

DOI <https://doi.org/10.30525/978-9934-26-277-7-200>

**USE OF ONTOLOGIES AND SEMANTIC MEDIAWIKI
FOR REPRESENTATION AND RETRIVAL OF SCIENTIFIC DATA
IN THE FAIR PARADIGM**

**ВИКОРИСТАННЯ ОНТОЛОГІЙ ТА SEMANTIC MEDIAWIKI
ДЛЯ ПОДАННЯ ТА ПОШУКУ НАУКОВИХ ДАНИХ
У ПАРАДИГМІ FAIR**

Rogushina Yu. V.

*Candidate of Physical and Mathematical
Sciences, Associate Professor,
Senior Researcher
Institute of Software Systems
of the National Academy of Sciences
of Ukraine
Kyiv, Ukraine*

Рогущина Ю. В.

*кандидат фізико-математичних
наук, доцент,
старший науковий співробітник
Інститут програмних систем
НАН України
м. Київ, Україна*

Одним з суттєвих факторів ефективності наукових досліджень є забезпечення відкритого доступу до попередніх результатів, що отримані іншими дослідниками у відповідній галузі. Це дозволяє запобігти повторного виконання складної роботи, зменшить час та ресурси, що потрібні для виконання різноманітних проектів. Крім того, можливість ознайомитися з напрямками досліджень різних організацій є основою для створення наукових спільнот, співпраці та впровадження колективних проектів.

Парадигма відкритої науки спрямована на поширення та більш ефективне використання наукових знань. Вона базується на використанні цифрових технологій з метою переходу від традиційних методів публікації результатів досліджень у наукових виданнях до спільної роботи дослідників шляхом вільного обміну загальнодоступними результатами досліджень [1]. Її можна розглядати як новий підхід до наукового процесу, який стосується технологій поширення наукових знань та підтримки колективної роботи територіально розподілених спільнот. Слід враховувати, що поняття відкритої науки має багато аспектів: відкритий доступ до публікацій, відкриті дані та відкрите рецензування наукових досліджень, відкриту методологію, відкриту освіту та відкрите програмне забезпечення, але основними передумовами її використання є розробка інфраструктури для доступу,

пошуку та використання наукових знань, до яких відносяться як наукові публікації, так і дані, що згенеровані в процесі відкритих досліджень. Європейська хмара відкритої науки (European Open Science Cloud – EOSC) – це мережа сховищ даних та сервісів, яка інтегрує дослідницькі інфраструктури для підтримки наукових досліджень [2]. Це дозволяє дослідникам знаходити наукові публікації та набори даних, забезпечуючи основу для створення нових знань.

Інтеграція України до цієї спільноти потребує низки як організаційних, так і технологічних рішень, які відображені Національному плані щодо відкритої науки (розпорядження Кабінету Міністрів України від 8 жовтня 2022 р.), в якому передбачаються заходи, що спрямовані на забезпечення відкритого доступу до наукових результатів та науково-технічної інформації, що утворюється в процесі виконання наукових досліджень. Серед цих заходів розглянуто забезпечення принципів належного управління науковими даними (FAIR – "FAIR Guiding Principles for scientific data management and stewardship") [3]. Акронім FAIR (Findable, Accessible, Interoperable, Reusable) був введений в 2016 р. для зручності ведення дискусії щодо принципів організації даних у відкритій науці. Згідно FAIR, функції пошуку, здобуття і представлення даних реалізують не користувачі, а інформаційна система. При цьому мова йде не тільки про самі дані і метадані, але і про алгоритми й інструменти керування ними. До розробки підходів до керування науковими даними залучаються всі зацікавлені сторони: науково-дослідні організації й окремі вчені; оператори баз даними і видання, що публікують наукові статті і результати експериментів; організації, що фінансують ці наукові дослідження; виробники програмного забезпечення й інструментів обробки даних; компанії, що надають послуги з аналізу й інтерпретації даних. Важливо, що в коло зацікавлених сторін також включаються самі обчислювальні системи (алгоритми обробки даних) як самостійний об'єкт – у залежності від їхнього рейтингу приймається рішення про включення обчислювального методу до конфігурації.

Крім широкого спектру спеціалізованих сервісів аналізу даних, які потрібні для розбудови відкритої науки, однією з базових функцій такої системи є *пошук інформації* на змістовному рівні, який має забезпечити повторне використання результатів раніше виконаних досліджень та експериментів.

