

---

# ScientoMiner ICR – moduł importu danych bibliograficznych z zasobów Crossref dla platformy Gephi

Anna Małgorzata Kamińska

*Instytut Bibliotekoznawstwa i Informatyki Naukowej  
Uniwersytet Śląski w Katowicach*

---

## Abstrakt

**Cel/Teza:** Autorka stawia tezę, że coraz bardziej powszechne wykorzystywanie systemu identyfikacji DOI oraz postępujące zjawisko upubliczniania przez poszczególnych wydawców informacji o cytowaniach artykułów otwierają nowe możliwości prowadzenia analiz sieci cytowań bibliograficznych i zapewnić mogą wyższą niż dotąd wiarygodność uzyskiwanych wyników. Na szczególną uwagę zasługują tutaj katalogujące usługi Crossref, umożliwiające pozyskanie ustrukturalizowanych informacji bibliograficznych (w tym informacji o cytowaniach) wszystkim zainteresowanym stronom. Celem artykułu jest przedstawienie opracowanego przez autorkę rozszerzenia platformy analitycznej Gephi o moduł importu danych opisujących cytowania z zasobów Crossref. Rozwiązanie to ułatwia prowadzenie analiz cytowań pobieranych z tego źródła danych.

**Koncepcja/Metody badań:** Implementacja zaproponowanego rozszerzenia wymagała analizy architektury informacyjnej zasobów gromadzonych i indeksowanych w bazie danych Crossref oraz sposobu ich udostępniania poprzez usługi sieciowe dostawcy. Opracowany został również grafowy model danych umożliwiający jednoznaczny identyfikację artykułów cytujących i cytowanych oraz relacji je łączących, tak aby rozbudowa tego modelu w przyszłości (np. o relacje autorstwa) nie uniemożliwiała korzystania z opracowanego modułu.

**Wyniki i wnioski:** Udostępniona funkcjonalność umożliwia importowanie skierowanego grafu cytowań z zasobów Crossref do platformy Gephi, dla wybranych przez użytkownika czasopism opisywanych identyfikatorem ISSN, z możliwością ograniczenia do zadanych przedziałów czasowych. Dla mniej wprawnych użytkowników oferuje ona również wstępne pokolorowanie wierzchołków grafu według ich typu. Posługiwanie się identyfikatorami DOI uwiarygodnia wyniki prowadzonych analiz.

**Oryginalność/Wartość poznawcza:** Platforma obliczeniowa Gephi, stanowiąc ciekawe narzędzie do analizowania struktur sieciowych, bywa również wykorzystywana, głównie przez badaczy zagranicznych, do realizacji badań bibliometrycznych i wizualizacji ich wyników. Od niedawna obserwować można rosnący wzrost zainteresowania tym narzędziem również badaczy krajowych, jednak wydaje się, że ciągle przegrywa ono konkurencję z prostszą w obsłudze aplikacją Pajek oraz ściśle przeznaczonymi do analiz bibliometrycznych środowiskami VOSviewer oraz CiteSpace. Choć te ostatnie narzędzia mają zdecydowanie mniejsze możliwości analityczne, to za ich wyborem przemawiają techniczne bariery zasilania platformy Gephi danymi bibliograficznymi. Autorka, korzystając z modularnej budowy tej platformy, opracowała rozszerzenie umożliwiające automatyczny import danych wprost z zasobów Crossref w sposób bardziej przyjazny w porównaniu ze środowiskiem VOSviewer. Według jej najlepszej wiedzy jest to również pierwsze rozszerzenie tej platformy przeznaczone do badań stricte bibliometrycznych.

## Słowa kluczowe

Analiza cytowań. Bibliometria. Bibliograficzne bazy danych. Crossref. DOI. Gephi. Indeks cytowań. Import danych. Mapowanie nauki. ScientoMiner ICR. Wizualizacja danych bibliograficznych.

Otrzymany: 5 lutego 2018. Zrecenzowany: 20 marca 2018. Poprawiony: 14 kwietnia 2018. Zaakceptowany: 22 maja 2018.

## 1. Wprowadzenie

Bibliometria, będąc jednym z podstawowych obszarów badań naukometrycznych, od dłuższego czasu cieszy się niemalejącym zainteresowaniem, zarówno w zakresie doskonalenia i rozwijania analitycznych metod obliczeniowych, jak również technik identyfikacji, składowania i udostępniania danych bibliograficznych, stanowiących przedmiot prac badawczych. Jednak stosowanie nawet najbardziej zaawansowanych metod analitycznych w odniesieniu do danych z błędami, niepełnych lub przeciwnie – zwielokrotnionych – prowadzi najczęściej do uzyskiwania wyników o niskiej wiarygodności.

Celem niniejszego artykułu jest przedstawienie propozycji zastosowania opracowanego przez autorkę modułu rozszerzającego platformę analityczną Gephi (Bastian et al., 2009) o możliwość importu danych bibliograficznych z ogólnodostępnych zasobów Crossref, do prowadzenia badań bibliometrycznych w zakresie analizy cytowań na podstawie danych, których jakość gwarantuje stosowanie systemu identyfikacji DOI (Digital Object Identifier). Użycie proponowanego modułu ułatwi więc akwizycję danych z Crossref oraz pozwoli na uzyskanie wiarygodnych wyników analiz.

Przedstawienie możliwości samego modułu poprzedzone zostanie wykazaniem zasadności jego stosowania do prowadzenia analiz bibliometrycznych poprzez:

- wskazanie przyczyn wybranych problemów stojących na przeszkodzie prowadzeniu wiarygodnych badań bibliometrycznych,
- uzasadnienie zastosowania systemu DOI jako mechanizmu rozwiązującego wcześniej wskazane problemy,
- wskazanie zasobów organizacji Crossref, jako cennego, ogólnodostępnego źródła danych bibliograficznych, które wykorzystuje system identyfikacji DOI,
- przedstawienie sposobu dostępu do usług Crossref oraz wybranej części zakresu udostępnianych danych.

Warunkiem koniecznym do prowadzenia wiarygodnych badań bibliometrycznych w zakresie analizy cytowań pomiędzy pracami naukowymi jest możliwość prawidłowej identyfikacji prac opublikowanych oraz dysponowanie skuteczną metodą jednoznacznego powiązania danej publikacji z pozycją bibliografii załącznikowej tych prac, które ją cytują. Choć samo zagadnienie wydaje się być trywialne, to zarówno wielorakość form wydawniczych, jak i złożoność struktury rekordu bibliograficznego oraz mnogość tzw. stylów cytowań (tj. stosowanych konwencji tworzenia opisu bibliograficznego) wymaganych przez poszczególne wydawnictwa powodują spore trudności przy wiązaniu jednostek bibliograficznych z odpowiadającymi im pozycjami bibliografii załącznikowych poszczególnych publikacji. Praprzyczyny tych problemów dopatrywać się można w złożonej identyfikacji jednostek bibliograficznych rozumianej *sensu stricto*. Klucz/identyfikator jednostki bibliograficznej jest kluczem złożonym, składającym się ze znacznej liczby atrybutów, których wystąpienie zależy od formy wydawniczej, np. inny zestaw atrybutów będzie wchodził w skład klucza identyfikującego artykuł z czasopisma, rozdział z monografii czy całą monografię. Niekonsekwentne stosowanie przez autorów stylów opisów bibliograficznych, umieszczanie w bibliografiach załącznikowych opisów niepełnych, błędy popełniane przez

operatorów systemów bibliograficznych baz danych podczas rejestracji rekordów bibliograficznych oraz błędne założenia architektoniczne niektórych baz są już tylko przyczynami wtórnymi, z wielokrotnie zwiększającymi skalę występowania nieprawidłowości w zarejestrowanych danych. Więcej informacji na temat typologii błędów występujących w bibliograficznych bazach danych oraz o metodzie gromadzenia danych pozwalającej na ich unikanie znaleźć można w (Kamińska, 2017c), natomiast o metodach detekcji zwielokrotnionych rekordów bibliograficznych i ich deduplikacji, uwzględniających złożoność struktury rekordu opisu bibliograficznego w (Kamińska, 2017b).

