



ISSN 2080-1807

TORUŃSKIE STUDIA BIBLIOLOGICZNE
2018, nr 2 (21)

Piotr Malak

Instytut Informacji Naukowej i Bibliotekoznawstwa
Uniwersytet Wrocławski
e-mail: piotr.malak@uwr.edu.pl

Veslava Osińska

Instytut Informacji Naukowej i Bibliologii
Uniwersytet Mikołaja Kopernika w Toruniu
e-mail: wiewo@umk.pl

Grzegorz Osiński

Instytut Informatyki
Wyższa Szkoła Kultury Społecznej i Medialnej w Toruniu
e-mail: grzegorz.osinski@wksim.edu.pl


Bożena Bednarek-Michalska

Biblioteka Uniwersytecka
Uniwersytet Mikołaja Kopernika w Toruniu
e-mail: Bozena.Bednarek-Michalska@bu.umk.pl

Metoda rozszerzonej oceny współpracy badawczej i produkcji naukowej jednostek. Raport z badań

DOI: <http://dx.doi.org/10.12775/TSB.2018.011>

STRESZCZENIE: Artykuł prezentuje propozycję i ocenę komplementarnej, w stosunku do punktacji za publikację, metody oceny dorobku naukowego. Proponowana metoda wykorzystuje osiągnięcia BigData w celu eksploracji lokalnych baz danych. Metoda jest skalowalna, można ją zastosować do ewaluacji aktywności badawczej indywidualnych naukowców, zespołów badawczych oraz jednostek, aż do skali uniwersytetu.



SŁOWA KLUCZOWE: naukometria, parametryzacja, data mining, textmining, wizualizacja informacji.

Cel badań

W obecnych czasach można dostrzec tendencje dysponentów środków na badania do coraz bardziej szczegółowego oceniania dorobku badawczego naukowców. Dla dysponentów środków finansowych na badania naukowe takie potraktowanie nauki jest wygodnym sposobem ich przydziału. Sprowadzenie dorobku badawczego do modelu punktów pozwala na automatyczną ocenę czy dany naukowiec, projekt lub instytucja są dobrze rokującymi adresatami dotacji finansowej. Sprowadzenie oceny wysiłku intelektualnego do kilku arbitralnych wskaźników oraz bezkrytyczne, nie uwzględniające kontekstu naukowego porównywanie różnych dyscyplin naukowych budzi zrozumiały sprzeciw. W Polsce przykładem takiego sprzeciwu może być m.in. termin *punktoza*, wprowadzony przez E. Kulczyckiego¹. Niezależnie od oceny zjawiska, nie da się uciec przed różnorodnymi ocenami prac badawczych i ich wyników, a ocena dorobku publikacyjnego jest jedną z prób wypracowania modelu uniwersalnego. W realiach ocen parametrycznych w interesie naukowców leży rejestrowanie swoich publikacji w bibliograficznych bazach danych, a tym samym dostarczanie nowych danych do analiz dorobku.

Z kwantytatywnego punktu widzenia wnioskowanie, także ocenianie, jest tym bardziej wiarygodne, im więcej danych reprezentujących różne dostępne kategorie zgromadzimy do analizy. Stąd wynika wysoka przydatność zasobów danych bibliograficznych, rejestrujących publikacje naukowe z przypisaniem do poszczególnych naukowców, jednostek czy dyscyplin naukowych. Przypisanie do danej kategorii organizacyjnej nie zawsze jest wyrażone wprost w danych bibliograficznych, ale w połączeniu z innymi kategoriami danych i rejestrami, z wykorzystaniem metod automatycznego przetwarzania danych, można wskazać zależności po-

¹ E. Kulczycki, *Punktoza jako strategia w grze parametrycznej w Polsce*, „Nauka i Szkolnictwo Wyższe” [online]. UAM 2017, nr 1(49) [dostęp 14 marca 2018]. Dostępny w World Wide Web: <https://pressto.amu.edu.pl/index.php/nsw/article/view/8036>.

między autorami prac, a ich jednostkami macierzystymi lub dyscyplinami naukowymi.

Wynikiem działalności naukowej, oprócz publikacji, są również konferencje. Zestawienia dotyczące organizowanych konferencji, prowadzone przez poszczególne jednostki naukowe, podają m. in. ich tematykę oraz organizatorów (identyfikowanych poprzez nazwy jednostek z różnych poziomów struktur organizacyjnych). Podobnie, jak w przypadku rejestracji bibliograficznej publikacji naukowych, informacje o konferencjach dostarczają wielowymiarowe dane, które można wykorzystać w procesach automatycznego klasyfikowania i grupowania, zaś wyniki tych procesów można zastosować m. in. do analiz zakresu oraz intensywności współpracy naukowej na różnych poziomach. Poziomy te mogą wskazywać współpracę indywidualną, pomiędzy zespołami, instytucjami czy też dziedzinami lub dyscyplinami naukowymi. Zasięg i częstość współpracy badawczej, poziom interdyscyplinarności prezentowany przez poszczególnych naukowców lub zespoły mogą m.in. zostać wykorzystane do urealnienia oceny parametrycznej dorobku naukowego.

Celem niniejszej pracy jest zaprezentowanie metody oraz wyników prac badawczych z zakresu naukometrii, wykorzystujących informacje o publikacjach, konferencjach oraz przynależności instytucjonalnej do danej dyscypliny naukowej. Badania, z wykorzystaniem metod nauczania maszynowego oraz wizualizacji informacji, przeprowadzono na danych rejestrowanych w Uniwersytecie Mikołaja Kopernika w Toruniu (dalej: UMK) w systemie Expertus (dla publikacji) oraz na stronie Konferencje. Celem podstawowym badań było sprawdzenie możliwości pozyskania informacji o współpracy badawczej z danych bibliograficznych o publikacjach naukowych oraz danych dotyczących zorganizowanych na Uniwersytecie konferencji naukowych. Głównym celem było opracowanie oraz ocena skalowalnej metody rozszerzonej oceny współpracy badawczej, poziomu interdyscyplinarności badań oraz transferu wiedzy pomiędzy poszczególnymi dyscyplinami. Nowością, na tle podobnych badań, jest włączenie do analiz informacji dotyczących organizowanych konferencji naukowych, a także skorelowanie wyników analiz dla konferencji oraz publikacji naukowych.



