślić, że cel ten został zrealizowany w trakcie analizy zbioru dokumentów zawierających teksty przygotowane zarówno w języku polskim, jak i angielskim. Najslabszą stroną zastosowanej metody jest konieczność oceny przez badacza wyodrębnionych tematów i ich powiązanie z rozpatrywanymi kompetencjami. Dalsze badania będą miały na celu opracowanie metody automatyzującej ten etap analizy.

Literatura

- Baron-Puda M., 2008, *Zapotrzebowanie na pracowników w przedsiębiorstwach produkcyjnych*, Zarządzanie Zasobami Ludzkimi, nr 3–4, s. 126–135.
- Blei D., Ng A., Jordan M., 2003, *Latent Dirichlet Allocation*, Journal of Machine Learning Research, no. 3, s. 993–1022.
- Delamare Le Deist F., Winterton J., 2005, *What is competence?*, Human Resource Development International, vol. 8, no. 1, s. 27–46.
- Filipowicz G., 2014, *Zarządzanie kompetencjami. Perspektywa firmowa i osobista*, Oficyna a Wolters Kluwer business, Warszawa.
- Grün B., Hornik K., 2011, *Topicmodels: An R Package for Fitting Topic Models*, Journal of Statistical Software, vol. 40(1), s. 1–30, <https://www.jstatsoft.org/index.php/jss/article/view/v040i13>.
- Heinrich G., 2004, *Parameter estimation for text analysis*, <http://www.arbylon.net/publications/text-est.pdf>.
- Jeruszka U., 2016, *Kompetencje. Aspekty teoretyczne i praktyczne*, Difin, Warszawa.
- Juchnowicz M., 2014, *Sektorowe badania podaży i popytu na kwalifikacje i kompetencje*, Edukacja Ekonomistów i Menedżerów. Problemy. Innowacje. Projekty, nr 2(32), s. 11–19.
- Katz R.L., 1974, *Skills of an effective administrator*, Harvard Business Review, vol. 52, no. 5, s. 90–102.
- Król H., 2006, *Podstawy koncepcji zarządzania zasobami ludzkimi*, [w:] *Zarządzanie zasobami ludzkimi*, red. H. Król, A. Ludwicyński, Wydawnictwo Naukowe PWN, Warszawa.
- Wickham H., 2016, *Package 'rvest'*, <https://cran.r-project.org/web/packages/rvest/rvest.pdf>.