

DOI <https://doi.org/10.15407/csc.2020.05.017>  
УДК 004.318

**О.В. ПАЛАГІН**, академік НАН України, доктор технічних наук, професор, заст. директора, Інститут кібернетики імені В.М. Глушкова НАН України, 03187, м. Київ, просп. Академіка Глушкова, 40, Україна, palagin\_a@ukr.net

**М.Г. ПЕТРЕНКО**, доктор технічних наук, старший науковий співробітник, провідний науковий співробітник, Інститут кібернетики імені В.М. Глушкова НАН України, 03187, м. Київ, просп. Академіка Глушкова, 40, Україна, petrng@ukr.net

## **ЗНАННЯ-ОРІЄНТОВАНИЙ ІНСТРУМЕНТАЛЬНИЙ КОМПЛЕКС ОБРОБКИ БАЗ ДАНИХ НАУКОВИХ ПУБЛІКАЦІЙ**

*Розглянуто архітектуру інструментального комплексу обробки баз даних наукових публікацій та алгоритми функціонування на прикладі наукових статей. При цьому широко застосовуються технології Semantic Web, такі як SPARQL, RDF і OWL, які дозволяють одночасно отримувати й аналізувати інформацію із різних джерел. Наведено приклад створення формального опису наукової статті в онторедакторі Protege 5.5.0, формування до неї запитів й отримання відповідей.*

**Ключові слова:** онтологія, технології Semantic Web, база даних наукових публікацій.

### **Вступ**

На сьогодні відомі численні додатки й інструментарій, які реалізують технології пошуку інформації у різних текстових джерелах у відповідності із заданими параметрами. Причому результати пошуку подаються користувачу за кожним параметром пошуку окремо і не зв'язані між собою. А застосування технологій *Semantic Web* з метою багатопараметричного та зв'язаного пошуку інформації в різних джерелах в Україні перебуває на початковій стадії розвитку. Окремою проблемою є мультимедійне подання результатів пошуку та порівняння їх із понятійною структурою предметної області (*Knowledge Domain*), яка цікавить, із метою екстракції нових знань. З цієї точки зору для наукових досліджень обробки наукових публікацій одного автора, авторів наукового підрозділу й академічного інституту загалом актуальним є застосування технологій *Semantic*

*Web*, мультимедійне подання інформації з ефективною підтримкою процесу екстракції нових знань.

### **Постановка проблеми**

Знання-орієнтований інструментальний комплекс обробки баз даних наукових публікацій (ЗОІК БД НП) є системою, яка реалізує технології *Information Retrieval* і *Knowledge Discovery in Databases* з акцентом на технології й інструментарій *Semantic Web* та когнітивної графіки. Остання технологія та відповідний інструментарій дають змогу створювати мультимедійне подання образно-понятійних структур, які описуються в наукових документах. Технології *Semantic Web* включають процеси створення й обробки сховищ *RDF*-документів наукових публікацій (НП), розробки локальних і/або віддалених прикінцевих точок, формування

та виконання *SPARQL*-запитів користувача. Із всієї широкої множини технологій *Semantic Web* слід виокремити *SPARQL*-технологію, яка надає науковцю (Нк) можливість створювати запити довільного рівня складності й отримувати відповіді на них, включно різного роду інформацію для:

- аналізу охоплення науковими дослідженнями області наукових інтересів Нк, наукового відділу, академічного інституту (у відповідності з опублікованими науковими матеріалами);
- аналізу поточного стану виконаних наукових досліджень (НД) та опублікованих результатів;
- аналізу статистичних показників НП;
- підтримки створення нових наукових документів (статей, доповідей, звітів з НДР, патентів та ін.);
- мультимедійного подання.

Метою статті є проектування архітектури й алгоритмів функціонування інструментального комплексу обробки баз даних наукових публікацій, а також розробка прикладів використання формального опису наукової статті з виконанням певних запитів.

### Проектування знання-орієнтованого інструментального комплексу БД НП

Архітектуру ЗОІК БД НП (далі ІК) спроєктовано на основі онтологічного підходу [1, 2] і показано на рис. 1.

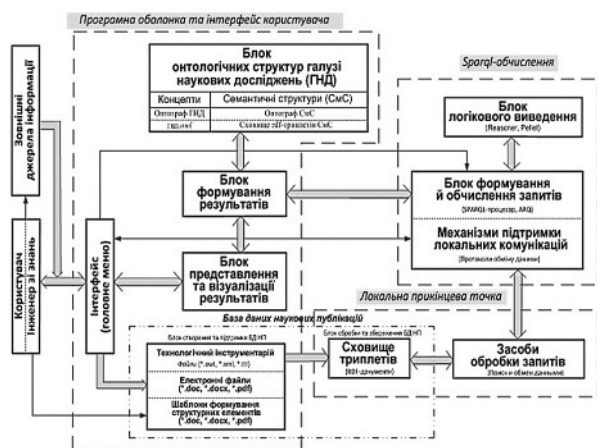


Рис. 1. Архітектура ЗОІК БД НП

ІК складається із трьох підсистем і блока зв'язку із зовнішнім середовищем, який забезпечує користувачу можливість завантажувати в базу даних (БД) нові наукові публікації. При цьому останні повинні пройти попередню обробку на відповідність прийнятим в системі вимогам подання НП в БД. ІК поєднує в собі різного роду інформаційні ресурси, оригінальні й вільно доступні програмні засоби, які, взаємодіючи між собою, реалізують сукупність алгоритмів автоматизованої побудови й використання БД наукових публікацій.

Підсистема *Програмна оболонка й інтерфейс користувача* включає інтерфейс (головне меню) користувача/інженера зі знань, блок онтологічних структур галузі наукових досліджень (ГНД), блок створення та підтримки БД НП, блок формування машинно-читаємих результатів виконання запитів і блок подання та візуалізації людино-читаємих результатів виконання запитів користувачів.

*Перший* компонент забезпечує взаємодію із системою користувача й інженера зі знань.

*Другий* компонент є інформаційним ресурсом, який створюється інженером зі знань і є набором онтографів галузі наукових досліджень і семантичних структур (СмС), а також включає відповідні файли їхніх формальних описів.

Особливістю подання й обробки наукових публікацій в ІК є узагальнення понятійних структур множини НП у вигляді онтографічного подання так званої галузі НД автора та деталізація кожної НП до рівня сукупності простих понятійних структур (або елементарних сенсів (ЕС)) і їхнього онтографічного подання. ЕС — це монада знань, неподільна одиниця. У лінгвістиці вона відповідає судженню чи двоскладному простому реченню. В основі ЕС лежить поняття (об'єкт), навколо якого групуються предикати, атрибути об'єктів та атрибути дій. Будь-яка система знань може бути подана як “згустком” ЕС, поданих у певній просторовій формі.

*Третій* компонент також створюється інженером зі знань і призначається для формування й підтримки в актуальному стані сховища



Рис. 2. Блок-схема шаблону для всіх типів НП

триплетів (*rdf*-документів) наукових публікацій. При цьому всі НП розбито на типи, такі як статті, монографії, енциклопедії, словники, препринти, доповіді, тези конференцій тощо. Для кожного типу НП розробляється свій шаблон, або метарівень онтографу.

Оскільки статті є найчисельнішим типом НП, далі буде розглядатися тільки цей тип публікацій (на першому етапі розробки). На рис. 2 подано блок-схему шаблону для всіх типів публікацій.

Важливою інформаційною компонентою в процесі перетворення електронної версії статті на її формальний опис в онторедаторі *Protege* є файл "Опис\_статті\_№N.docx". У ньому відображено текстові структури статті у відповідності з онтографом цієї статті.