Potrzeby wytworzenia mechanizmów stymulacji i ukierunkowywania rozwoju dziedzin naukowych w poszczególnych krajach spowodowały wzrost zainteresowania wskaźnikami bibliometrycznymi, które jednak, aby mogły być powszechnie akceptowane, powinny być wyliczane na podstawie globalnych i wiarygodnych źródeł. Budowanie tak dużych repozytoriów danych ze szczególną dbałością o wiarygodność gromadzonych informacji doprowadziło do komercjalizacji wielu takich przedsięwzięć, czego przykładem mogą być usługi Web of Science czy Scopus. Wzrost znaczenia analiz bibliometrycznych spowodował równocześnie dostrzeżenie wspomnianej wcześniej przyczyny problemów – złożonego charakteru identyfikatorów jednostek bibliograficznych. Potrzeba jednoznacznej identyfikacji tych jednostek, niezależnej od form wydawniczych, za pomocą pojedynczego atrybutu o globalnie unikalnej i niezmiennej w czasie wartości, doprowadziła ostatecznie do ugruntowania standardu DOI. Więcej o rozwoju metod identyfikacji prac naukowych znaleźć można w (Kamińska, 2017a).

Stosowanie systemu DOI, pośród wielu innych zalet, znakomicie upraszcza zagadnienia związane z opisem bibliograficznym i bibliografią załącznikową prac naukowych na każdym etapie jej przetwarzania – od sporządzania opisów przez autorów, poprzez ich weryfikację merytoryczną przez recenzentów, weryfikację syntaktyczną przez wydawców, po możliwość całkowicie automatycznego wiązania pozycji cytowanej z daną jednostką bibliograficzną na etapie zasilania bibliograficznych baz danych, z ewentualnym automatycznym uzupełnieniem wartości brakujących atrybutów na podstawie centralnej usługi katalogowej. Zakresy identyfikatorów (definiowane ich prefiksami) są udostępniane wydawcom poprzez specjalnie do tego powołane agencje (ang. DOI Registration Agency) w celu dalszego ich wykorzystania. Istnieje również możliwość bezpłatnego pozyskiwania identyfikatorów bezpośrednio przez autorów prac, co stanowi również istotny krok w kierunku upowszechniania koncepcji otwartej nauki. Prace oznakowane DOI nie muszą podlegać jakimkolwiek procedurom wydawniczym, recenzyjnym czy redakcyjnym, pozwalając na ostateczną ocenę pracy bezpośrednio społeczności naukowej<sup>1</sup>.

Choć organizacja sprawująca pieczę nad całym systemem DOI jest organizacją non profit (IDE, 2017a), to za pośrednictwem agencji rejestrowych pobierane są środki finansowe od wydawców celem pokrycia kosztów utrzymania całej infrastruktury. Warto zwrócić uwagę, że choć wspólnym mianownikiem systemu DOI jest centralna dystrybucja zakresów identyfikatorów oraz centralne repozytorium metadanych opisujących każdy z obiektów,

<sup>1</sup> Autorzy zamieszczający swoje prace bezpośrednio (tzn. bez uprzedniej formalnej publikacji w jakimkolwiek wydawnictwie) na naukowych platformach społecznościowych (np. ResearchGate.net) czy platformach udostępniania danych badawczych (np. Zenodo.org) mają możliwość automatycznego nadania identyfikatora publikowanym pracom całkowicie zgodnego z systemem DOI. Przykład takiej publikacji znaleźć można pod adresem: <https://doi.org/10.5281/zenodo.1042131>

którym identyfikator został przydzielony, to poszczególne agencje różnią się poziomem opłat, zasięgiem terytorialnym czy ofertą usług uzupełniających wspólną i zasadniczą część przedsięwzięcia. Kompletna lista agencji wraz z krótkim opisem oferowanych usług komplementarnych dostępna jest na stronie internetowej DOI (IDE, 2017b). Już pobieżna analiza tej listy wskazuje, że z punktu widzenia rozważań prowadzonych w niniejszym artykule największe znaczenie mają usługi Crossref, organizacji rejestrującej informacje o pracach naukowo-badawczych, publikowanych w czasopiśmie i wydawnictwach zwartych, informacje o bibliografiach załącznikowych oraz udostępniającej usługi przeszukiwania składnicy metadanych.

## 2. Usługi bibliograficzne Crossref

Usługi Crossref dają wydawcom możliwość przekazywania informacji o identyfikatorach nadanych poszczególnym publikacjom naukowym oraz wielu innych dodatkowych danych bibliograficznych, w tym i informacji o zawartościach bibliografii załącznikowych poszczególnych prac. W zależności od skali swojej działalności mogą skorzystać z różnych form dostarczania informacji – od prostych formularzy wypełnianych za pomocą przeglądarek internetowych, po całkowicie zautomatyzowane zasilanie z wykorzystaniem technologii usług sieciowych (ang. *web services*). Z przyczyn opisanych powyżej twórcy platformy aktywnie zachęcają wydawców do skrupulatnego rejestrowania informacji zawartych w bibliografiach załącznikowych, a nawet do rejestrowania wraz z poszczególnymi ich pozycjami numerów DOI (jeśli zostały wcześniej nadane). W tym celu Crossref oferuje usługi sieciowe usprawniające wyszukiwanie identyfikatora DOI na podstawie informacji bibliograficznych. Dane zgromadzone w ten sposób w zasobach Crossref są następnie udostępniane wszystkim zainteresowanym.

Trzeba jednak zauważyć, że chociaż podstawowe informacje o zarejestrowanych rekordach bibliograficznych są dostępne publicznie na każde żądanie zainteresowanego, to w przypadku informacji rozszerzonych o bibliografię załącznikową domyślnie są one zastrzeżone, a ich upublicznienie następuje na wniosek konkretnego wydawcy. Obecnie w dużej mierze dzięki staraniom I4OC (Initiative for Open Citations) – społeczności naukowców, wydawców i innych zainteresowanych, lobbujących na rzecz nieograniczonej dostępności danych o cytowaniach – publicznie dostępne są już dane wielu wydawców, w tym American Physical Society, Cambridge University Press, PLOS, SAGE, Springer, Nature oraz Wiley<sup>2</sup>. Publicznie dostępne dane stanowią obecnie połowę liczby rekordów o cytowaniach zgromadzonych w zasobach Crossref. Zgody na publiczne udostępnienie dostarczanych danych nie wyraziła jeszcze nieliczna grupa dużych wydawców, takich jak American Chemical Society, Elsevier, IEEE czy Wolters Kluwer Health, z czego czasopisma Elseviera pokrywają 65% wszystkich niedostępnych rekordów. Włączenie się tylko tego wydawcy do nurtu otwartych danych o cytowaniach spowodowałoby dostępność tych danych na poziomie prawie 83%. Powyższe dane zostały tu przytoczone za listem otwartym *Open citations: A letter from the scientometric community to scholarly publishers* (ISSI, 2017), który środowisko naukometryczne wystosowało do wydawców naukowych 5 grudnia 2017 r.,

<sup>2</sup> Pełna lista wydawców dostępna jest pod adresem: <https://i4oc.org/#publishers>

próbując przekonać ich do podjęcia decyzji o upublicznieniu dostarczanych przez nich danych. Uaktualniane listy wydawców z upublicznonymi danymi o cytowaniach oraz wydawców zabraniających dostępu do tych danych znaleźć można na platformie Crossref<sup>3,4</sup>.

