

УДК 002.513.5:004.75.05

**Катерина Лобузінa,**д-р наук із соціальних комунікацій, ст. наук. співробітник,  
зав. відділу Національної бібліотеки України імені В. І. Вернадського

## **Електронна наукова періодика відкритого доступу: семантичні веб-технології для бібліотек**

Розглянуто проблеми інтеграції бібліографічних даних та бібліотечних електронних ресурсів у наукові інтернет-комунікації на основі технологій семантичного вебу. Запропоновано рішення щодо інтеграції електронної бібліотеки наукових періодичних видань України у систему пошуку наукової інформації Google Академія. Проаналізовано перспективи використання та можливі шляхи реалізації технологій семантичного пошуку Веб 3.0 у бібліотеках.

К л ю ч о в і с л о в а: семантичний веб, семантичний пошук, Веб 3.0, системи пошуку наукової інформації, веб-технології у бібліотеках, наукова періодика відкритого доступу.

**Н**аявність величезних обсягів слабоструктурованої інформації у сучасному веб-середовищі унеможливила задоволення інформаційних потреб користувачів за допомогою простого навігаційного пошуку. У багатьох випадках засоби статичного перегляду веб-сторінок вкрай неефективні, або взагалі непридатні для вирішення завдань пошуку наукової інформації. Зі зміною поглядів на організацію інформації та пошукових систем глобальної мережі з'явилися нові технологічні моделі проектування і створення веб-ресурсів. Сьогодні ми можемо спостерігати розвиток технології Веб 3.0, яка ґрунтується на ідеях семантичного вебу. Останні зводяться до того, що додавання певної структури даних (метаданих) неструктурованому контенту глобальної мережі робить його більш зрозумілим пошуковим роботам, що, у свою чергу, дає можливість перетворити ресурси Інтернету в аналог бази даних. Ці процеси у середовищі веб-технологій спричинили посилення уваги до бібліотечних технологій, бо саме в них накопичений великий досвід упорядкування знань.

Питання про те, що може привнести технологічна модель Веб 3.0 у бібліотечну діяльність і як виглядатиме Бібліотека 3.0 або Семантична Бібліотека порушувались ще у 2007 р. на Генеральній конференції IFLA (Південна Африка) [12]. За узагальненим прогнозом, основною місією бібліотек у найближчому майбутньому буде упорядкування неорганізованих електронних ресурсів у систематичні категорії. Для цього використовуватимуться семантичні веб-технології, онтології та метадані. Великого значення набудуть електронні каталоги бібліотек як сховища метаданих документів різної природи. Набудуть чинності стандарти бібліотечного упорядкування на все веб-середовище; спостерігатиметься подальша інтеграція електронних каталогів бібліотек з наданням доступу до електронних інформаційних ресурсів; створюватимуться

бібліотечні портали, що надаватимуть доступ до всіх бібліотечних ресурсів у режимі «єдиного вікна». Запровадження цих технологій передбачається приблизно у 2016 р. [5; 6]. Зважаючи на широкі можливості, які відкривають технології семантичного вебу для бібліотек, на 77-му Конгресі ІФЛА в Пуерто-Рико (2011 р.) було утворено спеціальну групу за інтересами – Semantic Web Special Interest Group (SWSIG), орієнтовану на вивчення цього напрямку (<http://www.ifla.org/about-swsig>).

Сьогодні процеси інтеграції бібліотечних та веб-технологій особливо інтенсивно розвиваються у галузі організації доступу до наукової інформації. Набуття популярності у всьому світі ідеї відкритого доступу до наукової інформації, спричинилося до накопичення величезних обсягів наукових та освітніх ресурсів в електронному вигляді. Загальнодоступним інтегратором цих ресурсів стала глобальна пошукова система Google Академія (Google Scholar), призначена для виокремлення із глобального інформаційного потоку джерел наукової інформації. Останнім часом відбулись істотні зміни в наукометричних та бібліометричних можливостях системи. Зокрема, було удосконалено алгоритми екстрагування метаданих веб-сторінок та автоматичного виявлення цитування наукових публікацій, забезпечено можливість створення наукометричних профілів вчених та наукових фахових видань, що дає змогу авторам стежити за цитуванням своїх статей. Ці технологічні інновації в сервісних можливостях інформаційної системи Google Академія привернули увагу користувачів наукових інтернет-комунікацій. Якщо говорити про Україну, то тут слід відзначити розроблення інформаційного сервісу «Бібліометрика української науки», зорієнтованого на показники індексів цитування Google Академія [3].

