

INSTRUKCJA KORZYSTANIA Z APLIKACJI SCIENTIFIC VISUALIZER

OPRAC. DR HAB. MAŁGORZATA KOWALSKA
(INSTYTUT INFORMACJI NAUKOWEJ I BIBLIOLOGII UMK,
E-MAIL: KOMA@UMK.PL)

Wprowadzenie

Wizualizator dorobku naukowego – **Scientific Visualizer**, dostępny pod adresem http://visualizeme.umk.pl/, zwany dalej aplikacją, przygotowany został w ramach grantu Badanie struktury i dynamiki cyfrowych zasobów wiedzy za pomocą metod wizualizacji¹.

Jest on narzędziem, które oferuje analizę i wizualizację sumarycznych zestawień danych odnoszących się do dorobku publikacyjnego według kilku kryteriów: dynamiki produktywności naukowej w perspektywie czasowej, sieci współpracy autorskiej, geolokalizacji miejsc wydania poszczególnych prac, reprezentowanych dyscyplin nauki, sumarycznej liczby stron wszystkich prac naukowych opublikowanych w danym roku kalendarzowym oraz polskich i angielskich słów kluczowych.

Do wizualizacji danych – ładowanych do pamięci z systemu bazodanowego MongoDB – w aplikacji wykorzystano rozbudowaną bibliotekę skryptów D3.js², a w zakresie wskazywania położenia geograficznego – interfejs aplikacji Google Maps API.

Analiza dorobku naukowego z wykorzystaniem możliwości oferowanych przez aplikację składa się z trzech etapów:

- 1. przygotowania i pobrania pliku z danymi bibliograficznymi z systemu Expertus,
- 2. wczytania uprzednio pobranego pliku z danymi do aplikacji Scientific Visualizer,
- 3. analizy dorobku publikacyjnego w aplikacji Scientific Visualizer.

Poniżej w uproszczonej formie omówiono każdy z ww. etapów.

Do prezentacji możliwości systemu Expertus oraz aplikacji Scientific Visualizer wykorzystano dane bibliograficzne obrazujące dorobek naukowy autorki niniejszego opracowania.

-

¹ NCN nr 2013/11/B/HS2/03048.

² D3.js (Data-Driven Documents) jest biblioteką w języku JavaScript przeznaczoną do tworzenia zaawansowanych wizualizacji danych w rozmaitych aplikacjach internetowych. Umożliwia ona nie tylko generowanie wykresów, ale pozwala również na tworzenie map, interaktywnych diagramów, paneli kontrolnych dla danych, raportów itp.

ETAP 1:

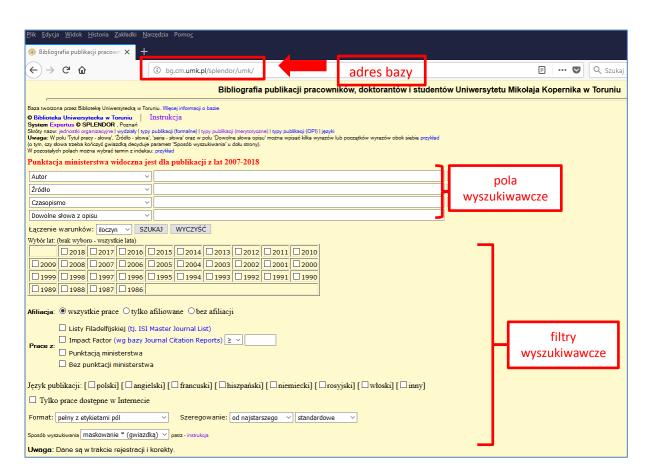
Przygotowanie i pobranie pliku z danymi z systemu Expertus

Aby przeprowadzić w aplikacji Scientific Visualizer analizę produktywności naukowej, konieczne jest posiadanie odpowiednio sformatowanego pliku z danymi.

Aktualnie aplikacja przetwarza dane pobrane z bibliograficznej bazy publikacji pracowników, doktorantów i studentów UMK, przygotowanej w systemie Expertus, dostępnej pod adresem: http://bg.cm.umk.pl/splendor/umk/

Aby przygotować plik z danymi należy wykonać następujące czynności:

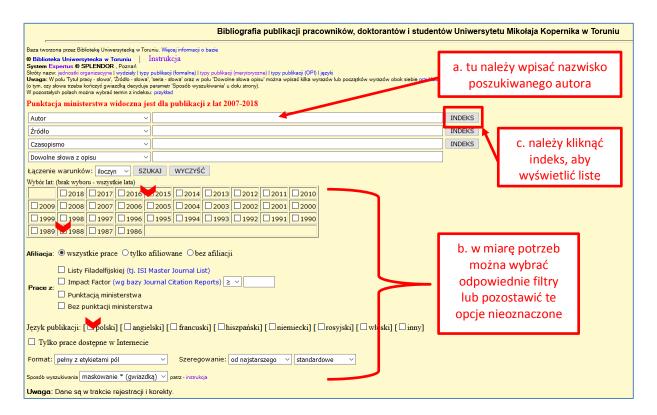
1. Za pomocą dowolnej przeglądarki internetowej wejść na stronę http://bg.cm.umk.pl/splendor/umk/ (zob. rys. 1)



