

DOI <https://doi.org/10.15407/csc.2022.03.011>
УДК 004.318

М.Г. ПЕТРЕНКО, доктор техн. наук, ст. наук. співробітник, пров. наук. співробітник, Інститут кібернетики імені В.М. Глушкова НАН України, 03187, м. Київ, просп. Академіка Глушкова, 40, Україна, ORCID: <https://orcid.org/0000-0001-6440-0706>, petrng@ukr.net

О.В. ПАЛАГІН, доктор техн. наук, академік НАНУ, професор, заст. директора, Інститут кібернетики імені В.М. Глушкова НАН України, 03187, м. Київ, просп. Академіка Глушкова, 40, Україна, ORCID: <https://orcid.org/0000-0003-3223-1391>, palagin_a@ukr.net

М.О. БОЙКО, наук. співробітник, Інститут кібернетики імені В.М. Глушкова НАН України, 03187, м. Київ, просп. Академіка Глушкова, 40, Україна, ORCID: <http://orcid.org/0000-0003-1723-5765>, swsfrmac@gmail.com

С.М. МАТВЕЙШИН, к.т.н., наук. співробітник, Інститут кібернетики імені В.М. Глушкова НАН України, просп. Акад. Глушкова, 40, Київ 03187, Україна, ORCID: <https://orcid.org/0000-0001-9329-8800>, sergmmm2507@gmail.com

ЗНАННЯ-ОРІЄНТОВАНИЙ ІНСТРУМЕНТАЛЬНИЙ КОМПЛЕКС ОБРОБЛЕННЯ БАЗ ДАНИХ НАУКОВИХ ПУБЛІКАЦІЙ ІЗ ЗАСТОСУВАННЯМ WEB-ТЕХНОЛОГІЙ

Інструментальний комплекс баз даних наукових публікацій (ІК БДНП) зорієнтовано на пришвидшення отримання науковим співробітником, у якого в наявності вже опубліковано від декількох десятків до сотень наукових праць, необхідної йому когнітивно-структурованої інформації зі своїх джерел. Нам невідомі пошукові системи, які змогли б у максимально стислий термін надати науковому співробітнику таку інформацію (тобто за одним запитом). Інструментальний комплекс реалізує технології Information Retrieval і Knowledge Discovery in Databases з акцентом на технології й інструментарії Semantic Web та когнітивної графіки. Розроблення такого ІК припускає три стадії: на першій стадії створюються інструментальні засоби реалізації системи, методики й алгоритми взаємодії системи «Користувач – Інженер зі знань – Віддалена прикінцева точка» та наповнення її даними; на другій стадії розв'язуються задачі мультимедійного подання образно-понятійних структур, що описані в наукових документах, і на третій стадії – вирішення проблеми добування нових знань.

Ключові слова: технології Semantic Web, онтологічний інжиніринг, база даних наукових публікацій, трансдисциплінарні наукові дослідження, інформаційно-аналітична підтримка наукового співробітника.

Вступ

Відомі численні застосунки, що реалізують технології пошуку інформації в різних базах даних (БД). Але вони здебільшого не враховують

когнітивного аспекту обробки інформації, необхідного для творчої діяльності, зокрема наукового співробітника (НР).

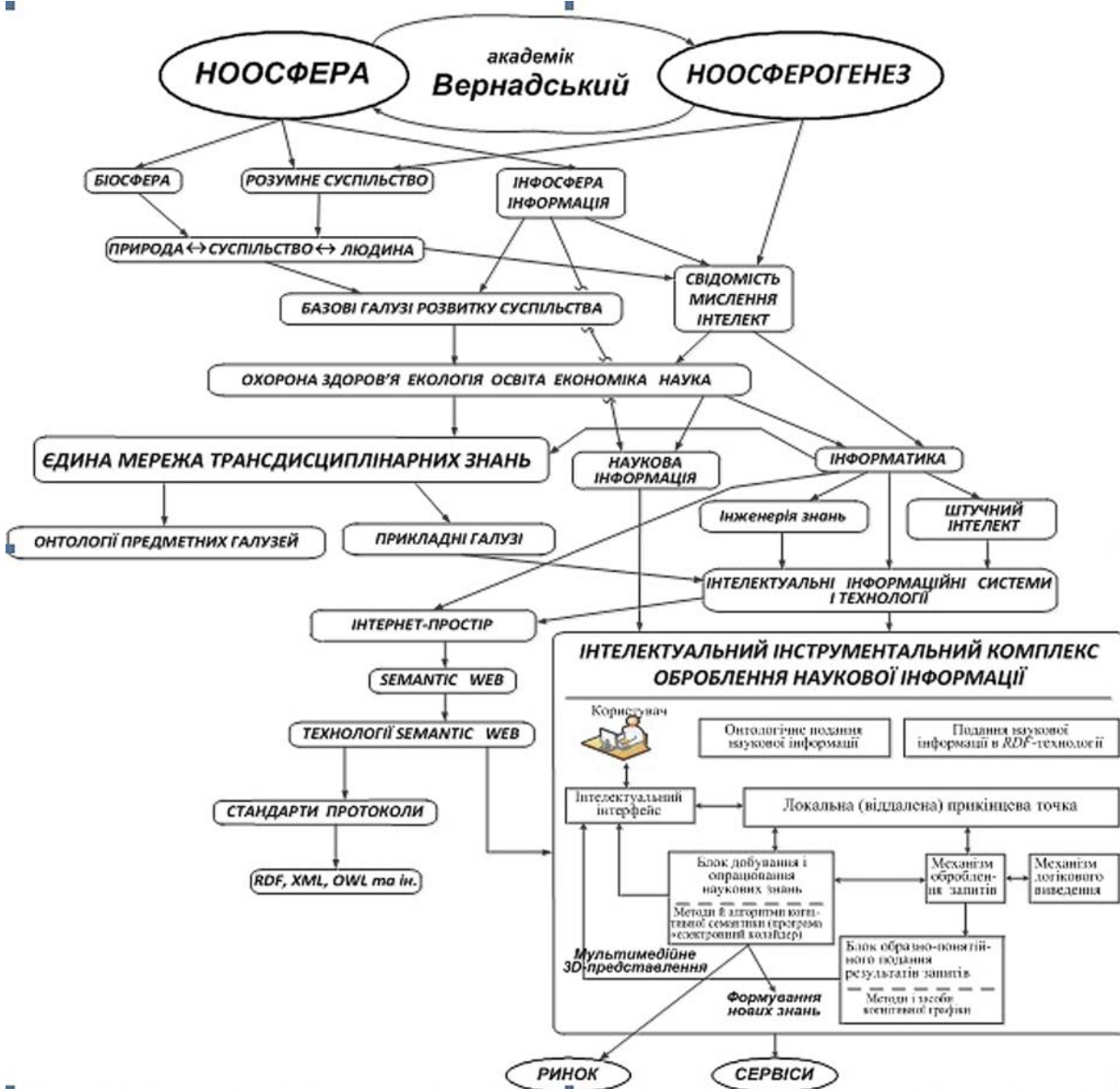


Рис. 1. Онтограф інтелектуального інструментального комплексу оброблення наукової інформації

Окремою проблемою є мультимедійне (образно-понятійне) подання результатів пошуку та порівняння їх із понятійною структурою предметної галузі (*Knowledge Domain*), яка є предметом зацікавлення, з метою добування нових знань. Під таким кутом зору для наукових досліджень оброблення наукових публікацій одного автора, авторів наукового підрозділу чи академічного інституту загалом із засто-

суванням технологій *Semantic Web*, мультимедійним поданням інформації й ефективною підтримкою процесу добування нових знань є актуальним.

Йдучи за онтологічною концепцією візуального моделювання тематики наукових досліджень предметних галузей (ПдГ) на рис. 1 представлено онтограф категорій і відповідних понять інтелектуального інструментального

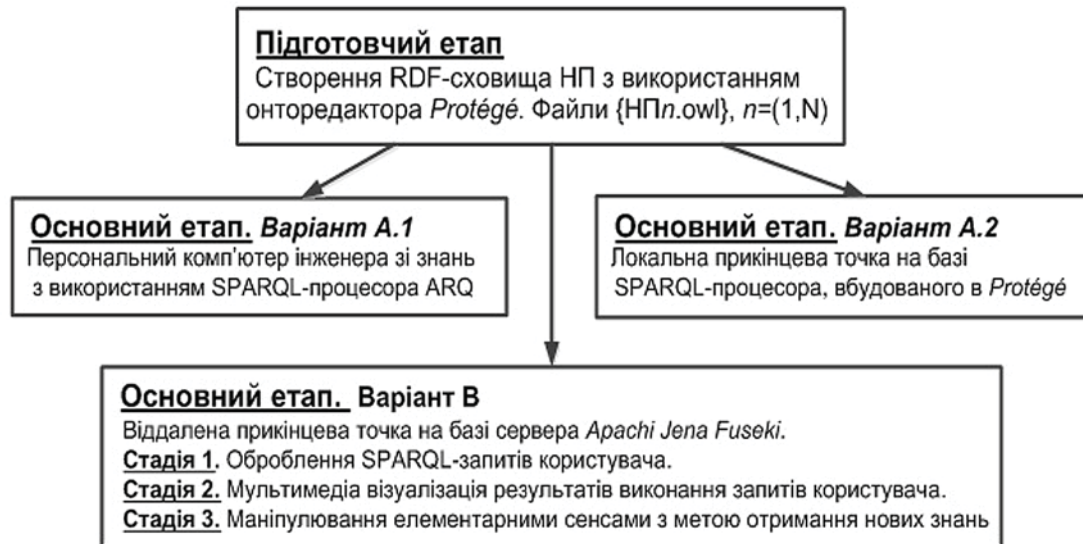


Рис 2. Узагальнена блок-схема розроблення ІК

комплексу оброблення наукової інформації, базова структура концептів якого взята за основу розроблення ІК БДНП.