Metody

Dane z obu analizowanych zakresów dostępne są na stronach WWW UMK. Informacje o publikacjach, wraz z ich oceną parametryczną, przechowywane i udostępniane są z poziomu systemu Expertus² pod adresem: <http://bg.cm.umk.pl/splendor/umk/>. Interfejs wyszukiwawczy bazy pozwala m. in. wybrać publikacje według kryterium nazwy wydziału lub jednostki organizacyjnej, w której zatrudniony jest autor publikacji. Dane dotyczące konferencji dostępne są poprzez wyszukiwarkę pod adresem: <https://www.umk.pl/badania/konferencje>. Interfejs wyszukiwawczy pozwala na wyświetlenie konferencji z wybranego miesiąca w danym roku.

Niestety organizacja udostępniania danych nie przewiduje pobierania zautomatyzowanego. Ani system Expertus, ani baza konferencji nie oferują dostępu do danych za pomocą interfejsu programowania aplikacji (ang. *API – application programming interface*). Brak takiego interfejsu jest poważnym utrudnieniem prac badawczych. Organizacja systemu wyszukiwania oraz prezentowania informacji w obu systemach narzuciła konieczność manualnego zbierania danych. W przypadku informacji o konferencjach było to wyświetlanie danych dla poszczególnych miesięcy w kolejnych analizowanych latach, zaś dla informacji o publikacjach było to wyświetlanie list publikacji dla poszczególnych wydziałów w określonym zakresie lat. Na potrzeby badawcze pobrano dane za lata 2004–2016, czyli pełen zakres konferencji zarejestrowanych w czasie prowadzenia badań, będących przedmiotem niniejszego artykułu. W badanym okresie na UMK zarejestrowano 1416 konferencji.

Wyniki zapytań z obu systemów zostały następnie przetworzone, oczyszczone oraz sformatowane według zadanych kategorii. W tym celu zastosowano system Google OpenRefine³, oferujący zestaw funkcji wstępnego oczyszczania danych oraz przetwarzania i organizowania danych za pomocą wyrażeń regularnych. Dalsze przetwarzanie danych, w tym

² Por. *Bibliografia publikacji pracowników UMK w Toruniu. Informacja o bazie* [online] [dostęp 14 marca 2018]. Dostępny w World Wide Web: <http://bg.cm.umk.pl/splendor/umk/obib.htm>.

³ *OpenRefine* [online] [dostęp 14 marca 2018]. Dostępny w World Wide Web: <http://openrefine.org/>.

przypisanie publikacji do poszczególnych dyscyplin naukowych oraz przygotowanie do analiz i wizualizacji, wykonano za pomocą skryptów i narzędzi języka Python.

Wzorce współpracy

Na podstawie informacji o afiliowaniu poszczególnych autorów przez konkretne wydziały oraz dotyczących organizatorów poszczególnych konferencji możliwe były analizy współpracy pomiędzy jednostkami organizacyjnymi. Sumaryczne wartości liczby publikacji stanowią podstawę do dalszych przeliczeń ewaluacyjnych i parametryzujących na różnych poziomach, od instytucjonalnego do indywidualnego. W przypadku konferencji naukowych, gdzie zawsze występuje komitet organizacyjny, nie ma możliwości przeprowadzenia ewaluacji na wszystkich poziomach szczegółowości dostępnych dla publikacji naukowych. Stąd konieczność wyboru odpowiedniego poziomu korelacji pomiędzy porównywanymi danymi. W tabeli 1. zaprezentowano porównanie ilościowe liczby publikacji oraz liczby konferencji na poziomie wydziałów UMK.

Tabela 1. Porównanie ilościowe liczby publikacji oraz konferencji na poziomie wydziałów UMK

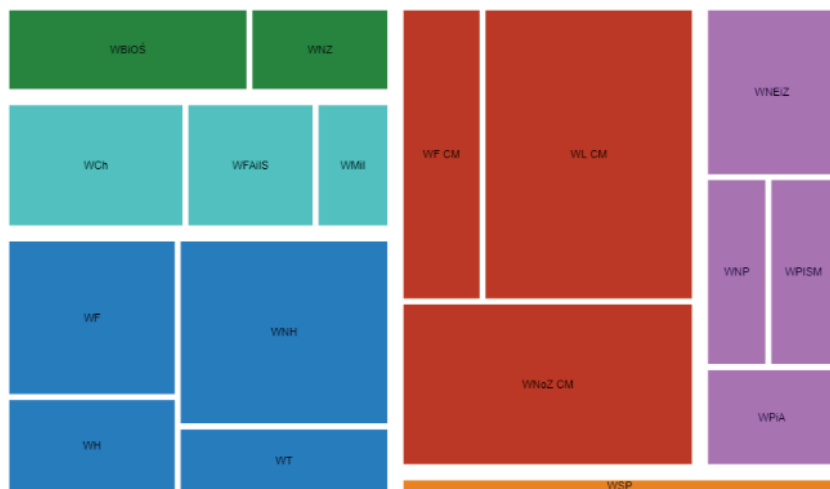
Wydział	Publikacje		Konferencje	
	liczba	udział%	liczba	udział%
WBiOŚ	4199	5.6	38	2.8
WCh	4442	6.0	42	3.1
WF	5144	6.9	222	16.6
WFAiIS	3229	4.3	79	5.9
WH	3168	4.3	79	5.9
WMiI	1856	2.5	56	4.2
WNEiZ	4424	6.0	84	6.3
WNH	7563	10.2	174	13.0
WNP	2433	3.3	53	4.0
WNEiZ	2343	3.2	61	4.6

Tabela 1. Porównanie ilościowe liczby publikacji (cd.)