Rys. 1. Strona startowa *Bibliografii publikacji pracowników, doktorantów i studentów UMK (system Expertus)* Źródło: http://bg.cm.umk.pl/splendor/umk/

2. Na stronie głównej:

- a. w polu wyszukiwawczym "Autor" ³ wpisać żądane nazwisko pracownika nauki, doktoranta lub studenta,
- w miarę potrzeb zaznaczyć interesujące filtry wyszukiwania (daty wydania publikacji, prace z afiliacją lub bez, prace punktowane, język publikacji, dostępność online, format danych, szeregowanie wyników itp.),
- c. **kliknąć na przycisk "Indeks"** znajdujący się obok pola "Autor" (zob. rys. 2) lub nacisnąć na klawiaturze ENTER.



Rys. 2. Wyszukiwanie konkretnych osób i ich dorobku w *Bibliografii publikacji pracowników, doktorantów i studentów UMK (system Expertus)*

⁻

^{*} Z względu na fakt, że dla potrzeb aplikacji konieczne jest utworzenie pliku zawierającego dorobek danego autora, w niniejszym tutorialu omówiona jest procedura tworzenia wykazu dorobku dla danego nazwiska badacza. W systemie poza wyszukiwanie wg nazwiska autora, wyszukiwać można także wg innych kryteriów, jak tytuł publikacji, jednostka, słowa kluczowe, typ publikacji, słowa z tytułu, miejsce wydania, tytuł czasopisma itd. Zasady postępowania są jednak identyczne jak w przypadku wyszukiwania wg hasła autorskiego.

- 3. Po wyświetleniu listy nazwisk (terminów z indeksu):
 - a. w przypadku wyszukiwania prac kilku autorów zaznaczyć kwadraty znajdujące się przy tych nazwiskach, a następnie zainicjować proces wyszukiwania w bazie poprzez wybór przycisku "Szukaj zaznaczonych" (zob. rys. 3),
 - w przypadku wyszukiwania prac jednego autora postąpić analogicznie albo kliknąć bezpośrednio na dane nazwisko lewym klawiszem myszy komputerowej.



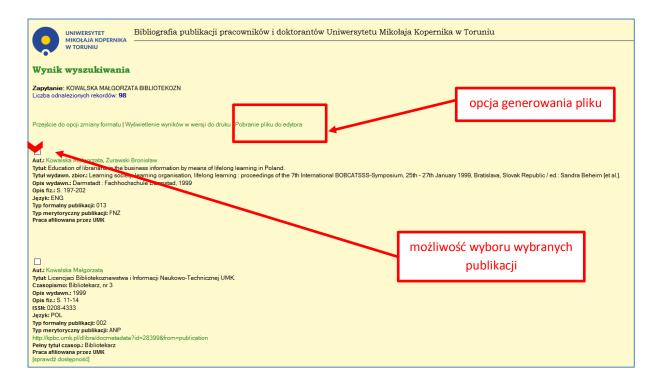
Rys. 3. Lista nazwisk wyświetlana po wywołaniu indeksu (system Expertus)

4. Po zakończeniu procesu wyszukiwania wyświetlona zostaje lista rezultatów wyszukiwania w postaci wykazu publikacji wskazanej osoby (zob. rys. 4).



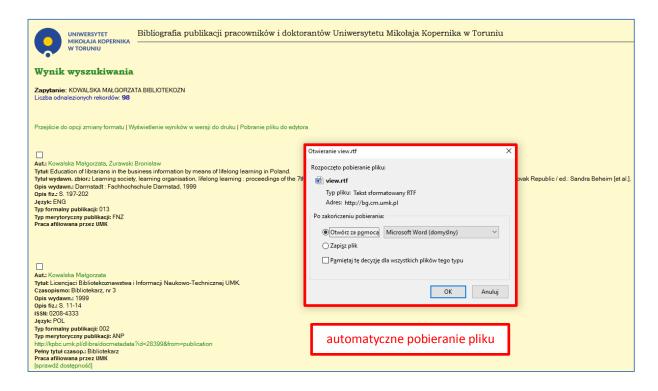
Rys. 4. Wyniki wyszukiwania (system Expertus)

5. Aby wygenerować plik z wybranymi publikacjami danej osoby należy zaznaczyć kwadraty znajdujące się nad opisem interesujących użytkownika publikacji, a następnie wybrać widniejącą nad wykazem publikacji opcję "Pobieranie pliku do edytora" (zob. rys. 5). W przypadku zamiaru pobrania listy wszystkich wyświetlonych publikacji, swoje działanie wystarczy ograniczyć do kliknięcia opcji "Pobieranie pliku do edytora".