Важливим показником ступеня інтеграції наукового е-ресурсу (бази даних, бази знань або інфор-

маційного порталу) в наукову комунікацію суспільства є OPENNESS (Відкритість). Цей показник входить до оцінок рейтингу Webometrics Ranking of World Universities, системи, яка визначає світовий рейтинг університетських сайтів відповідно до Берлінських принципів, визначених ЮНЕСКО (<http://www.webometrics.info>). Поняття відкритості передбачає глобальні зусилля зі створення цінних файлів (PDF, DOC, PPT), опублікованих та проіндексованих академічною пошуковою системою Google Академія. Під цінними файлами (Rich files) слід розуміти файли у форматах, які використовують автори для представлення та поширення власних робіт. Враховуються файли MS Word (doc), PowerPoint (ppt), Adobe Acrobat (pdf), а також деякі інші. Автори рейтингу вважають, що велика кількість таких документів засвідчує: на сайті зберігаються не лише адміністративна інформація, а й наукові матеріали.

Незмінною складовою міжнародних систем наукової інформації, які представляють в організованому інтерфейсі наукові публікації, стали також персональні профілі вчених. Вони уможливають введення та ідентифікацію персональної інформації за принципами інтерактивного обміну даними Веб 2.0. Самостійне формування спільнотою вчених наукових онлайн-мереж призвело до появи концепції та відповідного терміна наука 2.0 (Research 2.0, Science 2.0) або е-наука (e-Science) [2; 13]. У науковому інтернет-просторі персональний профіль вченого відіграє роль семантичної ланки між публікаціями та їх автором. Персональна інформація – необхідний атрибут метаданих, які повідомляються зовнішнім бібліографічним системам під час розміщення наукової публікації. Від даних персонального профілю слід відштовхуватися у процесі пошуку в Інтернеті робіт певного автора, підрахунку числа цитувань, дослідження наукових зв'язків. У такий спосіб персональна сторінка перетворюється на ключовий елемент інтернет-інфраструктури науки [7]. Технологічні засоби системи Google Академія уможливають також автоматизоване збирання бібліографічних даних та обчислення наукометричних показників, пов'язаних з персональним профілем вченого. Створивши свій персональний профіль, автор може дізнатися, хто посилається на його публікації, переглянути діаграму цитування та обчислити індекси цитування. Наявна також функція автоматичного оновлення списку статей та попереднього перегляду нових надходжень або додавання матеріалів власноруч.

Розуміння важливості відкритості результатів наукової роботи для визначення наукового рейтин-

гу призвело до масового створення електронних інституційних репозиторіїв, які інтегрують та надають у відкритому доступі здобутки науково-дослідних та навчальних установ. Стандартні платформи інституційних репозиторіїв мають готові рішення для постачання метаданих до систем-агрегаторів через протокол OAI-PMH. Цей протокол також дає змогу інтегрувати бібліографічні дані в інформаційні сервіси Google Академія. Правильне постачання якісних метаданих значно підвищує цитованість наукових робіт та покращує якість визначення індексів цитування.