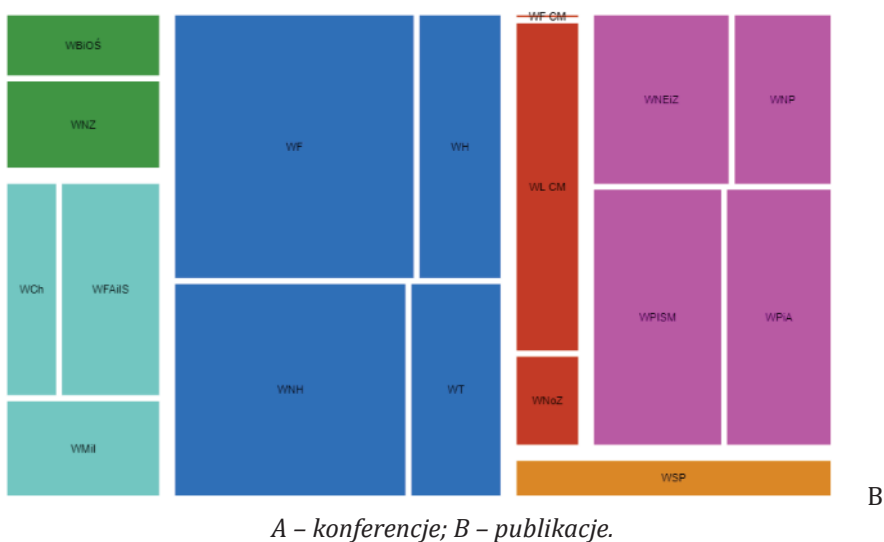
Wydział	Publikacje		Konferencje	
	liczba	udział%	liczba	udział%
WPiSM	2607	3.5	98	7.3
WPiA	2607	3.5	119	8.9
WSP	2186	2.9	54	4.0
WT	2745	3.7	70	5.2
WNoZ (CM)	9149	12.3	24	1.8
WL (CM)	11710	15.8	85	6.3
WF (CM)	4530	6.1	2	0.1

Źródło: opracowanie własne

Reprezentacja tabelaryczna wektoryzowanych wartości liczbowych dla 17 jednostek (wydziałów), z uwzględnieniem dwóch zmiennych oraz dwóch sposobów prezentacji danych, może być problematyczna ze względu na ograniczone możliwości sortowania i grupowania. Dla ułatwienia wizualnej analizy prezentowanych danych, zastosowano mapy cieplne (ang. *heatmap*), przedstawione na rys. 1.



A



Rysunek 1. Mapy ciepłe obrazujące dane dotyczące konferencji oraz publikacji, na poziomie wydziałów UMK⁴

Źródło: opracowanie własne

Mapy ciepłe są stosowane nie tylko do wizualizacji wartości liczbowych, ale również dla zobrazowania hierarchicznego danych⁵. Sąsiadujące ze sobą wykresy, pozwalają na równoległą analizę obu zbiorów danych. Władze UMK zgrupowały wszystkie wydziały w sześć kategorii w układzie podobnym do stosowanego w Web of Science.

Więzi współpracy

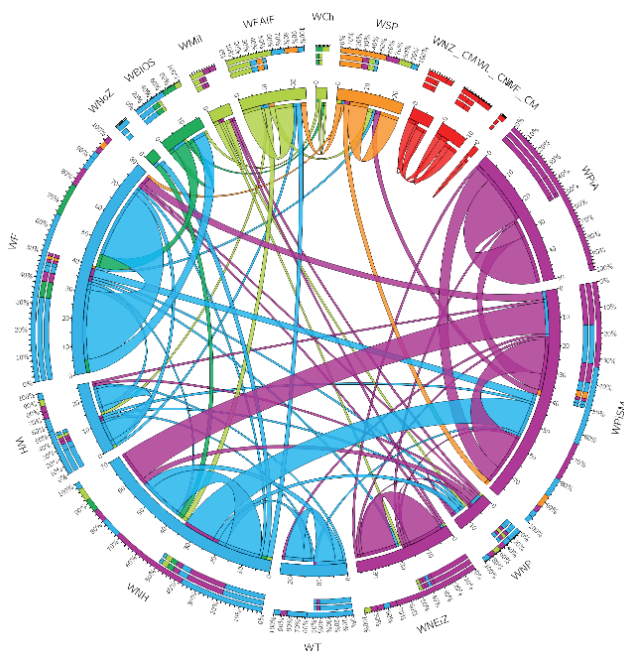
Współpraca pomiędzy jednostkami może zostać oszacowana na podstawie liczby wspólnych efektów pracy. Dla produkcji piśmienniczej będzie

⁴ Kolory przypisane poszczególnym wydziałom odpowiadają kolorom zdefiniowanym w *Systemie Identyfikacji Wizualnej UMK* [online] [dostęp 14 marca 2018]. Dostępny w World Wide Web: <https://www.umk.pl/siw/>.