Rys. 5. Tworzenie pliku z dorobkiem (system Expertus)

6. Plik z wykazem publikacji pobierany jest z systemu automatycznie i ma postać pliku tekstowego z rozszerzeniem .rtf. **Plik ten należy zapisać** w pamięci komputera lub na dysku pamięci.



Rys. 6. Pobieranie pliku z dorobkiem (system Expertus)

Etap 2:

Wczytanie pliku z danymi do aplikacji Scientific Visualizer

Aby w aplikacji rozpocząć proces analizy produktywności naukowej, należy wczytać plik z danymi. W tym celu na stronie głównej aplikacji pod adresem http://visualizeme.umk.pl/, z menu głównego (żółty pasek pod adresem, prawy róg) należy wybrać opcję "Wybierz plik" (zob. rys. 7) i za jej pomocą załadować (z pamięci komputera/dysku pamięci) pobrany uprzednio z bazy Expertus plik z danymi.



Rys. 7. Interfejs aplikacji

Źródło: http://visualizeme.umk.pl/

O właściwym załadowaniu pliku świadczy pojawienie się na pasku menu aplikacji nazwiska badacza oraz sumarycznych statystyk jego dorobku publikacyjnego (zob. rys. 8).

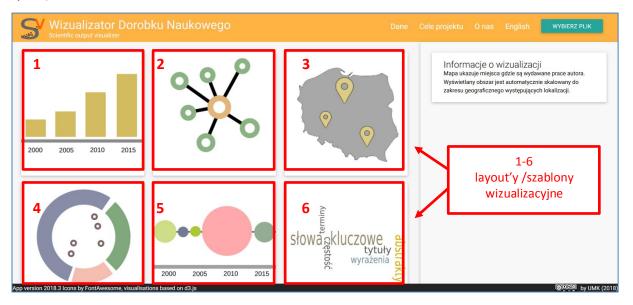


Rys. 8. Interfejs aplikacji po wczytaniu pliku z danymi

Etap 3:

Analiza dorobku naukowego w aplikacji Scienific Visualizer

Aplikacja Scientific Visualizer składa się z sześciu layotów/szablonów wizualizacyjnych (zob. rys. 9).



Rys. 9. Interfejs aplikacji po wczytaniu pliku z danymi Źródło: http://visualizeme.umk.pl/

Każdy z layout'ów/szablonów ma inną postać graficzną i spełnia inne zadanie:

- layout 1 (wykres słupkowy) dostarcza informacji o dynamice własnej oraz umożliwiać
 wychwycenie zmian w niej zachodzących,
- layout 2 (graf) ukazuje sieci współpracy, a ściślej współautorstwa publikacji oraz współredagowania monografii zbiorowych i periodyków,
- layout 3 (mapka) wskazuje geolokalizację miejsc wydania poszczególnych prac,
- layout 4 (wykres pierścieniowy mapa dziedzin) służy identyfikacji dyscyplin naukowych, w których publikuje dany uczony (aby umożliwić wykonanie tego zadania konieczne było zweryfikowanie m.in. na podstawie indeksów Web of Science list czasopism naukowych przypisanych do poszczególnych dyscyplin, a w przypadku monografii określenie ich tematu zasadniczego, a w dalszej kolejności przyporządkowanie ich do dziedzin wiedzy),
- layout 5 (wykres bąbelkowy) ilustruje sumaryczną liczbę (poniżej wykresu bąbelkowego, na wykresie liniowym, skumulowano dodatkowo liczbę wszystkich punktów przyznanych za publikacje wydane w danym roku),

 layout 6 (chmura słów kluczowych) – pozwala na zestawienie, a poprzez to porównanie tematyki podejmowanej przez pojedynczego badacza na łamach krajowych i zagranicznych periodyków.

Aby wyświetlić wyniki analizy w określonej postaci graficznej (wybrany layout), należy najechać wskaźnikiem myszy komputerowej na żądany layout, a następnie kliknąć w dowolnym miejscu jego obszaru lewym klawiszem myszy.

Aby wyjść z wybranego layout'u należy skorzystać z opcji "Wstecz/Powrót do poprzedniej strony" znajdującej na pasku menu przeglądarki internetowej lub strzałki widniejącej w menu górnym aplikacji (zob. rys. 10).





Rys. 10. Powrót do strony głównej aplikacji

Źródło: oprac. własne

Oprócz wbudowanych layout'ówi wizualizacyjnych, aplikacja wyposażona jest w menu główne (żółty pasek pod paskiem adresu strony WWW, zob. rys. 10), w którym pod pojedynczymi zakładkami przedstawiono cel tworzenia aplikacji, zespół projektantów oraz informacje na temat rodzaju przetwarzanych danych. By ułatwić użytkownikom pracę z aplikacją, w zakładce "Dane" zamieszczono przykładowy plik do analiz wizualnych oraz opis struktury danych – swego rodzaju wskazówki, jak należy przygotować dane wejściowe.