З розвитком інтеграційних процесів у сучасних наукових комунікаціях наукові бібліотеки опиняються перед певною технологічною ділемою: переходити на нову технологічну платформу електронних репозиторіїв [4], або адаптувати вже існуючі бібліотечні рішення до сучасних вимог представлення метаданих цитування. Кожне з цих рішень має як свої переваги, так і недоліки. Суттєвим недоліком першого рішення є необхідність повної перебудови всіх бібліотечних процесів, і, як наслідок, виокремлення репозиторія, що створюється із єдиної системи опрацювання та представлення електронних ресурсів бібліотеки. Друге рішення передбачає значне програмно-технологічне доопрацювання веб-інтерфейсу традиційних бібліотечних електронних каталогів, записи яких представлені у спеціалізованих бібліотечних форматах. Однак, як показав досвід, Національної бібліотеки Франції (НБФ) (The Bibliothèque nationale de France), введення до стандартних слабо-структурованих, з точки зору пошукових робіт, бібліографічних описів упорядкованої системи метаданих вирішує проблеми підвищення доступності бібліографічної інформації в інтернет-середовищі. Результати дослідження Н. Буле з впровадження різних схем метаданих паралельно з традиційним представленням бібліографічних описів представлені на інформаційному порталі «Data.bnf.fr» НБФ (<http://data.bnf.fr/>). Реалізовані в межах проекту технології семантичного вебу дають змогу зв'язувати дані з різних баз даних, каталогів і репозиторіїв, не змінюючи їх індивідуальної внутрішньої структури. Згідно з принципами відкритості, головною вимогою до організації даних є їх доступність і можливість багаторазового використання для різних інформаційних завдань. Реалізація цих принципів стосовно бібліотечних каталогів дає змогу розв'язати суперечність між наявними величезними масивами бібліотечних даних та незручністю їх використання у веб-середовищі. Як зазначають автори проекту, «Data.bnf.fr» є відкритим проектом даних на основі

семантичних веб-стандартів та інструментів. Він дає змогу інтегрувати в єдиній точці доступу інформацію різних каталогів НБФ, цифрової бібліотеки Галіка (Galica, <http://gallica.bnf.fr>) та інших веб-ресурсів [1].

Аналогічні проблеми виникли у Національній бібліотеці України імені В. І. Вернадського у процесі переведення загальнонаціонального порталу «Наукова періодика України» на нову технологічну платформу. Крім забезпечення стандартного бібліографічного опису наукових публікацій українських вчених, організації повнотекстового та інформаційного пошуку за метаданими, постало також завдання забезпечити інтеграцію української наукової періодики у відкриті наукові веб-ресурси.

Впровадження бібліотечної технології опрацювання ресурсів «Наукової періодики України» забезпечило включення електронної бібліотеки фахових періодичних видань до єдиного пошукового апарату наукової бібліотеки. Розроблені принципово нові технологічні рішення надали користувачам системи (авторам, редакціям, бібліотекарям, читачам) можливість:

- отримувати повну інформацію про публікацію: журнал, випуск, рік тощо;
- встановлювати термін затримки для надання відкритого доступу до повних текстів наукових публікацій;
- готувати правильний бібліографічний опис наукових публікацій відповідно до ДСТУ;
- вести пошук публікацій за «Авторським покажчиком» та «Покажчиком назв публікацій»;
- здійснювати оперативний добір всіх публікацій певного автора за гіперпосиланням у форматі перегляду знайденого бібліографічного опису;
- проводити розширений пошук за ключовими словами;
- робити перехід за гіперпосиланням від бібліографічного опису наукової публікації до повного опису журналу для з'ясування його фахової спрямованості та спеціальності ДАК України;
- робити перехід за гіперпосиланням від бібліографічного опису наукової публікації до змісту журналу для перегляду всіх публікацій випуску;
- попередньо переглядати повний текст публікації без його завантаження на комп'ютер користувача;
- інтегрувати інформацію «Наукової періодики України» та реферативної бази даних «Україніка наукова»: вести пошук повних текстів через реферативну інформацію та наведення наявних рефератів для бібліографічних описів повних текстів наукових публікацій;
- здійснювати повнотекстовий пошук наукових публікацій безпосередньо в інтерфейсі електронного каталогу наукової періодики.

Отже, створення професійної електронної бібліотеки періодичних видань вирішило завдання обліку та наукового опису електронних публікацій, забезпечило стабільне щоденне бібліографічне опрацювання 700–1000 електронних текстів окремих статей [8].