Пошукові сервіси мають не тільки забезпечувати доступ до різних типів інформації (природномовних документів і мультимедійних даних) і підтримувати здобуття знань зі знайденої інформації,

але й надавати можливість знаходження складних інформаційних об'єктів, що відповідають вимогам користувачів. Прикладами таких об'єктів є:

- різноманітні обчислювальні засоби та ресурси, що можуть бути отримані в тимчасове користування,
- наукові колективи територіально розподілених дослідників, які об'єднуються для виконання спільного проекту;
- створення колективних монографій або тематичних видань, публікація статей у найбільш релевантних виданнях.

Тому важливою часткою досліджень є створення системи метаданих для опису не тільки інформаційних ресурсів відкритої науки, але й усіх її об'єктів та суб'єктів.

Передумовами для цього є створення порталних версій науково-технічних ресурсів з відкритим доступом – електронних бібліотек, галузевих довідників та енциклопедій, архівів та наукометричних баз даних. Але, крім наявності доступу та підтримки загальноприйнятих стандартів метаданих (таких як Dublin Core), ефективне використання таких ресурсів потребує також більш специфічних засобів навігації та пошуку в цих ресурсах, які дозволили б задовольнити персоніфіковані інформаційні потреби дослідників, які описувалися б у терміно-системах відповідних галузей знань та наукових напрямків. Великий потенціал для цього мають семантично розмічені Wiki-ресурси, які можуть інтегруватися із зовнішніми онтологіями тих предметних областей (PrO), що цікавлять користувачів та відповідають напрямкам їх наукових досліджень. Онтології мають надати спільну терміносистему, однозначно визначивши зміст використаних понять та відношень між ними, та зробити спілкування членів наукового співтовариства більш ефективним. Одним з аспектів цієї роботи має бути побудова онтологічних моделей конкретних спільнот, проектів та задач, яка дозволить створювати та аналізувати релевантні набори Big Data.

Онтології як основа семантизації інформаційних ресурсів

Семантизація інформаційних ресурсів полягає у зв'язуванні елементів контенту з формально описаними поняттями певної PrO. На сьогодні у Web-орієнтованих інтелектуальних система найчастіше для цього використовуються онтології, подання яких базується на відкритих стандартах Semantic Web. Створення семантичної розмітки наукових інформаційних ресурсів України на основі таких онтологій не тільки забезпечує інтеграцію у світове/європейський науковий інформаційний простір, але й забезпечує розширені сервіси

семантичного пошуку у різних галузях науки (приклади такого пошуку демонструє портална версія Великої української енциклопедії e-VUE – vue.gov.ua). Саме на основі досвіду розробки цього порталу ми пропонуємо створити семантичний Wiki-довідник, в якому розробниками будуть забезпечені шаблони для введення типових об'єктів наукових досліджень (статті, монографії, проекти тощо), а контент зможуть вводити й уточнювати самі науковці.

У середовищі Semantic MediaWiki онтології можуть не тільки використовуватися як базис семантичної розмітки, але й поповнюватися на основі сторінок з такою розміткою відповідно до запитів користувачів. Це дозволяє генерувати персоніфіковані онтології, що відображають уявлення про світ різних наукових спільнот та окремих дослідників. Щоб інтегрувати ці персоніфіковані уявлення, доцільно розробляти методи та програмні засоби співставлення та вирівнювання таких онтологій зі спільною онтологією наукового співтовариства. Це дозволить, з одного боку, явно представляти відмінності поглядів різних наукових груп та напрямків, а з іншого – забезпечить ефективний пошук потрібної інформації незалежно від того, яку саме терміносистему обирають для метаданих розробники окремих ресурсів.

Семантичні Wiki-технології та FAIR

Технологія Wiki та її семантичне розширення відповідають базовим принципам FAIR. Більш детальне обґрунтування цього, що базується на практичному досвіді розробки e-VUE, наведено у дослідженні [4] "FAIR Guiding Principles for scientific data management and stewardship" (Findable, Accessible, Interoperable, Reusable).

Семантичні розширення Wiki-технологій, наприклад, на основі Semantic MediaWiki, – це потужне рішення для спільного редагування даних та їх метаописів, створення різних довільних наборів властивостей в шаблонах цих метаописів, з одночасним поданням їх як в машинно-оброблюваній формі, так і формі придатній для розуміння людиною, що в результаті надає можливість оперувати цими даними, автоматизовано керувати, проводити аналіз, публікувати. Вбудовані можливості Semantic MediaWiki забезпечують завантаження файлів різного формату і додавання до них метаданих з різним набором атрибутів, які можливо змінювати, доповнювати. Інформаційні ресурси, які будуються в цьому середовищі, відповідають усім вимогам FAIR до відкритих даних великого обсягу.

