**Проскудіна Галина Юріївна,**

ORCID <https://orcid.org/0000-0001-9094-1565>,

наукова співробітниця,

відділ теорії комп’ютерних обчислень,

Інститут програмних систем Національної академії наук України,

Київ, Україна

e-mail: [guproskudina@gmail.com](mailto:guproskudina@gmail.com)

**Інформаційно-пошукові можливості**

**Харвестера відкритої науки Національної академії наук** **України**

Розглянуто пошукові функції, що надаються користувачам Харвестера відкритої науки НАН України: простий і розширений пошук, пошук окремих слів, пошук по багатьом словам, використання групових символів, пошук по близькості звучання слів, пошук за фразами, пошук з використанням відстані між словами, пошук за важливістю слів або фраз, уточнення результатів пошуку тощо. Описано збереження та поширення пошукових запитів, фасетний і авторитетний пошук, виявлення подібних за тематикою робіт.

*Ключові слова:* простий пошук*,* розширений пошук, фасетний пошук, авторитетний пошук.

В рамках виконання цільового науково-технічного проєкту НАН України «Створення й впровадження інфраструктури відкритої науки в НАН України (ОРЕNS)» був розроблений Харвестер відкритої науки НАН України (Харвестер) [https://harvester.nas.gov.ua/] – система інтеграції наукових інформаційних ресурсів установ НАН України, мета якого полягає у сприянні більш ефективному використанню отриманих в НАН України наукових результатів завдяки суттєвому полегшенню процесу доступу до наукових інформаційних ресурсів з використанням засобів їх інтеграції та надання єдиної точки їх пошуку та доступу. Одне з ключових завдань цієї розробки – система повинна мати потужні пошукові засоби, має бути реалізована наскрізна пошукова система з підтримкою фільтрації, індексації, навігації та роботи з різними форматами даних.

У Харвестері механізм пошуку реалізований на пошуковій системі Apache Solr, яка в свою чергу основана на дослідницькій бібліотеці Lucene, яку на сьогоднішній день використовують більшість електронних бібліотечних систем. Зокрема, Наукова бібліотека періодичних видань НАНУ, яка функціонує з 2007 р., про пошукові можливості якої ми неодноразово розповідали в своїх роботах.

*Пошук у Харвестері* (Search) – це основна функція, що надається користувачам. Він дозволяє користувачам знаходити інформацію в бібліотечних колекціях і ресурсах. Процес включає в себе пошук книг, журналів, статей, електронних ресурсів, цифрових архівів і багатьох інших видів даних. Цей сервіс індексує всі наявні записи всіх бібліотек, підключених до системи, що дає змогу реалізувати простий і розширений пошук.

Пошук здійснюється з використанням інтуїтивних інтерфейсів, що робить його доступним і зручним для користувачів усіх рівнів. Користувачі можуть виконувати пошук за ключовими словами, авторами, темами та іншими критеріями, що допомагає їм знайти потрібні ресурси швидко і точно.

*Виявлення* (Discover) у Харвестері означає, що система надає не тільки результати пошуку, але й активно допомагає користувачам відкривати нові, цікаві для цього користувача ресурси. Це може включати в себе рекомендаційні системи, пов’язані ресурси, аналоги тощо. Наприклад, система може запропонувати користувачеві пов’язані статті або ресурси, засновані на його запиті, що сприяє глибшому дослідженню теми.

*Перегляд за фасетами* (Search with Faceted Results) у Харвестері є потужним інструментом для уточнення і навігації за результатами пошуку. Фасети (або фасетні поля) являють собою категорії, які групують результати пошуку за певними атрибутами або метаданими, наприклад, автор, організація, рік публікації, тема, тип ресурсу. Цей підхід дає змогу користувачам точніше визначити параметри, що їх цікавлять, і звузити результати пошуку.

*Авторитетний пошук* (Authority Search) широко застосовується в бібліотечних системах і наукових ресурсах, зокрема і в Харвестері, де стандартизація та ідентифікація авторів, термінів і тем є критично важливою для точного дослідження. Ось декілька прикладів: використання авторитетних файлів (Authority Files), нормалізація імен і термінів для дотримання стандартів і запобігання неоднозначності, уніфікація результатів запиту, тобто надання стандартизованих посилань на відповідні авторитетні файли, використання унікальних ідентифікаторів, наприклад, ORCID для дослідників.

*Дослідити зв’язки у VuFind* (Discover Relationships) – це спеціальний інструмент, модуль візуалізації зв’язків, який дає змогу знаходити й аналізувати зв’язки між різними ресурсами для визначення схожості, цитування, авторства, приналежності та інше. Зв’язки будуються на основі метаданих: авторів, тем, ключових слів, пов’язаних видань. Система надає інтерфейс для перегляду й аналізу зв’язків, які можуть бути налаштовані під конкретні потреби і завдання. Наприклад, дослідник може відстежити, як розвивалася певна наукова ідея, переглядаючи зв’язки між публікаціями різних років.