Постановка проблеми

Узагальнену блок-схему розробки інструментального комплексу оброблення бази даних наукових публікацій представлено на рис 2.

Блок-схема включає: блок підготовчого етапу та блоки основного етапу з варіантами А і В. Підготовчий етап на описовому рівні наведено в [1], далі наведемо його *UML*-діаграми, які значно деталізують процеси та виконання відповідних процедур. Варіант А1 основного етапу інженером зі знань (ІЗ) розглядається як етап відпрацювання його творчих задумів. Тому цей етап не розглядається. Для етапу А2 буде наведено короткий опис та основні *UML*-діаграми. Він розглядається як експериментальна перевірка функціонування ІК в локальній мережі, напрацювання списку запитів природною мовою (ПМ), узгодження роботи інтерфейсів користувачів та ІЗ, а також вирішення питань автоматизації створення *owl*-файлів НП. А основну увагу в цьому матеріалі зосереджено на описі етапу В, стадія 1 (В1).

Наявна персоніфікована база знань науковця-дослідника, в якій задекларовано сукупність функціональних можливостей, що підтримують процеси наукової та творчої діяльності [2–3]. Така персоніфікована база знань є:

- засобом підтримки наукових досліджень і одним з центральних напрямів розвитку практичної інформатики [4, 5];
- розвитком системи знань НР й отримання нових знань (або впорядкування наявних, перевірка їх на коректність і суперечності тощо) [6–9];
- однією з основних підсистем сучасної системи дослідного проектування [10], автоматизованого робочого місця НР [4];
- одним з основних елементів конструювання сталого канонічного знання [11] і підтримки функціонування знання-орієнтованої інформаційної системи [12].

Відомо, що є тісний зв'язок між технологіями *Semantic Web* і *UML*-технологіями, зокрема, зв'язок між *OWL*-синтаксисом і візуальним моделюванням *UML*-діаграмами. Мова *UML* подається як загальноцільова мова візуального моделювання, яку розроблено для специфікації, візуалізації, проектування та документування компонентів програмного забезпечення, бізнес-процесів та інших систем. Мова *UML* одночасно слугує простим і



Рис. 3. Діаграма варіантів використання ІК на підготовчому етапі (варіант А)



Рис. 4. Діаграма компонентів (підготовчий етап)

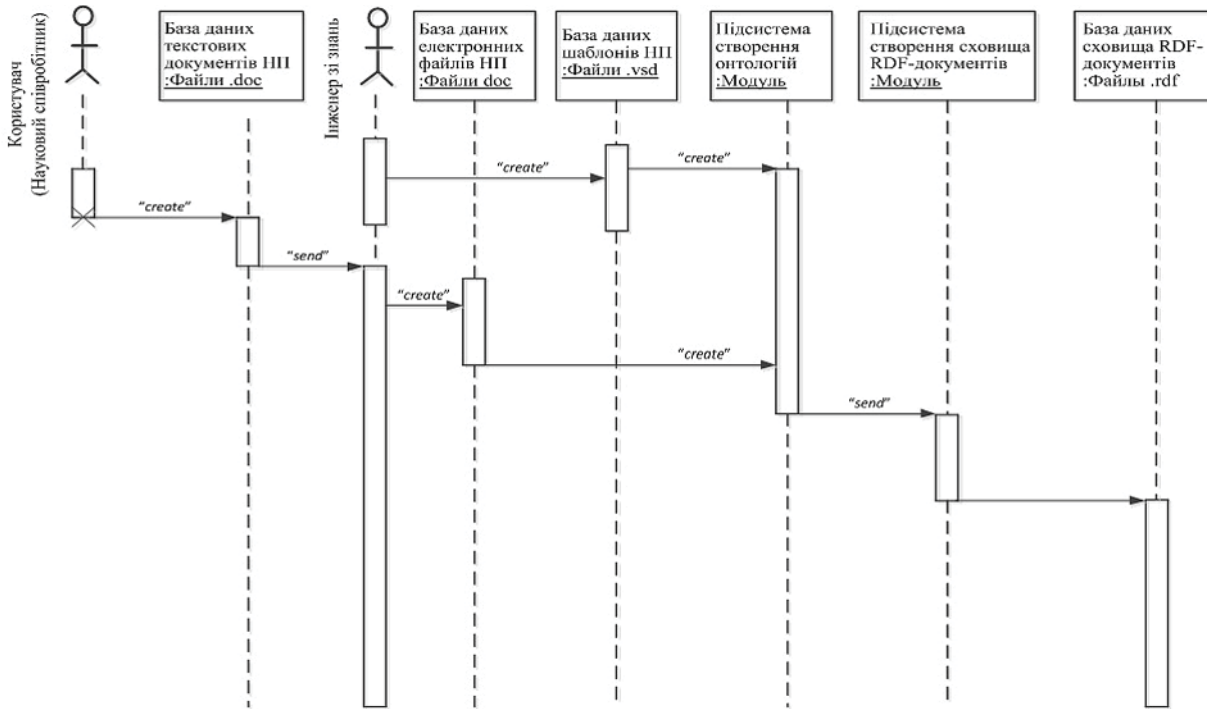


Рис. 5. Діаграма послідовності (підготовчий етап)

потужним засобом моделювання, який може бути ефективно використаний для побудови концептуальних, логікових і графічних моделей складних систем найрізнішого цільового призначення. Ця мова увібрала в себе найкращі якості методів програмної інженерії, які з успіхом використовувалися протягом багатьох років при моделюванні великих і складних систем [13].

Візуальне моделювання в *UML* можна представити як певний процес порівневого спуску від найзагальнішої та найабстрактнішої концептуальної моделі вихідної системи до логікової, а потім і до фізичної моделі відповідної програмної системи. Для досягнення цих цілей спочатку будується модель у формі так званої діаграми варіантів використання, що описує функціональне призначення системи або, інакше кажучи, те, що система виконуватиме в процесі свого функціонування. Діаграма варіантів використання слугує вихідним концептуальним поданням або концептуальною моделлю системи в процесі її проектування та розроблення [14, 15].

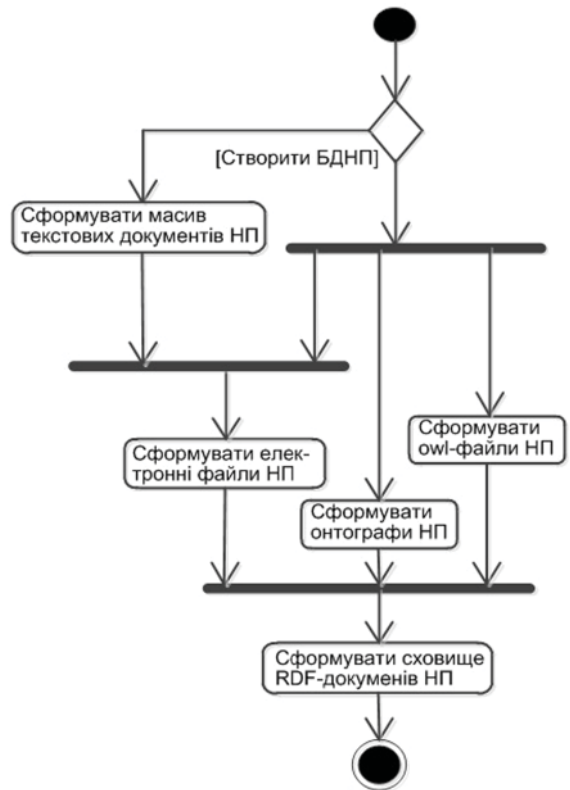


Рис. 6. Діаграма діяльності (підготовчий етап)

ІК «База даних наукових публікацій» зорієнтовано на такого автора, який активно займається підготовкою та продукуванням різних видів нових НП. Звісно, що пошук по своїм НП можливо виконати і вручну (що здебільшого й робиться), але за допомогою ІК він може бути виконаний набагато швидше. До того ж отриману інформацію можна автоматично структурувати у відповідні шаблони майбутніх НП.

