

А. М. КАМІНСЬКА, С. А. НАЗАРОВЕЦЬ

CROSSREF ЯК ДЖЕРЕЛО НАУКОМЕТРИЧНИХ ДАНИХ ДЛЯ СОЦІАЛЬНИХ ТА ГУМАНІТАРНИХ НАУК

Сучасна наукометрія успішно використовує низку аналітичних методів для дослідження наукових публікацій, проте, застосування навіть найбільш розвинених аналітичних методів до помилкових або неповних даних призводить до хибних результатів. **Предметом** дослідження в статті є сучасні підходи та технології отримання бібліографічних даних з надійних відкритих джерел наукової інформації для проведення наукометричних досліджень. **Метою** даної роботи є презентація розробленого модулю ScientoMiner ICR, що розширює функціональні можливості аналітичної платформи Gephi та дозволяє імпортувати бібліографічні дані з загальнодоступних записів Crossref для проведення наукометричних досліджень у сфері аналізу цитувань, якість яких гарантовано застосуванням системи ідентифікації DOI. У статті вирішуються наступні **завдання**: оцінити перспективи розвитку Initiative for Open Citations для створення некомерційних відкритих інструментів відслідковування наукових цитувань; обґрунтувати використання системи DOI та бази даних реєстраційного агентства Crossref, як надійного джерела бібліографічних даних і механізму отримання метаданих наукових видань; розкрити можливості модулю ScientoMiner ICR для відкритої платформи Gephi для імпорту бібліографічних даних. Використані **методи дослідження** – візуалізація та аналіз мережі наукових цитувань. **Висновок** – запропонований модуль ScientoMiner ICR для платформи Gephi спрощує процес автоматичного експорту даних Crossref безпосередньо в Gephi, зокрема, для потреб проведення наукометричного аналізу. Отримані **результати** можуть бути використані в процесі створення нових некомерційних інструментів відстежування цитованості наукових робіт, насамперед, у галузях соціальних та гуманітарних наук, що недостатньо добре представлені в базах даних комерційних наукометричних продуктів.

Ключові слова: наукометрія; DOI; Crossref; Open Citations; Gephi; ScientoMiner.

Вступ

Необхідною умовою для проведення вірогідних наукометричних досліджень щодо аналізу цитувань між науковими роботами є можливість їх вірної ідентифікації для однозначного їх пов'язання з роботами, які вказано в списках бібліографії в інших наукових роботах. Хоча, на перший погляд, дана проблема видається тривіальною, однак складність структури бібліографічних записів, технічні помилки, що трапляються під час реєстрації бібліографічних записів та велика кількість стилів цитувань створюють значні труднощі в процесі встановлення подібних зв'язків між науковими документами [1]. У багатьох країнах наукометричні показники використовують для потреб прогнозування розвитку наукових досліджень, проте, для того, щоб такі дані можна було використовувати в процесі прийняття таких важливих рішень, вони повинні бути розраховані на основі надійних джерел. Створення великих бібліографічних баз даних з особливою увагою до надійності зібраної інформації призвело до комерціалізації подібних проектів. Найбільш відомими прикладами таких продуктів стали бази даних платформи Web of Science та Scopus. Ці реферативні бази даних забезпечують широке охоплення наукової літератури, що дозволяє проводити наукометричні дослідження у галузях природничих наук. Однак, результати досліджень показують, що використання згаданих ресурсів для оцінювання продуктивності та впливовості видань, вчених, підрозділів та установ, що працюють у багатьох галузях соціальних та гуманітарних наук, має свої суттєві обмеження.

Представлення тез та матеріалів конференцій і наукових монографій у Web of Science та Scopus далеко не повне, а для багатьох дисциплін у галузі соціальних та гуманітарних наук монографії

посідають дуже важливе місце в процесі наукової комунікації. Результати наукових досліджень у природничих дисциплінах, як правило, публікуються англійською мовою у рецензованих журналах та орієнтовані на міжнародну аудиторію вузькоспеціалізованих фахівців. Значна ж частина результатів у суспільних та гуманітарних науках присвячена регіональній тематиці, скерована на національну аудиторію, відтак результати соціогуманітарних досліджень часто публікуються у регіональних журналах та монографіях різними мовами, а не лише англійською, і доступні для розуміння ненауковій громадськості. Результати соціогуманітарних досліджень, які отримані в одній країні, можуть бути некорисними для дослідників з інших країн через певні суспільні та культурні відмінності [2]. При цьому, соціогуманітарні науки не є однорідними, деякі дисципліни подібні до природничих наук в плані частоти публікацій та цитувань, але для частини права, лінгвістики, літературознавства, соціології, державного управління та політології згадані особливості є надзвичайно суттєвими [3]. Таким чином, практика офіційного використання наукометричних показників у соціальних і гуманітарних науках відіграє не таку важливу роль, як в природничих науках, і може суттєво відрізнятись у різних країнах. Так, результати опитування вчених-гуманітаріїв в австралійських та шведських університетах щодо використання ними наукометричних показників показали, що 62% австралійських науковців зверталися до наукометричних метрик у своїх грантових заявках та резюме, водночас, для шведських вчених цей показник становив лише 14% [4].

Дослідження публікаційної активності вчених, які працюють у галузях соціальних та гуманітарних наук у неангломовних європейських країнах показало,

що частка їх журнальних англомовних публікацій зростає, при чому в країнах, де на державному рівні вже тривалий час використовуються формалізовані підходи до оцінки наукової продуктивності, англійською мовою публікується переважна більшість соціогуманітарних досліджень країни – понад 60% у Фінляндії, Данії і Норвегії та майже 80% – у Фландрії [5]. Водночас, виявилось, що публікаційна поведінка вчених у соціогуманітарних дисциплінах тісно пов'язана з культурною та історичною спадщиною країни. Так, попри суттєве збільшення кількості англомовних публікацій у Польщі, їх частка становить лише 17,2% від загальної кількості соціогуманітарних публікацій, і більшість польських вчених у даних галузях надають перевагу польській мові для представлення результатів своїх досліджень. Автори дослідження стверджують, що чеський та польський досвід засвідчив легкість впровадження змін у публікаційні практики науковців шляхом надання відповідних стимулів, проте наукова політика в неангломовних країнах, яка спрямована на збільшення кількості англомовних наукових публікацій, обов'язково повинна розроблятися з урахуванням поточних моделей публікацій, які функціонують у країні, адже надання таких стимулів може мати й негативні наслідки для розвитку соціальних та гуманітарних наук.