⁵ B. Shneiderman, M. Wattenberg, *Ordered Treemap Layouts* [online], [w:] *Proceedings of the IEEE Symposium on Information Visualization 2001 INFOVIS'01*, San Diego,

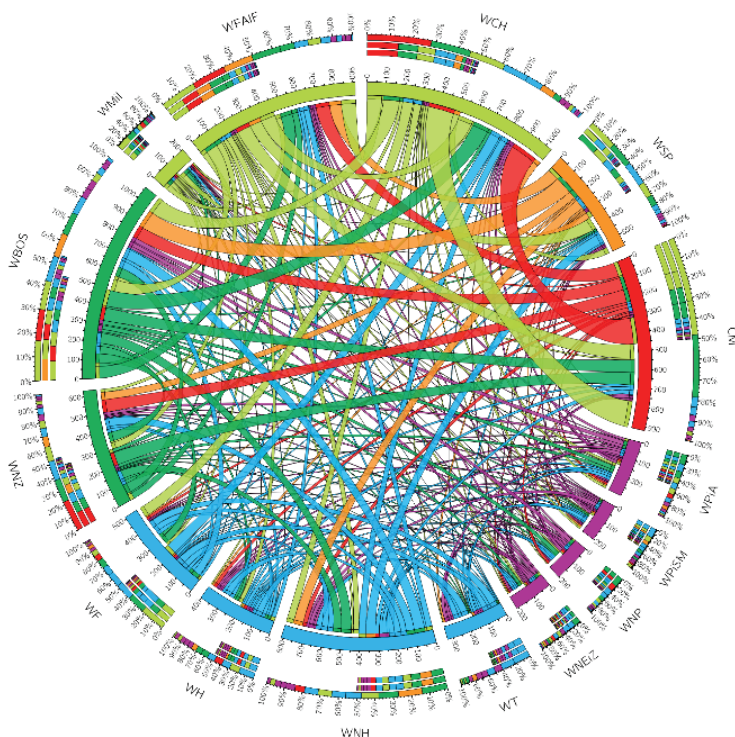
to liczba publikacji, których autorzy są afiliowani przy różnych wydziałach, zaś dla konferencji – liczba wydarzeń współorganizowanych. Na potrzeby prezentowanych badań przyjęto założenie, że współpraca na obu płaszczyznach była zrównoważona, to znaczy nie można wskazać dominującego autora wśród współautorów publikacji wspólnych oraz dominującego wydziału wśród współorganizatorów konferencji.

Relacje współpracy pomiędzy wydziałami UMK w zakresie organizacji konferencji oraz publikowania zostały zaprezentowane za pomocą wykresów przedstawionych na rys. 2 i 3. Macierze podobieństwa posłużyły do wygenerowania dwóch wykresów sieciowych. W wykresach tych również odzwierciedlono oficjalną kolorystykę SIW UMK.



Rysunek 2. Sieci współpracy pomiędzy wydziałami UMK – organizacja konferencji

Źródło: opracowanie własne



Rysunek 3. Sieci współpracy pomiędzy wydziałami UMK – współautorstwo publikacji

Źródło: opracowanie własne

Analiza zaprezentowanych wykresów pozwala na wysnucie wniosków na temat współpracy, wymiany idei oraz interdyscyplinarności badań prowadzonych na UMK. Dwa wydziały: Nauk Historycznych (dalej: WNH) oraz Politologii i Studiów Międzynarodowych (dalej: WPiSM), oznaczone fioletową i niebieską linią, wykazują największy zakres współpracy przy organizowaniu konferencji.⁶ Współpraca organizacyjna pomiędzy pozostałymi wydziałami Uniwersytetu nie jest aż tak intensywna. Z kolei w zakresie współpracy autorskiej siatka obrazująca powiązania pomiędzy

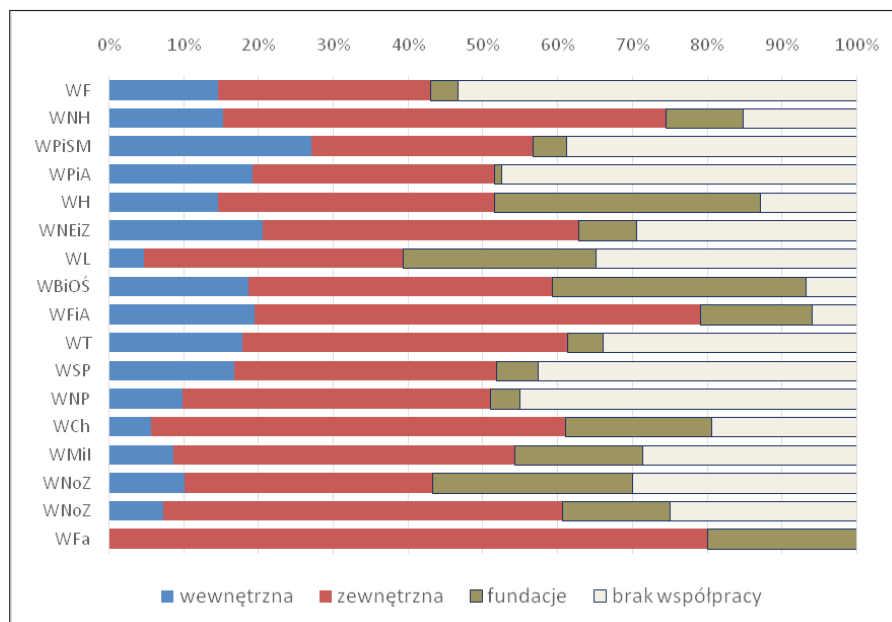
⁶ Skala i zakres tej współpracy wymaga dodatkowych badań kalibrujących, ponieważ WPiSM, jako jeden z młodszych wydziałów UMK, powstał w dużej mierze z kadry WNH.

wydziałami jest zagęszczona, co z jednej strony wskazuje na intensywność współautorstwa, ale z drugiej utrudnia bardziej szczegółowe analizy.

Próba oszacowania korelacji pomiędzy oboma aspektami wzajemnej współpracy, przeprowadzona za pomocą testu Mantela, wskazuje na brak takiej korelacji na poziomie 0,01($r(AB)=0,025$; $p=0,631$)⁷.

Geografia współpracy przy organizowaniu konferencji

Pobrane dane pozwoliły również na analizę współpracy zewnętrznej poszczególnych wydziałów UMK przy organizowaniu konferencji. Rozkład zakresu współpracy prezentuje rys. 4.