Взаємозв'язки між відповідними вершинами онтографів шаблонів типів НП і самих НП (на рисунках показані онтографи тільки для статей) є такими:

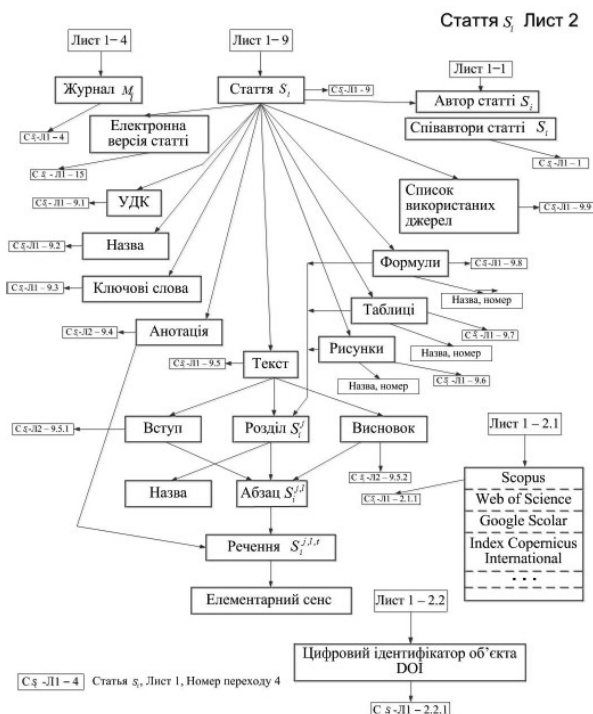


Рис. 3. Онтограф шаблону для наукових статей

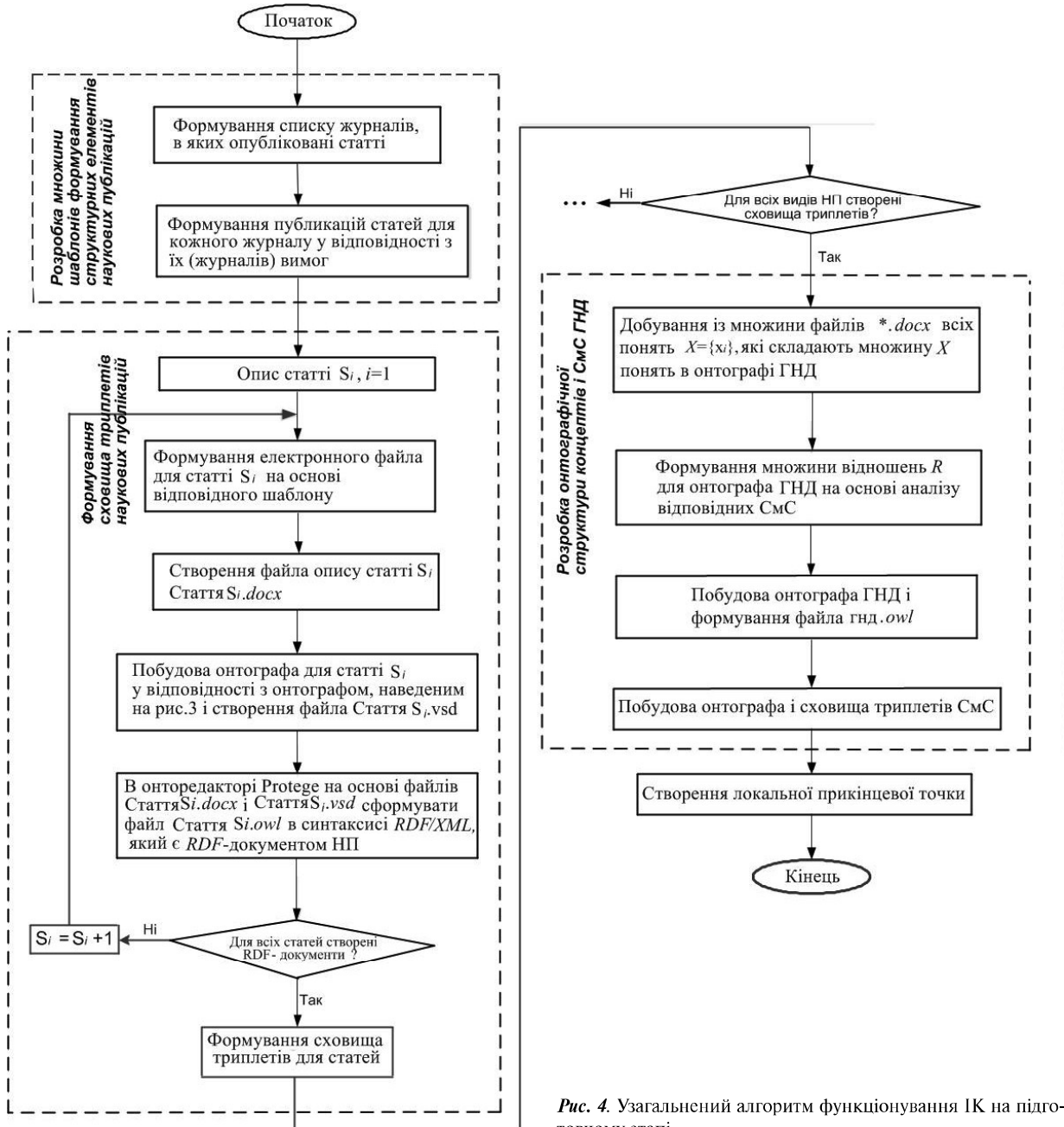


Рис. 4. Узагальнений алгоритм функціонування ІК на підготовчому етапі

- всі онтографи мають наскрізну нумерацію;
- онтограф шаблону для всіх типів НП подано на аркуші 1 (рис. 2);
- онтограф шаблону для статей подано на аркуші 2 (рис. 3);
- всі атрибути статті  $S_i$  включено в перехід з листа 1 “Стаття  $S_i$ , Аркуш 2 – 9”;

- інші атрибути, безпосередньо не пов’язані зі статтею  $S_i$ , пронумеровано як:
  - “Автор, співавтор – Стаття  $S_i$ , Аркуш 2 – 1”;
  - “База даних цитування – Стаття  $S_i$ , Аркуш 2 – 2.1”;
  - “Ідентифікатор цифрового об’єкта DOI – Стаття  $S_i$ , Аркуш 2 – 2.2”;

– “Журнал – Стаття  $S_i$  Аркуш 2 – 4”;

- для типів НП онтографи нумеруються як “Наукова публікація №... Аркуш 1”, “Наукова публікація №... Аркуш 2 (продовження)” тощо. Наприклад, для статей онтографи пронумеровано як “Стаття... № Аркуш 1”, “Стаття... № Аркуш 2 (продовження)”.

Слід також зазначити, що коли в НП який-небудь атрибут відсутній, то відповідне поняття (клас) в онтографі не буде мати індивідів.

*Четвертий* блок формує машинно-читаємі результати виконання запитів користувача разом із онтологічними структурами ГНД та їх формальними описами.

Нарешті *п’ятий* блок призначено для людино-машинного подання обчислених результатів користувачу, в тому числі і їхньої візуалізації в графічній формі.

Вільно доступними засобами реалізації ІК є поліфункціональний інструментарій *Protege 5.5.0*, який включає онторедактор, *Sparql*-процесор і різонер (або механізм логічного виведення). До нього можна підключити різного роду *plug in* для розширення функціональних можливостей. Зокрема, в ІК в ролі *Sparql*-процесора підключено *ARQ*-процесор, а в ролі різонера — *Pellet*. Останні є розповсюдженими інструментальними засобами (в тому числі для створення локальних прикінцевих точок). Вони підтримують версію рекомендацій *W3C SPARQL 1.1*.

У підсистемі *Програмна оболонка та інтерфейс користувача* блоки онтологічних структур ГНД і база даних НП реалізовано в *Protege 5.5.0*, а блоки формування машинно-читаємих результатів виконання запитів і подання й візуалізації людино-читаємих результатів є оригінальними програмами.