Стрімка зміна комунікаційних практик у соціогуманітарних науках на користь англомовних журналів, що представлені в базах Web of Science та Scopus, відбувається і в інших країнах-сусідах України. Як правило, дані зміни відбуваються внаслідок орієнтації наукової політики країн, що розвиваються на міжнародні рейтинги університетів, глобалізаційні процеси, покращення інформаційного та культурного обміну після падіння "залізної завіси", зниження ролі мовного фактору в процесі наукової комунікації та краще проведення процедури рецензування в англомовних міжнародних виданнях у порівнянні з більшістю наукових видань країн колишнього соцтабору [6].

В Україні ж впровадження ініціатив щодо зміни прийнятих публікаційних практик і використання наукометричних метрик у соціогуманітарних науках відбувається непослідовно і без проведення відповідних попередніх досліджень. Відтак, очікувано, дані ініціативи викликають опір з боку значної частини української наукової спільноти. Так, у квітні 2018 року до Міністерства освіти і науки України звернувся Науковий комітет Національної ради з питань розвитку науки і технологій з ініціативою внести зміни до наказу МОН від 17 жовтня 2012 року № 1112 "Про опублікування результатів дисертацій на здобуття наукових ступенів доктора і кандидата наук". Серед запропонованих Науковим комітетом змін була вимога щодо наявності у здобувачів наукового ступеня публікацій у виданнях, які включені до баз Scopus, або Web of Science Core Collection. Враховуючи те, що Науковий комітет – це робочий колегіальний орган Національної ради з питань розвитку науки і технологій, що

представляє інтереси наукової громадськості, Міністерство освіти і науки України запропонувало для громадського обговорення наданий Науковим комітетом проект наказу Міністерства освіти і науки України "Про затвердження змін до наказу Міністерства освіти і науки України від 17 жовтня 2012 року № 1112".

У травні 2018 року з'явився відкритий лист представників українських соціогуманітарних дисциплін до Міністерства освіти і науки України щодо змін до наказу про оцінювання результатів досліджень (<https://www.petitions247.net/203462>), де було сказано, що "якість публікацій у соціогуманітарних науках не пов'язується з критерієм їх міжнародної цитованості – тут діють інші чинники, зокрема спрямованість вітчизняної науки на вирішення національних завдань і пріоритетів, її загостреність на національну проблематику. <...> Механізм оцінювання якості має бути продуманий і вдосконалений з урахуванням специфіки соціогуманітарної сфери, її принципової відмінності від точних і природничих наук". Автори даного звернення запропонували Міністерству освіти і науки України відхилити запропоновані Науковим комітетом зміни до оприлюднення результатів наукових досліджень і активніше працювати над усуненням чинників, які призводять до імітації науки. Невдовзі Колегія МОН прийняла рішення створити Робочу групу із залученням провідних представників соціогуманітарних дисциплін для доопрацювання змін до наказу про оцінювання результатів наукових досліджень.

Постановка проблеми

Використання провідних наукометричних ресурсів для потреб аналізу українських рецензованих видань має свої додаткові обмеження, які пов'язані не тільки з особливостями соціогуманітарних дисциплін, а й зі слабкою представленістю українських журналів у загаданих базах. Станом на липень 2018 року в Scopus та Web of Science Core Collection представлено 99 українських наукових журналів. При цьому, в Journal Citation Reports вказано лише 15 назв, а переважна більшість українських журналів індексується у базі Emerging Sources Citation Index (ESCI), яка знаходиться ще на етапі вироблення чітких правил щодо включених наукових видань. Також, внаслідок різних причин, індексування нових випусків певних українських видань у Scopus та Web of Science вже припинена, тобто реальна кількість активних вітчизняних назв у провідних реферативних базах є ще меншою.

Водночас, на порталі "Наукова періодика України" Національної бібліотеки України імені В.І. Вернадського, що покликаний представляти електронні копії періодичних друкованих наукових видань України, станом на липень 2018 року було представлено 2633 назв українських наукових журналів. Проте, технічні можливості порталу дозволяють користувачам використовувати його дані

хіба що для проведення найпростіших кількісних досліджень, без встановлення зв'язків між науковими документами. Також невивченим залишається питання щодо критеріїв наповнення, оновлення та використання реферативної бази даних "Україніка наукова" і українського реферативного журналу "Джерело". У нещодавньому дослідженні європейських та ізраїльських реферативних баз даних, які використовують вчені у галузях соціальних та гуманітарних наук, було виявлено 21 національну базу даних, проте у даній роботі було сказано, що в Україні немає жодної бібліографічної бази даних [7, С. 4]. Така помилка вказує не лише на прорахунки авторів у процесі вибору респондентів дослідження, але й на невідомість українських реферативних наукових продуктів у світі, відсутність відповідного функціоналу, який би міг задовольнити потреби сучасних дослідників, приміром, надання даних у машиночитному форматі за допомогою відкритого API.

Отримати певні інформативні дані щодо продуктивності українських вчених, установ, журналів, встановити зв'язки між публікаціями українських вчених дозволяє інформаційно-аналітична система "Бібліометрика української науки", що розроблена та підтримується відділом бібліометрії та наукометрії Служби інформаційно-аналітичного забезпечення Національної бібліотеки України імені В. І. Вернадського і яка покликана надавати суспільству цілісну картину стану українського академічного середовища на основі даних отриманих з бібліометричних профілів учених в Google Scholar [8]. Проте, в даній системі частково залишилися проігнорованими усі недоліки та небезпеки використання даних Google Scholar. Вміст бази Google Scholar – непрозорий, за її наповнення відповідає комп'ютерний алгоритм, і достеменно не відомо, чому певні ненаукові документи представлені в базі, натомість відсутні деякі рецензовані джерела. Також, дуже часто Google Scholar вважає дублікати різними документами, особливо, коли на сайті видання бібліографічний опис документу представлений багатьма мовами [9]. Механізм підрахунку цитувань Google Scholar цілковито незахищений від маніпуляцій, і дослідники можуть керувати даними Google Scholar шляхом створення фейкових публікацій, які міститимуть велику кількість самоцитованим вченим [10]. Деякі дослідники переконують, що Google Scholar можна використовувати для наукометричних оцінок за умови ретельного попереднього очищення даних [11], проте таку попередню роботу доволі складно реалізувати для масштабного аналізу, оскільки Google Scholar надає обмежений доступ до своїх даних.