Мета статті — розроблення ІК, який дасть змогу значно пришвидшити отримання автором необхідної йому інформації із своєї БД НП, візуального подання концептів НП та відповідної ПдГ і реалізації алгоритмів формування й отримання нових знань [7, 9].

У статті основну увагу буде приділено розробці архітектурно-структурних компонент і *UML*-діаграм функціонування ІК із застосуванням віддаленої прикінцевої точки *Apache Jena Fuseki*, а також прикладам використання формального опису наукових статей із виконанням сукупності запитів, узятих з початкових запитів користувача (наведених далі).

Проектування знання-орієнтованого інструментального комплексу БД НП (варіант А)

Як зазначено в [1], на підготовчому етапі всі процеси та процедури виконує інженер зі знань, користувач тільки надає йому масив текстових документів НП. Створення сховища *RDF*-документів НП виконується на персональному комп'ютері (ПК) ІЗ, за його проектом (див. рис. 3) та у відповідності з методичними матеріалами [16, 17]. Також слід додати, що основний варіант А припускає оброблення за запитом інформації, взятої тільки з однієї НП, тобто обчислення можливе тільки для однієї наукової статті. Як буде показано далі, варіант В припускає обчислення з будь-якої кількості НП.

На рис. 3 наведено *UML*-діаграму використання ІК для реалізації підготовчого етапу, на рис. 4 — *UML*-діаграму компонентів, на рис. 5 — *UML*-діаграму послідовності і на рис. 6 — *UML*-діаграму діяльності. Це основні *UML*-діаграми підготовчої фази етапу А2 функціонування ІК. Як видно з рисунку, програмні блоки у підсистемі створення/підтримки БДНП

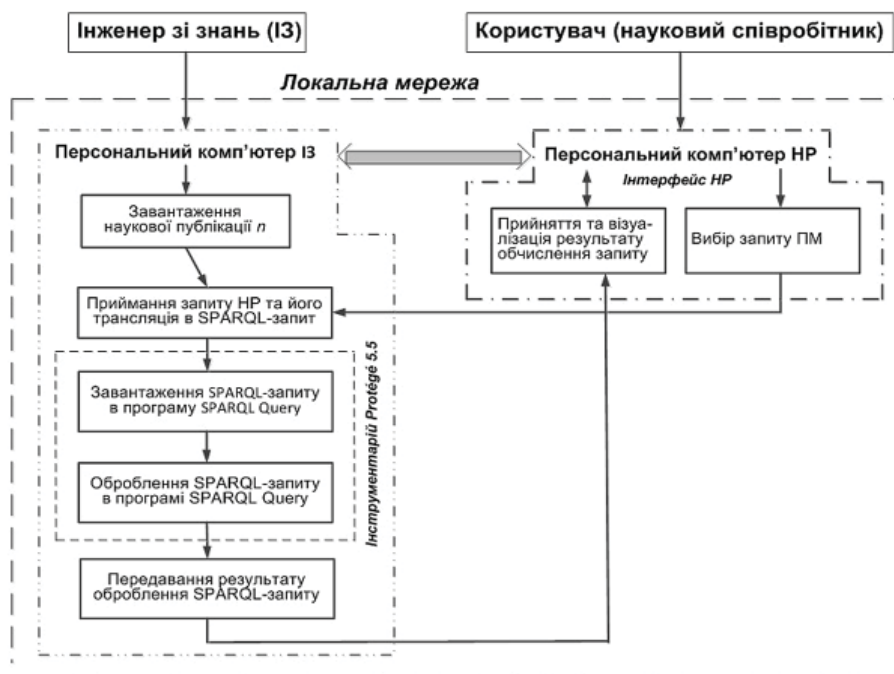


Рис. 7. Блок-схема ІК БДНП (основний етап А2)

реалізовано в онторедакторі *Protégé*, а бази даних НП — на жорсткому диску ПК ІЗ.

Функціонування ІК на основному етапі А2

На рис. 7 наведено блок-схему функціонування ІК.

Як видно з рисунка, він функціонує у локальній мережі «Інженер зі знань ↔ Користувач (один або декілька)». На ПК користувача відображається список запитів на ПМ з можливістю вибору відповідних аргументів та візуалізується результат обчисленого запиту. Ця інформація передається до ПК інженера зі знань. Як видно з рис. 7, у ньому функціонують блоки інтерфейсу ІЗ та інструментарію *Protégé*, в якому встановлено відповідні *SPARQL*-процесор та *Reasoner*. Опис відповідних процесів і процедур наведено на рис. 8–11.

UML-діаграми функціонування ІК на основному етапі А2

На рис. 8 наведено діаграму варіантів використання, на рис. 9 — діаграму компо-

нентів, на рис. 10 — діаграму послідовності і на рис. 11 — діаграму діяльності.

Розроблення інструментального комплексу БДНП, варіант В1

ІК для цього варіанта функціонує як віддалена прикінцева точка на базі сервера *Apache Jena Fuseki* і складається з трьох стадій: стадія 1 — оброблення *SPARQL*-запитів користувача; стадія 2 — мультимедіа візуалізація результатів виконання запитів користувача або побудова та використання образно-понятійних структур галузі наукових досліджень; стадія 3 — маніпулювання елементарними сенсами з метою отримання нових знань.

На рис. 12 наведено блок-схему ІК для варіанта В1.

Спочатку інженеру зі знань необхідно розгорнути сервер *Apache Jena Fuseki* як віддалену прикінцеву точку [18, 19] й завантажити у сховище наукових публікацій файли *RDF*-графів наукових статей, що сформовані на попередньому етапі.

Користувачу в його інтерфейсі надається список можливих запитів природною мовою

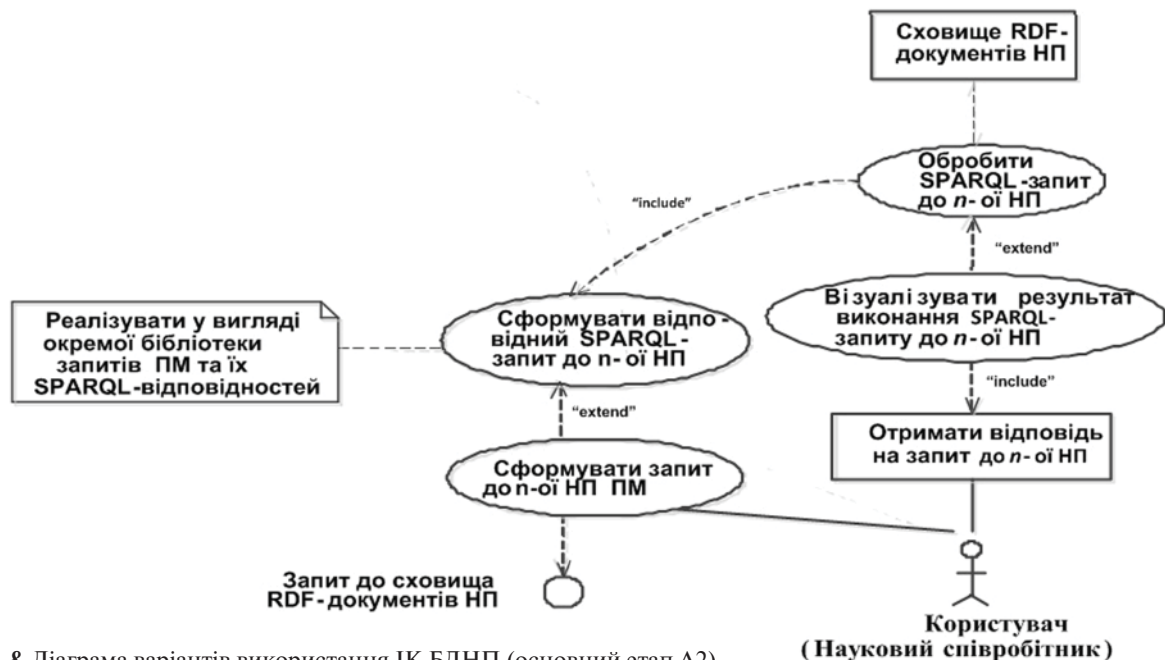


Рис. 8. Діаграма варіантів використання ІК БДНП (основний етап А2)