Харвестер надає розвинені функції пошуку та перегляду інформаційних ресурсів на рівні стандарту Apache Lucene та їх маніпулювання. Результати пошуку можуть бути роздруковані, відправлені по пошті, експортовані та збережені для подальшого використання.

Незареєстрованим (анонімним) користувачам надають усі можливості з пошуку, які передбачені Apache Lucene:

* + звуження простору пошуку: репозитарієм, автором, мовою, діапазоном дат, типом інформаційного ресурсу;
  + надання умов пошуку на значення метаданих, відсікання слів, використання в словах групових символів, пошук за близькістю звучання слова, пошук за фразами, пошук з використанням відстані між словами, пошук за важливістю слів або фраз, обов’язковою наявність слів або фраз, будування логічних пошукових виразів;
  + результати пошуку можна переглянути, зберегти, надрукувати або експортувати в інші бібліографічні формати, помітити тегами для подальшого використання.

Зареєстровані користувачі можуть робити все те ж саме, що й анонімні користувачі, а також їм надаються додаткові функції:

* + переглядати історію своїх пошукових запитів і результатів, а також зберігати їх під своїм обліковим записом;
  + створювати і редагувати списки вибраних документів, а також ділитися ними з іншими користувачами;
  + додавати або вибирати свої власні теги для каталогізації чи пошуку ресурсів у Харвестері;
  + залишати відгуки, коментарі та оцінки до документів, а також читати відгуки інших користувачів;
  + підписуватися для отримання новин про нові документи за темами, що цікавлять;
  + використовувати різні інструменти для роботи з документами, як-от експорт у формати BibTeX, EndNote, RefWorks, Zotero тощо;
  + надсилати електронною поштою, друкувати, цитувати тощо.

Загальна схема пошуку та перегляду полягає у наступному:

* + спочатку звужується простір пошуку з використанням фільтрів по: тематичним рубрикам, репозитарію, типу ресурсу, автору, мові та діапазону дат, потім
  + задається умова пошуку та ініціюється пошук;
  + результат пошуку надається у стислому вигляді, який можна розгорнути (при бажанні можна продивитись повний текст отриманого ресурсу);
  + отримані результати можна експортувати, друкувати, відправити по електронній пошті та зберегти;
  + більш детальне формулювання умов пошуку можна виконати на сторінці розширеного пошуку.

Розширений пошук дає змогу вказати точніші критерії для пошуку ресурсів, як-от тип пошуку, назви полів метаданих та логічний оператор. Це допомагає уникнути нерелевантних результатів і звузити область пошуку. Надається можливість комбінувати різні критерії для пошуку ресурсів, використовуючи кілька рядків пошуку і фасети. Це дає змогу створювати складні та різноманітні запити, які задовольняють різні потреби користувачів. Розширений пошук надає зручний інтерфейс для введення критеріїв пошуку, що спрощує вибір потрібних параметрів і позбавляє від необхідності запам’ятовувати синтаксис запитів.

Реалізація функцій підказок і рекомендацій спрямована на те, щоб зробити пошук більш інтуїтивним, персоналізованим і ефективним. Для цього використовується Solr SuggestComponent – спеціальний модуль, який індексує часто використовувані терміни і фрази. А також Solr MoreLikeThis, який аналізує текстові поля (наприклад, анотації, заголовки) і пропонує схожі ресурси. Рекомендації користувача здійснюються на основі історії пошуку або збережених записів користувача.

Таким чином в систему Харвестера закладено цілий ряд корисних інформаційно-пошукових функцій, які реалізовані на сучасних пошукових програмах. Більшість розглянутих функцій реалізовано, але деякі з них, наприклад, авторитетний пошук, все ще знаходяться на стадії налаштування.

**Halyna Proskudina**,

ORCID <https://orcid.org/0000-0001-9094-1565>,

Researcher,

Department of Computer Science Theory,

Institute of Software Systems of the National Academy of Sciences of Ukraine,

Kyiv, Ukraine

e-mail: [guproskudina@gmail.com](mailto:guproskudina@gmail.com)

**Information and search capabilities of the Harvester of Open Science**

**of the National Academy of Sciences of Ukraine**

The search functions provided to users are considered, namely: simple and advanced search, search for single words, search for many words, use of group symbols, search by word sounding similarity, search by phrases, search using the distance between words, search by importance of words or phrases, mandatory presence of words or phrases, improvement of search results. Described how to save or share a search query, faceted and authority search, identification of similar works.

*Keywords*: simple search, advanced search, faceted search, authority search.