Dane bibliograficzne udostępniane są z poziomu platformy Crossref za pomocą usług sieciowych przekazujących komunikaty o zakresie zdefiniowanym we wcześniej sformułowanym zapytaniu i zawierające dane bibliograficzne w formacie JSON (JavaScript Object Notation). Format ten zarówno może być poddany parsowaniu maszynowemu, jak i jest zrozumiały dla człowieka nawet w oryginalnej i nieprzetworzonej postaci. Jednym z parametrów wyszukiwania może być numer ISSN, który odnaleźć można za pomocą wyszukiwarki udostępnionej na platformie (Rys. 1).

← Back to the main Crossref website

**Crossref**

title list

A B C D E F G H I J K L M  
N O P Q R S T U V W X Y Z

SELECT A LETTER TO BROWSE THE LIST OF TITLES THAT BEGIN WITH THAT LETTER.

Genre:  all  Journals  books  conf proc

Search:  title  issn/isbn  subject  publisher

chemistry Search Clear

This page provides an alphabetically indexed list of journals for which CrossRef has article metadata. By clicking on the first letter of a title you will be presented with the portion of the list of journals that begin with that letter. Along with the journal title, the ISSN (for both the print and electronic versions of the journal, if available) and the start year are included.

Please note that CrossRef does not host the full text of any journals. CrossRef enables linking to and from the journals included

NOTE: Click the ▼ icon to see the Title coverage information for journals. Click the ▼ icon to see additional title information. Click the » icon to request a missed-conflict report.

A full comma-separated journal coverage list is available here (caution: this file is large ~5MB).

your search for [chemistry] returned 750 matches.  
Pages: 1 2 3 4 5 6 7 8 9 10 [ > ] [ >> ]

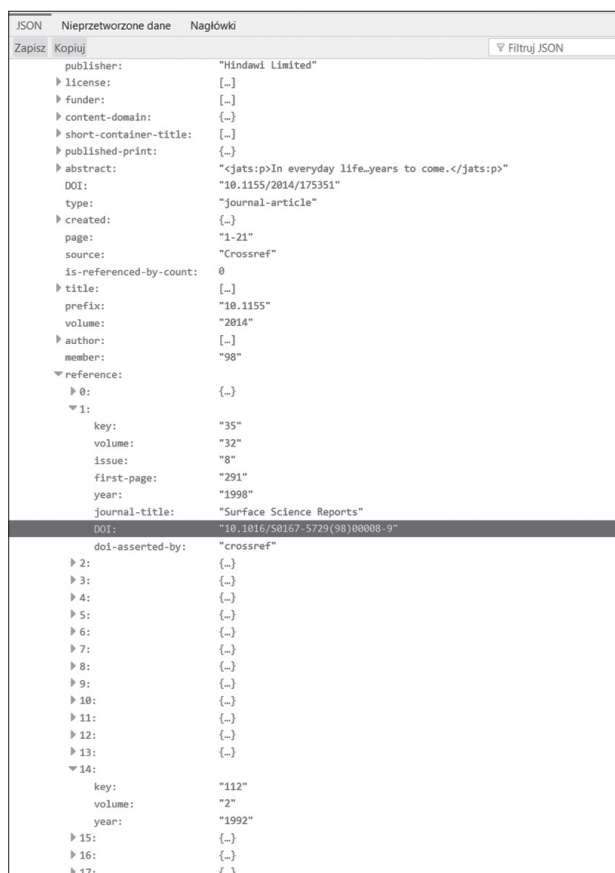
title ( journal / book / conf proc )	publisher	print issn / isbn	electronic issn / isbn	DOI
▼▼ ACS Earth and Space <b>Chemistry</b> ⓘ	American Chemical Society	2472-3452	2472-3452	»
▼ ACS Medicinal <b>Chemistry Letters</b> ⓘ	American Chemical Society	1948-5875	1948-5875	»
▼ ACS Sustainable <b>Chemistry &amp; Engineering</b> ⓘ	American Chemical Society	2168-0485	2168-0485	»
▼ Acta Crystallographica Section C Structural <b>Chemistry</b> ⓘ	International Union of Crystallography	2053-2296	2053-2296	»

Rys. 1. Okno wyszukiwania informacji o czasopismach i wydawnictwach zwartych dostępnych na platformie Crossref

<sup>3</sup> Lista podmiotów wyrażających zgodę na upublicznienie informacji o cytowaniach dostępna jest pod adresem: <https://www.crossref.org/reports/members-with-open-references/>

<sup>4</sup> Lista podmiotów, które nie wyraziły zgody na upublicznienie informacji o cytowaniach dostępna jest pod adresem: <https://www.crossref.org/reports/members-with-closed-references/>

Szczegółowe omówienie poszczególnych komunikatów w formacie JSON, stanowiących odpowiedzi na żądania dostępu do danych, zawiera dokumentacja<sup>5</sup>. Wynika z niej, że każdy rekord opisujący jednostkę bibliograficzną jest zawsze identyfikowany przez DOI, natomiast w przypadku rekordów opisujących bibliografię załącznikową DOI jest atrybutem opcjonalnym. Choć zakres informacyjny opisów bibliograficznych może być bardzo szeroki, to wiele z atrybutów jest nieobligatoryjnych, co powoduje, że w praktyce wydawcy mogą przekazywać dane w bardzo różnych zakresach (zob. przykład z Rys. 2). Warto zwrócić uwagę, że w przypadku rekordów opisu bibliografii załącznikowych jedynym wymaganym atrybutem jest identyfikator – dokumentacja natomiast nic nie wspomina na temat jego pochodzenia oraz możliwej interpretacji.



Rys. 2. Fragment przykładowej zawartości komunikatu zwrotnego z usługi sieciowej Crossref dla zapytania o artykuły z roku 2014 z czasopisma identyfikowanego ISSN o wartości 2356–6612 (komenda wpisana w pasek adresu przeglądarki: <http://api.crossref.org/works?filter=issn:2356-6612,from-pub-date:2014-01-01,until-pub-date:2014-12-31&rows=1000>)

<sup>5</sup> Dokumentacja udostępniona jest pod adresem: [https://github.com/CrossRef/rest-api-doc/blob/master/api\\_format.md](https://github.com/CrossRef/rest-api-doc/blob/master/api_format.md)

Rysunek 2 przedstawia przykład konkretnego artykułu z bibliografią załącznikową, gdzie dla jej pozycji oznaczonej numerem 1 zarejestrowano artykuł z czasopisma identyfikowany poprzez DOI, natomiast dla pozycji oznaczonej na rysunku numerem 14 przekazany zestaw atrybutów jest zbyt wąski na podjęcie jakichkolwiek prób identyfikacji tej pracy naukowej.

W wytycznych dla wydawców<sup>6</sup>, dotyczących sposobu konstruowania informacji, opisujących poszczególne rekordy bibliografii załącznikowej, przedstawiono trzy podejścia:

- identyfikację poprzez DOI zawsze, gdy tylko jest to możliwe,
- „rozbiór strukturalny” opisu i przekazanie informacji w postaci atomowych metadanych opisujących poszczególne atrybuty zależne od formy wydawniczej (np. nazwę czasopisma, autorów, tytuł artykułu itp.),
- przekazanie oryginalnego opisu bez jego interpretacji jako pojedynczej metadanej klasy „unstructured\_citation” (czyli w postaci pojedynczego łańcucha znakowego), pozostawiając próby identyfikacji źródłowej jednostki bibliograficznej mechanizmom serwisu.

Doświadczenia autorki w pracy z danymi pozyskanymi z platformy Crossref wskazują, że w praktyce jakość przekazywanych informacji o bibliografiach załącznikowych bywa różna, a w bardzo rzadkich przypadkach ogranicza się jedynie do przekazania atrybutu identyfikatora. Podczas realizacji badań nad omówionym dalej modułem ScientoMiner zaobserwowano jednak poprawę jakości tych danych w nowszych rekordach (wprowadzonych do zasobów Crossref w ostatnich kilku latach). Dodatkowo, ciągłe upowszechnianie się korzystania przez wydawców z DOI<sup>7</sup> pozwala spoglądać z optymizmem na wiarygodność wyników analiz bibliometrycznych, które będą prowadzone na podstawie zasobów tego coraz bardziej kompletnego źródła danych.