Підсистема *Sparql-обчислення* представляє собою *Sparql*-процесор і різонер *Hermit*, вбудовані в *Protege 5.5.0*, а також *plug-in*-процесор *ARQ* і різонер *Pellet*. Також в цю підсистему включено блок механізмів підтримки локальних комунікацій, який забезпечує передачу різного роду інформації у відповідності зі стандартизованими протоколами консорціуму *W3C*.

Підсистема *Локальна прикінцева точка* підтримує протокол *Sparql*-запитів і надає користувачу змогу звертатися до бази даних НП й отримувати від неї інформацію. Включає сховище *rdf*-документів НП, створене інженером зі знань, а також засоби обробки запитів. Останні також виконують стандартизовані функції передачі та перетворення інформації. Функціонування ІК включає два етапи — підготовчий і основний (користувацький).

## Алгоритми функціонування ІК

На підготовчому етапі інженеру зі знань необхідно виконати такі дії:

- розробити множину шаблонів формування структурних елементів НП. При цьому в *Microsoft Word* формуються файли: “Шаблон\_стаття.docx”, “Шаблон\_доповідь.docx”, “Шаблон\_звіт.docx”, “Шаблон\_патент.docx” та інші. Вихідною інформацією для розробки є вимоги та рекомендації відповідних видавництв наукової продукції;
- сформувати базу електронних файлів НП, які надаються користувачем (автором статей), у відповідності з розробленими шаблонами;
- сформувати базу даних НП у відповідності з онтографом, верхній рівень якого (загальний для всіх видів наукових публікацій) наведено на рис 2. При цьому для кожної НП в онторедакторі *Protege* формується файл із розширенням *\*.owl*;
- сформувати сховище триплетів *rdf*-документів НП;
- створити локальну прикінцеву точку збереження й обробки *rdf*-документів НП;
- поповнювати сховище триплетів в режимі надходження нових НП;
- у взаємодії з користувачами розширювати функціональні можливості ІК (створювати нові запити);
- розробити онтологічну структуру концептів галузі наукових досліджень. При цьому для ГНД в онторедакторі *Protege* розробляється онтограф і формується файл із розширенням *\*.owl*. Припускається, що вказана

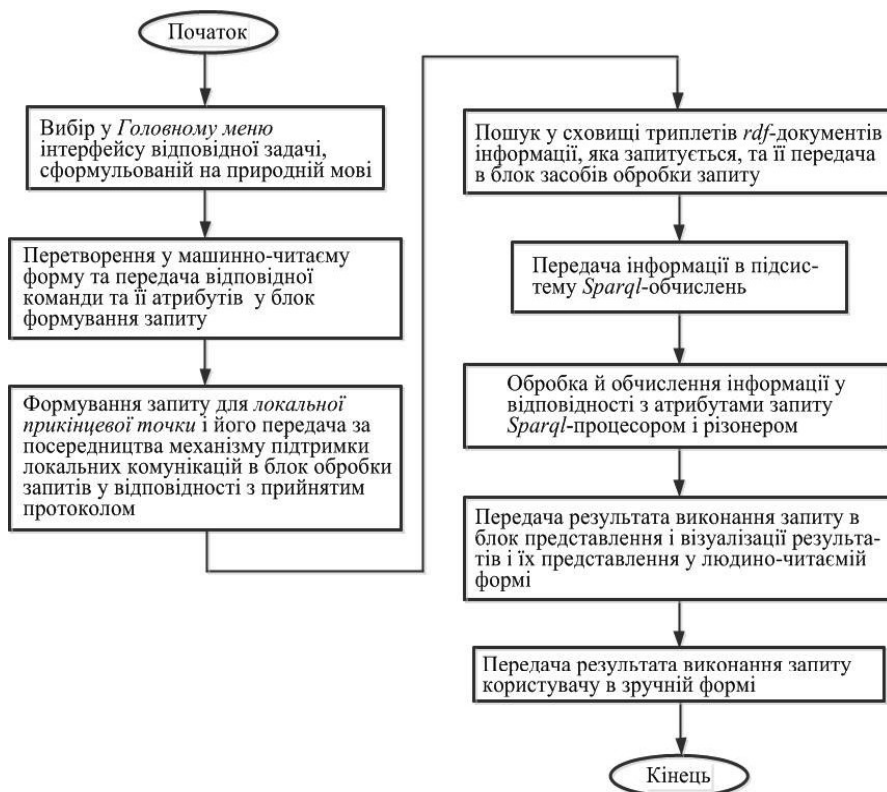


Рис. 5. Узагальнений алгоритм виконання задач I

онтологічна структура концептів формується на основі множини електронних версій НП у відповідності з методикою системної інтеграції й інструментарієм, описаними в [1];

- сформувати сукупність семантичних структур ГНД і в онторедаторі *Protege* розробити онтограф СМС галузі НД і сховище *rdf*-триплетів СМС.

Узагальнений алгоритм функціонування ІК на підготовчому етапі подано на рис. 4. Алгоритм наведено для формування БД статей, для інших НП відповідні процедури процесу розробки є аналогічними. Як уже зазначалося, всі процедури на цьому етапі виконуються інженером зі знань. Припускається, що авторами статей попередньо передано повнотекстові електронні файли статей інженеру зі знань. В основному режимі функціонування ІК користувачу на основі запитів (список яких сформовано в головному меню інтерфейсу зв'язку) надається можливість рішення наступних задач (передбачається розширення списку запитів і кількісно, і якісно).

#### I. Задачі інформаційного пошуку в наукових публікаціях:

- формування списку публікацій (за назвами) в заданому журналі (збірнику);
- формування списку статей за певною тематикою;
- формування списку статей за роками і датами;
- формування списку статей автора із заданим співавтором;
- формування списку статей за бібліографічними і наукометричними показниками;
- формування тексту (рисуноків, формул, таблиць, списку літературних джерел) заданої статті;
- формування списку анотацій (висновків, назв розділів тощо) заданих статей;
- формування списку видань (видавництва, журналів тощо) за заданими критеріями;
- пошук у НП інформації за заданими критеріями:
  - за ключовим словом;