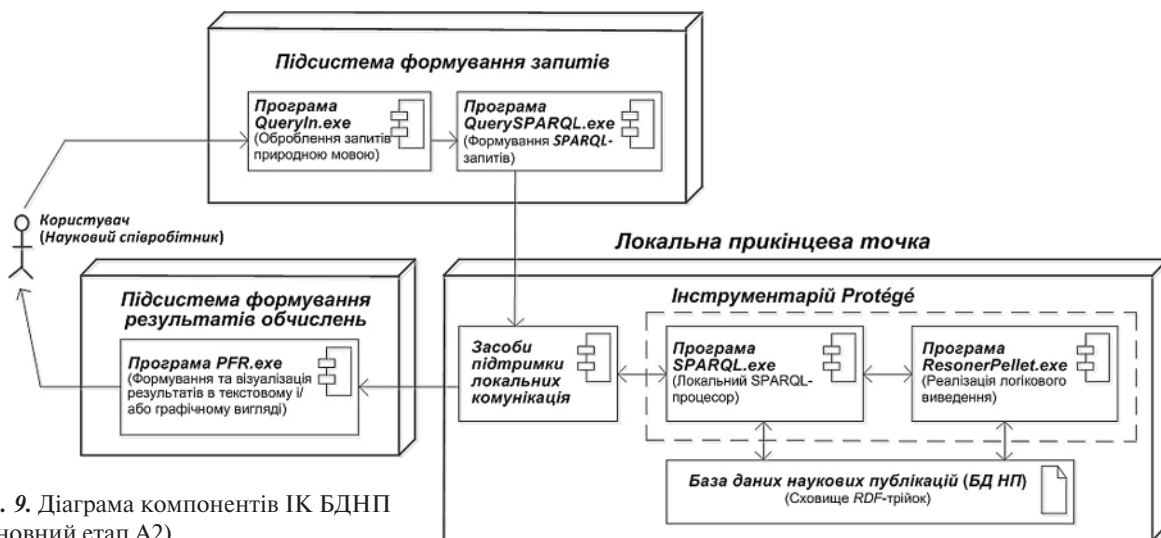


Рис. 9. Діаграма компонентів ІК БДНП (основний етап А2)

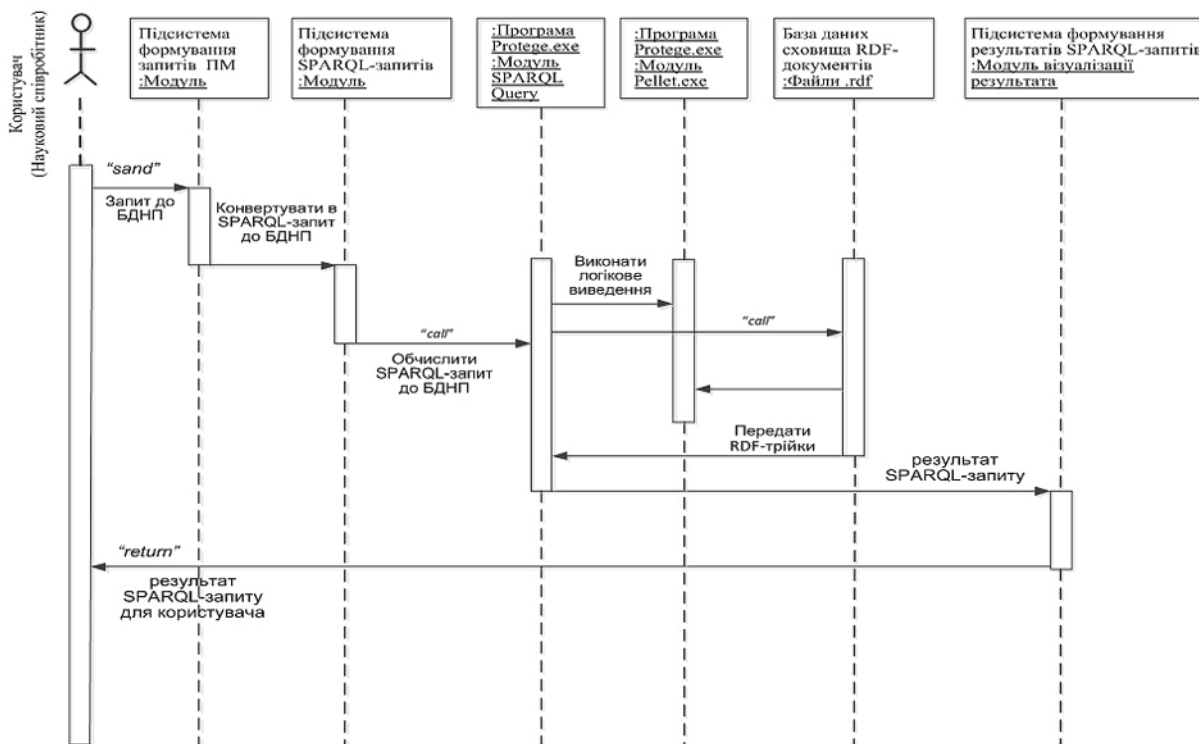


Рис. 10. Діаграма послідовності (основний етап А2)

до його БДНП. Із цього списку він поступово вибирає по одному запиту, який передається через локальну мережу в персональний комп'ютер ІЗ, тобто крок за кроком деталізує необхідну йому інформацію. Причому, в

одному запиту користувач має можливість вказати декілька статей, з яких йому потрібно отримати інформацію.

На підготовчому етапі інженер зі знань (враховуючи побажання користувачів) формує

початковий список запитів природною мовою. Далі наведено варіант такого початкового списку.

Запити користувача початкові

Вихідна інформація.

Науковий співробітник у своїй базі даних має опублікованими у відомих наукових журналах N наукових статей. Множина N описується у такий спосіб:

$$N = 1, 2, \dots, m_1, \dots, m_1+1, \dots, m_2, \dots, m_2+1, \dots, m_k, m_k+1, \dots, N-1.$$

Припускається, що автором є перший у публікації співавтор, або той, кому належить база даних НП.

1 *Вивести* назви всіх N статей:

- 1.1 у порядку зростання дати публікації;
- 1.2 після ___ року публікації;
- 1.3 без співавторів.

2. *Вивести* назви статей $m_1, m_1+b \dots m_2, \dots, m_2+c, m_k, \dots, m_k+d, \dots$

змінні параметри $m_1, b, m_2, c, m_k, d \dots$ задаються користувачем.

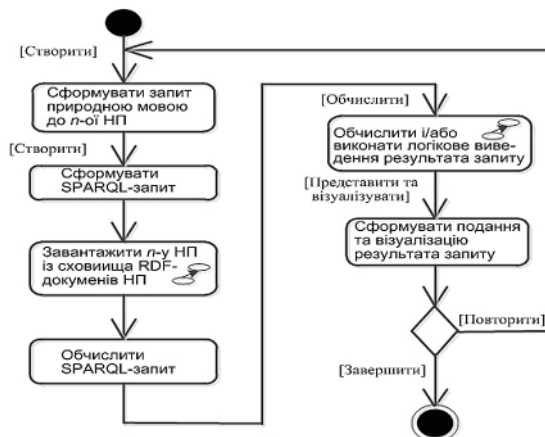


Рис. 11. Діаграма діяльності (основний етап А2)

3. *Вивести* ПІБ співавторів статей $m_1, m_1+b \dots m_2, \dots, m_2+c, m_k, \dots, m_k+d, \dots$

7. *Вивести* анотації статей $m_1, m_1+b \dots m_2, \dots, m_2+c, m_k, \dots, m_k+d, \dots$

13. *Вивести* назви статей на тему «трансдисциплінарність».