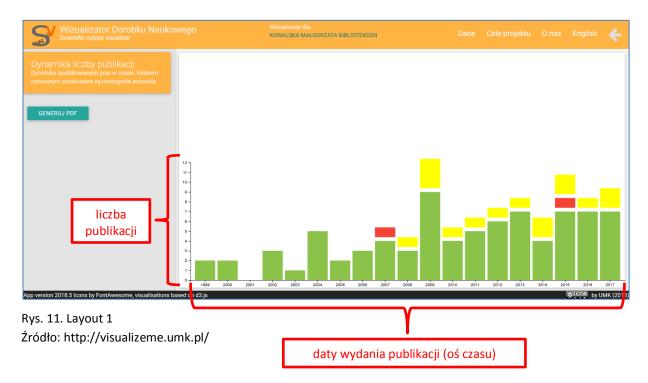
Poniżej szczegółowo omówiono zawartość każdego z layoutów oraz uruchomione w ich ramach opcje nawigacyjne.

Layout 1 (zob. rys 11) ma postać wykresu słupkowego. Wyświetla się na nim sumaryczna liczba prac opublikowanych przez danego autora. Wszystkie prace uporządkowane są tu według roku publikacji i rozłożone na osi czasu.

Kolorem <mark>zielonym</mark> oznaczone są artykuły z czasopism i prac zbiorowych,

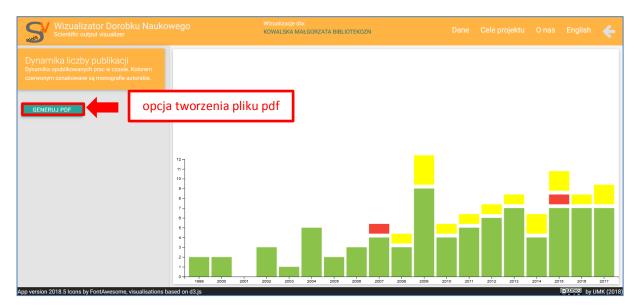
Kolorem <mark>żółtym</mark> – redakcje wydawnictw zbiorowych i redakcje czasopism,

Kolorem <mark>czerwonym</mark> – monografie.



Najechanie wskaźnikiem myszy komputerowej na pojedyncze kolumny powoduje wyświetlenie nad nimi etykiet danych, tj. sumarycznej liczby prac za dany rok kalendarzowy wraz ze wskazaniem przygotowanych w tym czasie liczby monografii i prac zbiorowych.

Efekty analizy danych uwidocznione na layout'cie 1 można wygenerować do pliku .pdf poprzez wybór opcji "Generuj pdf" (zob. rys. 12).

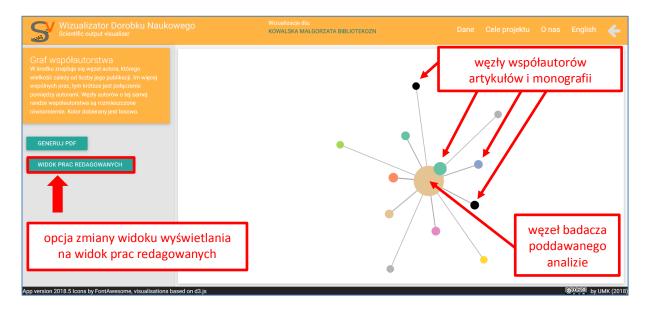


Rys. 12. Layout 1 (generowanie pliku pdf)

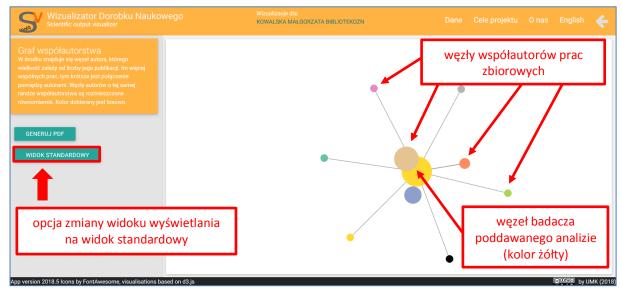
Layout 2 ma postać grafu. Ukazuje on sieć współautorstwa artykułów naukowych i monografii oraz współredagowania wydawnictw zbiorowych i periodyków.

Ze względu na różne typy publikacji, sieć współpracy autorskiej można analizować w widoku standardowym i w widoku prac redagowanych. W pierwszym z nich wyświetlani są współautorzy artykułów naukowych i monografii (zob. rys. 13), w drugim – współredaktorzy prac zbiorowych i wydawnictw ciągłych (czasopism) (zob. rys. 14).