### 3. Moduł ScientoMiner ICR

Prowadzenie badań bibliometrycznych realizowane jest z wykorzystaniem różnych platform i narzędzi analitycznych, z czego do badań ilościowych cytowań oraz wizualizacji powiązań bibliograficznych wybierane są najczęściej narzędzia umożliwiające operowanie na strukturach grafowych. Spośród tych ostatnich największą popularnością cieszą się: Pajek, VOSviewer, CiteSpace, CytoScape, CitNetExplorer, HistCite, Sci2, Gephi (Van Eck & Waltman, 2014).

Pierwsza z wymienionych aplikacji, choć równocześnie najstarsza, nadal jest chętnie używana. Popularność zyskała dzięki prostocie obsługi oraz przejrzystemu formatowi danych, jakkolwiek przez to ograniczającemu zakres danych, które mogą podlegać analizom. Dwie kolejne aplikacje specjalizują się w analizie danych pochodzących z usługi WoS, jednak autorzy aplikacji VOSviewer, dostrzegając rodzący się potencjał nowego bibliograficznego źródła informacji, w jej najnowszej wersji (1.6.6), wydanej 23 października 2017 r., dodali możliwość analizowania danych pochodzących z platformy Crossref<sup>8</sup>. Aplikacja umożliwia

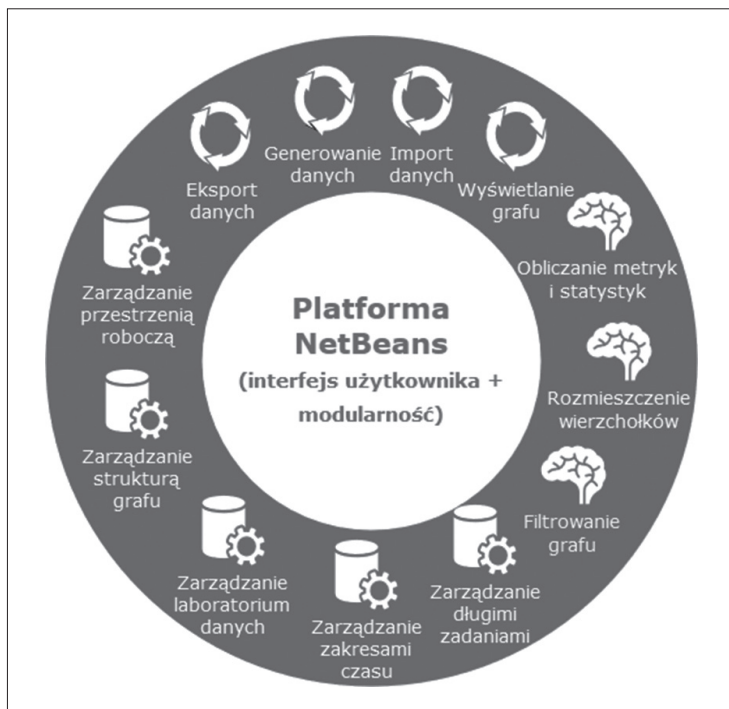
<sup>6</sup> Zob. <https://support.crossref.org/hc/en-us/articles/215578403>

<sup>7</sup> Zob. <https://www.crossref.org/pdfs/annual-report-2016.pdf>

<sup>8</sup> Szczegółowy opis tej funkcjonalności znaleźć można pod adresem: <https://www.cwts.nl/blog?article=n-r2r294>

wczytywanie plików w formacie JSON (w przypadku obszernych wolumenów danych jest to wiele plików zawierających po 1000 rekordów), wyeksportowanych uprzednio z tego serwisu za pomocą przeglądarki internetowej.

Platforma Gephi (<https://gephi.org/>), której dotyczy dalsza część artykułu, nie powstała z myślą o zastosowaniach *stricte* bibliometrycznych. Stanowi ona narzędzie o szerszych możliwościach, wykorzystywane do analizy i wizualizacji struktur sieciowych, reprezentujących sieci społecznościowe, biologiczne, semantyczne i inne. Posiada ona otwartą architekturę, umożliwiającą tworzenie modułów rozszerzających jej podstawowe funkcjonalności. Zbudowana jest na platformie NetBeans, która, oferując funkcjonalności bogatego szkieletu interfejsu graficznego użytkownika oraz zarządzania modułami implementującymi poszczególne funkcjonalności, umożliwia łatwe zapewnienie spójność wyglądu, zachowania i instalacji rozszerzeń aplikacji ją wykorzystujących. Na platformie tej wytworzono już wiele programów<sup>9</sup> charakteryzujących się przyjaznym interfejsem użytkownika, a aplikacja Gephi stanowi jeden z ich ciekawszych przykładów, gdyż od samego początku jej powstawania projektowana była jako łatwo rozszerzalna. Możliwość jej rozszerzania w kierunku poszczególnych funkcjonalności przedstawia rysunek 3.



Rys. 3. Ogólna architektura funkcjonalna platformy Gephi

<sup>9</sup> Najciekawsze z aplikacji zbudowanych z wykorzystaniem platformy Netbeans wraz z krótkim omówieniem ich funkcjonalności znaleźć można pod adresem: <https://netbeans.org/features/platform/showcase.html>



Omówienie każdego z zakresów funkcjonalnych Gephi wykracza poza tematykę niniejszego artykułu, jednak warto zwrócić uwagę, że każda z funkcjonalności dostępnych w aplikacji Gephi zaraz po jej zainstalowaniu zaimplementowana jest również w postaci modułów, które można odinstalować. Twórca platformy nie stawia więc siebie na uprzywilejowanej pozycji, ale przede wszystkim dba o implementację dobrze zdefiniowanych interfejsów, z których sam korzysta i które udostępnia innym zainteresowanym, aby wykorzystali je do opracowywania kolejnych rozszerzeń.

Autorka, zainspirowana wcześniej wspomnianymi działaniami na rzecz otwartej nauki, a w szczególności nowymi możliwościami ewaluacji nauki w zakresie analiz bibliometrycznych, opracowała własny moduł ScientoMiner<sup>10</sup> Import Crossref (ScientoMiner ICR), służący do importu danych bibliograficznych opisujących zawartość czasopism, których wydawcy deponują dane na platformie Crossref i wyrażają zgodę na upublicznienie danych opisujących bibliografie załącznikowe. Według najlepszej wiedzy autorki jest to równocześnie pierwsze rozszerzenie platformy Gephi profilujące jej wykorzystanie w kierunku analiz bibliometrycznych.

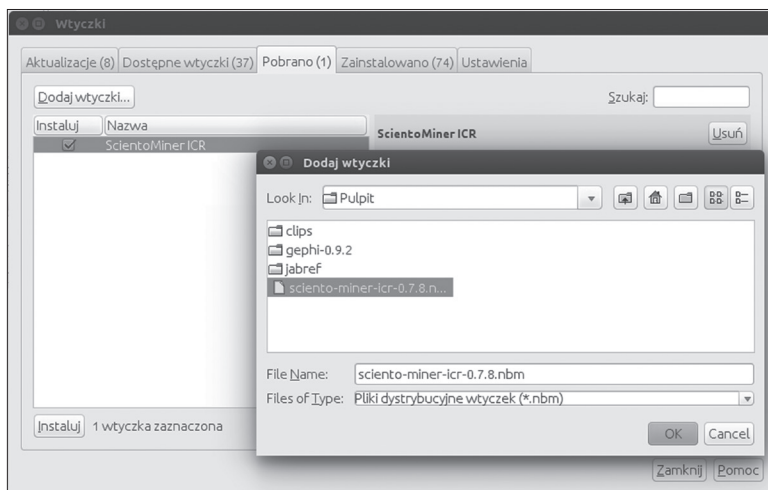
Moduł ScientoMiner ICR wykorzystuje następujące interfejsy funkcjonalne, udostępnione z poziomu aplikacji Gephi:

- zarządzanie strukturą grafu – w celu stworzenia grafowej struktury danych, gdzie wierzchołki wraz z ich atrybutami odzwierciedlają informacje bibliograficzne o publikacjach pobranych z platformy Crossref, a krawędzie o łączących je relacjach cytowań; dodatkowo w celu wstępnego rozmieszczenia poszczególnych węzłów na widoku obszaru roboczego aplikacji (Overview) i opcjonalnego wstępnego nadania im kolorów,
- zarządzania długimi zadaniami – w celu uruchamiania procesu pobierania danych w trybie asynchronicznym z pełnym zachowaniem współbieżności z innymi procesami, w tym z procesami odpowiedzialnymi za obsługę graficznego interfejsu użytkownika; dzięki temu użytkownik nie ma wrażenia „zawieszenia się aplikacji” podczas pobierania danych, a działanie procesu jest sygnalizowane przez animację paska postępu aplikacji Gephi (wspólnego dla wszystkich zadań), zaś sam proces może zostać w dowolnym momencie przerwany przez użytkownika.