Таким чином, в Україні наразі відсутня надійна реферативна база даних, яку б можна було використати для проведення повноцінних наукометричних досліджень, при цьому представлення українських наукових видань у Scopus та Web of Science є дуже обмеженим, а процес

створення та наповнення нової національної бази даних та інструменту для відстеження наукових цитувань, за принципом роботи авторитетних комерційних продуктів, є доволі тривалим, коштовним і потребує значних витрат для подальшої підтримки. Виходом з цієї складної ситуації може стати пристосування для цих завдань інших спеціалізованих ресурсів та інструментів, які використовують сучасні наукові видавництва, зокрема системи DOI та баз даних реєстраційних агентств DOI, які зберігають метадані наукових документів. Мета даної статті – презентація розробленого модулю ScientoMiner ICR для імпорту бібліографічних записів з бази даних реєстраційного агентства DOI задля проведення наукометричних досліджень, зокрема в галузі соціальних та гуманітарних наук.

У цьому дослідженні розглядаються наступні питання:

- обґрунтування використання системи DOI, як механізму для отримання метаданих наукових видань;
- індикація ресурсів реєстраційного агентства CrossRef, як цінного публічного джерела бібліографічних даних;
- перспективи розвитку Initiative for Open Citations для створення некомерційних відкритих інструментів відслідковування наукових цитувань;
- можливості використання модулю ScientoMiner ICR відкритого інструменту Gephi для імпорту бібліографічних даних з CrossRef, візуалізації та аналізу мереж наукових цитувань.

Система DOI

Впровадження системи DOI значно спростило авторам, рецензентам, видавцям, бібліотекарям та упорядникам реферативних баз даних роботу з бібліографічним записами на кожному етапі їх використання. Цифровий ідентифікатор DOI (Digital Object Identifier) – унікальний цифровий ідентифікатор об'єкта, що складається з цифр та літер і пов'язаний з посиланням на веб-сторінку, де знаходиться саме цей об'єкт, або інформація про нього. DOI-об'єкт – це дані, що складаються з унікального ідентифікатора, URL та певного набору метаданих. Ідентифікатор DOI реєструється для об'єкта одноразово і залишається незмінним, а метадані та URL-адреси об'єкта можуть змінюватись за запитом власника [12]. Діапазони ідентифікаторів (префікси) надаються видавцям спеціальними установами DOI Registration Agency. Хоча організація, яка підтримує систему DOI, є неприбутковою організацією, реєстраційні агенції збирають від видавців кошти, щоб покрити витрати на підтримку інфраструктури. Хоча усі реєстраційні агенції пропонують своїм клієнтам унікальні ідентифікатори, спектр їх додаткових послуг суттєво різниться. Повний перелік реєстраційних агенцій, а також короткий опис пропонованих ними додаткових послуг, доступний на офіційному веб-сайті DOI (https://www.doi.org/registration_agencies.html), і навіть побіжного аналізу даного переліку достатньо, щоб

зрозуміти, що саме агенція Crossref має першочергове значення для видавців наукової літератури.

Реєстраційна агенція CrossRef не лише присвоює унікальні ідентифікатори, але й дозволяє видавцям автоматично передавати бібліографічні дані публікацій різним бібліографічним базам даних, зокрема, списки використаної літератури, які містяться у цих роботах, що значно покращує представлення видань у світовому інформаційному просторі. Також, агенція Crossref пропонує своїм клієнтам взяти участь у безкоштовній програмі Cited-by (<https://www.crossref.org/services/cited-by/>), що дозволяє довідатись та підрахувати посилання між різними науковими документами з DOI, а також підтримати ініціативу Open Citations (I4OC), яка має на меті зробити дані про наукові цитування структурованими, відокремленими та відкритими для усіх (<https://i4oc.org>) [13].

Програма Cited-by від Crossref допомагає з'ясувати, хто процитував публікацію, підрахувати кількість цитувань та отримати докладну інформацію про документ, що цитує дану публікацію. Члени Crossref надають метадані про документи процитовані у публікаціях, які вони реєструють і, водночас, вони можуть дізнатися, які публікації цитують їхні видання. Таким чином, члени Crossref допомагають читачам легко знаходити потрібні дослідження, зрозуміти, які роботи виявились найбільш впливовими в науковій спільноті, та досліджувати зв'язки між науковими публікаціями. Для того, щоб взяти участь у програмі Cited-by, видавці повинні виконати три прості кроки: 1) Видавці подають у Crossref списки використаної літератури, як частину реєстрації звичайного контенту; 2) Crossref порівнює отримані метадані з посиланнями на DOI, щоб встановити зв'язки між документами в своїй базі; 3) Видавці та онлайн-платформи отримують метадані щодо цитування своїх публікацій та можуть вказувати їх кількість на веб-сайтах видань.

Якщо установа використовує сервіс Cited-by від Crossref, то вона також може долучитися до Initiative for Open Citations і відкрити всі свої посилання та інші метадані, що надсилаються у Crossref. Для цього видавцеві потрібно встановити відповідний елемент метаданих для кожного депозиту DOI (`<reference_distribution_opt> to "any"`), чи просто звернутися до служби підтримки Crossref. Як правило, дані щодо цитувань не представлені в машинозчитуваній формі і знаходяться "під захистом" непослідовних ліцензій. Initiative for Open Citations – це співпраця між науковими видавництвами, дослідниками та іншими зацікавленими сторонами задля сприяння необмеженій доступності даних про цитування в науковій літературі [14]. Дану ініціативу підтримали такі відомі академічні видавці, як American Geophysical Union, Association for Computing Machinery, BMJ, Cambridge University Press, Cold Spring Harbor Laboratory Press, EMBO Press, Royal Society of Chemistry, SAGE Publishing, Springer Nature, Taylor & Francis, Wiley. Також, частина метаданих бібліографічних посилань Crossref доступні у

OpenCitations Corpus (<http://opencitations.net>) – відкритому репозитарії, який спеціально створено для розміщення та збереження даних наукових цитувань, що систематично збирає ці дані з Crossref та інших джерел [15].