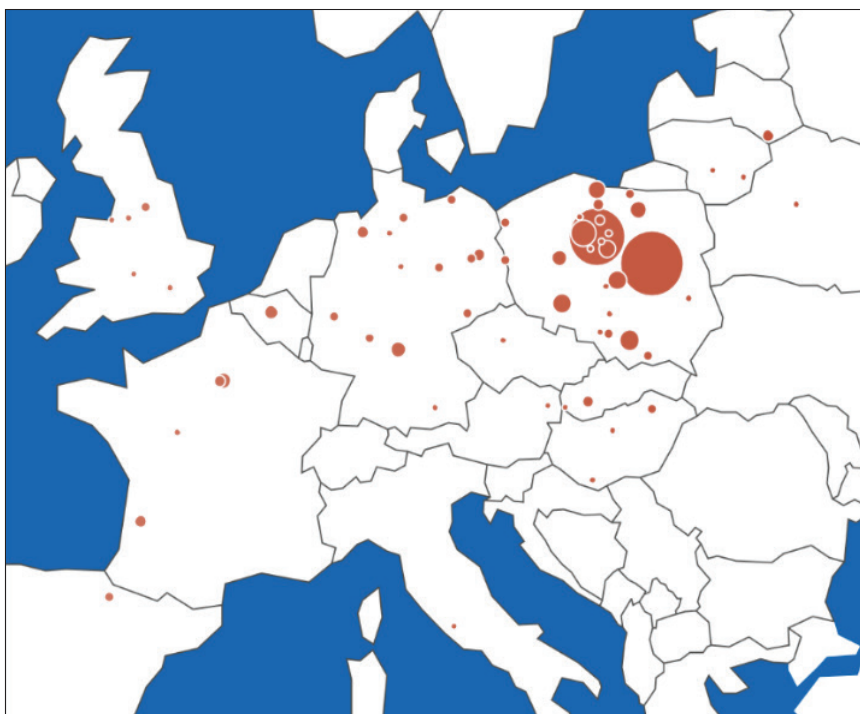


Rysunek 4. Zakresy współpracy wydziałów UMK przy organizowaniu konferencji

Źródło: opracowanie własne

⁷ N. Mantel, *A technique of disease clustering and a generalized regression approach*, „Cancer Research” 1967, nr 27, s. 209–220.

Można zaobserwować wysoki współczynnik współpracy z organizatorami zewnętrznymi oraz fundacjami w przypadku nauk farmaceutycznych oraz nauk o zdrowiu, czyli w dwóch dziedzinach z tradycyjnie wysokim udziałem interesariuszy zewnętrznych. Z kolei na rys. 5 zaprezentowana została dystrybucja geograficzna współorganizatorów konferencji UMK, bez podziału na typ współpracy. Zaznaczają się dwa główne obszary lokalizacji współorganizatorów: region najbliższy oraz region stołeczny. Z około 1400 konferencji zorganizowanych na UMK w przeciągu 13 lat, około 20% (275) było współorganizowanych z partnerami zagranicznymi.



Rysunek 5. Rozmieszczenie geograficzne współorganizatorów konferencji UMK

Źródło: opracowanie własne

W przypadku publikacji, system Expertus nie udostępnia, niestety, danych o afiliacjach współautorów spoza UMK. Pośrednio można dane te

pozyskać eksplorując bazę Scopus. W tabeli 2. prezentujemy szacunkowe dane dotyczące współpracy zewnętrznej UMK przy tworzeniu publikacji naukowych.

Tabela 2. Współpraca zagraniczna UMK w zakresie publikacji naukowych

Region	Liczba współpracujących instytucji	Liczba wspólnych publikacji
Świat	1294	2546
Afryka	27	86
Azja i region Pacyfiku	163	291
Europa	738	2270
ŚrodkowyWschód	54	138
Ameryka Północna	267	480
Ameryka Południowa	45	56

Źródło: opracowanie własne

Wyniki analizy wykazały, że zarówno współudział w organizowaniu konferencji, jak i współautorstwo publikacji odnotowują znaczny wpływ ośrodka centralnego, jakim są instytucje i uczelnie wyższe znajdujące się w Warszawie – stolicy Polski. Wpływ ośrodków warszawskich wynosi około 20% i jest to wartość bardzo znacząca. Może to świadczyć o dużej koncentracji ośrodków akademickich w stolicy oraz o ich dużej aktywności badawczej i organizacyjnej. Może też świadczyć o dużej atrakcyjności ośrodków centralnych, jednakże interpretacja tych wartości wymaga dalszych badań.

W ujęciu współpracy międzynarodowej ogranicza się ona w zasadzie do Niemiec, Francji i Wielkiej Brytanii. Udział ośrodków z innych krajów jest raczej znikomy. Nie pozostaje to bez wpływu na dynamikę wzrostu aktywności naukowej. Jest to czynnik, który winien być w najbliższych latach znacznie rozszerzony w celu utworzenia produktywnej i aktywnej sieci współpracy z międzynarodowymi ośrodkami.