Proponowany moduł paradoksalnie nie wykorzystuje natomiast funkcjonalności importu danych, która w środowisku Gephi jest zaprojektowana tak, że wszystkie importowane dane trafiają najpierw do specjalnego bufora, następnie aplikacja przedstawia ich charakterystykę (liczba wierzchołków, krawędzi, ewentualne napotkane problemy, itp.) i ostatecznie użytkownik decyduje czy dane z bufora mają zostać zaimportowane jako nowy graf w nowej przestrzeni roboczej. Mimo że tak realizowana funkcja importu ma wiele zalet, to jej wykorzystanie w module ScientoMiner ICR uniemożliwiłoby rozbudowywanie bieżącego grafu o kolejne partie informacji opisane nowymi przedziałami dat czy identyfikatorami ISSN.

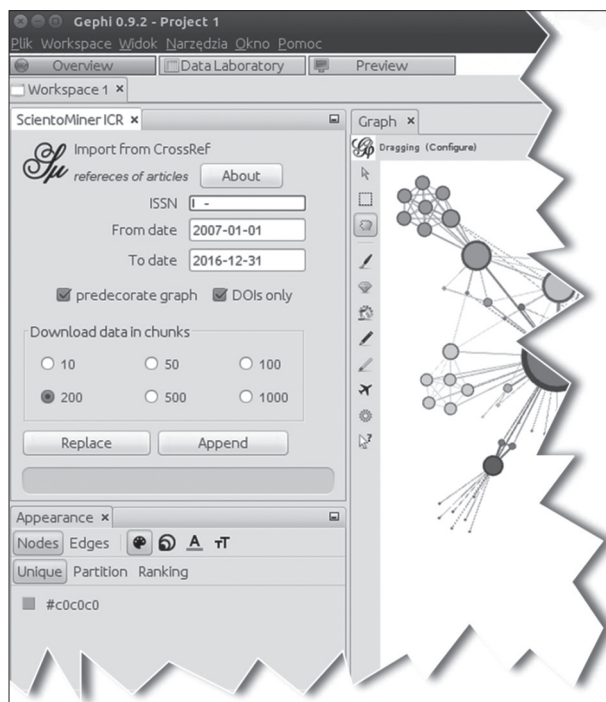
Moduł ScientoMiner ICR dostępny jest dla najnowszej wersji platformy o oznaczeniu 0.9.2, a jego instalacja w środowisku analitycznym sprowadza się do pobrania pojedynczego pliku i dodania go jako modułu w sposób zaprezentowany na rysunku 4 (czyli z wykorzystaniem funkcjonalności instalacji modułu dostarczonego jako plik o rozszerzeniu .nbm).

<sup>10</sup> Aktualna wersja modułu dostępna jest pod adresem: <https://doi.org/10.5281/zenodo.1215008>

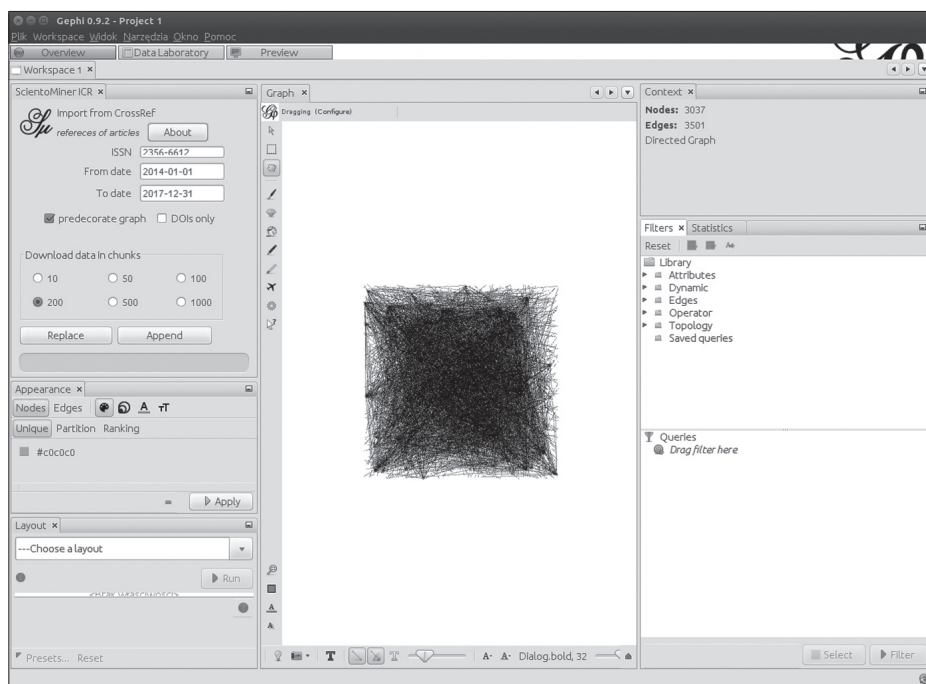


Rys. 4. Okno platformy Gephi umożliwiające dodawanie modułów z plików zewnętrznych

Po prawidłowym zainstalowaniu modułu, w oknie roboczym aplikacji pojawia się panel przedstawiony na rysunku 5, implementujący funkcjonalność importowania danych bezpośrednio z platformy CrossRef.