Bez względu na rodzaj widoku, w centralnym punkcie grafu (środek) znajduje się węzeł badacza, którego dorobek jest przedmiotem analizy w aplikacji. Jego wielkość zależy od sumarycznej liczby jego publikacji. Dookoła węzła badacza widnieją (z reguły nieco mniejsze) węzły współautorów. Im więcej wspólnych prac pomiędzy badaczem będącym głównym podmiotem analizy a innymi autorami, tym krótsze połączenia między węzłami. Węzły autorów o tej samej randze współautorstwa (ta sama liczba wspólnie zredagowanych prac) są rozmieszczone równomiernie. Kolor węzłów dobierany jest losowo.



Rys. 13. Layout 2 (widok standardowy) Źródło: http://visualizeme.umk.pl/

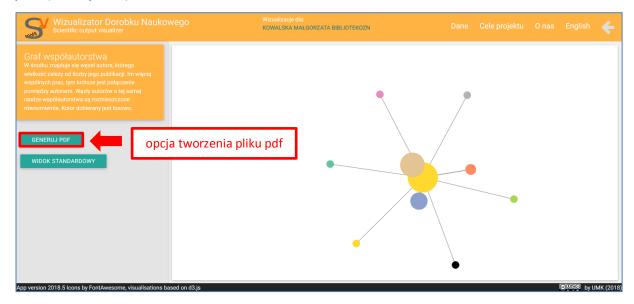


Rys. 14. Layout 2 (widok prac redagowanych)

Źródło: http://visualizeme.umk.pl/

Najechanie wskaźnikiem myszy komputerowej na pojedyncze węzły powoduje wyświetlenie nad nimi etykiet danych, tj. nazwisk i imion współautorów oraz sumarycznej liczby prac przygotowanych we współpracy z autorem, którego dorobek poddawany jest analizie (znajdującym się w centralnym punkcie grafu). Z kolei kliknięcie w dany węzeł lewym klawiszem myszy oraz dłuższe przytrzymanie go, daje możliwość przeciągnięcia go w inne miejsce grafu.

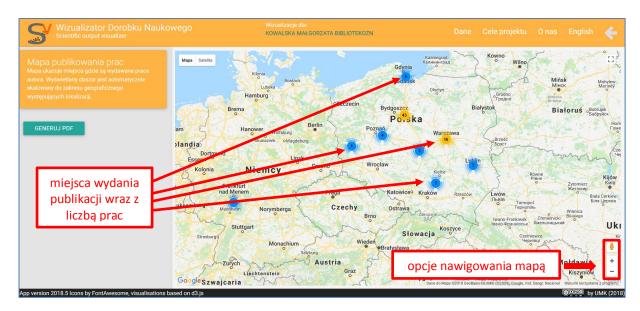
Efekty analizy danych uwidocznione na layout'cie 2 (zarówno w widoku standardowym, jak i widoku prac redagowanych) można wygenerować do pliku .pdf poprzez wybór opcji "Generuj pdf" (zob. rys. 15).



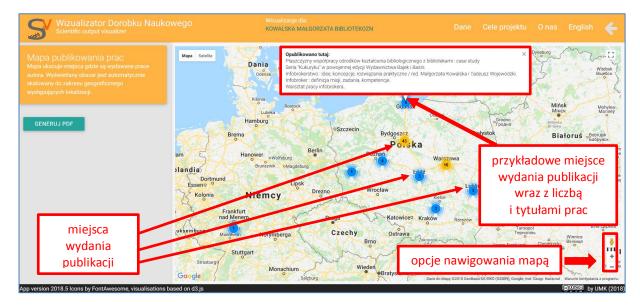
Rys. 15. Layout 2 (generowanie pliku pdf)

Layout 3 ma postać mapy. Przedstawia ona miejsca, w których wydawane były publikacje analizowanego autora. Wyświetlany obszar jest automatycznie skalowany do zakresu geograficznego występujących lokalizacji. Mapę można powiększać, pomniejszać oraz przesuwać (opcja "rączka").

Wyświetlające się na mapie cyfry i liczby wskazują liczbę wydanych w danym miejscu prac. Kolorem niebieskim zaznaczono miejsca, w których ukazało się od 1–9 publikacji, kolorem żółtym – powyżej 10 publikacji (zob. rys. 16). Najechanie wskaźnikiem myszy komputerowej na określoną liczbę oraz kliknięcie w nią powoduje wyświetlenie się tytułów prac opublikowanych w danym miejscu (zob. rys. 17).