Для вирішення другої частини завдання – інтеграції наукової періодики України у глобальні наукові інформаційні системи, у першу чергу, потрібно було з'ясувати стандарти метаданих цитування (meta-citations) необхідні для формування пошукового образу наукової публікації. Стандартний список метаданих цитування Google Академія включає такі елементи [11]:

1.	Назва журналу	Citation_journal_title
2.	Видавець	Citation_publisher
3.	ISSN	Citation_issn
4.	Дата публікації	Citation_publication_date
5.	Том	Citation_volume
6.	Випуск	Citation_issue
7.	Перша сторінка	Citation_firstpage
8.	Остання сторінка	Citation_lastpage
9.	Назва публікації	Citation_title
10.	Автор(и)	Citation_author
11.	Інтернет-адреса анотації	Citation_abstract_html_url
12.	Інтернет-адреса повного тексту	Citation_pdf_url

Наявність на новій технологічній платформі повноцінних бібліографічних записів для кожної наукової публікації забезпечила автоматичне формування метаданих, необхідних для коректного індексування пошуковими роботами. Всі 12 наведених елементів присутні у будь-якому стандартному аналітичному описі статті. Для автоматичного опрацювання роботами Google Академія (Google Scholar Crawler) інформації про нові надходження наукових публікацій було також організовано щоденне оновлення спеціального XML-файла в стандарті протоколу OAI-PMH. На сьогодні Google Scholar Crawler проіндексував бібліографічні дані приблизно 350 тис. повних текстів публікацій «Наукової періодики України» із домену НБУВ. При опрацюванні наукової інформації інтелектуальними засобами системи Google Академія важливим елементом представлення наукових публікацій є наявність повного тексту у відкритому доступі. Це

створює передумови для автоматичного аналізу роботом пристатейної бібліографії (References). Аналіз списку бібліографічних посилань наукової публікації надалі використовується системою для автоматичного обчислення індексів цитування.

Індексування пошуковими роботами Google Академія величезного інтегрованого масиву наукової періодики України забезпечило постачання до глобальної наукової інформаційної системи якісних метаданих, необхідних для аналізу публікаційної активності та наукових зв'язків вітчизняних вчених. Попередня оцінка результатів індексування дає змогу стверджувати, що персональні профілі вчених, які публікуються у журналах, котрі перебувають у вільному доступі, на порталі НБУВ, значно збагатилися бібліографічною інформацією (кількість бібліографічних посилань зростає у 3–4 рази). Спостерігається також зростання показників цитування публікацій та виявлення раніше недоступних семантичних зв'язків представлених результатів наукової діяльності українських учених.

Набутий досвід показав, що бібліографічний опис наукових публікацій у стандартному інтерфейсі автоматизованої бібліотечної системи містить критично необхідний набір метаданих для їх коректного цитування. Так, стаття «Патон Б. Є. Пріоритети Національної стратегії теплозабезпечення населених пунктів України / Б. Є. Патон, А. А. Долінський, В. М. Геєць, В. П. Кухар, Б. І. Басок, Є. Т. Базеєв, Р. З. Подолець // Вісник Національної академії наук України. – 2014. – № 9. – С. 29–47», розміщена у «Науковій періодиці України» буде відображена у Google Академія зі всіма необхідними бібліографічними атрибутами та посиланням до повного тексту (рис. 1).

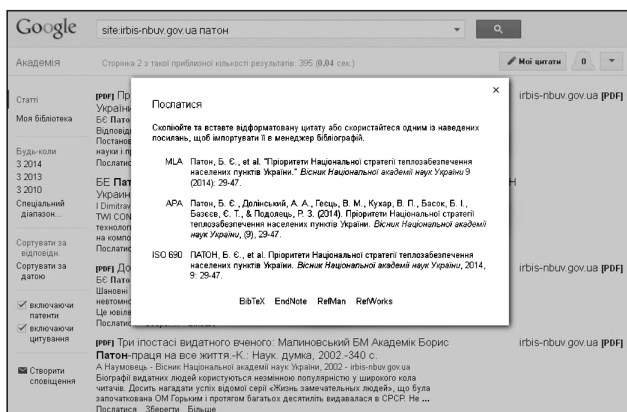


Рис. 1. Метадані цитування наукових публікацій українських вчених, представлені у форматах системи Google Академія

Інформація про публікацію буде також доступна через персональні профілі всіх співавторів публікації, якщо такі профілі наявні у системі.