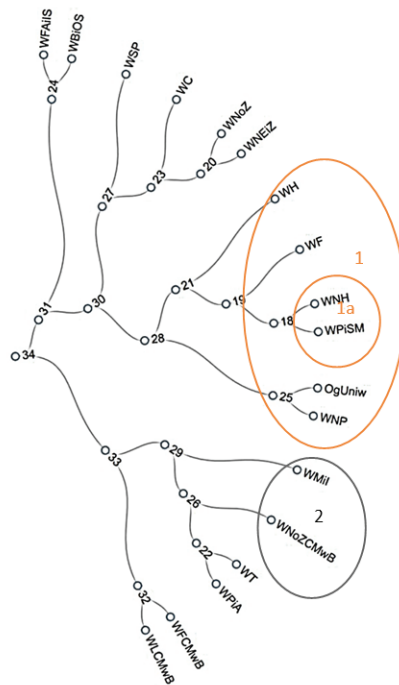
Grupowanie tekstu – podobieństwo semantyczne

Zaprezentowane dotychczas analizy oparte były na danych liczbowych. Analizy ilościowe są stosunkowo łatwe do przeprowadzenia. Większym wyzwaniem są analizy danych tekstowych. W przypadku, gdy materiałem porównawczym nie są wartości liczbowe a ciągi znaków, zazwyczaj dokonuje się konwersji metody reprezentacji danych. Teksty, dokumenty przedstawiane są wtedy jako wektory wielowymiarowe, gdzie wartości przypisane poszczególnym wymiarom wektora odzwierciedlają frekwencje słów tworzących dany tekst. Wartości te oscylują od prostych zero-jedynkowych, rejestrujących wystąpienie danego słowa w tekście lub jego brak, poprzez liczbowe oddające liczbę powtórzeń, aż do znormalizowanych według jednej z wielu dostępnych metod ujednociania. W opisywanych badaniach, dokonano analizy metodą grupowania tekstu dla tytułów i opisów konferencji organizowanych na UMK. Teksty zostały zaprezentowane w postaci wektorów, gdzie wartościami były frekwencje lematów, wygenerowanych ze słów występujących w zapisie oryginalnym. Do wskazania podobieństwa pomiędzy poszczególnymi elementami zbioru zastosowano algorytm analizy skupień Warda⁸. Służy on do porządkowania obiektów w strukturę hierarchiczną z założeniem, że obiekty należące do tej samej grupy są do siebie jak najbardziej podobne, a jednocześnie są jak najmniej związane z obiektami z pozostałych grup. Wyniki grupowania według podobieństwa tekstów, wyznaczonego z użyciem funkcji częstości słów w dokumentach, pokazuje dendrogram na rys. 3. Pomiedzy gałęziami podana jest wartość odległości pomiędzy tekstami.

Na podstawie analizy opisów tematyki konferencji można wskazać podobieństwa intuicyjne oraz takie, które nie zgadzają się z intuicją. Podobieństwo intuicyjne czyli takie, którego można spodziewać się intu-

⁸ J. H. Ward, *Hierarchical grouping to optimize an objective function*, „Journal of the American Statistical Association” ASA US 1963, nr 58, s. 236. Por. F. Murtagh, P. Legendre, *Ward’s Hierarchical Clustering Method: Which Algorithms Implement Ward’s Criterion?* „Journal of Classification” Springer 2014, nr 31, s. 274–295 oraz F. Murtagh, P. Legendre, *Ward’s Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm* [online]. Repozytorium ArXiv 2011 [dostęp 14 marca 2018]. Dostępny w World Wide Web: <https://arxiv.org/pdf/1111.6285.pdf>.

icyjnie, występuje pomiędzy Wydziałami Lekarskim a Farmaceutycznym Collegium Medicum. Te dwa wydziały tworzą klastery oznaczony jako 3 na prezentowanym dendrogramie. Zaskakujące jest natomiast umiejscowienie Wydziału Nauk o Zdrowiu we wspólnym klastrze z Wydziałem Teologicznym oraz Wydziałem Prawa i Administracji. Bliskość semantyczna WPiA i WT oraz zgrupowanego razem z nimi WNoZ wymaga dalszych, bardziej szczegółowych analiz słownictwa użytego w opisach konferencji tych trzech wydziałów.



Rysunek 6. Podobieństwo wydziałów na podstawie analizy opisów konferencji

Źródło: opracowanie własne

Środkowa grupa oznaczająca gałąź humanistyczną, jest najbardziej reprezentatywna dla szeroko pojmowanej humanistyki, bowiem skupia wydziały: Humanistyczny, Filologiczny oraz bliźniacze, wyrastające z jed-

nego środowiska i odnoszące się do studiów historycznych wydziały: WNH oraz WPiSM.

Podsumowanie i dyskusja

W wyniku analizy tekstowej, udało się wykazać pewne podobieństwo semantyczne opisów aktywności naukowców, pochodzących z różnych wydziałów uczelni. Porównując te wyniki z wykazanymi korelacjami w dynamice czasowej liczby publikacji i organizowanych konferencji na poszczególnych wydziałach, możemy wyciągnąć wnioski dotyczące całościowej charakterystyki uczelni. Najważniejsze wydaje się wyodrębnienie większej kategorii nauk humanistyczno-społecznych, które charakteryzują się odmiennymi parametrami dynamicznymi i semantycznymi, niż klasa nauk przyrodniczo-społecznych i medycznych. Metoda pozwalająca w sposób automatyczny oddzielić abstrakty humanistyczne z poszczególnych, szerszych kategorii naukowych, nie została dotychczas wdrożona, szczególnie dla prac w języku polskim. Wynika to ze specyfiki nie tylko gramatycznej, ale również semantyczno-leksykalnej języka polskiego. Wcześniejsze próby nie dały tak jednoznacznych wyników⁹ i nie pozwoliły na opracowanie praktycznych implementacji. Rezultat jest ważny, szczególnie dla przyszłego rozwoju polskiej humanistyki cyfrowej, uwzględniającej uwarunkowania leksykalne.

Otrzymane wyniki klasyfikacyjne można z powodzeniem skonfrontować ze strukturą wzajemnych powiązań aktywności naukowej poszczególnych wydziałów za pomocą metod wizualizacyjnych. Standardowo w takich przypadkach stosowana jest wizualizacja w postaci wykresu aluwialnego. Możemy zaobserwować, że kategoria humanistyki jest semantycznie obszerniejsza, niż pozostałe klastry naukowe. Poza tym widzimy, że nauki społeczne nie posiadają własnej, oddzielnej reprezentacji w analizowanej aktywności konferencyjnej. Nie można ich odróżnić od nauk humanistycznych na analizowanym poziomie treści. Natomiast klasteryzacja Warda wykazuje istotność zagadnień związanych z naukami

⁹ V. Osińska, P. Malak, *Maps and Mapping in Scientometrics*, [w:] *Metody i narzędzia badań piśmiennictwa cyfrowego i jego użytkowników*, pod red. Małgorzaty Góralskiej i Agnieszki Wandel, Wrocław 2016, s. 59–72.