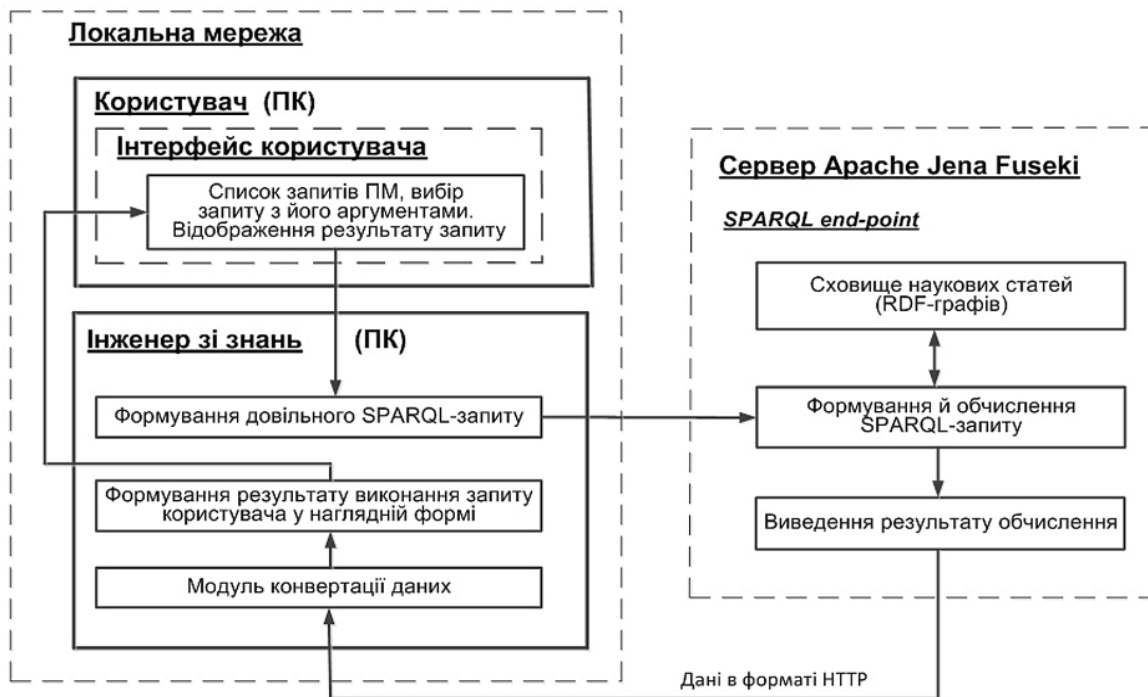


Рис. 12. Узагальнена блок-схема ІК БДНП

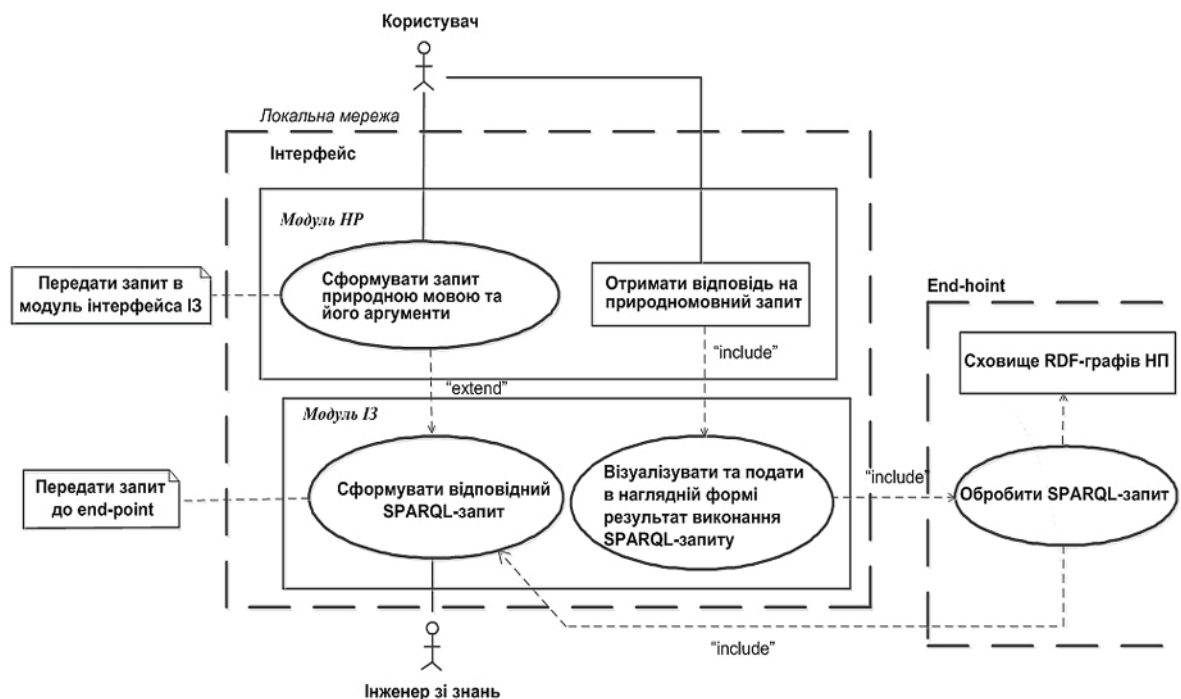


Рис. 13. Діаграма варіантів використання ІК БДНП

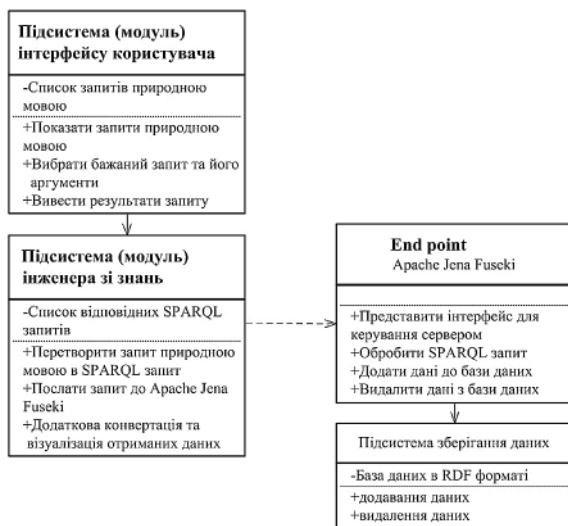


Рис. 14. Діаграма класів ІК БДНП

15. *Вивести* назви розділів статей $m_1, m_1+b... m_2, \dots, m_2+c, m_k, \dots, m_k+d, \dots$

16. *Вивести* текст розділу n статті m .

.....

20. *Вивести* назви рисунків у статтях $m_1, m_1+b... m_2, \dots, m_2+c, m_k, \dots, m_k+d, \dots$

.....

23. *Вивести* тексти абзаців в статтях $m_1, m_1+b... m_2, \dots, m_2+c, m_k, \dots, m_k+d, \dots$, у яких є фраза $\{f_i\}$. Фраза для перевірки запиту – «*трандисциплінарні наукові дослідження*», «*онтологічний*» та інші.

...

25. *Вивести* назви статей, в яких автором чи співавтором є «ПІБ».

26. *Вивести* назви статей, які опубліковані в журналі $\{g_k\}$, де g_k – назва журналу.

Слід зазначити, що у ПМ-запиті слово «*Вивести*» не слід трактувати як ключове, воно може бути замінене будь-яким його синонімом.

UML-діаграми функціонування ІК для варіанта В1

Розгляньмо UML-діаграми, які розкривають суть функціонування ІК для варіанта В1.

На рис. 13 представлено діаграму варіантів використання, на рис. 14 — діаграму класів, на рис. 15 — діаграму компонентів і на рис. 16 — діаграму послідовності.

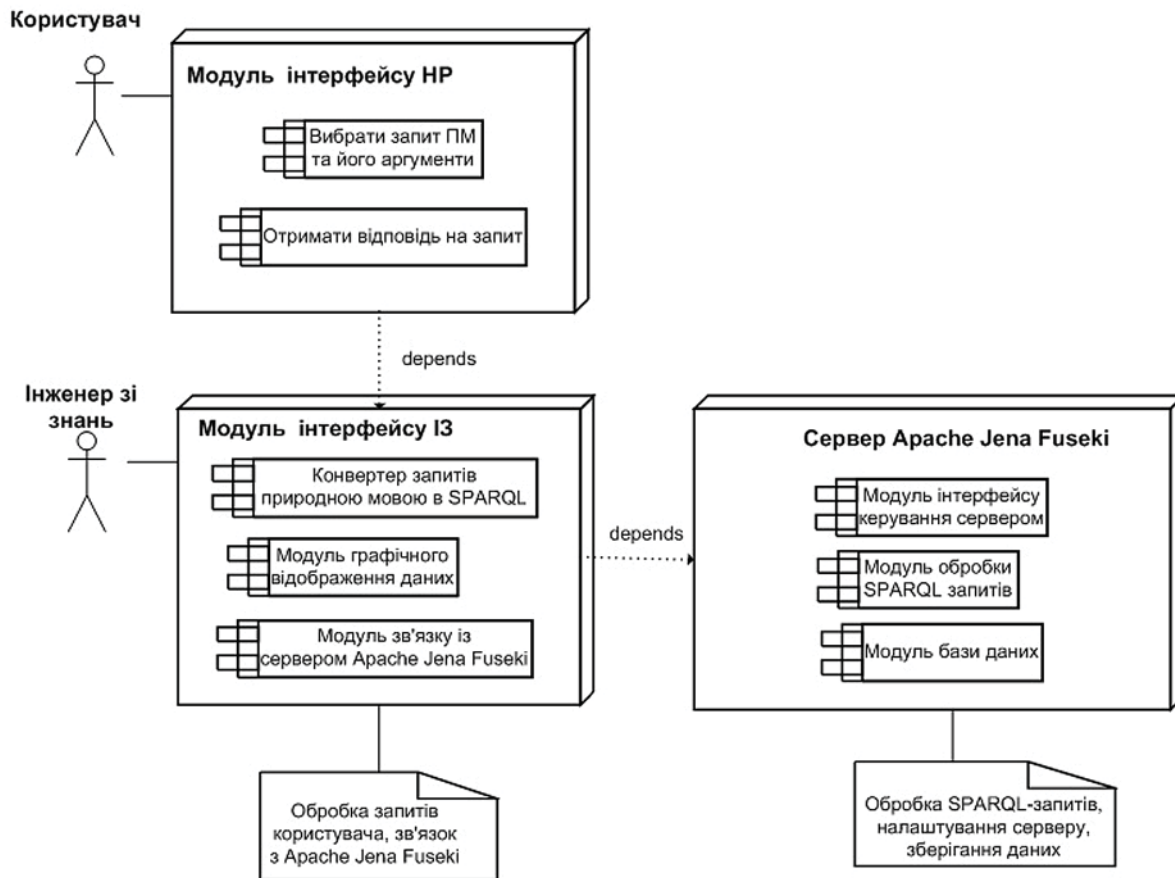


Рис. 15. Діаграма компонентів ІК БДНП

До локальної мережі (ЛМ), адміністратором якої є інженер зі знань, підключено певну кількість НР. Роботу мережі ми розглянемо для одного користувача, для кількох користувачів запит виконується в аналогічний спосіб.