Даний підхід апробовано в процесі створення бази знань порталної версії Великої української енциклопедії (vue.gov.ua),

яка є джерелом мультигалузових інтероперабельних знань, що придатні для повторного використання в інших інтелектуальних застосуваннях [5].

Висновки

Пропонується доповнити забезпечення відкритого доступу до наукових результатів та науково-технічної інформації створенням семантичних Wiki-довідників наукових установ як окремих країн, так і спеціалізованих напрямків досліджень, об'єднання яких дозволить доповнити пошук та навігацію у публікаціях не тільки за формальними ознаками, а й на змістовному рівні певної галузі знань або наукової школи, знання яких представлені у вигляді комп'ютерної онтології.

Важливою особливістю запропонованого підходу є те, що, хоча розробка такої онтології виконується спеціалістами з онтологічного інженірингу у співпраці з експертами предметних областей, але використання технології Wiki дозволяє вільно застосовувати цю модель широким колам користувачів (як для розмітки, так і для пошуку інформації) без спеціальних знань та без потреби у встановленні додаткового програмного забезпечення. Створення семантичного Wiki-довідника наукових ресурсів України є не альтернативою забезпечення відкритого доступу до наукових публікацій та результатів досліджень, а його доповненням для розширення функціоналу пошуку в цих ресурсах з використанням знань щодо окремих областей досліджень.

Література:

1. Draft recommendation on open science – UNESCO General Conference September 8, 2021. URL: https://unesdoc.unesco.org/ark:/48223/pf0000378841_rus (дата звернення: 20.10.2022).
2. Budroni, P., Claude-Burgelman, J., & Schouppe, M. (2019). Architectures of knowledge: the European open science cloud. *ABI Technik*, 39(2), 130–141.
3. FAIR_data. URL: https://en.wikipedia.org/wiki/FAIR_data (дата звернення: 25.10.2022)
4. Rogushina, Y. V., Grishanova, I. J. (2022). Study of principles, models and methods of FAIR paradigm of scientific data management for analysis for BIG data metadata. *Problems in programming*, (4), 26–35.
5. Rogushina J. V., Grishanova I. J. Ontological methods and tools for semantic extension of the media WIKI. Proc. of the 12th International Scientific and Practical Conference of Programming (UkrPROG 2020),

CEUR Workshoop Proceedings. 2021. Vol. 2866. P. 61–73.
URL: http://ceur-ws.org/Vol-2866/ceur_61-73Rogushina6.pdf (дата
звернення: 20.10.2022).

DOI <https://doi.org/10.30525/978-9934-26-277-7-201>

SYSTEM OF SEARCHING FOR VAGUE DUPLICATES IN ELECTRONIC TEXTS

СИСТЕМА ПОШУКУ НЕЧІТКИХ ДУБЛІКАТІВ В ЕЛЕКТРОННИХ ТЕКСТАХ

Rozlomii I. O.

*PhD,
Senior Lector at the Department
of Information Technologies
Bohdan Khmelnytsky
National University of Cherkasy
Cherkasy, Ukraine*

Розломій І. О.

*кандидат технічних наук,
старший викладач кафедри
інформаційних технологій
Черкаський національний
університет
імені Богдана Хмельницького
м. Черкаси, Україна*

Veretelnik V. V.

*PhD,
Docent at the Department
of Information Technologies
Bohdan Khmelnytsky
National University of Cherkasy
Cherkasy, Ukraine*

Веретельник В. В.

*кандидат технічних наук, доцент,
доцент кафедри
інформаційних технологій
Черкаський національний
університет
імені Богдана Хмельницького
м. Черкаси, Україна*

Hrushovyi V. O.

*Master at the Department
of Information Technologies
Bohdan Khmelnytsky
National University of Cherkasy
Cherkasy, Ukraine*

Грушовий В. О.

*магістр кафедри
інформаційних технологій
Черкаський національний
університет
імені Богдана Хмельницького
м. Черкаси, Україна*

Сучасний розвиток інформаційних технологій та мережі Інтернет надав широким колам користувачів доступ до великих об'ємів