Додатково до персонального представлення бібліографічної інформації, завдяки присутності у кожному описі наукової публікації елементу метаданих «citation\_journal\_title» – «Назва журналу», є також можливість створювати наукометричні профілі наукових журналів України. Проіндексовані Google Академія публікації з наукового видання можна переглянути за гіперпосиланням від опису журналу в «Науковій періодиці України» (рис. 2).

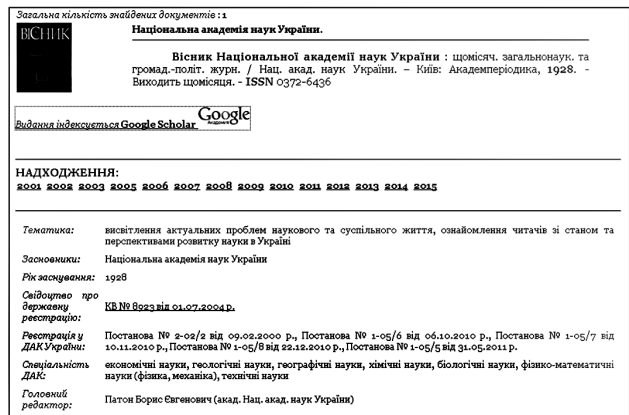


Рис. 2. Наукове фахове видання у пошуковому інтерфейсі інформаційної системи «Наукова періодика України»

Так, журнал «Вісник Національної академії наук України» сьогодні представлений у Google Академія 1110 публікаціями, розміщеними на порталі НБУВ. Є можливість добору відповідних публікацій за роками та перегляду цитувань у коректному вигляді з повним набором необхідних метаданих (рис. 3).

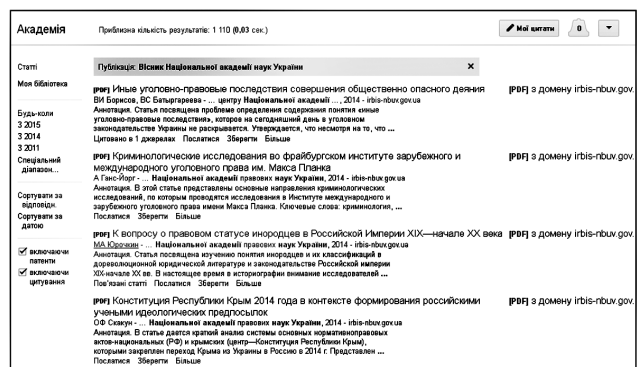


Рис. 3. Результати пошуку наукових публікацій у Google Академія за назвою періодичного видання

Система метаданих цитування, «зрозуміла» пошуковим машинам Інтернету, у поєднанні з розвинутим інформаційним апаратом бібліотечних каталогів забезпечила реалізацію роботи інструментарію технологій семантичного вебу – пошук інформації Веб 3.0. Це означає, що науковий пошук інформації може здійснюватися через інтерфейс пошукових машин Інтернету, таких як Google, на основі семантично пов'язаних даних [9; 10; 14]. Проіндексовані метадані

відсилають користувача не просто до повного тексту публікації, а до сторінки наукової публікації на порталі «Наукова періодика України», де він може продовжити інформаційний пошук за автором, знайти повний бібліографічний опис статті та її вихідні відомості. Крім того, інтерактивний взаємозв'язок бібліографічних записів та проіндексованих Google Академія метаданих дає змогу скористатися гіперпосиланнями для перевірки індексу цитування знайденої публікації та цитованості авторів наукової роботи, переглянути їх персональні профілі (рис. 4).