ПК користувача функціонує як модуль НР загального інтерфейсу ЛМ. В ньому, з однієї сторони відображаються всі запити на ПМ, з яких НР може вибрати один з відповідними аргументами, а з іншої сторони — відображається результат виконання запиту у наглядній формі.

Другу частину інтерфейсу складає модуль ІЗ. У ньому формується відповідний *SPARQL*-запит і передається за *HTTP*-протоколом до *end-point*. На сервері *Apache Jena Fuseki* виконується *SPARQL*-запит і відповідь на нього

за *HTTP*-протоколом передається в ЛМ, інтерфейс, модуль ІЗ.

Процеси формування, оброблення запитів користувача та отримання ним відповідей детально показано на основних *UML*-діаграмах на рис.13–16.

Приклади виконання *SPARQL*-запитів та їхні результати

Слід зазначити, що: на діаграмах не показано процедури вибору аргументів НР та перетворення їх у номери статей у БД; ІК зорієнтовано на оброблення НР, поданих українською чи російською мовами, а тому у *SPARQL*-запитах ключі пошуку надаються двома мовами.

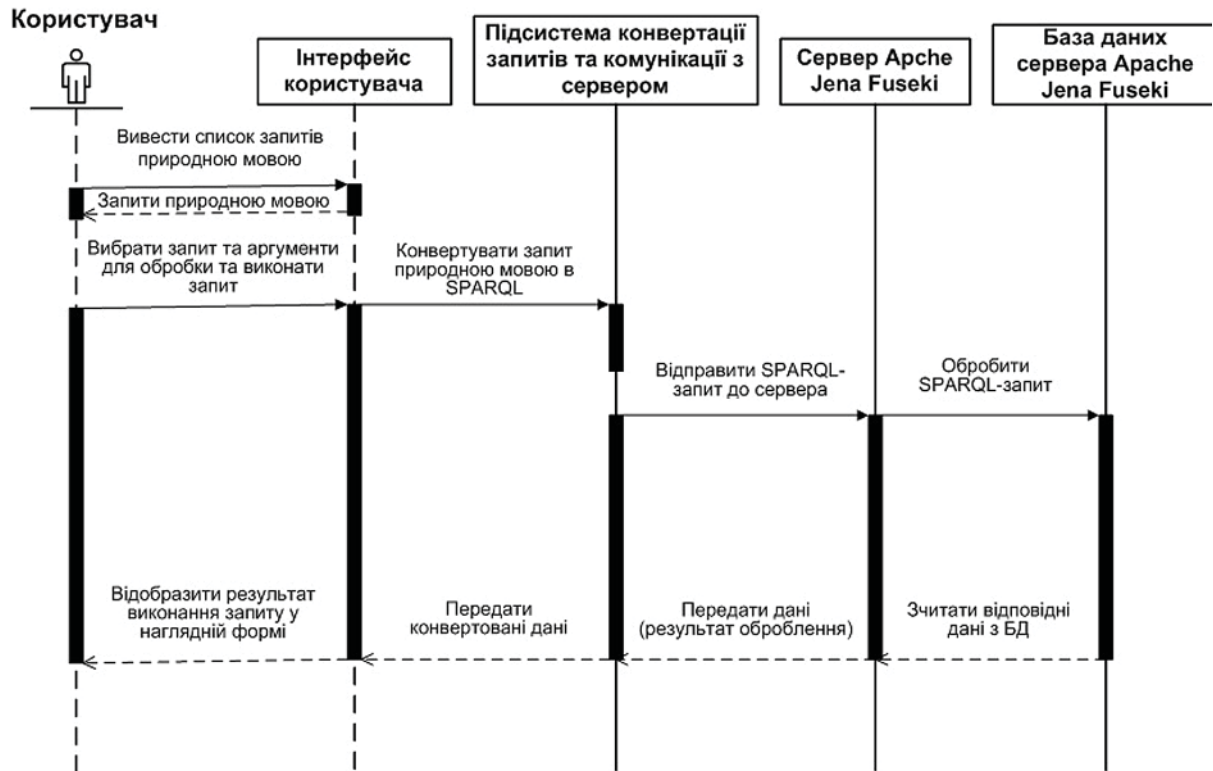


Рис. 16. Діаграма послідовності ІК БДНП

1. Запит на ПМ.

Вивести назви статей на тему «онтологічний».

Відповідний SPARQL-запит.

PREFIX : <http://www.semanticweb.org/ніколай/ontologies/2020/5/19/untitled-ontology-36#>

SELECT DISTINCT ?номер_статті ?назва_статті

{
GRAPH ?номер_статті {?s1 :Назва_статті ?назва_статті.
FILTER REGEX(?назва_статті, "онтолог", "i")
 }

Результат виконання.

№	Номер статті	Назва статті
1.	<http://test.ulif.org.ua:51089/articles/data/article5>	«Про деякі особливості побудови онтологічних моделей предметних областей»

№	Номер статті	Назва статті
2.	<http://test.ulif.org.ua:51089/articles/data/article7>	«Введение в класс трансдисциплинарных онтологоуправляемых систем исследовательского проектирования»
3.	<http://test.ulif.org.ua:51089/articles/data/article8>	«Онтологическая концепция информатизации научных исследований»
4.	<http://test.ulif.org.ua:51089/articles/data/article10>	«Архитектура онтологоуправляемых компьютерных систем»
5.	<http://test.ulif.org.ua:51089/articles/data/article16>	«К вопросу системно-онтологической интеграции знаний предметной области»
6.	<http://test.ulif.org.ua:51089/articles/data/article19>	«Знание-ориентированные информационные системы с обработкой естественноязыковых объектов: онтологический подход»
7.	<http://test.ulif.org.ua:51089/articles/data/article21>	«Системно-онтологический анализ предметной области»

2. Запит на ПМ.

Вивести назви статей на тему «трансдисциплінарність».

Відповідний SPARQL-запит.

PREFIX : <http://www.semanticweb.org/николай/ontologies/2020/5/19/untitled-ontology-36#>

```
SELECT ?номер_статті ?назва_статті
{
  GRAPH ?номер_статті {?s1 :Название_статьи
?назва_статті.
  FILTER REGEX (?назва_статті, «трансдисципл», «і»)}
}
```

Результат виконання.

№	Номер статті	Назва статті
1.	<http://test.ulif.org.ua:51089/articles/data/article1>	«Методологические основы развития, становления и IT-поддержки трансдисциплинарных исследований»
2.	<http://test.ulif.org.ua:51089/articles/data/article7>	«Трансдисциплинарность, информатика и развитие современной цивилизации»
3.	<http://test.ulif.org.ua:51089/articles/data/article6>	«Проблемы трансдисциплинарности и роль информатики»
4.	<http://test.ulif.org.ua:51089/articles/data/article7>	«Введение в класс трансдисциплинарных онтологически управляемых систем исследовательского проектирования»

Результат виконання.

Номер статті	Назва статті	Абзац	Текст речення
1.	<http://test.ulif.org.ua:51089/articles/data/article2>	Абзац_4.11.1.17	«Во-первых, формальные компьютерные онтологии являются одновременно и результатом развития, и инструментом <i>knowledge engineering</i> .»
2.	<http://test.ulif.org.ua:51089/articles/data/article2>	:Абзац_4.11.1.15	«Таким каркасом может служить совокупность формальных компьютерных онтологий конкретных предметных областей [5].»
3.	<http://test.ulif.org.ua:51089/articles/data/article2>	:Абзац_4.11.1.15	«Формально онтологию можно представить четверкой множеств:»

3. Запит на ПМ.

Вивести абзаци в розділах статей, у яких розглядається «формальна комп'ютерна онтологія».

Відповідний SPARQL-запит.

PREFIX : <http://www.semanticweb.org/николай/ontologies/2020/5/19/untitled-ontology-36#>

```
SELECTDISTINCT ?номер_статті ?назва_статті ?paragraph ?текст_речення
FROM NAMED <http://test.ulif.org.ua:51089/articles/data/article1>
FROM NAMED <http://test.ulif.org.ua:51089/articles/data/article2>
FROM NAMED <http://test.ulif.org.ua:51089/articles/data/article4>
{
  GRAPH ?номер_статті {?s1 :Название_статьи
?назва_статті
    {?paragraph :Иметь_Предложение
?речення_абзацу}
    {?речення_абзацу :Иметь_Текст
?текст_речення}
  FILTERREGEX (?текст_речення, «формальн.* онтолог.*», «і»)
}
```


Номер статті	Назва статті	Абзац	Текст речення
4. <http://test.ulif.org.ua:51089/articles/data/article4>	«Информационная технология и инструментальные средства поддержки процессов исследовательского проектирования <i>Smart</i> -систем»	:Абзац_4.11.2.2	« извлечение из множества ТД знаний, релевантных заданной ПдО, их системно-онтологическая структуризация и формально-логическое представление на одном или нескольких из общепринятых языков описания онто-логий (<i>Knowledge Representation</i>);»
5. <http://test.ulif.org.ua:51089/articles/data/article1>	«Методологические основы развития, становления и ИТ-поддержки трансдисциплинарных исследований»	Абзац_3.15.3	«В ее недрах зародилась онтологическая концепция, сущность которой состоит в формальном онтологическом описании предметных областей и научной картины мира в целом (процесс только начал).»
6. <http://test.ulif.org.ua:51089/articles/data/article1>	«Методологические основы развития, становления и ИТ-поддержки Т»	Абзац_4.11.3.1	«Формально онтологию можно представить чет-веркой множеств:»

4. Запит на ПМ.