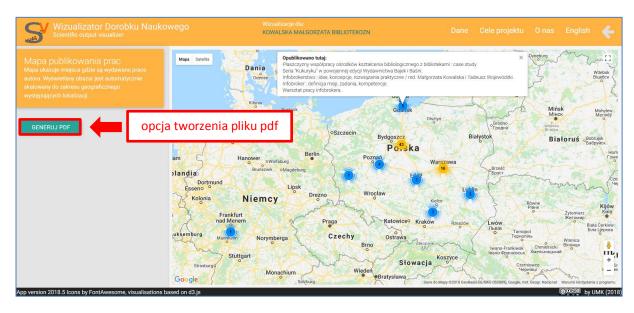
Rys. 16. Layout 3 (miejsca wydania publikacji wraz z ich liczbą)



Rys. 17. Layout 3 (miejsca wydania publikacji wraz z ich liczbą oraz przykładowe miejsce wydania prac z wymieniona ich liczbą i tytułami)

Źródło: http://visualizeme.umk.pl/

Efekty analizy danych uwidocznione na layout'cie 3 można wygenerować do pliku .pdf poprzez wybór opcji "Generuj pdf" (zob. rys. 18).



Rys. 18. Layout 3 (generowanie pliku pdf)

Layout 4 to mapa dziedzin w postaci wykresu pierścieniowego. Wskazuje ona na przynależność dziedzinową opublikowanych przez danego autora artykułów naukowych. Artykuły te rozmieszczane są kole według zasad geometrii z uwzględnieniem przynależności dziedzinowej czasopism, w których zostały one opublikowane. Aby tego typu zestawienie było możliwe wcześniej dokonano przypisania czasopism naukowych do poszczególnych dyscyplin i dziedzin nauki, korzystając z indeksów światowych baz Web of Science oraz polskiej bazy danych czasopism ARIANTA.

Koło dziedzin w aplikacji Scientific Visualizer podzielone jest segmenty, którym przypisano kolory według wzorca globalnej mapy nauki UCSD autorstwa K. Boyacka⁴ (dziedziny nauki są w nim podzielone na dyscypliny, z których każda ma określone położenie na kole)⁵. Poszczególnymi kolorami oznaczono tu (zob. rys. 19):

- kolor czerwony nauki medyczne i o zdrowiu,
- kolor jasno szary nauki humanistyczne,
- kolor seledynowy nauki przyrodnicze,
- kolor granatowy nauki społeczne,
- kolor ciemno szary nauki ścisłe,
- kolor różowy nauki techniczne,
- kolor zielony nauki rolnicze, leśne, weterynaryjne,
- kolor żółty sztukę.

Umieszczone wewnątrz pierścienia punkty symbolizują artykuły naukowe. Najechanie wskaźnikiem myszy komputerowej na określony punkt powoduje wyświetlenie informacji o obszarze nauki do jakiego należy publikacja oraz tytule czasopisma, na łamach którego została wydana. Znajdujące się w bezpośrednim sąsiedztwie czy też nakładające się na siebie punkty tego samego koloru wskazują z kolei, że prace reprezentowane przez te punkty zostały opublikowane w tym samym źródle (czasopiśmie).

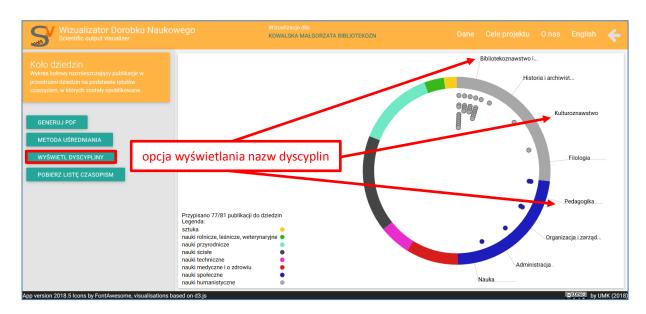
⁴ https://scitech-strategies.com/

⁵ Trwają prace nad dołączeniem wzorca do aplikacji.



Rys. 19. Layout 4 (obszary nauki) Źródło: http://visualizeme.umk.pl/

Layout 4, poza możliwością przyporządkowywania publikacji do obszarów nauki, umożliwia także wyświetlenie konkretnych dyscyplin, reprezentowanych przez te publikacje. Służy do tego opcja "Wyświetl dyscypliny" (zob. rys. 20). Jej wybór pociąga za sobą wyświetlenie nazw dyscyplin po zewnętrznej stronie koła dziedzin.



Rys. 20. Layout 4 (obszary nauki + etykiety zawierające nazwy dyscyplin) Źródło: http://visualizeme.umk.pl/