Бібліографічні дані, які доступні в базі CrossRef, можна отримати за допомогою відповідного запиту у текстовому форматі обміну даними JSON (JavaScript Object Notation), який може бути прочитаним як комп'ютером, так і людиною, навіть без додаткового опрацювання. Параметром такого запиту може слугувати назва видання, видавництва, ISSN, префікс DOI ітд. (докладна специфікація доступна за посиланням: https://github.com/CrossRef/rest-api-doc/blob/master/api_format.md). Хоча видавці можуть передавати в Crossref доволі багато різноманітної бібліографічної інформації, багато атрибутів у записах не є обов'язковими, відтак повнота метаданих різних видавництв та журналів може суттєво різнитися. Варто окремо відзначити, що ідентифікатор – єдиний обов'язковий атрибут для пристатейних списків літератури, проте в документації нічого не сказано про його походження, чи можливість інтерпретації. Авторський досвід роботи з даними CrossRef показує, що якість бібліографічних даних буває різною, проте під час проведення цього дослідження нами спостерігалось покращення якості та повноти метаданих для нових документів, що дозволяє з оптимізмом дивитися на перспективи використання даних CrossRef для потреб проведення надійного бібліометричного аналізу.

Методологія

Ми експортували з бази Crossref метадані двох наукових журналів видавництва, що підтримали Initiative for Open Citations і які станом на липень 2018 р. мають високі значення CiteScore серед видань у категорії Library and Information Sciences згідно з даними Scopus: European Journal of Information Systems (ISSN: 0960-085X) видавництва Springer Nature та Information Communication and Society (E-ISSN: 1369-118X) видавництва Taylor & Francis. Дані було експортовано за 2016 р. за допомогою спеціального модулю ScientoMiner ICR відкритого програмного інструменту Gephi (<https://gephi.org>), що призначений для візуалізації та аналізу мереж [16, 17]. Модуль ScientoMiner ICR доступний безкоштовно для використання з некомерційною метою за адресою <https://doi.org/10.5281/zenodo.1215008>.

Обсягом імпортованих даних у модулі ScientoMiner ICR можна керувати шляхом вказання ідентифікатора ISSN, а також встановлення відповідного діапазону дат для інформації, що аналізується. Користувачі, які не мають великого досвіду роботи з платформою Gephi, можуть скористатися опцією "predecorate graph", яка автоматично встановлює різні кольори та розміри вершин графа, що репрезентують цитовані документи і документи, які посилаються на інші документи. Дані

атрибути згодом можуть бути довільно модифіковані під час подальших досліджень.

За замовчуванням у модулі ScientoMiner ICR увімкнено опцію "DOIs only", що обмежує імпорт записів з бази Crossref лише тими документами, які ідентифіковано за допомогою DOI. Використовувати дану опцію потрібно обережно, попередньо перевіривши, чи видавець аналізованого журналу подбав про унікальну ідентифікацію документів, які згадані в списку використаної літератури та передав відповідні метадані в реєстраційне агентство.

Опція "Download data in chunks" модулю ScientoMiner ICR дозволяє задати розмір одноразових комунікативів, які буде отримано з бази Crossref – якщо вказати менший розмір, то час імпортування буде довшим, проте заданий більший розмір даних при невисокій пропускній здатності, або обмежених обчислювальних можливостях системи, може призвести до помилок з'єднання з базою даних Crossref. Загальний вигляд робочого інтерфейсу модулю ScientoMiner ICR для відкритої платформи Gephi показано на рисунку 1.