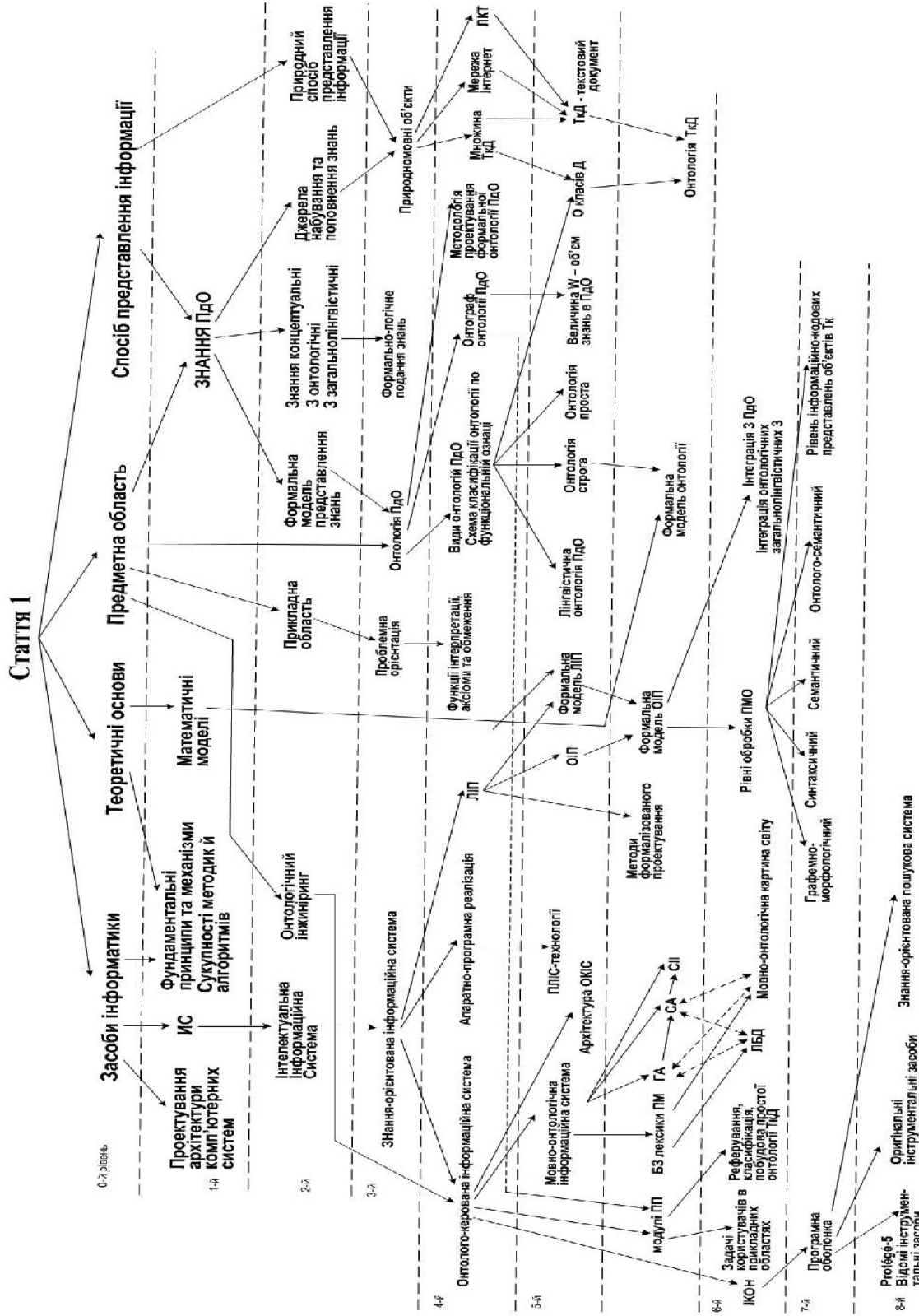


Рис. 6, а. Онтограф онтології об'єктів статті 17

ПДО – предметна область, З – знання, ЛП – логіко-інформаційний підхід, ОП – онто-інформаційний підхід, Тдд – тематичний довідник, Д – документ, ЛКТ – лінгвістичний курсив, ПП – предметний процес, ІЗ – база знань, ГА, СА – алгоритми графової і семантичної аналітики, СІ – семантико-інформаційний інтегратор, КОС – інструментальний комплекс обробки баз даних наукових публікацій, ЛБД – лексикографічна БД



- за групою ключових слів;
- за фрагментом тексту;
- комбінований пошук інформації у декількох базах даних НП (наступний етап розробки ІК).

II. *Задачі побудови онтографів і візуалізації інформації, яка міститься у відповідях на запити:*

- побудова та візуалізація онтографа наукової дисципліни;
- побудова та візуалізація онтографа наукової публікації;
- побудова та візуалізація онтографа заданого фрагмента тексту із НП;
- побудова та візуалізація онтографа за ключовими словами;

- обчислення та візуалізація у графічному вигляді статистичних показників НП.

III. *Задачі роботи із семантичними структурами (елементарними сенсами):*

- побудова та візуалізація онтографа ЕС наукової публікації (фрагмента НП).

Задачі користувача II, III потребують спільної обробки результатів виконання запитів у підсистемах *Локальна прикінцева точка та Sparql-обчислення* й інформації, яка міститься в блоці онтологічних структур ГНД підсистеми *Програмна оболонка й інтерфейс користувача*.

Узагальнений алгоритм виконання задач I представлено на рис. 5.



## Практична реалізація етапів попереднього проектування інформаційних компонентів ІК

Інформаційними компонентами ІК є блок онтологічних структур галузі наукових досліджень автора підсистеми “Програмна оболонка та інтерфейс користувача” та сховище *RDF*-документів підсистеми “Локальна прикінцева точка” (рис. 1). Перший інформаційний компонент (онтограф ГНД) розробляється у відповідності з методологією проектування *CRF*-моделі онтології предметної області [1] (*CRF* — *Concept Relation Interpretation Function*), а другий (*owl*-модель НП) — у відповідності з методиками, детально описаними в [3–6].

### Розробка *CRF*-моделі онтології статті 17

На рис. 6 наведено онтограф для статті №17 “К проектуванню онтолого-управляемой информационной системы с обработкой естественно-языковых объектов” [7]. Після розробки онтографів для кожної наукової публікації виконується їхня системна інтеграція [1], в результаті чого отримуємо загальний онтограф ГНД.

У статті 17 описуються теоретичні основи проектування онтолого-керованої інформаційної системи з обробкою природномовних об’єктів, сам процес проектування та наводиться вирішення конкретної задачі. Тому, у відповідності з методологією розробки *CRF*-моделі онтології онтограф подається сукупністю трьох компонент: онтології об’єктів, онтології процесів та онтології задач. Вказані онтографи подано на рис. 6, а–в.

### Розробка *owl*-моделі онтології структурних компонентів статті 17

Вихідний онтограф структурних компонентів статті 17 розроблено на основі узагальненого шаблону для наукових статей, файлу “Опис\_статті\_№17.docx” і подано на рис. 7, а, б.

Далі в онторедаторі *Protege* 5.5.0 вводяться онтографи (рис. 8, а–в) у відповідності зі зв’язками, показаними на рис. 8, г. При цьому періодично необхідно перевіряти на логікову сумісність введених класів та індивідів ризонером.

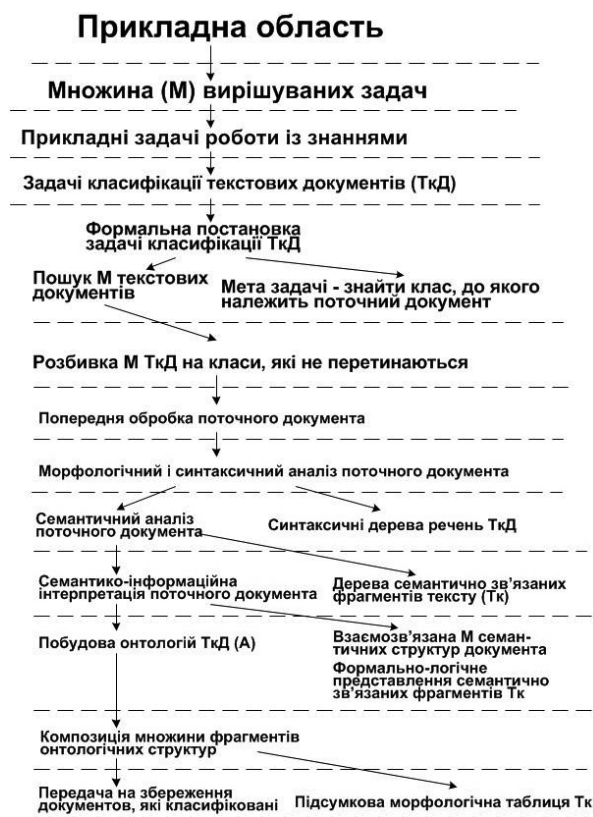


Рис. 6, в. Онтограф онтології задач статті 17

На рис. 9 наведено онтограф класів *owl*-онтології статті 17, сформований на вкладці *OntoGraf*. На рис. 10 подається вікно онторедатора, в якому показано вкладки, що відображають основні кроки розробки *owl*-онтології статті 17.

## Практичне використання *owl*-онтології статті 17

Нижче подано декілька запитів до змісту та реквізітам статті 17 і відповіді на них.

### Запит 1.

Показати анотацію статті 17.