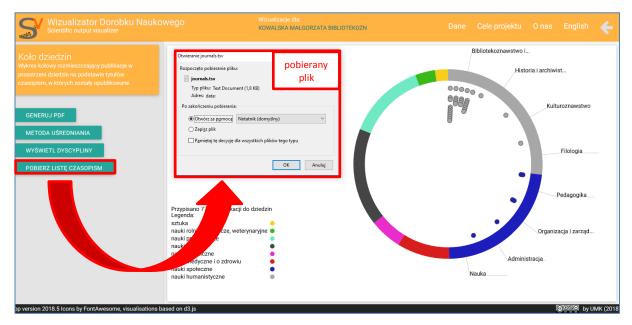
Na layout'cie 4 dodatkowych wyjaśnień wymagają dwie opcje: "Metoda uśredniania" oraz "Pobierz listę czasopism". Pierwsza z wymienionych pozwala doprecyzować aspekt danej publikacji. Jest to istotne zwłaszcza w przypadku czasopism należących jednocześnie do kilku obszarów nauki (np. nauk humanistycznych i społecznych). W takim przypadku metoda

uśredniania pozwala na wskazanie, który z obszarów nauki ma w danej publikacji charakter wiodący, a który – poboczny. Stwarza ona także możliwość wykrywania pokrewieństw czy też punktów stycznych między różnymi obszarami nauki (zob. rys. 21). W porównaniu do widoku standardowego (por. rys. 20) wyraźnie widać tu przesunięcie punktów (publikacji) do dziedzin/dyscyplin drugorzędnych.



Rys. 21. Layout 4 (metoda uśredniania) Źródło: http://visualizeme.umk.pl/

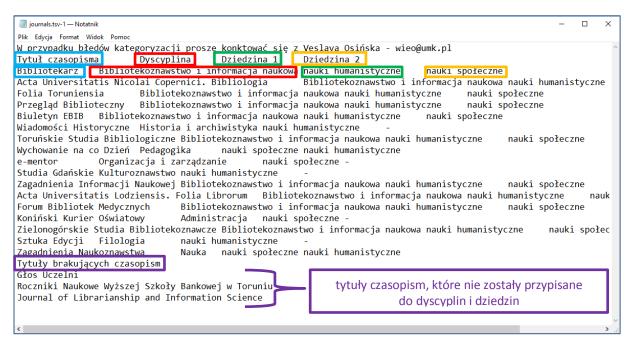
Druga z opcji – "Pobierz listę czasopism" umożliwia pobranie pliku zawierającego tytuły czasopism, które zostały zidentyfikowane w dorobku danego badacza (zob. rys. 22). Chociaż generowany automatycznie plik ma postać dokumentu tekstowego (.tsv) został przygotowany tak, by można było na nim przeprowadzać operacje także w programie do tworzenia arkuszy kalkulacyjnych (Excel). W tym celu plik ten należy zaimportować do takiego programu jako plik tekstowy.



Rys. 22. Layout 4 (pobieranie pliku z listą czasopism, w jakich publikował analizowany autor) Źródło: http://visualizeme.umk.pl/

Warto zauważyć, że w pliku z czasopismami wymienione zostały tytuły czasopism, w których publikuje dany autor. Obok tytułu czasopisma umieszczono nazwę dyscypliny przypisanej do danego periodyku oraz nazwę dziedziny, do której dane czasopismo należy. W przypadku przynależności do kilku dyscyplin i dziedzin zastosowano opis naprzemienny: dyscyplina 1/dziedzina 1; dyscyplina2/dziedzina 2 itd. Na końcu listy wymieniono tytuły tych czasopism, których nie udało się przypisać do określonych dyscyplin (m.in. ze względu na brak występowania danego tytułu w bazach danych, jego nienaukowy charakter, nieprawidłowości występujące w procesie wydawniczym – zmiana wydawcy, zawieszenie lub zmiana tytułu) należy (zob. rys. 23).

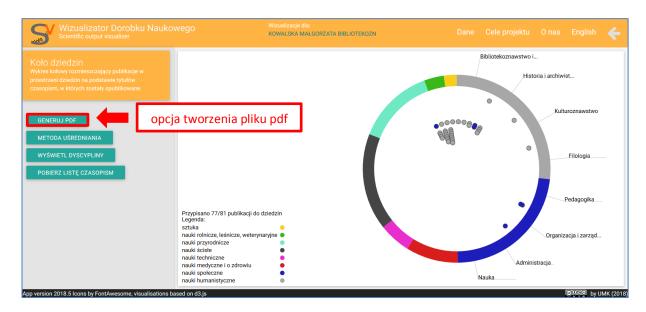
W przypadku dostrzeżenia braków w tym zakresie lub też stwierdzenia niezgodności danych dla pozostałych czasopism, Naukowcy dokonujący analizy własnego dorobku publikacyjnego, proszeni są o kontakt z **dr hab. Veslavą Osińską na adres wieo@umk.pl**. Ewentualne zgłoszenia dotyczące uzupełniania lub korekty danych mieć następującą postać: tytuł czasopisma, nazwa dyscypliny 1, dziedzina 1, nazwa dyscypliny 2, dziedzina 2...).



Rys. 23. Layout 4 (struktura pliku z czasopismami)

Źródło: http://visualizeme.umk.pl/

Efekty analizy danych uwidocznione na layout'cie 4 można wygenerować do pliku .pdf poprzez wybór opcji "Generuj pdf" (zob. rys. 24).