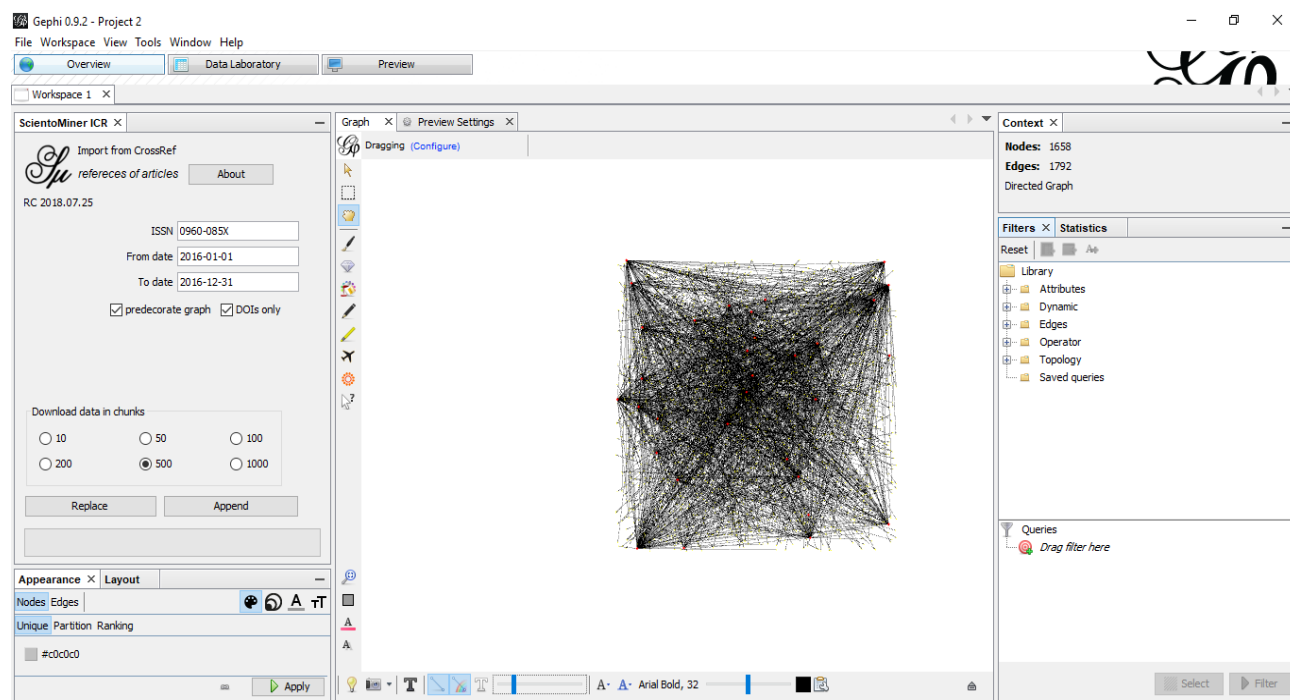


Рис. 1. Приклад робочого інтерфейсу модулю ScientoMiner ICR для Gephi

Процес імпорту даних у ScientoMiner ICR розпочинається за допомогою кнопки "Replace", яка запускає видалення усіх попередніх даних з робочої області, або кнопки "Append", що запускає розширення існуючого графу в робочій області новими даними, які імпортуються з бази Crossref. Друга опція, насамперед, потрібна для проведення аналізу спільних даних, які походять з різних видань, зокрема, для аналізу мережі цитувань. Оскільки граф цитувань є орієнтованим графом, то слід окремо зазначити, що імпорт даних можливий лише для проектів, які базуються на орієнтованих, або мішаних графах. По звершенні імпорту даних з'явиться граф з вершинами, які розміщено випадково на поверхні квадрату, і подальший аналіз якого обмежений лише уявою дослідника.

Експорт метаданих для потреб даного дослідження проводився 28 липня 2018 р. У модулі ScientoMiner ICR були встановлені відповідні часові інтервали та двічі було проведено імпорт даних Crossref, із вказанням ISSN обраних журналів та використанням параметра Append, що дозволило розмістити метадані обидвох проаналізованих журналів на одному графі. Також, для кращої візуалізації даних, нами було використано один з

алгоритмів Gephi для автоматичного розміщення вершин графа та відкидання вершин з незначною кількістю цитувань.

Результати

Інструмент Gephi та модуль ScientoMiner ICR дозволили провести аналіз і візуалізацію експортованих даних двох обраних наукових журналів. Результат цієї візуалізації показано у вигляді графа на рисунку 2.

Мережу цитувань журналів European Journal of Information Systems та Information Communication and Society представлено різними кольорами (червоним та зеленим), де кулі – цитовані статті, які ідентифіковано завдяки ідентифікаторам DOI, що вказані у списках використаної літератури, при цьому, чим більший розмір кулі, тим частіше були цитовані дані публікації.

Цифрова версія даного рисунку, а також усі аналізовані дані представлено авторами у відкритому доступі за адресою: <https://doi.org/10.5281/zenodo.1324030>. Окрім швидкого встановлення мережі цитувань між науковими документами, модуль ScientoMiner ICR

дозволяє користувачам переглянути усі експортовані дані у зручній табличній формі, швидко відшукати

ідентифікатор DOI потрібної публікації та отримати інформацію про даний документ на сторінці видавця.

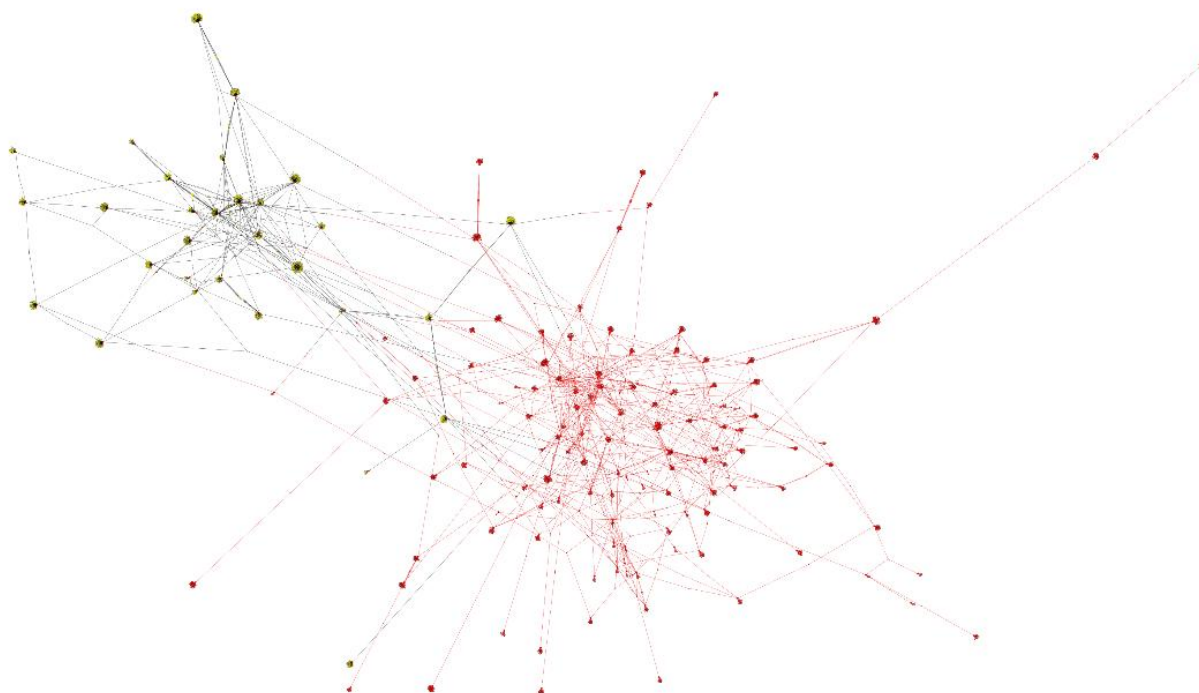


Рис. 2. Приклад простого графу наукових цитувань статей обраних журналів (<https://doi.org/10.5281/zenodo.1324030>)

Візуалізація даних двох обраних наукових журналів з галузі бібліотекознавства проведено для демонстрації базових функціональних можливостей спеціального модулю ScientoMiner ICR виключно з ілюстративною метою і без наміру робити будь-які подальші висновки на підставі результатів даного аналізу.

Висновки та обговорення

Платформа Gephi є відмінним відкритим інструментом для дослідження структур графів, а також надійною базою для проведення наукометричних досліджень. Попередні обмеження щодо використання Gephi, на які звертали увагу інші дослідники, полягали у відсутності інтеграції платформи з будь-яким джерелом бібліографічних даних [18]. Дану проблему вдалось успішно вирішити в цьому дослідженні шляхом використання спеціального модулю ScientoMiner ICR. Реалізована можливість автоматичного експорту даних Crossref безпосередньо в Gephi повинна значно спростити і полегшити клопітливий початковий етап проведення наукометричного аналізу, що включає експортування та форматування бібліографічних даних.