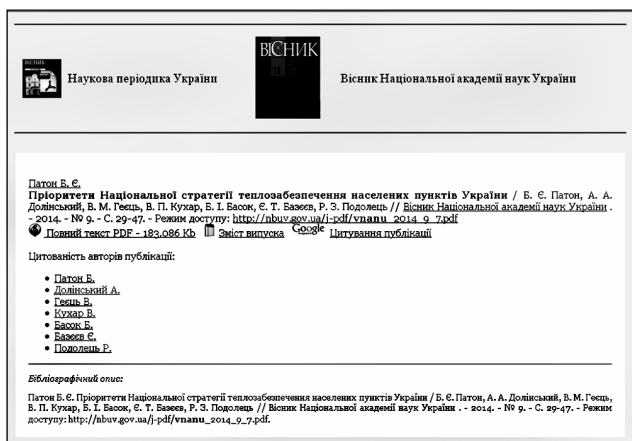


Рис. 4. Веб-сторінка наукової публікації в інформаційній системі «Наукова періодика України»

Здійснена робота з усією очевидністю показала перспективність використання бібліотечних технологій для упорядкування та структурування наукових веб-ресурсів, сприяла вирішенню проблеми інтеграції наукових фахових періодичних видань України в пошуковий та аналітичний інтерфейси глобальних наукових комунікацій. Інтелектуальні можливості безкоштовної наукометричної системи Google Академія дають змогу українським вченим інтегрувати свою діяльність у сучасні наукові веб-технології, покращити відкритість результатів наукової діяльності та персональну присутність у глобальному науковому співтоваристві. Ці технології також забезпечують можливість в оперативному режимі слідкувати за цитуванням наукових публікацій, виявляти тематично пов'язані напрями наукової діяльності та визначати коло колег, котрі активно працюють у споріднених галузях науки.

Подальші перспективи розвитку інформаційної системи «Наукова періодика України» пов'язані з удосконаленням представлення інформації про нові надходження, введенням тематичного упорядкування колекції електронних публікацій, розширенням аналітичних та бібліометричних можливостей розробленої платформи відповідно до вимог сучасних

наукових інформаційних систем. Ці рішення на базі бібліотечно-бібліографічних баз даних, за умов певного доопрацювання, можуть також використовуватися для створення сайтів наукових фахових видань з автоматичним постачанням метаданих до професійних наукометричних систем.

### Список використаних джерел

1. Буле В. Информационная среда, ориентированная на потребности пользователя: будущее за структурированными данными / В. Буле // Электронная библиотека. Вып. 4 : Научные и организационно-технологические основы интеграции цифровых информационных ресурсов : сб. научных трудов. – СПб, 2014. – С. 179–214.
2. Журавлёва Е. Ю. Научно-исследовательская инфраструктура Интернет // Вопр. философии. – 2010. – № 8. – С. 155–166.
3. Костенко Л. Бібліометрика української науки: інформаційно-аналітична система / Л. Костенко, О. Жабін, О. Кузнецов, Є. Кухарчук, Т. Симоненко // Бібліотечний вісник. – 2014. – № 4. – С. 8–12.
4. Костенко Л. Стратегії технологічної модернізації наукових бібліотек / Л. Костенко, Т. Симоненко // Бібліотечний вісник. – 2012. – № 6. – С. 53–55.
5. Лобузін К. В. Бібліотека 3.0: знання, сховища даних, експерти / К. В. Лобузін // Бібліотекознавство. Документознавство. Інформологія. – 2012. – № 1. – С. 26–35.
6. Лобузін К. В. Технології організації знань у бібліотечно-інформаційній діяльності : монографія / Катерина Лобузін ; відп. ред. О. С. Онищенко ; НАН України, Нац. б-ка України ім. В. І. Вернадського. – К., 2012. – 252 с.
7. Полилова Т. А. Персональные веб-страницы в научном сообществе [Электронный ресурс] // Труды Международной конференции «Научный сервис в сети Интернет: эксафлопное будущее» (19–21 сентября 2011 г., г. Новороссийск). – Новороссийск, 2011. – С. 476–479. – Режим доступа : <http://agora.guru.ru/abrau2011/pdf/476.pdf>.
8. Самохіна Н. Організація повнотекстових електронних ресурсів у Національній бібліотеці України імені В. І. Вернадського / Н. Самохіна // Бібліотечний вісник. – 2014. – № 6. – С. 13–17.
9. Hendler J. The Dawn of Semantic Search / J. Hendler // Computer. – 2010. – Vol. 43, № 1. – P. 111–113.
10. Gonzales B. M. Linking Libraries to the Web: Linked Data and the Future of the Bibliographic Record [Electronic resource] / B. M. Gonzales // Information Technology and Libraries. – 2014. – Vol. 33, № 4. – Mode of access : <http://ejournals.bc.edu/ojs/index.php/ital/article/view/5631>.
11. Google Scholar Help : Indexing Guidelines [Electronic resource]. – Mode of access : <http://scholar.google.com/intl/en/scholar/inclusion.html#indexing>.
12. Saw G. Library 3.0: where art our skills? / G. Saw, H. Todd // World library and information congress : 73rd IFLA general conference and council (19–23 August 2007, Durban, South Africa). – 2007. – P. 1–15.