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

PREFIX owl: <<http://www.w3.org/2002/07/owl#>>

Стаття 17, Лист 1

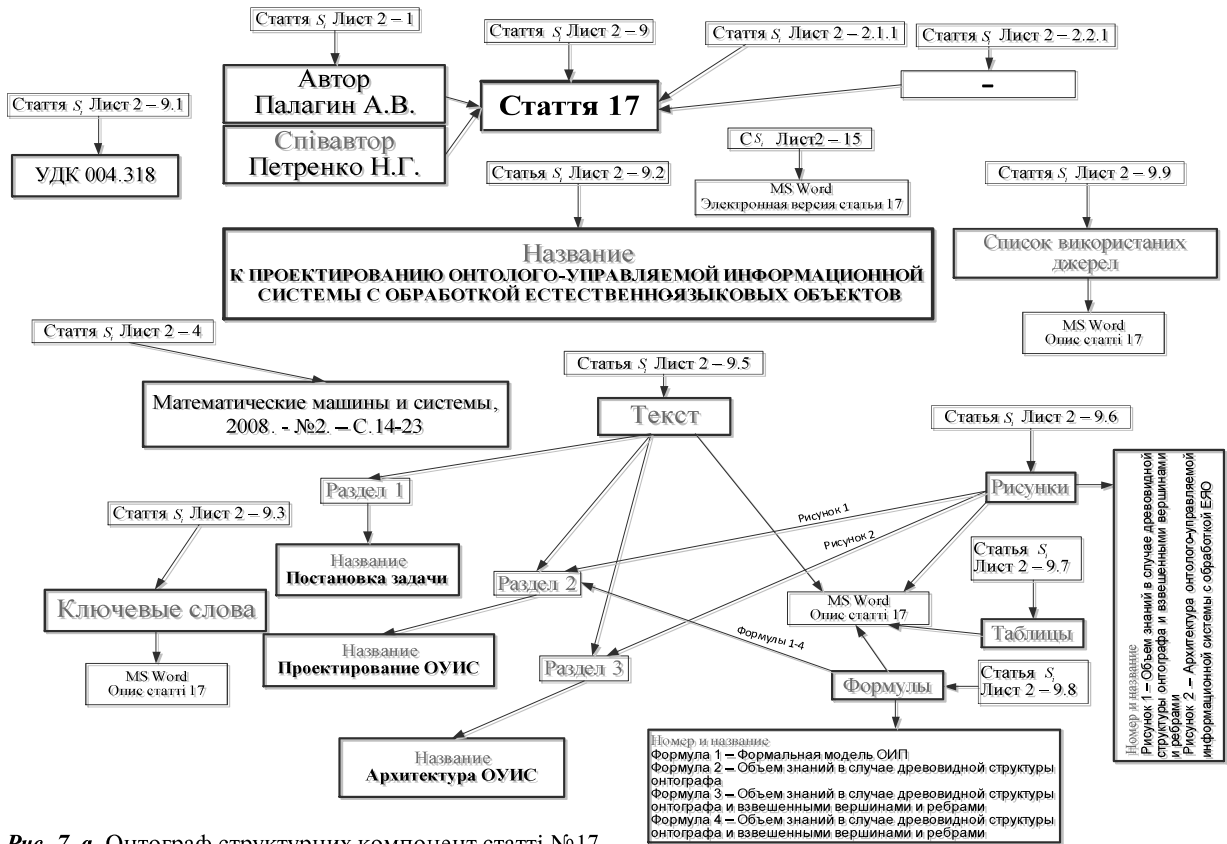


Рис. 7, а. Онтограф структурних компонент статті №17

Стаття 17, Лист 2

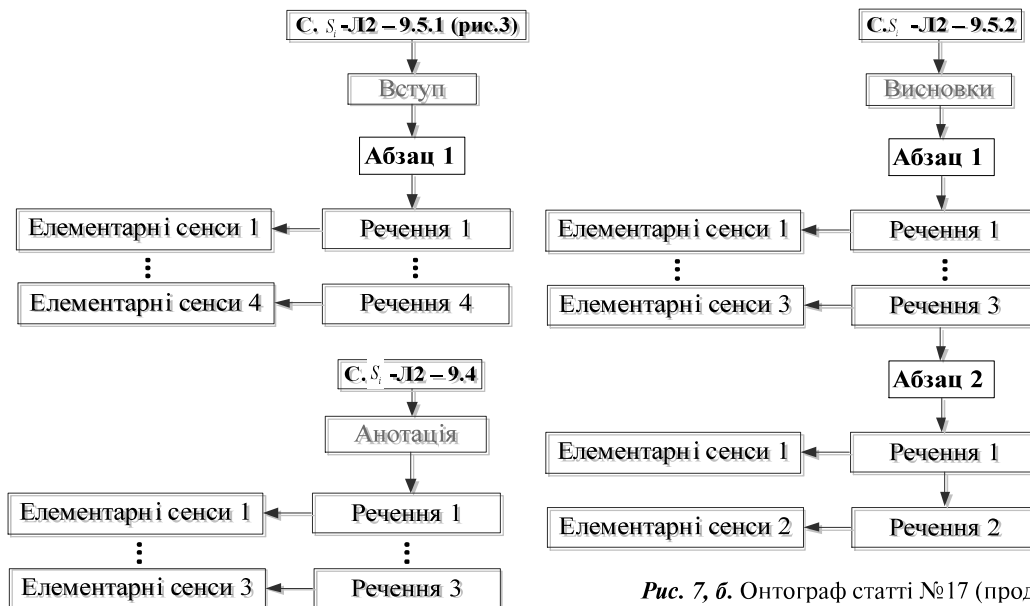
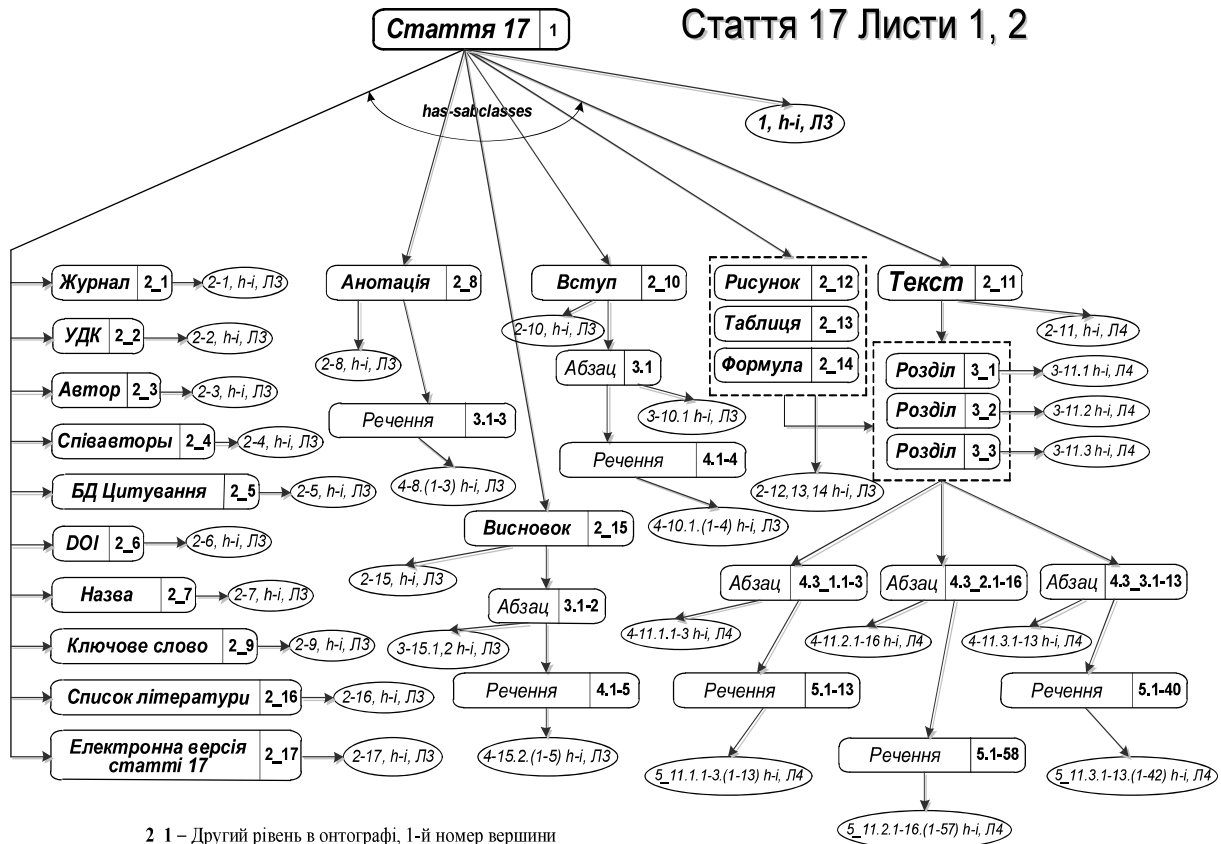


Рис. 7, б. Онтограф статті №17 (продовження).