Rys. 5. Panel modułu ScientoMiner ICR



Rys. 6. Roboczy widok zaimportowanego grafu

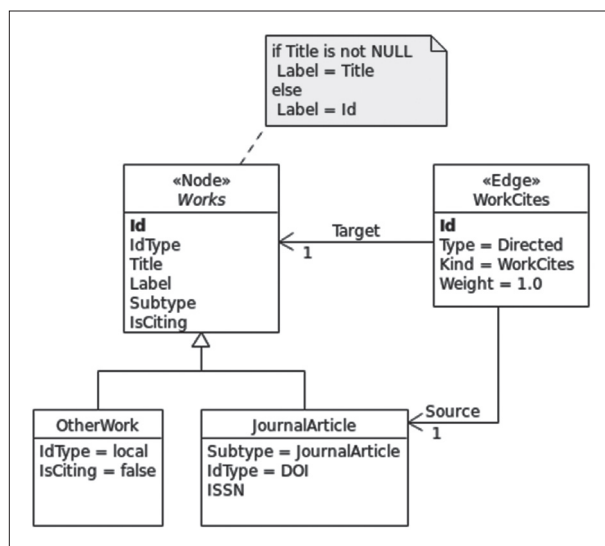
Zakres importowanych danych sterowany jest ustaleniem identyfikatora ISSN (określonego na przykład z wykorzystaniem usługi przedstawionej na rysunku 1) oraz przedziałem dat definiujących ramy czasowe dla analizowanych informacji. Użytkownicy mniej doświadczeni w korzystaniu z platformy Gephi skorzystać mogą z opcji „predecorate graph”, ustawiającej automatycznie różne barwy i rozmiary wierzchołków reprezentujących jednostki cytujące i cytowane. Atrybuty te mogą zostać później dowolnie zmodyfikowane przez użytkownika. Domyślnie włączona jest również opcja „DOIs only”, ograniczająca importowane rekordy do takich, które są identyfikowane poprzez DOI. Odznaczenie tej opcji powinno być stosowane z pewną ostrożnością, po upewnieniu się, że wydawcy danego czasopisma dbają o unikalną identyfikację poszczególnych pozycji bibliografii załącznikowych w ramach przekazywanych zbiorów danych. Zdarzają się bowiem przypadki, kiedy różnym jednostkom bibliograficznym wydawcy nadają identyfikatory tej samej wartości. Opcja „Download data in chunks” pozwala na zdefiniowanie rozmiaru jednorazowych komunikatów odbieranych z usługi Crossref. Mniejsze wartości skutkują dłuższym czasem importu danych, natomiast większe, w przypadku łączy internetowych o małych przepustowościach czy środowisk analitycznych o niewielkiej mocy obliczeniowej, mogą powodować błędy komunikacji z usługą Crossref. Rozpoczęcie procesu importu danych możliwe jest za pomocą przycisku „Replace”, co spowoduje usunięcie istniejących wcześniej danych z bieżącego obszaru roboczego lub przycisku „Append”, co spowoduje rozszerzenie grafu istniejącego w bieżącym obszarze roboczym o nowe dane zaimportowane z usługi Crossref – opcja ta jest w szczególności przydatna w razie potrzeby analizy wspólnych danych pochodzących z różnych czasopism. Warto tutaj zwrócić uwagę, że

graf cytowań jest grafem skierowanym, tak więc import danych możliwy będzie tylko dla projektów bazujących na grafie skierowanym bądź mieszanym – więcej informacji na temat wykorzystania struktur grafowych do analiz bibliometrycznych znaleźć można w innym artykule autorki (Kamińska, 2018b).

Po zaimportowaniu danych pojawi się graf z wierzchołkami rozmieszczonymi losowo na powierzchni kwadratu (Rys. 6), którego dalsze analizowanie ograniczone jest jedynie wyobraźnią badacza<sup>11</sup>.

Praktyczne przykłady analiz realizowanych w środowisku obliczeniowym Gephi na podstawie struktur grafowych znaleźć można w (Kamińska, 2018a) oraz (Kamińska, 2017d).

Możliwości analizowania danych zależą nie tylko od funkcjonalności konkretnych narzędzi (oferujących np. mechanizmy wyliczania różnorodnych miar czy operacje filtrowania, dzielenia i łączenia struktur grafowych), ale również od wyjściowego modelu danych, na których narzędzia te pozwalają operować. Platforma Gephi jako narzędzie uniwersalne, oferuje dużą elastyczność w zakresie implementowania własnych modeli danych, sprzyjających konkretnym profilom analiz. Moduł ScientoMiner ICR importuje dane zgodnie z opracowanym przez autorkę modelem, przedstawionym przy pomocy diagramu klas języka UML (Rys. 7).



Rys. 7. Model grafu zaimplementowany w module ScientoMiner IRC

<sup>11</sup> W najnowszej wersji platforma Gephi implementuje blisko 30 różnych algorytmów rozmieszczania wierzchołków, z których większość posiada wiele parametrów modyfikujących sposób ich działania. Przy opracowaniu danej wizualizacji można korzystać jednocześnie z kilku z tych algorytmów i to w różnej kolejności. Dodatkowo istnieje możliwość wpływu na sposób prezentacji poszczególnych wierzchołków (rozmiary, barwy i inne), ich etykiet oraz łączących je krawędzi. Otwiera to bardzo szerokie możliwości w tworzeniu wymownych i przekonujących wizualizacji naukometrycznych, których opracowywanie wymaga jednak znajomości funkcjonowania poszczególnych algorytmów, ich przeznaczenia i doświadczenia samego badacza.

Platforma Gephi przechowuje informacje o analizowanych strukturach w układzie relacyjnym, dzieląc je na informacje dotyczące wierzchołków (ang. *Node*) oraz krawędzi (ang. *Edge*). Wierzchołkami opracowanego modelu są prace/artykuly (ang. *Works*) zawierające informacje o identyfikatorze, typie identyfikatora, tytule, etykiecie (ang. *Label* – atrybut szczególnego przeznaczenia w systemie Gephi, służący merytorycznemu, ale skrótowemu opisowi wierzchołka; moduł importu zapewnia przypisanie mu wartości tytułu jeśli informacje o tytule dostępne są w zasobach Crossref, a w przeciwnym razie przypisuje mu wartość identyfikatora), podtypie oraz atrybucie mówiącym czy dana praca jest pracą cytującą inne prace (nie wyklucza to możliwości, że jest ona jednocześnie pracą cytowaną). Warto zwrócić uwagę, że prace/artykuly są klasą abstrakcyjną, tj. taką, która nie może wystąpić w systemie bez dalszego doprecyzowania, co konkretnie reprezentuje dany wierzchołek. Ponieważ funkcjonalność modułu obejmuje import zbioru prac wydanych w ramach czasopism określonych identyfikatorem ISSN, a usługi Crossref identyfikują te prace za pomocą DOI, jedną z klas potomnych jest artykuł z czasopisma (ang. *JournalArticle*), który zawsze posiada określoną wartość ISSN, a typem jego identyfikatora zawsze jest DOI. Drugą klasą potomną jest natomiast inna praca naukowa (ang. *OtherWork*), która jest identyfikowana lokalnym systemem identyfikacji (w ramach danego wydawcy) i która nie może być pracą cytującą (w subiektywnym rozumieniu importowego z Crossref zbioru danych).

Krawędzie importowane przez moduł ScientoMiner ICR opisane są typem cytowania pracy (ang. *WorkCites*) i są to krawędzie skierowane o początkowo jednostkowo ustalonej wartości (o sposobach prowadzenia analiz, gdzie każdej krawędzi może być przypisana inna waga przeczytać można np. w (Kamińska, 2018b)). Krawędzie są opisane również identyfikatorem wierzchołka źródłowego (ang. *Source*) oraz identyfikatorem wierzchołka docelowego (ang. *Target*). Jak wynika z modelu, wierzchołkiem źródłowym (czyli cytującym) może być jedynie artykuł z czasopisma (w subiektywnym rozumieniu importowanego z Crossref zbioru danych), natomiast wierzchołkiem docelowym (czyli cytowanym) może być zarówno artykuł z czasopisma, jak i inna praca naukowa.

Warto zauważyć, że tak zdefiniowany model określa grafy unimodalne (czyli o wierzchołkach jednakowego typu, którymi są tutaj prace/artykuly), co umożliwia obliczanie dla nich miar implementowanych przez platformę Gephi i bezpośrednią interpretację uzyskanych tą drogą wyników. W tej wersji modułu celowo zaniechano importu danych o autorach publikacji, gdyż skutkowałoby to koniecznością budowy grafu bimodalnego z dodatkowym typem wierzchołków reprezentujących autorów. Do wyników analiz przeprowadzonych dla sieci wielomodalnych należy podchodzić z ostrożnością<sup>12</sup>, gdyż przykładowo dla grafu skonstruowanego zgodnie z modelem przedstawionym na rysunku 7 wartość stopnia wierzchołka reprezentującego publikację cytowaną przedstawiać będzie jej liczbę cytowań, ta sama miara wyliczona dla wspomnianego wyżej grafu bimodalnego przedstawiałaby sumę liczby cytowań i liczby autorów, której możliwości interpretacji wydają się wątpliwe. Warto też zwrócić uwagę na łatwą rozszerzalność zaproponowanego modelu o kolejne rodzaje prac naukowych czy ich systemów identyfikacji, gdyż model opisuje strukturę danych importowanych przez moduł ScientoMiner ICR, zaś nie ogranicza danych, np. importowanych z innych źródeł do tej struktury. Umożliwia to, na przykład, implementację funkcji importu danych z innych źródeł, gdzie prace cytujące nie są ograniczone jedynie do artykułów z czasopism,

<sup>12</sup> Szerzej o powodach tej ostrożności zob. np. w Kamińska, 2018b.

a mimo to źródło pochodzenia poszczególnych prac jest jednoznacznie określone i wiedza ta może zostać wykorzystana do analiz o wybranych przekrojach.