Все більше і більше українських наукових видавців, як у галузях природничих, так і соціальних та гуманітарних наук, передають свої бібліографічні дані в Crossref, відповідно, перспективи комп'ютерного аналізу цієї інформації повинні зацікавити дослідників і управлінців, які займаються питаннями стану та розвитку науки в Україні.

А, враховуючи слабше представлення вітчизняних соціогуманітарних публікацій у світових реферативних базах даних у порівнянні з природничими науками, насамперед, спеціалісти, які займаються проблемами оцінювання наукової продуктивності у соціальних та гуманітарних дисциплінах, повинні активно впроваджувати нові можливості використання платформи Crossref для отримання наукометричних даних. Зі свого боку, фахівці Crossref активно заохочують видавців скрупульозно вносити інформацію, що міститься в списках використаної літератури, вказувати DOI у списках використаних джерел, щоб потім дана інформація стала доступною для всіх зацікавлених сторін.

Зростання підтримки ініціативи надання постійного відкритого доступу до наукових цитувань серед видавців наукової літератури призвів до появи нових інструментів відстежування цитованості наукових робіт, що дозволяє сподіватися на зменшення залежності вчених-наукометристів від комерційних продуктів у майбутньому. Так, наприклад, у липні 2017 р. запущено Crossref Open Citation Index (COCI) – перший відкритий індекс наукових цитувань, який вдалось створити завдяки поширенню Initiative for Open Citations і який містить понад 300 мільйонів відкритих цитувань DOI-to-DOI з Crossref. До кінця 2018 р. Міністерство освіти та науки України, також планує запустити власний відкритий український індекс наукового цитування, що повинен використовувати метадані журналів та відкриті наукові цитування платформи Crossref [19].

Водночас, попри значний приріст прихильників ідеї поширення відкритих цитувань, такі великі наукові видавництва як American Chemical Society, Elsevier, IEEE та Wolters Kluwer Health досі не підтримали Initiative for Open Citations, відповідно, дані щодо цитувань видань згаданих видавництв є недоступні, що суттєво ускладнює проведення якісного наукометричного аналізу. Також, повнота переданих метаданих членів Crossref поки може суттєво різнитися. Проте, виявлене авторами зростання якості метаданих нових публікацій, найімовірніше, свідчить про те, що попередні прогалини в представленні бібліографічної інформації пов'язані з опануванням працівниками редакцій нових редакційно-видавничих процесів, відтак, якість метаданих видання покращується з набуттям редакцією відповідного досвіду.

Доступ громадськості до бібліографічних даних, зокрема тих, що містять дані щодо наукових цитувань, відкривають нові можливості, як для оцінки науки, так і для перевірки результатів опублікованих раніше досліджень, що були проведені за використання комерційних наукометричних інструментів. Прозорість процесу оцінювання продуктивності, впливовості науковців та відтворюваність результатів

досліджень надзвичайно важливі, особливо у випадку використання методів, що вимагають встановлення певних граничних значень параметрів, таких як метод класифікації, чи групування даних [20]. Доступність даних Crossref, зрозуміла модель їх створення, поширення та ліцензування для подальшого безперешкодного повторного використання робить Crossref перспективним джерелом даних для проведення наукометричних досліджень.

Зростання попиту на подібні види досліджень повинно сприяти розширенню функціональних можливостей середовища Gephi, зокрема інтеграції з іншими джерелами бібліографічних даних. Враховуючи той факт, що українські наукові видання, насамперед, у галузях соціальних та гуманітарних наук слабо представлені у міжнародних реферативних базах даних, редакціям варто докласти зусиль щодо створення та поширення метаданих своїх видань задля покращення їх представлення в інформаційному просторі. Водночас, національні реферативні бази даних наукової літератури повинні забезпечити користувачам можливість експорту бібліографічної інформації у популярних форматах для її подальшого аналізу, зокрема, й засобами платформи Gephi.

Список літератури

1. Kamińska, A. M. (2017), "ProBIT – perspektywna metoda tworzenia trawersowalnych indeksów cytowań a współczesne problemy organizacji przestrzeni informacji w tradycyjnych bibliograficznych bazach danych", *Zagadnienia Informacji Naukowej*, No. 1 (55), P. 66–82.
2. Moed, H. F., Luwel, M., Nederhof, A. J. (2002), "Towards indicators of research performance in the humanities", *Library Trends*, No. 50, P. 498–520.
3. Nederhof, A. J. (2006), "Bibliometric monitoring of research performance in the social sciences and the humanities : A review", *Scientometrics*, No. 1 (66), P. 81–100. DOI: <http://doi.org/10.1007/s11192-006-0007-2>.
4. Hammarfelt, B., Haddow, G. (2018), "Conflicting measures and values: How humanities scholars in australia and sweden use and react to bibliometric indicators", *Journal of the Association for Information Science and Technology*, No. 7 (69), P. 924–935. DOI: <http://doi.org/10.1002/asi.24043>.
5. Kulczycki, E., Engels, T. C. E., Pölonen, J. et al. (2018), "Publication patterns in the social sciences and humanities : evidence from eight European countries", *Scientometrics*, No. 1 (116), P. 463–486. DOI: <http://doi.org/10.1007/s11192-018-2711-0>.
6. Стерлигов И. Отражение и преобразование, или гуманитарии под воздействием наукометрии. *Измерения науки*. 2018. Т. 6, № 1. С. 1–4.
7. Sile L., Pölonen, J., Sivertsen, G. et al. (2018), "Comprehensiveness of national bibliographic databases for social sciences and humanities: Findings from a European survey", *Research Evaluation*. DOI: <http://doi.org/10.1093/reseval/rvy016>.
8. Симоненко Т. Глобальна бібліометрика: концептуальна модель. *Вісник Книжкової палати*. 2016. № 6. С. 12–14.
9. Halevi, G., Moed, H., Bar-Ilan, J. (2017), "Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation – Review of the Literature", *Journal of Informetrics*, No. 3 (11), P. 823–834. DOI: <http://doi.org/10.1016/j.joi.2017.06.005>.
10. Delgado López-Cózar, E., Robinson-García, N., Torres-Salinas, D. (2014), "The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators", *Journal of the American Society for Information Science and Technology*, No. 3 (65), P. 446–454. DOI: <http://doi.org/10.1002/asi.23056>.
11. Bornmann, L., Thor, A., Marx, W. et al. (2016), "The application of bibliometrics to research evaluation in the humanities and social sciences: An exploratory study using normalized Google Scholar data for the publications of a research institute", *Journal of the Association for Information Science and Technology*, No. 11 (67), P. 2778–2789. DOI: <http://doi.org/10.1002/asi.23627>.
12. Соловяненко Д. Цифровий ідентифікатор об'єкта (DOI) : "ISBN суспільства знань". *Бібл. вісник*. 2009. № 4. С. 3–15.
13. Shotton, D. (2013), "Open citations", *Nature*, No. 7471 (502), P. 295–297. DOI: <http://doi.org/10.1038/502295a>.
14. Peroni, S., Dutton, A., Gray, T. et al. (2015), "Setting our bibliographic references free: towards open citation", *Journal of Documentation*, No. 2 (71), P. 253–277. DOI: <http://doi.org/10.1108/JD-12-2013-0166>.
15. Kamińska, A. M. (2017), "OpenCitations (I4OC) – otwarty indeks cytowań publikacji naukowych", *Biuletyn EBIB*, No. 6 (176), P. 1–9.
16. Bastian, M., Heymann, S., Jacomy, M. (2009), "Gephi: an open source software for exploring and manipulating networks", *Proceedings of the Third International AAAI Conference on Weblogs and Social Media*, P. 361–362.
17. Kamińska, A.M. (In press), "ScientoMiner ICR – moduł importu danych bibliograficznych z zasobów Crossref dla platformy Gephi", *Zagadnienia Informacji Naukowej*.
18. Van Eck, N. J., Waltman L. (2014), "Visualizing bibliometric networks", *Measuring scholarly impact : Methods and practice*, Springer, P. 285–320. DOI: http://doi.org/10.1007/978-3-319-10377-8_13.