2\_1 – Другий рівень в онтографі, 1-й номер вершини

Абзац 4.3\_1.1-3 – Абзаци, 4 рівень, розділ 3\_1, з 1 по 3-ій

Речення 5.1-13 – Речення Абзаців 4.3\_1.1-3, 5 рівень, з 1 по 13-е

Рис. 8, а. Онтограф класів статті 17

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

PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

PREFIX art17: <http://www.semanticweb.org/ніколай/ontologies/2020/3/Стаття17#>

SELECT ?Анотація?Речення\_N

WHERE {

?annot rdfs:type art17:Анотація .

?annot art17:Мати\_Речення ? Речення\_N .

? Речення\_N art17:Иметь\_Текст ?senttext\_str .

BIND (str(?senttext\_str) AS ?Анотація) .

}

Відповідь.

"В работе рассмотрен подход к формализованному проектированию онтолого-управляемой информационной системы с обработкой естественно-языковых объектов."

<http://www.semanticweb.org/ніколай/ontologies/2020/3/Стаття17#П4.8.1>

"Кратко рассмотрены основные задачи анализа и синтеза, решаемые на всех этапах проектирования." <http://www.semanticweb.org/ніколай/ontologies/2020/3/Стаття17#П4.8.2>

"Также рассмотрен пример решения задачи классификации текстовых документов." <http://www.semanticweb.org/ніколай/ontologies/2020/3/Стаття17#П4.8.3>.

**Запит 2.**

Показати назви розділів тексту.

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

PREFIX owl: <http://www.w3.org/2002/07/owl#>

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

PREFIX xsd: <http://www.w3.org/2001/XML-

Стаття 17, Лист 3

1, h-1, П1 – перехід з Листа 1, вершини 1-го рівня, відношення has-Individual

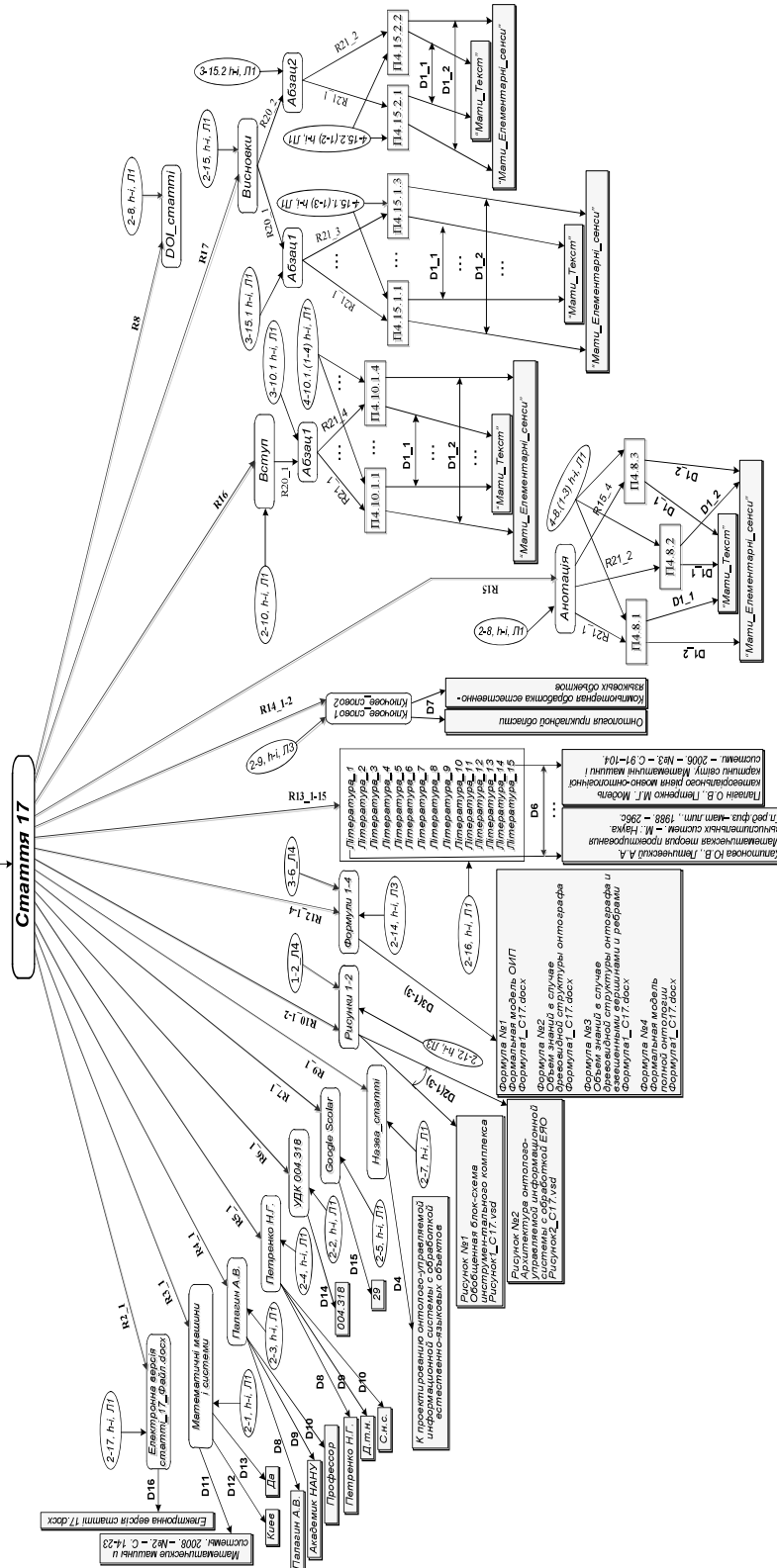
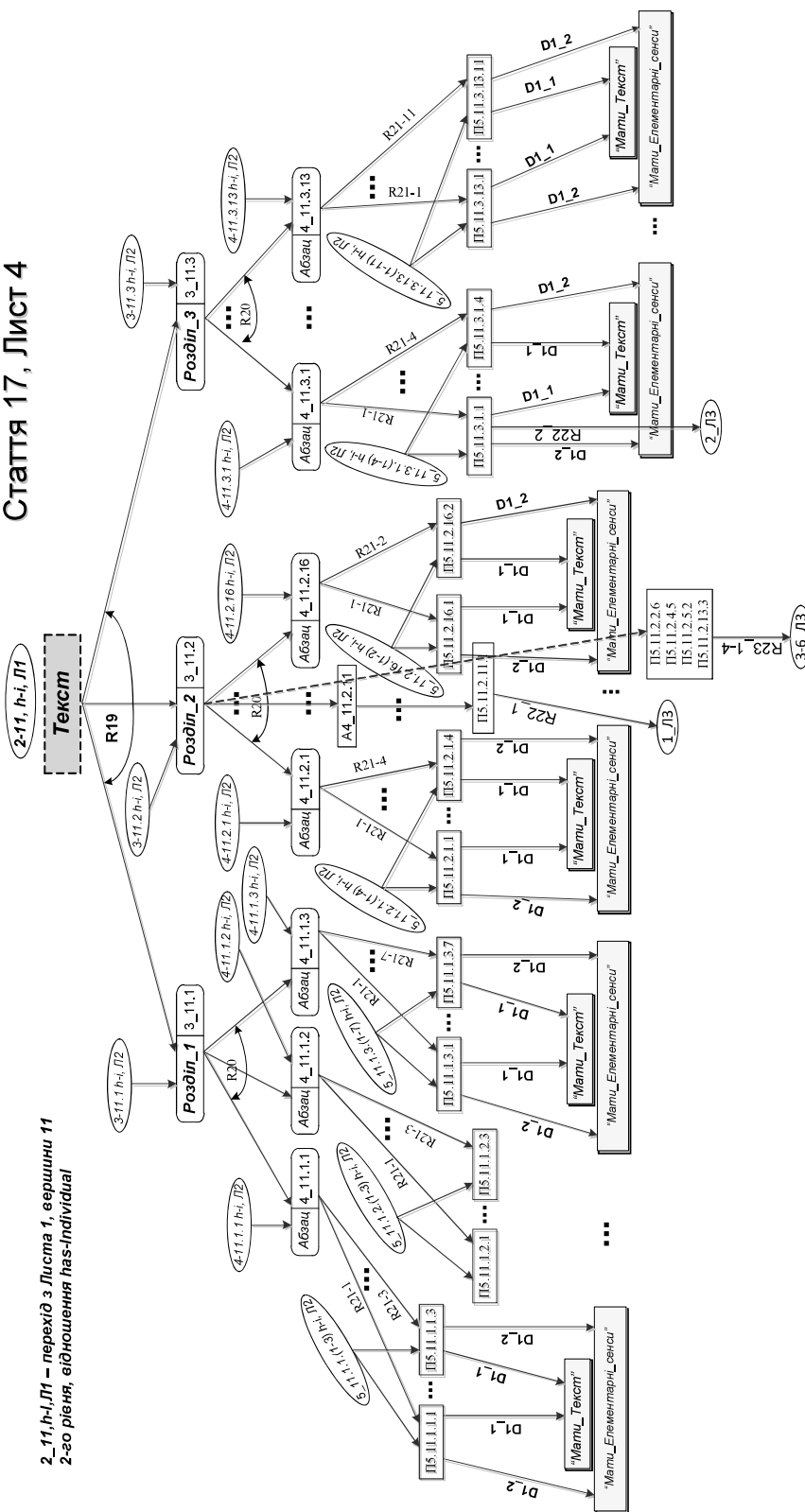