Вивести анотації статей 2, 6, 13.

Відповідний SPARQL-запит.

PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>

PREFIX : <http://www.semanticweb.org/ніколай/ontologies/2020/5/19/untitled-ontology-36#>

SELECT ?номер_статті ?назва_статті (group_concat(?анотация) as ?анотация_повна)

FROM NAMED <http://test.ulif.org.ua:51089/articles/data/article2>

FROM NAMED <http://test.ulif.org.ua:51089/articles/data/article6>

FROM NAMED <http://test.ulif.org.ua:51089/articles/data/article13>

```
{
  GRAPH ?номер_статті {?s1 :Название_статьи
?назва_статті.
```

```
?subclass rdfs:subClassOf :Анотация
{?shortname rdf:type ?subclass}
{?shortname :Иметь_Текст ?анотация}
}
```

groupby ?номер_статті ?назва_статті

Результат виконання.

Номер статті	Назва статті	Анотация повна
1. <http://test.ulif.org.ua:51089/articles/data/article6>	«Проблемы трансдисциплинарности и роль информатики»	«Предложены и проанализированы некоторые методы и средства информационно-технологической поддержки трансдисциплинарных научных исследований в их становлении и развитии. »
2. <http://test.ulif.org.ua:51089/articles/data/article2>	«Трансдисциплинарность, информатика и развитие современной цивилизации»	«Перспективы и проблемы развития человеческой цивилизации всегда волновали общество. Особенно острый интерес к ним возникает, как правило, на крутых виражах истории, в периоды общественных потрясений и техногенных катаклизмов. ...»

Номер статті	Назва статті	Анотація повна
3. < http://test.ulif.org.ua:51089/articles/data/article13 >	«Міждисциплінарні наукові дослідження: оптимізація системно-інформаційної підтримки»	«Епоху аналітизму й властиву йому диференціацію науки завершено. Реальні проблеми, що постають перед людством, набагато складніші за наукові. Сучасна наука не в змозі їх кардинально вирішити. Одна з причин — роз'єднаність наукових дисциплін, а отже, нескоординованість роботи наукових установ, їхньої тематики. ...»

5. Запит на ПМ.

Вивести назви статей із співавтором «Кургаєв».

Відповідний *SPARQL*-запит.

PREFIX: <<http://www.semanticweb.org/николай/ontologies/2020/5/19/untitled-ontology-36#>>

SELECT ?номер_статті ?назва_статті ?ПІБ_співавтора
{

GRAPH ?номер_статті {?s :Назва_статті ?назва_статті.

?s1 :Иметь_Соавтора ?співавтор_статті.

?співавтор_статті :Иметь_ФИО ?ПІБ_співавтора

FILTER REGEX(?ПІБ_співавтора, "Кургаєв О.П.", "i")
}

Результат виконання.

Номер статті	Назва статті	ПІБ співавтора
1. < http://test.ulif.org.ua:51089/articles/data/article9 >	«Ноосферна парадигма розвитку науки та штучний інтелект»	«Кургаєв О.П.»
2. < http://test.ulif.org.ua:51089/articles/data/article13 >	«Міждисциплінарні наукові дослідження: оптимізація системно-інформаційної підтримки»	«Кургаєв О.П.»

Висновки

Метою наших досліджень було подальше розроблення інструментального комплексу оброблення баз даних наукових публікацій, який дає науковцю змогу значно пришвидшити отримання необхідної когнітивно-структурованої інформації із своїх джерел.

У цій статті було запроваджено й описано архітектурно-структурну організацію ІК, який включає локальну мережу із ПК користувачів, ПК адміністратора-інженера зі знань та віддалену прикінцеву точку на базі сервера *Apache Jena Fuseki*, основні *UML*-діаграми функціонування ІК та приклади виконання запитів користувача.

ЛІТЕРАТУРА

1. Палагін О.В., Петренко М.Г. Знання-орієнтований інструментальний комплекс обробки баз даних наукових публікацій. *Control systems and computers*. 2020. №5. С. 17–33. <https://doi.org/10.15407/usim.2020.05.003>.
2. Palagin A.V., Petrenko N.G., Velychko V.Yu., Malakhov K.S., 2014. Development of formal models, algorithms, procedures, engineering and functioning of the software system “Instrumental complex for ontological engineering purpose”. In: *Proceedings of the 9th International Conference of Programming UkrPROG. CEUR Workshop Proceedings*, 1843. Kyiv, Ukraine, May 20–22, 2014. URL: <http://ceur-ws.org/Vol-1843/221-232.pdf> (дата звернення: 23.06.2022).
3. Палагін А.В., Крывий С.Л., Петренко Н.Г. *Онтологические методы и средства обработки предметных знаний*. Луганск: изд-во ВНУ им. В. Даля. 2012. 323 с. URL: <http://www.aduis.com.ua/Monography.pdf>.

4. Палагін О.В., Малахов К.С., Величко В.Ю., Щуров О.С. Проектування та програмна реалізація підсистеми створення та використання онтологічної бази знань публікацій наукового дослідника. Проблеми програмування. 2017. №2. С. 72–81.
5. Palagin O.V., Velychko V.Yu., Malakhov K.S., Shchurov O.S., 2020. Distributional semantic modeling: a revised technique to train term/word vector space models applying the ontology-related approach. In: Proceedings of the 12th International Scientific and Practical Conference of Programming UkrPROG 2020. CEUR Workshop Proceedings 2866. Kyiv, Ukraine, September 15-16, 2020. URL: http://ceur-ws.org/Vol-2866/ceur_342-352palagin34.pdf (дата звернення: 23.06.2022).
6. Palagin A.V., Petrenko N.G. Methodological foundations for development, formation and it-support of transdisciplinary research. Journal of automation and information sciences. 2018. Vol. 50, no. 10. P. 1–17. <https://doi.org/10.1615/jau-tomatinfscien.v50.i10.10>.
7. Палагін А.В. Проблемы трансдисциплинарности и роль информатики. Кибернетика и системный анализ. 2013. № 5. С. 3–13.
8. Петренко Н.Г., Зеленцов Д.Г. О практическом использовании онтологических моделей предметных областей. Компьютерное моделирование: анализ, управление, оптимизация. 2019. №2 (6). С. 58–73. DOI: 10.32434/2521-6406-2019-6-2-58-73.
9. Палагін А.В. Архитектура онтолого-управляемых компьютерных систем. Кибернетика и системный анализ. 2006. № 2. С.111–124.
10. Палагін А.В. Введение в класс трансдисциплинарных онтолого-управляемых систем исследовательского проектирования. Управляющие системы и машины. 2016. № 6. С. 3–11. <https://doi.org/10.15407/usim.2016.06.003>.
11. Палагін О., Кургаєв О. Міждисциплінарні наукові дослідження: оптимізація системно-інформаційної підтримки. Вісн. НАН України. 2009. № 3. С.14–25.
12. Палагін А.В., Петренко Н.Г., Кривый С.Л. О построении знание-ориентированных компьютерных систем для научных исследований. Управляющие системы и машины. 2015. № 2. С. 64–73.
13. Booch G., Rumbaugh J., Jacobson I. The Unified Modeling Language User Guide. Addison-Wesley. Reading. MA. 2005. 475 p.
14. Шмüller Д. Освой самостоятельно UML 2 за 24 часа. Практическое руководство. Sams Teach Yourself UML in 24 Hours, Complete Starter Kit. М.: Вильямс, 2005. 416 с.
15. Леоненков А. В. Самоучитель UML 2. СПб.: БХВ-Петербург, 2007. 576 с. ISBN 978-5-94157-878-8.
16. OWL 2 Web Ontology Language Primer. URL: <http://www.w3.org/TR/2012/REC-owl2-primer-20121211/> (дата звернення: 23.06.2022).
17. Horridge M. A Practical Guide To Building OWL Ontologies Using Prot g 4 and CO-ODE Tools. Edition 1.3. Copyright The University Of Manchester. March 24, 2011. 107p.
18. Apache Jena Fuseki. URL: <https://jena.apache.org/documentation/fuseki2/> (дата звернення: 23.06.2022).
19. Bob DuCharme. Learning SPARQL. Querying and Updating with SPARQL 1.1 (Second edition), O'Reilly Media, All rights reserved, August 2013. 367p.