Moduł ScientoMiner ICR jest gotowym i przetestowanym produktem. Jego testy obejmowały weryfikację wydajności, czyli w tym przypadku narzutu czasowego na zaimportowanie danych z zewnętrznego źródła oraz weryfikację konsumpcji pamięci operacyjnej, czyli jak duże wolumeny danych możliwe są do zaimportowania. Wnioski z obserwacji funkcjonowania modułu pokryły się z przewidywaniami: czas wczytywania danych praktycznie zależy jedynie od szybkości łącza internetowego, zaś rozmiar danych – od wbudowanych mechanizmów platformy, która stosuje własną pamięciową reprezentację grafów oraz dostępności pamięci operacyjnej dla środowiska wirtualnej maszyny Javy, na którym uruchamiana jest cała aplikacja.

#### 4. Przykład zastosowania

W celu zaprezentowania funkcjonalności oferowanej przez proponowany moduł posłużono się przykładem analizy artykułów z dziedziny górnictwa. Przykład ten opracowano jedynie w celach poglądowych, bez zamiaru wyciągania na podstawie uzyskanych wyników jakichkolwiek dalszych wniosków o charakterze ocen naukometrycznych.

Zbadano zbiór artykułów opublikowanych w trzech czasopismach w latach 2014–2017:

- *International Journal of Coal Science & Technology* – ISSN: 2095–8293, wydawane przez SpringerOpen<sup>13</sup>,
- *International Journal of Mining, Reclamation and Environment* – ISSN: 1748–0930, wydawane przez Taylor & Francis<sup>14</sup>,
- *Archives of Mining Sciences* – ISSN: 1689–0469 wydawane przez Polską Akademię Nauk w ramach De Gruyter Open<sup>15</sup>.

W module ScientoMiner ICR ustawiono przedziały lat i przeprowadzono trzykrotny import z zasobów Crossref ze wskazaniem kolejnych ISSN i z użyciem opcji Append, co spowodowało umieszczenie na jednym grafie danych o artykułach pochodzących ze wszystkich trzech czasopism. Po zastosowaniu jednego z algorytmów automatycznego rozmieszczania wierzchołków grafu oraz odrzuceniu wierzchołków o znikomej liczbie cytowań (odrzucono wierzchołki cytowane mniej niż dwa razy, pozostawiając jednak wszystkie wierzchołki cytujące choć jedną publikację) uzyskano strukturę przedstawioną na rysunku 8.

Pod adresem wskazanym w źródle rysunku 8 wyświetlić można jego kolorowy obraz wyjściowy. Ciemniejsze krawędzie (oryginalnie wybarwione błękitem) wskazują na cytowania publikacji z *International Journal of Coal Science & Technology*, natomiast jaśniejsze (oryginalnie wybarwione zielenią) na *International Journal of Mining, Reclamation and Environment*. Nieliczne cytowane artykuły z czasopisma *Archives of Mining Sciences* (oryginalnie wybarwione czerwienią) można dostrzec jedynie w cyfrowej wersji obrazu.

Korzystając z prostej metryki zliczającej krawędzie wchodzące do danego wierzchołka (ang. *indegree*) zaimplementowanej w aplikacji Gephi obliczyć można liczbę cytowań

<sup>13</sup> <https://link.springer.com/journal/40789>

<sup>14</sup> <http://www.tandfonline.com/loi/nsme20#.VbuxENLbLGg>

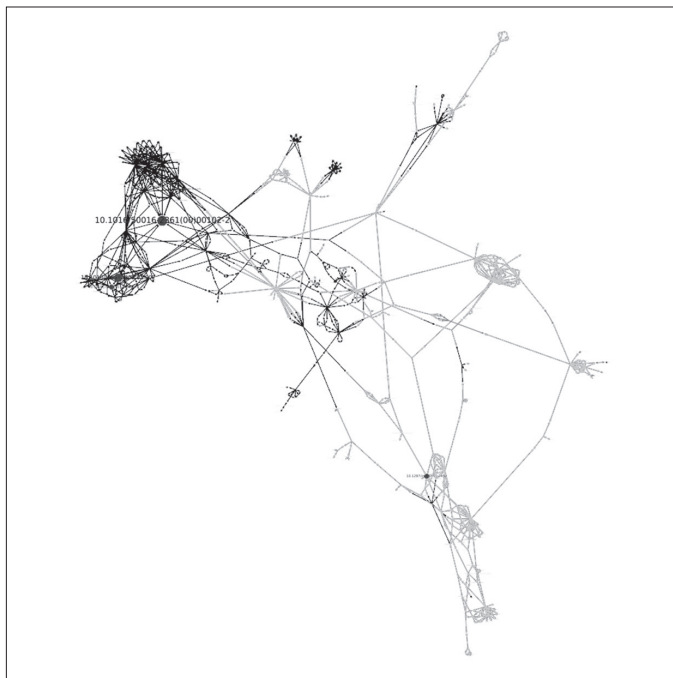
<sup>15</sup> <http://www.mining.archives.pl/>

danego artykułu. Najczęściej cytowane artykuły (o liczbie cytowań odpowiednio 11, 9, 8, 8) w analizowanym zbiorze danych to:

- *The new inertinite classification (ICCP System 1994)*<sup>16</sup>,
- *Estimations of Clarkes for Carbonaceous biolithes: World averages for trace element contents in black shales and coals*<sup>17</sup>,
- *Geochemistry of trace elements in Chinese coals: A review of abundances, genetic types, impacts on human health, and industrial utilization*<sup>18</sup>,
- *A Review of Operations Research in Mine Planning*<sup>19</sup>,

Spośród tych artykułów tylko ostatni nie był opublikowany przez wydawnictwo Elsevier. Artykuły z czasopisma *Archives of Mining Sciences*, reprezentowane stosunkowo nielicznie w analizowanym zbiorze cytowały natomiast najczęściej prace Mirosławy Bukowskiej.

Poza jednoznaczna identyfikacją, warto odnotować dalsze korzyści płynące z posługiwania się oznaczeniami DOI: czytelnik, znając identyfikator artykułu, może w każdej chwili odnaleźć jego dane bibliograficzne, gdyż usługa doi.org, na podstawie podanego oznaczenia DOI, automatycznie przekieruje przeglądarkę internetową czytelnika na aktualną źródłową stronę wydawcy.



Rys. 8. Przykład prostego grafu cytowań artykułów z wybranych czasopism  
(źródło: <https://doi.org/10.5281/zenodo.1214952>)

<sup>16</sup> [https://doi.org/10.1016/S0016-2361\(00\)00102-2](https://doi.org/10.1016/S0016-2361(00)00102-2)

<sup>17</sup> <https://doi.org/10.1016/j.coal.2009.01.002>

<sup>18</sup> <https://doi.org/10.1016/j.coal.2011.02.003>

<sup>19</sup> <https://doi.org/10.1287/inte.1090.0492>

## 5. Wnioski

Publiczny dostęp do danych bibliograficznych, w szczególności danych opisujących bibliografie załącznikowe prac naukowych, otwiera nowe możliwości zarówno ewaluacji nauki, jak i weryfikacji wyników publikowanych wcześniej czy ogłaszanych przez operatorów komercyjnych baz bibliograficznych. Transparentność ewaluacji nauki i powtarzalność realizowanych badań wydaje się być bardzo pożądana, szczególnie w przypadku stosowania metod wymagających ustalania pewnych granicznych wartości parametrów, jak na przykład metody klasyfikacji czy grupowania danych. Zostało to już dostrzeżone i zasygnalizowane przez Jochena Gläsera, Wolfganga Glänzela i Andreę Scharnhorst (Gläser et al., 2017).