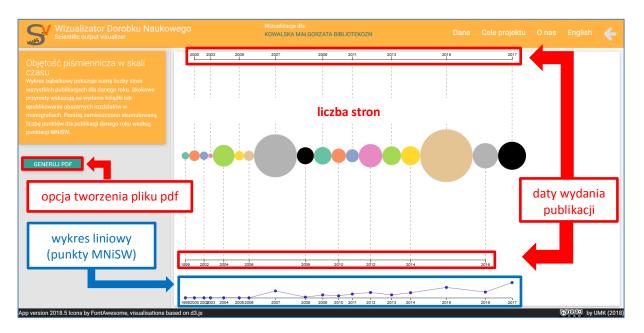


Rys. 24. Layout 4 (generowanie pliku pdf)

Layout 5 ma postać wykresu bąbelkowego. Ilustruje on sumaryczną liczbę stron wszystkich prac naukowych opublikowanych w danym roku kalendarzowym (zob. rys. 25). Im większe koło, tym większa liczba stron. Kolor kół dobierany jest losowo. Najechanie wskaźnikiem myszy komputerowej na pojedyncze koła widniejące na wykresie powoduje wyświetlenie nad nimi etykiet danych, tj. oznaczenia roku wraz z sumaryczną liczbą stron. Skokowe przyrosty wskazują na wydanie monografii lub opublikowanie obszernych rozdziałów w wydawnictwach zbiorowych.

Poniżej wykresu bąbelkowego, na wykresie liniowym ze znacznikami, skumulowano liczbę wszystkich punktów (MNiSW) uzyskanych za publikacje wydane w danym roku kalendarzowym. Najechanie wskaźnikiem myszy komputerowej na pojedyncze znaczniki widniejące na wykresie powoduje wyświetlenie nad nimi etykiet danych, tj. oznaczenia roku wraz z liczbą uzyskanych punktów.

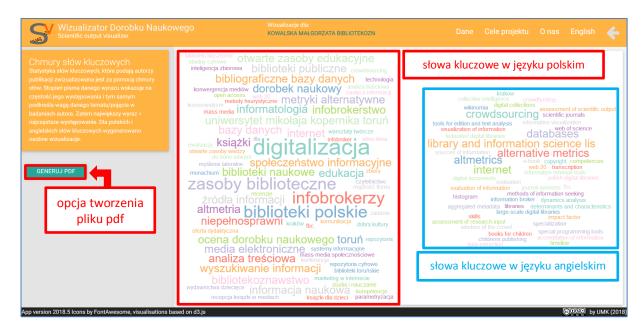
Efekty analizy danych uwidocznione na layout'cie 5 można wygenerować do pliku .pdf poprzez wybór opcji "Generuj pdf".



Rys. 25. Layout 5

Layout 6 ma postać dwóch chmur słów kluczowych: po stronie lewej znajdują się słowa kluczowe w języku polskim, po stronie prawej – w języku angielskim. Ten rodzaj wizualizacji służy zestawieniu lub porównaniu tematyki podejmowanej przez pojedynczego badacza na łamach krajowych i zagranicznych periodyków. Wielkość i pogrubienie pisma poszczególnych słów kluczowych zależne są od ważności lub częstotliwości występowania danego słowa w publikacjach analizowanego autora. Kolor czcionki poszczególnych terminów dobierany jest losowo (zob. rys. 26).

Efekty analizy danych uwidocznione na layout'cie 6 można wygenerować do pliku .pdf poprzez wybór opcji "Generuj pdf".



Rys. 26. Layout 6

Podsumowanie

Celem omówionej w niniejszym tutorialu aplikacji – wizualizatora dorobku naukowego – jest

wzbogacanie indywidualnej przestrzeni informacyjnej współczesnych badaczy, tj. ułatwianie

im oceny aktywności publikacyjnej, podnoszenie świadomości konieczności monitorowania

procesu jego rejestracji, pomaganie w odkrywaniu zależności między podejmowanymi

zagadnieniami badawczymi, inspirowanie do poszerzania horyzontów i sieci współpracy.

Wizualizator może pełnić także role narzędzia wspomagającego działalność administracyjną

(ocena pracowników, postępowania awansowe, akredytacja, powoływanie ekspertów,

rozdzielanie funduszy).

Zadaniem niniejszego opracowania jest pomoc użytkownikom w zakresie odczytywania

skomplikowanych i nietradycyjnych wykresów.

We wszystkich kwestiach związanych z działaniem aplikacji (m.in. wykrywanie błędów

kategoryzacji dziedzin nauki i czasopism oraz zgłaszanie sugestii wprowadzenia

funkcjonalnych usprawnień) należy kontaktować się z inicjatorką aplikacji:

dr hab. Veslava Osińska

Instytut Informacji Naukowej i Bibliologii

Uniwersytet Mikołaja Kopernika w Toruniu

ul. Bojarskiego 1, 87-100 Toruń

e-mail: wieo@umk.pl

25