19. МОН планує запустити відкритий український індекс наукового цитування до кінця року. Міністерство освіти та науки України : Веб-сайт. URL : <https://mon.gov.ua/ua/news/mon-planuye-zapustiti-vidkritij-ukrayinskij-indeks-naukovogo-cituvannya-do-kincy-a-roku>.

20. Gläser, J., Glänzel, W., Scharnhorst, A. (2017), "Same data–different results? Towards a comparative approach to the identification of thematic structures in science", *Scientometrics*, No. 2 (111), P. 981–998. DOI: <http://doi.org/10.1007/s11192-017-2296-z>.

References

- Kamińska, A. M. (2017), "ProBIT – perspektywna metoda tworzenia trawersowalnych indeksów cytowań a współczesne problemy organizacji przestrzeni informacji w tradycyjnych bibliograficznych bazach danych", *Zagadnienia Informacji Naukowej*, No. 1 (55), P. 66–82.
- Moed H. F., Luwel, M., Nederhof, A. J. (2002), "Towards indicators of research performance in the humanities", *Library Trends*, No. 50, P. 498–520.
- Nederhof, A. J. (2006), "Bibliometric monitoring of research performance in the social sciences and the humanities : A review", *Scientometrics*, No. 1 (66), P. 81–100. DOI: <http://doi.org/10.1007/s11192-006-0007-2>.
- Hammarfelt, B., Haddow, G. (2018), "Conflicting measures and values : How humanities scholars in australia and sweden use and react to bibliometric indicators", *Journal of the Association for Information Science and Technology*, No. 7 (69), P. 924–935. DOI: <http://doi.org/10.1002/asi.24043>.
- Kulczycki, E., Engels, T.C.E., Pölönen, J. et al. (2018), "Publication patterns in the social sciences and humanities: evidence from eight European countries", *Scientometrics*, No. 1 (116), P. 463–486. DOI: <http://doi.org/10.1007/s11192-018-2711-0>.
- Sterligov, I. (2018), "Reflection and Transformation, or Humanities under the Influence of Scientometrics" ["Otrazhenie i preobrazhenie, ili gumanitarii pod vozdeystviem naukometrii"], *Measurements of Science*, No. 1 (6), P. 1–4.
- Sile L., Pölönen, J., Sivertsen, G. et al. (2018), "Comprehensiveness of national bibliographic databases for social sciences and humanities: Findings from a European survey", *Research Evaluation*. DOI: <http://doi.org/10.1093/reseval/rvy016>.
- Symonenko, T. (2016), "Global bibliometrics: conceptual model" ["Hlobal'na bibliometryka: kontseptual'na model'"], *Bulletin of the Book Chamber*, No. 6, P. 12–14.
- Halevi, G., Moed, H., Bar-Ilan, J. (2017), "Suitability of Google Scholar as a source of scientific information and as a source of data for scientific evaluation – Review of the Literature", *Journal of Informetrics*, No. 3 (11), P. 823–834. DOI: <http://doi.org/10.1016/j.joi.2017.06.005>.
- Delgado López-Cózar, E., Robinson-García, N., Torres-Salinas, D. (2014), "The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators", *Journal of the American Society for Information Science and Technology*, No. 3 (65), P. 446–454. DOI: <http://doi.org/10.1002/asi.23056>.
- Bornmann L., Thor A., Marx W. et al. (2016), "The application of bibliometrics to research evaluation in the humanities and social sciences: An exploratory study using normalized Google Scholar data for the publications of a research institute", *Journal of the Association for Information Science and Technology*, No. 11 (67), P. 2778–2789. DOI: <http://doi.org/10.1002/asi.23627>.
- Solovyanenko, D. (2009), "The Digital Object Identifier (DOI) : ISBN of Knowledge Society" [Cyfrovjy identyfikatory ob'jekta (DOI) : ISBN suspil'stva znan'], *Library Bulletin*, No. 4, P. 3–15.
- Shotton, D. (2013), "Open citations", *Nature*, No. 7471 (502), P. 295–297. DOI: <http://doi.org/10.1038/502295a>.
- Peroni, S., Dutton, A., Gray, T. et al. (2015), "Setting our bibliographic references free: towards open citation", *Journal of Documentation*, No. 2 (71), P. 253–277. DOI: <http://doi.org/10.1108/JD-12-2013-0166>.
- Kamińska, A. M. (2017), "OpenCitations (I4OC) – otwarty indeks cytowań publikacji naukowych", *Biuletyn EBIB*, No. 6 (176), P. 1-9.
- Bastian, M., Heymann, S., Jacomy, M. (2009), "Gephi: an open source software for exploring and manipulating networks", *Proceedings of the Third International AAAI Conference on Weblogs and Social Media*, P. 361–362.
- Kamińska, A. M. (In press), "ScientoMiner ICR – moduł importu danych bibliograficznych z zasobów Crossref dla platformy Gephi", *Zagadnienia Informacji Naukowej*.
- Van Eck, N. J., Waltman L. (2014), "Visualizing bibliometric networks", *Measuring scholarly impact : Methods and practice*, Springer, P. 285–320. DOI: http://doi.org/10.1007/978-3-319-10377-8_13.
- The Ministry of Education and Science of Ukraine plans to launch an Open Ukrainian Scientific Citation Index by the end of the year [online], available at : <https://mon.gov.ua/ua/news/mon-planuye-zapustiti-vidkritij-ukrayinskij-indeks-naukovogo-cituvannya-do-kincy-a-roku> (Accessed: 30 July 2018).
- Gläser, J., Glänzel, W., Scharnhorst, A. (2017), "Same data–different results? Towards a comparative approach to the identification of thematic structures in science", *Scientometrics*, No. 2 (111), P. 981–998. DOI: <http://doi.org/10.1007/s11192-017-2296-z>.