Рис. 8, б. Онтограф індивідів статті 17 (без індивідів класу "Текст")

Стаття 17, Лист 4



2\_11, h-i, Л1 – перехід з Листа 1, вершини 11  
2-го рівня, відношення has-Individual

Рис. 8, в. Онтограф індивідів класу "Текст" статті 17

**ObjectProperties**

- R2 – Мати електронну версію  
R2\_1 – Мати\_електронну\_версію
- R3 – Мати назву журналу  
R3\_1 – Мати\_НЖ
- R4 – Мати Автора  
R4\_1 – Мати\_Автора
- R5 – Мати Співавторів  
R5\_1 – Мати\_Співавтора
- R6 – Мати УДК  
R6\_1 – Мати\_УДК1
- R7 – Мати БД цитування  
R7\_1 – Мати\_БДЦ
- R8 – Мати DOI  
R8\_1 – Мати\_DOI
- R9 – Мати назву статті  
R9\_1 – Мати\_НС
- R10 – Мати\_рисунок
- R11 – Мати\_таблиці
- R12 – Мати\_Формули
- R13 – Мати\_Літературу
- R14 – Мати\_КС
- R15 – Мати\_Анотацію
- R16 – Мати\_Вступ
- R17 – Мати\_Висновки
- R18 – Мати\_Текст
- R19 – Мати\_Розділ
- R20 – Мати\_Абзац
- R21 – Мати\_Речення
- R22 – Посилатися\_на\_Рисунок
- R23 – Посилатися\_на\_Формулу

**Стаття 17, Лист 5**

**DataProperties**

- D1 – Реквізити\_речення  
D1\_1 – Мати\_Текст  
D1\_2 – Мати\_ЕС
- D2 – Реквізити\_рисунок  
D2\_1 – Назва\_рисунок  
D2\_2 – Номер\_рисунок  
D2\_3 – Файл\_рисунок
- D3 – Реквізити\_формули  
D3\_1 – Назва\_формули  
D3\_2 – Номер\_формули  
D3\_3 – Файл\_формули
- D18 – Реквізити\_таблиці  
D18\_1 – Назва\_таблиці  
D18\_2 – Номер\_таблиці  
D18\_3 – Файл\_таблиці
- D4 – Назва\_статті
- D5 – Назва\_розділу
- D6\_1 – Назва\_літератури
- D6\_2 – Кількість\_літератури
- D7 – Назва\_КС
- D8 – Мати\_ПІБ
- D9 – Мати\_вчений\_ступінь
- D10 – Мати\_вчене\_звання
- D11 – Мати\_Повні\_реквізити\_в\_Ж
- D12 – Мати\_місто\_видання\_журналу
- D13 – Бути\_СИ\_по\_ТН
- D14 – Мати\_значення\_УДК
- D15 – Мати\_кількість\_посилань
- D16 – Мати\_Ім'я\_Розширення
- D17 – Мати\_значення\_DOI

WHERE {

```
?text_ind rdf:type art17:Текст .
?text_ind art17:Мати_Розділ ?Розділ_статті .
?Розділ_статті art17:Назва_розділ ?nazv_str .
BIND (str(?nazv_str) AS ?Назва) .
}
```

Відповідь.

"Постановка задачі"

[http://www.semanticweb.org/ніколай/ontologies/2020/3/Стаття17#Раздел\\_1](http://www.semanticweb.org/ніколай/ontologies/2020/3/Стаття17#Раздел_1)

"Проектирование ОУИС"

[http://www.semanticweb.org/ніколай/ontologies/2020/3/Стаття17#Раздел\\_2](http://www.semanticweb.org/ніколай/ontologies/2020/3/Стаття17#Раздел_2)

"Архитектура ОУИС"

[http://www.semanticweb.org/ніколай/ontologies/2020/3/Стаття17#Раздел\\_3](http://www.semanticweb.org/ніколай/ontologies/2020/3/Стаття17#Раздел_3)

Запит 3.

Показати найменування наукометричних БД і кількість цитувань.