Надійшла 18.07.2022

REFERENCES

1. Palagin, A.V., Petrenko, M.G., 2020. “Knowledge-Oriented Tool Complex Processing Databases of Scientific Publications”. Control systems and computers. N 5, pp. 17–33. <https://doi.org/10.15407/usim.2020.05.003> (In Ukrainian).
2. Palagin, A.V., Petrenko, N.G., Velychko, V.Yu., Malakhov, K.S., 2014. “Development of formal models, algorithms, procedures, engineering and functioning of the software system. Instrumental complex for ontological engineering purpose”. In Proceedings of the 9th International Conference of Programming UkrPROG. CEUR Workshop Proceedings 1843. Kyiv, Ukraine, May 20–22, 2014. [Online]. <<http://ceur-ws.org/Vol-1843/221-232.pdf>> (date of access: 23.06.2022) (In Russian).
3. Palagin, A.V., Kryvyy, S.L., Petrenko, N.G., 2012. “Ontological methods and means of processing subject knowledge”. Lugansk: V.I. Dal East Ukr. Nac. University. [Online]. <<http://www.aduis.com.ua/Monography.pdf>> (date of access: 20.06.2022).
4. Palagin, A.V., Malakhov, K.S., Velychko, V.Yu., Shchurov O.S., 2017. “Design and software implementation of the sub-system for creating and using the ontological knowledge base of the researcher’s publications”. Programming problems, N 2, pp. 72–81 (In Ukrainian).

5. *Palagin, O.V., Velychko, V.Yu., Malakhov, K.S., Shchurov, O.S.*, 2020. “Distributional semantic modeling: a revised technique to train term/word vector space models applying the ontology-related approach”. In Proceedings of the 12th International Scientific and Practical Conference of Programming UkrPROG 2020. CEUR Workshop Proceedings 2866. Kyiv, Ukraine, September 15–16, 2020. [Online]. <http://ceur-ws.org/Vol-2866/ceur_342-352palagin34.pdf> (date of access: 23.06.2022).
6. *Palagin, A.V., Petrenko, N.G.*, 2018. “Methodological foundations for development, formation and it-support of transdisciplinary research”. Journal of automation and information sciences. Vol. 50, no. 10. pp. 1–17. <https://doi.org/10.1615/jautomatinfscien.v50.i10.10>.
7. *Palagin, A.V.*, 2013. “Problemy` transdisciplinarnosti i rol` informatiki”. Kibernetika i sistemny`j analiz. N 5. pp. 3–13.
8. *Petrenko, N.G., Zelentsov, D.G.*, 2019. “On the practical use of ontological models of subject areas”. Computer modeling: analysis, management, optimization. N 2(6). pp. 58–73. <https://doi.org/10.32434/2521-6406-2019-6-2-58-73>.
9. *Palagin, A.V.*, 2006. “Arkhitektura ontologo-upravlyaemy`kh komp`yuterny`kh system”. Kibernetika i sistemny`j analiz. N 2. pp. 111–124.
10. *Palagin, A.V.*, 2016. “An Introduction to the Class of the ransdisciplinary Ontology-controlled Research Design Systems”. Upravlausie sistemy i masiny. N 6, pp. 3–11 <https://doi.org/10.15407/usim.2016.06.003> (In Russian).
11. *Palahin, O., Kurhaiev, O.*, 2009. “Mizhdystsyplynarni naukovi doslidzhennia: optymizatsiia systemno-informatsiinoi pidtrymky”. Visn. NAN Ukrainy. N 3. pp. 14–25. (In Russian).
12. *Palagin, A.V., Petrenko, N.G., Kryvyj, S.L.*, 2015. “To the Question of Constructing Knowledge-Oriented Computer Systems for Scientific Research”. Upravlausie sistemy i masiny. N 2. pp. 64–73.
13. *Booch, G., Rumbaugh, J., Jacobson, I.*, 2005. The Unified Modeling Language User Guide. Addison-Wesley. Reading. MA. 475 p.
14. *Shmuller, D.*, 2005. Osvoj samostoyatel`no UML 2 za 24 chasa. Prakticheskoe rukovodstvo.Sams Teach Yourself UML in 24 Hours, Complete Starter Kit. M.: Vil`yams, 416 p. ISBN 0-672-32640-X.
15. *Leonenkov, A.V.*, 2007. Samouchitel UML 2. SPb.: BKhV-Peterburg, 576 p. ISBN 978-5-94157-878-8 (In Russian).
16. OWL 2 Web Ontology Language Primer. [Online]. <<http://www.w3.org/TR/2012/REC-owl2-primer-20121211/>> (date of access: 23.06.2022).
17. *Horridge, M.* A Practical Guide to Building OWL Ontologies Using Prot g 4 and CO-ODE Tools. Edition 1.3. Copyright The University Of Manchester. March 24, 2011. 107 p.
18. Apache Jena Fuseki. [Online]. <<https://jena.apache.org/documentation/fuseki2/>> (date of access: 23.06.2022).
19. *DuCharme, B.*, 2013. Learning SPARQL. Querying and Updating with SPARQL 1.1 (Second edition), O`Reilly Media, All rights reserved, August 2013. 367 p.

Received 18.07.2022

M.G. Petrenko, Doctor of Technical Sciences, Leading Researcher, Microprocessor Technology Department, V.M. Glushkov Institute of Cybernetics, The National Academy of Sciences of Ukraine, Glushkov ave., 40, Kyiv, 03187, Ukraine, ORCID: <https://orcid.org/0000-0001-6440-0706>, petrng@ukr.net

O.V. Palagin, Academician of NAS of Ukraine (Computer Science), Professor, Doctor of Technical Sciences (Computers, Systems and Networks), deputy director of V.M. Glushkov Institute of Cybernetics, The National Academy of Sciences of Ukraine, Glushkov ave., 40, Kyiv, 03187, Ukraine, ORCID: <https://orcid.org/0000-0003-3223-1391>, palagin_a@ukr.net

M.O. Boyko, Research Fellow, Microprocessor Technology Department, V.M. Glushkov Institute of Cybernetics, The National Academy of Sciences of Ukraine, Glushkov ave., 40, Kyiv, 03187, Ukraine, ORCID: <http://orcid.org/0000-0003-1723-5765>, swsfrmac@gmail.com

S.M. Matveyshyn, PhD (Eng.), Research Fellow, Microprocessor Technology Department, V.M. Glushkov Institute of Cybernetics, The National Academy of Sciences of Ukraine, Glushkov ave., 40, Kyiv, 03187, Ukraine, ORCID: <https://orcid.org/0000-0001-9329-8800>, sergmmm2507@gmail.com.

KNOWLEDGE-ORIENTED TOOL COMPLEX FOR DEVELOPING DATABASES OF SCIENTIFIC PUBLICATIONS AND TAKING INTO ACCOUNT SEMANTIC WEB TECHNOLOGY

Introduction. The development of theories, methods and algorithms for the discovery and formation of new knowledge has always occupied one of the central places for any researcher, especially if he is actively working on the creation of new scientific publications. It is known that there is no universal language for formal description of concepts (knowledge) and systemology of transdisciplinary scientific research. And therefore, scientists face a number of priority problems, including the problem of significantly accelerating the receipt by a researcher of the cognitively structured information he needs from his sources. The tool complex for processing databases of scientific publications is oriented in this way to a researcher who has published from several tens to hundreds of scientific papers. We are not aware of search engines that could provide such information to a researcher in the shortest possible time. The toolkit implements *Information Retrieval and Knowledge Discovery* in *Databases* technologies with an emphasis on *Semantic Web* and cognitive graphics technologies and tools. The development of such a tool complex involves three stages: at the first stage, tools for implementing the complex, methods and algorithms for the interaction of the “*User ↔ Knowledge Engineer ↔ Remote Endpoint*” system and filling it with data are created; at the second stage, the tasks of multimedia representation of figurative-conceptual structures are solved, which are described in scientific documents, and at the third stage — the solution of the problem of extracting new knowledge.

Purpose. The purpose of our research was to further develop a tool complex for processing databases of scientific publications, which allows a scientist to significantly speed up the receipt of the necessary cognitively structured information from his sources.

Methods. The methods and models used in the work are based on the information technologies of the *Semantic Web* and ontological engineering.

Results. A tool complex for processing databases of scientific publications based on a remote endpoint based on the *Apache Jena Fuseki* server, basic *UML* diagrams of functioning and examples of executing user requests have been developed.

Conclusion. The article introduced and described the architectural and structural organization of the tool complex, which includes a local network from the user's *PC* and the *PC* of the administrator-knowledge engineer and a remote endpoint based on the *Apache Jena Fuseki* server, the main *UML* diagrams of the tool complex functioning and examples of executing user requests.

Keywords: *semantic Web technologies, ontological engineering, database of scientific publications, transdisciplinary scientific research, information and analytical support for a researcher.*