Надійшла (Received) 10.08.2018

Відомості про авторів / Сведения об авторах / About the Authors

Камінська Анна Малгожата – Інститут бібліотекознавства і наукової інформації, Сілезький університет у Катовицях, ад'юнкт кафедри бібліотекознавства, Польща, e-mail: anna.kaminska@us.edu.pl; ORCID ID: <https://orcid.org/0000-0001-5411-5426>.

Каминская Анна Малгожата – Институт библиотековедения и научной информации, Силезский университет в Катовицах, ад'юнкт кафедри бібліотекознавства, Польща.

Kamińska Anna Malgorzata – Institute of Library and Information Science, University of Silesia in Katowice, Assistant Professor at the Department of Library Science, Poland.

Назаровець Сергій Андрійович – кандидат наук із соціальних комунікацій, Державна науково-технічна бібліотека України, заступник директора з наукової роботи, м. Київ, Україна; e-mail: nazarovets@gntb.gov.ua; ORCID ID: <https://orcid.org/0000-0002-5067-4498>.

Назаровец Сергей Андреевич – кандидат наук по социальным коммуникациям, Государственная научно-техническая библиотека Украины, заместитель директора по научной работе, г. Киев, Украина.

Nazarovets Serhii – PhD (Social Communication Sciences), State Scientific and Technical Library of Ukraine, Deputy Director for Research, Kyiv, Ukraine.

CROSSREF КАК ИСТОЧНИК НАУКОМЕТРИЧЕСКИХ ДАННЫХ ДЛЯ СОЦИАЛЬНЫХ И ГУМАНИТАРНЫХ НАУК

Современная наукометрия успешно использует ряд аналитических методов для исследования научных публикаций, однако, применение даже самых продвинутых аналитических методов к ошибочным или неполным данным приводит к ошибочным результатам. **Предметом** исследования в статье являются современные подходы и технологии получения библиографических данных из надежных открытых источников научной информации для проведения наукометрических исследований. **Целью** данной работы является презентация разработанного модуля ScientoMiner ICR, который расширяет функциональные возможности аналитической платформы Gephi и позволяет импортировать библиографические данные из общедоступных записей Crossref для проведения наукометрических исследований в области анализа цитирований, качество которых гарантировано применением системы идентификации DOI. В статье решаются следующие **задачи**: оценены перспективы развития Initiative for Open Citations для создания некоммерческих открытых инструментов отслеживания научных цитирований; обосновано использование системы DOI и базы данных регистрационного агентства Crossref, как надежного источника библиографических данных и механизма получения метаданных научных изданий; раскрыты возможности модуля ScientoMiner ICR для открытой платформы Gephi для импорта библиографических данных. Используются **методы исследования** – визуализация и анализ сети научных цитирований. **Вывод** – предложенный модуль ScientoMiner ICR для платформы Gephi упрощает процесс автоматического экспорта данных Crossref непосредственно в Gephi, в частности, для нужд проведения наукометрического анализа. Полученные **результаты** могут быть использованы в процессе создания новых некоммерческих инструментов отслеживания цитируемости научных работ, прежде всего, в области социальных и гуманитарных наук, которые недостаточно хорошо представлены в базах данных коммерческих наукометрических продуктов.

Ключевые слова: наукометрия; DOI; Crossref; Open Citations; Gephi; ScientoMiner.

CROSSREF AS A SOURCE OF SCIENTOMETRIC DATA FOR SOCIAL & HUMAN SCIENCES

Modern scientometrics successfully uses a number of analytical methods for studying scientific publications, however, the application of even most advanced analytical methods to erroneous or incomplete data leads to false results. The **subject** matter of the article is modern approaches and technologies for obtaining bibliographic data from reliable open sources of scientific information for scientometric research. The **goal** of this work is to present the developed module ScientoMiner ICR, which extends the functionality of the analytical platform Gephi and enables importing bibliographic data from the public records of Crossref to conduct scientometric research in the field of citation analysis, the quality of which is guaranteed by using the DOI identification system. The following **tasks** are solved in the article: prospects for the development of the Initiative for Open Citations to create non-commercial open tools for tracking scientific citations; the use of the DOI system and the database of the Crossref registration agency is justified as a reliable source of bibliographic data and a mechanism for obtaining metadata of scientific publications; the capabilities of the ScientoMiner ICR module for the open Gephi platform for importing bibliographic data are revealed. The following **methods of research** are used - visualization and the analysis of the network of scientific citations. **Conclusion** – the suggested ScientoMiner ICR module for the Gephi platform simplifies the process of automatic export of Crossref data directly to Gephi, in particular, for the needs of scientometric analysis. The obtained **results** can be used in the process of creating new non-commercial tools for tracking the rate of scientific paper citation, especially in the social and human sciences, which are not presented in the databases of commercial scientometric products well enough.

Keywords: scientometrics; DOI; Crossref; Open Citations; Gephi; ScientoMiner.