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

PREFIX owl: <<http://www.w3.org/2002/07/owl#>>

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

PREFIX xsd: <<http://www.w3.org/2001/XMLSchema#>>

PREFIX art17: <<http://www.semanticweb.org/ніколай/ontologies/2020/3/Стаття17#>>

Рис. 8, з. Зв'язки між індивідами класів статті 17

Schema#>

PREFIX art17: <<http://www.semanticweb.org/ніколай/ontologies/2020/3/Стаття17#>>

SELECT ?Розділ\_статті ?Назва

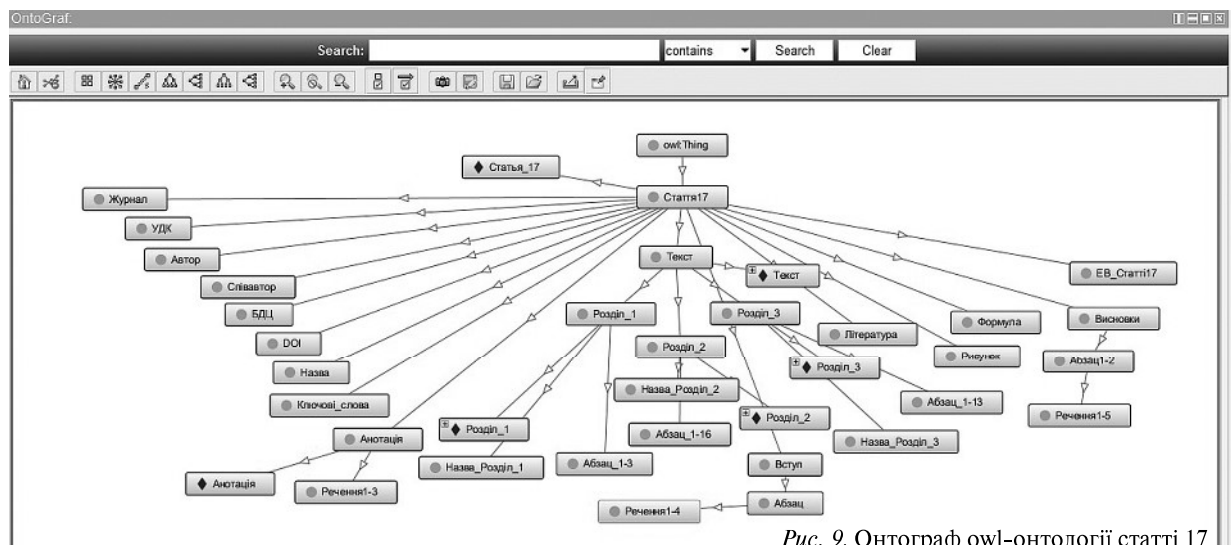


Рис. 9. Онтограф owl-онтології статті 17



Рис. 10. Вікно онторедатора з основними вкладками розробки owl-онтології статті 17

```
SELECT ?БД_цитування ?Кількість_цитувань
WHERE {
?БД_цитування rdf:type art17:БДЦ .
?БД_цитування art17:Мати_кількість_
посилань ?kolssyl_str .
BIND(str(?kolssyl_str)AS?Кількість_цитувань).
}
```

**Відповідь.**

**"Google\_Scholar"**

[http://www.semanticweb.org/ніколай/ontologies/2020/3/Стаття17#БД\\_цитування](http://www.semanticweb.org/ніколай/ontologies/2020/3/Стаття17#БД_цитування)  
**"29"**  
[http://www.semanticweb.org/ніколай/ontologies/2020/3/Стаття17#Кількість\\_цитувань](http://www.semanticweb.org/ніколай/ontologies/2020/3/Стаття17#Кількість_цитувань).

## Висновки

У статті розглянуто архітектуру інструментального комплексу обробки баз даних наукових публікацій та алгоритми його функціонування на підготовчому й основному етапах. Детально розглянуто кроки підготовчого етапу, які реалізуються інженером зі знань. При цьому зазначено створення двох моделей онтології наукової статті із поданням відповідних онтографів: *CRF*-модель описує поняття,

які містяться в статті, і *OWL*-модель описує структурні компоненти статті. Насамкінець наведено приклади запитів до БД НП, які демонструють працездатність інструментального комплексу. Надалі слід розширити використання в розробці інструментального комплексу технологій, таких як когнітивні семантика та графіка, мультимедійне подання інформації, орієнтовані на ефективну підтримку процесів екстракції і/або генерації нових знань.

## ЛІТЕРАТУРА

1. Палагін А. В., Кривий 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>.
2. Палагін О. В., Малахов К. С., Величко В. Ю., Шуров О. С. Проектування та програмна реалізація підсистеми створення та використання онтологічної бази знань публікацій наукового дослідника. Проблеми програмування. 2017. № 2. С. 72–81.
3. 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.
4. *OWL 2 Web Ontology Language Primer*. W3C. 2nd ed. URL: <http://www.w3.org/TR/2012/REC-owl2-primer-20121211/>.
5. Horridge M. *A Practical Guide To Building OWL Ontologies Using Protégé 4 and CO-ODE Tools*. Edition 1.3. The University Of Manchester, 2011. 107 p.
6. DuCharme B. *Learning SPARQL. Querying and Updating with SPARQL 1.1*. 2nd ed. O'Reilly Media. 2013. 367 p.

Надійшла 24.06.2020

## REFERENCES

1. Palagin A.V., Kryvyy S.L., Petrenko N.G., 2012. Ontologicheskiye metody i sredstva obrabotki predmetnykh znaniy [Ontological methods and means of processing subject knowledge], VNU im. V. Dalya V.I. Dal East Ukr. Nac. University, Lugansk. [online] Available at: <<http://www.aduis.com.ua/Monography.pdf>>. (In Russian).
2. Palagin A. V., Malakhov K. S., Velichko V. Yu., Shchurov O. S., 2017. "Proyektivannya ta prohramna realizatsiya pidsystemy stvorennya ta vykorystannya ontolohichnoyi bazy znan publikatsiy naukovoho doslidnyka" ["Design and software implementation of the subsystem for creating and using the ontological knowledge base of the researcher's publications"], Programming problems, 2, pp. 72–81. (In Ukrainian).
3. Petrenko N.G., Zelentsov D.G., 2019. "On the practical use of ontological models of subject areas", Computer modeling: analysis, management, optimization, 2 (6), pp. 58–73. DOI: 10.32434/2521-6406-2019-6-2-58-73. (In Russian).
4. *OWL 2 Web Ontology Language Primer*, W3C, 2nd ed. [online] Available at: <<http://www.w3.org/TR/2012/REC-owl2-primer-20121211/>>.
5. Horridge M. A., 2011. *Practical Guide To Building OWL Ontologies Using Protege 4 and CO-ODE Tools*, Edition 1.3, The University Of Manchester, 107 p.
6. DuCharme B. 2013. *Learning SPARQL. Querying and Updating with SPARQL 1.1*, 2nd ed., O'Reilly Media. 367 p.

Received 24.06.2020



*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, palagin\_a@ukr.net

*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, petrng@ukr.net

## KNOWLEDGE-ORIENTED TOOL COMPLEX PROCESSING DATABASES OF SCIENTIFIC PUBLICATIONS

**Introduction.** Nowadays, numerous applications and tools are known that implement information retrieval technologies in various text sources in accordance with specified parameters. Moreover, the search results are provided to the user for each search parameter individually and not related to each other. And the application of Semantic Web technologies for the purpose of multi-parameter and related information retrieval in various sources in Ukraine is at the initial stage of development. A separate problem is the multimedia presentation of search results and their comparison with the conceptual structure of the domain of interest (Knowledge Domain) with the goal of extracting new knowledge. From this point of view, it is relevant for scientific research to process the scientific publications of one author, authors of a scientific unit and the academic institute as a whole, using the Semantic Web technologies, multimedia presentation of information, and effective support for the process of extracting new knowledge.

**Purpose.** Designing the architecture and functioning algorithms of the instrumental complex for processing databases of scientific publications, as well as developing examples of using a formal description of a scientific article with a number of queries.

**Methods.** The methods and models used in this work are based on Semantic Web information technologies focused on the development and use of subject ontologies. Ontologies are the basic components of these technologies both for conducting scientific research and creating large data-bases, including scientific publications of the authors.

**Results.** The architecture of the instrumental complex for processing databases of scientific publications and the algorithms for its functioning at the preparatory and main stages have been developed. Examples of queries to the database of scientific publications that demonstrate the performance of IR are given.

**Conclusion.** The article discusses the architecture of the instrumental complex for processing databases of scientific publications and the algorithms for its functioning at the preparatory and main stages. The steps of the preparatory phase, which are implemented by the knowledge engineer, are examined in detail. At the same time, the creation of two ontology models of the scientific article with the presentation of the corresponding ontographs was highlighted: the CRF-model describes the concepts contained in the article, and the OWL-model describes the structural components of the article. In conclusion, examples of queries to the databases of scientific publications are presented, demonstrating the performance of the instrumental complex.

Further, it is necessary to expand the use in the development of IR technologies, such as cognitive semantics and graphics, multimedia presentation of information, focused on the effective support of the processes of extraction and/or generation of new knowledge.

**Keywords:** *ontology, Semantic Web technologies, database of scientific publications.*