13. *Ullmann T. D.* Components of a Research 2.0 Infrastructure / T. D. Ullmann, F. Wild, P. Scott, E. Duval, B. Vandeputte, G. Parra, W. Reinhardt, N. Heinze, P. Kraker, A. Fessel // Sustaining TEL : From Innovation to Learning and Practice. 5th European Conference on Technology Enhanced Learning, EC-TEL 2010, Barcelona, Spain, September 28 –

October 1, 2010. Proceedings – Berlin, Heidelberg : Springer-Verlag Heidelberg, 2010. – P. 590–595.

14. *Yadagiri N.* Semantic Web and the Libraries: An Overview / N. Yadagiri, P. Ramesh //International Journal of Library Science. – 2013.– Vol. 7, № 1. – P. 80–94.

Стаття надійшла до редакції 03.04.2015.

UDC 002.513.5:004.75.05

**Catherine Lobuzina,**

Head of Department of Information and Communication Technologies, The Vernadsky National Library of Ukraine

**OPEN ACCESS ELECTRONIC SCIENTIFIC JOURNAL: SEMANTIC WEB-TECHNOLOGY FOR LIBRARIES**

The problems of integration of bibliographic data and library electronic resources in scientific Internet communication based on Semantic Web technologies are considered. Solutions of integration of scientific periodicals of Ukraine electronic library in scientific information system Google Scholar are proposed. The prospects of the use and possible ways to implement semantic search technology Web 3.0 in libraries are analyzed.

**К е у в о р д с:** Semantic Web, semantic search, Web 3.0, scientific search engine, web-technology in libraries, open access scientific journal.

УДК 002.513.5:004.75.05

**Екатерина Лобузина,**

зав. отделом Национальной библиотеки Украины имени В. И. Вернадского

**ЭЛЕКТРОННАЯ НАУЧНАЯ ПЕРИОДИКА ОТКРЫТОГО ДОСТУПА: СЕМАНТИЧЕСКИЕ ВЕБ-ТЕХНОЛОГИИ ДЛЯ БИБЛИОТЕК**

Рассмотрены проблемы интеграции библиографических данных и библиотечных электронных ресурсов в научные интернет-коммуникации на основе технологий семантического веба. Предложены решения интеграции электронной библиотеки научных периодических изданий Украины в систему поиска научной информации Google Академия. Проанализированы перспективы использования и возможные пути реализации технологий семантического поиска Веб 3.0 в библиотеках.

**К л ю ч е в ы е с л о в а:** семантический веб, семантический поиск, Веб 3.0, системы поиска научной информации, веб-технологии в библиотеках, научная периодика открытого доступа.