Platforma Gephi, będąc otwartym i rozszerzalnym narzędziem do analizowania struktur grafowych, stanowi również dobrą bazę do realizacji wybranych analiz bibliometrycznych. Podstawowe jej ograniczenie w tej dziedzinie, zauważone również przez innych badaczy (Van Eck & Waltman, 2014), które zostało wyeliminowane za pomocą omówionego w tym artykule modułu ScientoMiner ICR, polegało na braku jej integracji z jakimkolwiek źródłem danych bibliograficznych. Możliwość całkowicie automatycznego pozyskiwania podzbioru danych udostępnianych przez platformę Crossref wprost do aplikacji Gephi powinno bardzo uprościć żmudne początki analiz bibliometrycznych składające się z ekstrakcji i formatowania danych. Warto również zauważyć, że coraz więcej polskich wydawców, zarówno dziedzin technicznych jak i humanistycznych, przekazuje swoje rekordy bibliograficzne do zasobów Crossref, co powinno sprawić, że analizy bazujące na tym źródle będą interesujące również dla badaczy ukierunkowanych na badanie rozwoju nauki w Polsce.

Wraz ze wzrostem zainteresowania tego typu badaniami możliwe będzie również dalsze rozszerzanie funkcjonalności środowiska Gephi o inne obszary związane z prowadzeniem analiz bibliometrycznych, jak na przykład integracja z kolejnymi źródłami danych lub implementacja miar specyficznych dla zagadnień bibliometrii.

## Bibliografia

- Bastian, M., Heymann, S., Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. *International AAAI Conference on Weblogs and Social Media*, 8, 361–362.
- Gläser, J., Glänzel, W., Scharnhorst, A. (2017). Same Data-Different Results? Towards a Comparative Approach to the Identification of Thematic Structures in Science. *Scientometrics*, 111(2), 981–998.
- IDF (2017a). *DOI Handbook* [online]. International DOI Foundation [28.01.2018], [https://www.doi.org/doi\\_handbook/7\\_IDF.html](https://www.doi.org/doi_handbook/7_IDF.html)
- IDF (2017b). *Registration Agencies – Areas of Coverage* [online]. International DOI Foundation [28.01.2018], [https://www.doi.org/RA\\_Coverage.html](https://www.doi.org/RA_Coverage.html)
- ISSI (2017). *Open Citations: A Letter from the Scientometric Community to Scholarly Publishers* [online]. ISSI – International Society for Informetrics and Scientometrics [28.01.2018], <http://issi-society.org/open-citations-letter>
- Kamińska, A. M. (2017a). Dobre praktyki publikowania danych badawczych [online]. *Biuletyn EBIB*, 7(177), [28.01.2018], <http://open.ebib.pl/ojs/index.php/ebib/article/view/569/752>
- Kamińska, A. M. (2017b). Miary podobieństw łańcuchów znakowych a deduplikacja rekordów w bibliograficznych bazach danych. *Przegląd Biblioteczny*, 4, 477–495.
- Kamińska, A. M. (2017c). ProBIT – perspektywna metoda tworzenia trawersowalnych indeksów cytowań a współczesne problemy organizacji przestrzeni informacji w tradycyjnych bibliograficznych bazach danych. *Zagadnienia Informacji Naukowej*, 55(1), 66–82.



- Kamińska, A. M. (2017d). Wizualizacje wybranych wskaźników bibliometrycznych na przykładzie bibliograficznej bazy danych CYTBIN. *Toruńskie Studia Bibliologiczne*, 2 (19), 163–187. DOI: 10.12775/TSB.2017.022
- Kamińska, A. M. (2018a). Visualizations of the GRUBA Bibliographic Database: From Printed Sources to the Maps of Science. In: V. Osinska & G. Osinski (eds.). *Information Visualization Techniques in the Social Sciences and Humanities* (151–174). Hershey, PA: IGI Global, 151–174. DOI: 10.4018/978-1-5225-4990-1.ch009
- Kamińska, A. M. (2018b). Zastosowanie struktur grafowych do analiz bibliometrycznych i webometrycznych. Modele i metody. *Nowa Biblioteka. Usługi, Technologie Informacyjne i Media* [w druku].
- Van Eck, N.J.; Waltman, L. (2014) Visualizing Bibliometric Networks. In: Y.; R. Rousseau & D. Wolfram (eds.). *Measuring Scholarly Impact* (285–320). Springer, Cham. DOI: 10.1007/978-3-319-10377-8\_13
- 

## ScientoMiner ICR – The Plug-in for Importing Bibliographic Data from Crossref Resources to the Gephi Platform

### Abstract

**Purpose/Thesis:** Presenting what causes problems in reliable bibliometric research using citation network analysis, the author indicates that the growing use of the DOI identification system and the growing phenomenon of publishing information on article citations by individual publishers result in new possibilities for such research. Particularly noteworthy here are Crossref services that enable downloading structured bibliographic information (including citations) for all interested parties. The author proposes the extension of the Gephi analytical platform with a data import plug-in that imports data describing citations from Crossref services, thus facilitating citation analysis by all interested parties in this source of data.

**Approach/Methods:** The implementation of the proposed plug-in required the analysis of the information architecture of resources collected and indexed in the Crossref database and the way they are shared as web services. One of the important parts of the research was also the development of a graph data model that allows unambiguous identification of citing and cited papers and the relations between them, so that the evolution of this model in the future (e.g. adding co-authorship relations) would not prevent the usage of the developed plug-in.

**Results and conclusions:** The developed functionality allows to import a directed citation graph from Crossref resources to the Gephi platform, for user-defined journals identified by the ISSN, with the possibility of limiting the data set to given time intervals. For less experienced users, it also offers preliminary coloring of the graph's vertices by their type. The use of DOI identifiers makes the results of the analyzes more reliable than in the case of analyses where matching of titles and authors of bibliographic units with records of references section is performed.

**Originality/Value:** The Gephi computing platform, being an interesting tool for analyzing network structures, is also used, mainly by foreign researchers, to implement bibliometric research and visualize their results. Recently, there has been a growing interest in this tool among local researchers, but it seems that it is still losing competition with the easier to use Pajek application and strictly dedicated to the bibliometric analysis VOSviewer platform. Although the latter tools have less analytical capabilities, the technical barriers in loading the Gephi platform with bibliographic data limits its usage. The author, using the modular construction of this platform, has developed a plug-in enabling automatic data import directly from the Crossref resources in a more friendly way compared to the

VOSviewer environment. To the best of her knowledge, it is also the first extension of this platform dedicated to the support of strictly bibliometric research.

### **Keywords**

Bibliometrics. Bibliographic databases. Citation analysis. Citation index. Crossref. DOI. Data import. Gephi. Science mapping. ScientoMiner ICR. Visualization of bibliographic data.

---

*Dr ANNA MAŁGORZATA KAMIŃSKA jest adiunktem w Instytucie Bibliotekoznawstwa i Informatyki Uniwersytetu Śląskiego w Katowicach oraz pracownikiem Biblioteki Głównej Politechniki Śląskiej w Gliwicach. W 2016 r. obroniła rozprawę doktorską pt. Informacja naukowa o górnictwie w świetle wydawnictw ciągłych uczelni technicznych w Polsce (1945–1989) na Wydziale Filologicznym Uniwersytetu Śląskiego w Katowicach. Jej zainteresowania naukowe skupiają się wokół trzech uzupełniających się obszarów: informatologia, graficzne języki komunikacji oraz wizualizacja informacji.*

*Kontakt z autorką:*

*anna.kaminska@us.edu.pl*

*Instytut Bibliotekoznawstwa i Informatyki Naukowej*

*Uniwersytet Śląski w Katowicach*

*pl. Sejmu Śląskiego 1*

*40-032 Katowice*