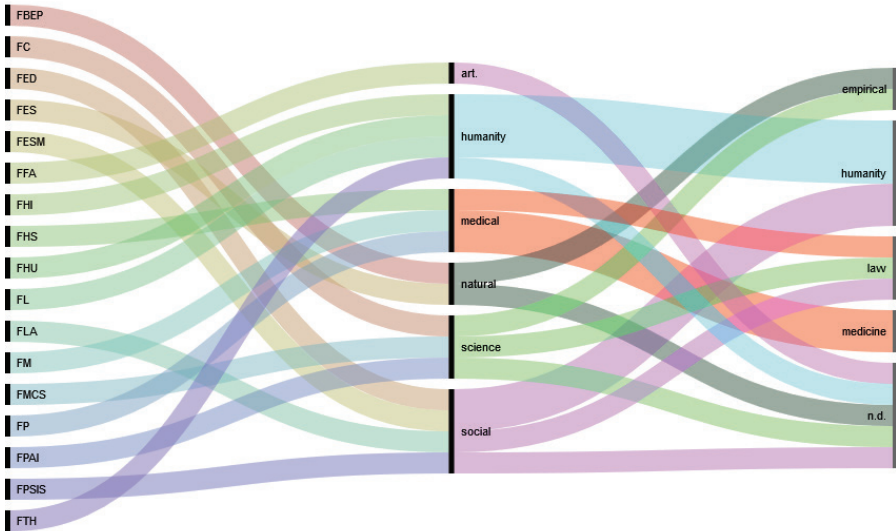
prawnymi we wszystkich analizowanych dyscyplinach naukowych. To raczej oczywiste spostrzeżenie, bowiem aspekty prawne, szczególnie dotyczące problematyki innowacji naukowych, patentów, praw autorskich i samej ustawy dotyczącej szkolnictwa wyższego, są naturalnym wewnętrznym elementem dotyczącym klasycznej aktywności naukowej.

W wizualizacji możemy również zidentyfikować klaster, który nie jest jednoznacznie określony, ponieważ łączy różne dyscypliny zgrupowane wokół problematyki artystycznej. Zagadnienia dotyczące „sztuk pięknych” na UMK realizowane są w zakresie konserwacji i restauracji dzieł sztuki. Władze uczelni planują utworzenie nowego gmachu specjalistycznych pracowni i budowę Toruńskiego Centrum Konserwacji Zabytków. Wyniki naszej analizy jednoznacznie wykazują, że taka inicjatywa jest jak najbardziej uzasadniona. Dlatego stosowana sztywna hierarchia wydziałowa może stanowić pewną przeszkodę w restrukturyzacji uniwersytetu, ponieważ kategoria Sztuka jest bardzo szerokim zagadnieniem i powinna być dalej rozwijana w paradygmacie badań interdyscyplinarnych. Wzajemne sprzężenia pomiędzy naukami przyrodniczymi i technicznymi a sztuką, choć na „pierwszy rzut oka” zupełnie abstrakcyjne, stają się obecnie niezwykle ważnym elementem rozwoju naukowego. Badania dynamiki naukowej tego wydziału również wykazują znaczne fluktuacje, związane zapewne z bardzo aktywną współpracą międzydziedzinową. Otwartym pozostaje pytanie, jak zagadnienia interdyscyplinarne mają się zmieścić w sztywnej strukturze wydziałowej polskich uniwersytetów? Organizacja pracy naukowej na uczelni, w odniesieniu do wykazanych relacji pomiędzy tematami badawczymi, wymaga zatem pewnej zmiany struktury. Zarządzanie jednostkami wymaga większej elastyczności, w celu stymulowania dynamiki zmian w stosunku do mierzonych osiągnięć naukowych.

Problem ten nie dotyczy tylko Wydziału Sztuk Pięknych. W ogólności struktura wydziałowa wydaje się niedopasowana zarówno do polityki finansowej, jak i aktywności naukowej. Badania prowadzone w strukturach międzywydziałowych wykazują większą dynamikę publikacyjną. Jednak jednoznaczne wnioski na ten temat będzie można wyciągnąć dopiero po szczegółowej analizie współpublikowania i współcytowania pracowników różnych wydziałów. Takie badania są planowane w przyszłości i będą one uwzględniały wyniki opisywane w niniejszej pracy. Jednak już dzisiaj możemy stwierdzić, że aby usprawnić przepływ pomysłów i koncepcji naukowych pomiędzy pracownikami różnych wydziałów,



należy połączyć wysiłki humanistów i przedstawicieli nauk społecznych. W analizie globalnej należy zatem stworzyć poszerzoną grupę humanistyczno-społeczną, natomiast klastery sztuk pięknych można rozszerzyć o czynnik dużej współpracy międzywydziałowej.



Rysunek 7. Zmiany strukturalne organizacji wydziałów na podstawie analizy semantycznej

Źródło: opracowanie własne

Konferencje na wydziałach medycznych często współorganizują komercyjne firmy medyczne i farmaceutyczne, nie ma to jednak bezpośredniego wpływu na aktywność współpracy naukowej, ale zapewnia odpowiedni poziom finansowania badań. Z analizowanych danych wynika, że zagraniczni partnerzy biorą udział w organizacji konferencji jako wykładowcy (key-speakers), członkowie rad naukowych i innych gremiów reprezentacyjnych. W ujęciu badania realnej współpracy naukowej nie są to czynniki najważniejsze. Z punktu widzenia zwiększenia aktywności naukowej, dużo ważniejszy jest realny udział naukowców z innych ośrodków w projektach naukowych na różnych poziomach organizacyjnych. Być może brak aktywnej współpracy międzynarodowej w zakresie organizacji konferencji uzupełnia dość duży udział organizacji non-profit i fundacji,

jednak dokładna ocena tych wpływów nie jest możliwa bez analizy danych pochodzących z zewnętrznych źródeł. Możemy jednak domniemywać, iż UMK nie jest wyjątkiem na arenie międzynarodowej i jego pracownicy aktywnie angażują się we współpracę z organizacjami non-profit.

W opisanych badaniach udało się wypracować metodę oceny potencjału naukowego uczelni. Zaprezentowana metoda, wykorzystująca dane zbierane i udostępniane lokalnie przez uczelnię, może pełnić funkcje komplementarne do innych, stosowanych dotychczas, metod parametryzacji. Przeprowadzone badania oraz ich wyniki, zaprezentowane w postaci wizualizacji informacji, potwierdzają skalowalność opracowanej metody. Za jej pomocą można ewaluować dorobek poszczególnych naukowców, jednostek czy wydziałów w ramach jednej uczelni. Dalsze badania przewidują ocenę metody w skali kilku uniwersytetów. Wskazane byłoby również opracowanie współczynników, określających wartość dorobku wynikającego ze współpracy na poziomie wewnętrznym oraz zewnętrznym.

Podziękowania

Badania były realizowane w ramach projektu NCN nr 2013/11/B/HS2/03048 pt. „Badanie struktury i dynamiki rozwoju cyfrowych zasobów wiedzy przy pomocy metod Wizualizacji”.

Bibliografia:

Publikacje drukowane:

- Mantel Nathan, *A technique of disease clustering and a generalized regression approach*, „Cancer Research” 1967, nr 27, s. 209–220.
- Murtagh Fionn, Legendre Pierre, *Bibliografia publikacji pracowników UMK w Turunii. Informacja o bazie „Journal of Classification”* Springer 2014, nr 31, s. 274–295.
- Osińska Veslava, Malak Piotr, *Maps and Mapping in Scientometrics*, [w:] *Metody i narzędzia badań piśmiennictwa cyfrowego i jego użytkowników*, pod red. Małgorzaty Góralskiej i Agnieszki Wandel, Wrocław 2016, s. 59–72.

Ward Joe H., *Hierarchical grouping to optimize an objective function*, „Journal of the American Statistical Association” ASA US 1963, nr 58, s. 236.

Publikacje elektroniczne:

- Bibliografia publikacji pracowników UMK w Toruniu. Informacja o bazie* [online]. Uniwersytet Mikołaja Kopernika w Toruniu. Collegium Medicum w Bydgoszczy 2017 [dostęp 14 marca 2018]. Dostępny w World Wide Web: <http://bg.cm.umk.pl/splendor/umk/obib.htm>.
- Expertus – the list of installations. Resource document.* [online] [dostęp 14 marca 2018]. Dostępny w World Wide Web: <http://www.splendor.net.pl/e0500001.htm>.
- Kulczycki Emanuel, *Punktoza jako strategia w grze parametrycznej w Polsce*, „Nauka i Szkolnictwo Wyższe” [online]. UAM 2017, nr1(49). [dostęp 14 marca 2018]. Dostępny w World Wide Web: <https://pressto.amu.edu.pl/index.php/nsw/article/view/8036>.
- Murtagh Fionn, Legendre Pierre, *Ward’s Hierarchical Clustering Method: Clustering Criterion and Agglomerative Algorithm* [online]. Repozytorium ArXiv 2011. [dostęp 14 marca 2018]. Dostępny w World Wide Web: <https://arxiv.org/pdf/1111.6285.pdf>.
- Murtagh Fionn, Legendre Pierre, *Ward’s Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward’s Criterion?* [online]. „Journal of Classification” 2014, nr 31, s. 274–295. [dostęp 14 marca 2018]. Dostępny w World Wide Web: <https://pdfs.semanticscholar.org/2f42/0d40bbef29f47005c1fc22c2aff98605f659.pdf>.
- OpenRefine. Open source rozwijany przez społeczność* [online] [dostęp 14 marca 2018]. Dostępny w World Wide Web: <http://openrefine.org/>.
- Osińska Veslava, Osiński Grzegorz, Tomaszewski Wojciech, *Modelling of Scientific Collaboration based on Graphical Analysis* [online], [w:] *Proceedings of 15th International Conference on Scientometrics and Informetrics*. Bożazići University Printhouse 2015, s. 1257 [dostęp 14 marca 2018]. Dostępny w World Wide Web: <https://repository.uantwerpen.be/docman/irua/.../ae3eb5f8.pdf>.
- Shneiderman Ben, Wattenberg Martin, *Ordered Treemap Layouts* [online], [w:] *Proceedings of the IEEE Symposium on Information Visualization 2001 INFOVIS’01*, San Diego, CA, USA, October 22–23, 2001. IEEE Computer Society,

s. 73. [dostęp 14 marca 2018]. Dostępny w World Wide Web: <https://dblp1.uni-trier.de/db/conf/infovis/infovis2001.html>.

Splendor® Information Systems. *Resource document* [online] [dostęp 14 marca 2018]. Dostępny w World Wide Web: <http://www.splendor.net.pl/english.html>.

A method of increased documentation of production and production of a scientific unit. Research report

ABSTRACT: Article presents a proposition and evaluation of a new method of scientific output evaluation. Presented method may be considered as complementary to publication evaluation. It uses up-to-date findings of BigData in order to explore local databases. Proposed method is scalable, it may be successfully applied for scientific evaluation of individual scholars, research teams and units, up to university size.

KEYWORDS: scientometrics, scientific evaluation, data mining, text mining, information visualization.