

МАТЕМАТИЧНІ ТА ІНФОРМАЦІЙНІ МОДЕЛІ В ЕКОНОМІЦІ

УДК 004.9+556

<https://orcid.org/0000-0003-1946-0202>

<https://orcid.org/0000-0002-1229-1322>

<https://orcid.org/0000-0001-7153-5983>

<https://orcid.org/0000-0003-0211-520X>

В.Б. МОКІН, І.І. ОВЧАРЕНКО, А.М. ЛУЧКО, О.М. ДАВИДЮК

ПОБУДОВА МАСШТАБОВАНОЇ ІНФОРМАЦІЙНО-ПОШУКОВОЇ СИСТЕМИ ДЛЯ УПРАВЛІННЯ РІЧКОВИМ БАСЕЙНОМ НА ОСНОВІ РЕЄСТРІВ ТА ОНТОЛОГІЧНИХ МОДЕЛЕЙ

***Анотація.** Досліджувались варіанти підвищення швидкості масштабованого багатокритеріального пошуку релевантної інформації в інформаційно-пошуковій системі з уніформатними підходами для парсингу вхідних даних на прикладі системи для управління річковим басейном з можливістю пошуку даних за ключовими словами опису предметної області та просторовою і часовою прив'язкою цих даних. Запропоновано удосконалення принципів формалізації та моделі даних для таких систем, які дозволяють швидше здійснювати пошук даних та формування їх датасетів, готових для оброблення, після імпорту та формалізації інформації з різноформатних реєстрів та онтологічних моделей, за рахунок оперування окремими даними, а не складеними з ними типами даних. Кожне з таких даних ув'язується з усіма можливими семантичними мережами системи. Описана робота пілотної версії веб-системи, побудованої за запропонованими принципами, на прикладі роботи з даними моніторингу якості вод у басейні річки Південний Буг.*

***Ключові слова:** онтологічна модель, семантична мережа, інформаційно-пошукова система, план управління річковим басейном, просторові дані, парсер, багатокритеріальний пошук даних.*

DOI: 10.35350/2409-8876-2019-15-2-46-57

Вступ

Одним із найскладніших завдань підписаної Угоди про асоціацію України з ЄС є імплементація основних вимог Водної Рамкової Директиви ЄС. Фактично ця вимога потребує повної зміни усієї водної політики в країні. Головним у ній є гідрографічне районування країни, виділення річкових басейнів, створення Планів управління річковими басейнами (ПУРБ), які передбачають детальний аналіз

© В.Б. Мокін, І.І. Овчаренко, А.М. Лучко, О.М. Давидюк, 2019
стану водних екосистем, антропогенних впливів на них, стану водоохоронних

територій, економічний аналіз водокористування, створення програм заходів для усунення виявлених проблем тощо [1, 2]. Системна багаторічна робота Державного агентства водних ресурсів України, за участі провідних науковців та євроекспертів, спільно з іншими відомствами та за сприяння Верховної Ради України і Кабінету Міністрів України дозволила за відносно стислий термін напрацювати і затвердити законодавчу та нормативно-правову базу переходу на європейські принципи управління річковими басейнами [1–3]. Але в інформаційному плані ситуація набагато складніша.

Звичайно, вже створено нові лабораторії, нові органи управління та ін., які найближчим часом зберуть і дозволять опрацювати інформацію за новими вимогами і підходами, але для задач управління складними природними та соціально-техногенними системами важливим є максимальне врахування релевантної інформації, зібраної в минулому: даних про водокористування, гідрометеоінформації, історичних даних про стан водних екосистем. Ефективні довгострокові прогнози можна робити на основі аналізу не тільки поточного стану – треба враховувати ретроспективні дані.

У той же час у попередні роки так і не був створений єдиний реєстр річок, не було єдиного погодженого між відомствами кодування водних об'єктів водного кадастру. Досі єдиним затвердженим реєстром річок вважається їх перелік у роботі 1983 р. [4]. Однак з вісімдесятих років ХХ ст. існує електронний реєстр спецводокористування 2-ТП (водгосп), з початку дев'яностих – в електронному вигляді, з 1860 р. Український гідрометцентр накопичує гідрометеоінформацію, різними суб'єктами моніторингу вод зібрано величезний обсяг інформації про стан водних екосистем та якість вод у басейнах річок України, багато років ведеться облік режимів експлуатації гідротехнічних об'єктів – всю цю інформацію варто ув'язати з новими підходами до реалізації водної політики, до формалізації екологічної та водогосподарської інформації, з даними, які будуть збиратись вже за новими стандартами – без цього неможливим є розроблення та впровадження ефективних ПУРБ [2, 5, 6]. Отже, актуальним є удосконалення принципів і підходів побудови банку даних інформаційної системи для управління річковим басейном, які дозволять здійснювати швидкий та ефективний пошук цих даних з можливістю масштабування, тобто різного ступеня агрегації: наприклад, знайшов інформацію про певну річку, одразу цікаво – а що в річці вищого порядку, куди ця річка впадає, або – а що на ділянках цієї річки чи на водогосподарських об'єктах на ній? Слід розробити алгоритми наповнення такої системи і ретроспективними даними, часто погано формалізованими і систематизованими, і сучасними, зібраними вже на основі єдиних реєстрів, у т.ч. узгоджених з європейськими.

Побудована за такими принципами система дозволить зробити вибірку даних за заданими критеріями, застосувати потужний інструментарій математичного моделювання, методів системного аналізу, методів прогнозування та теорії прийняття рішень до заданого водного або водогосподарського об'єкта, водної екосистеми чи їх частин, що дозволить здійснити наукове обґрунтування економічно ефективних рішень з управління господарством відповідно до принципів сталого розвитку та з урахуванням принципів інтегрованого та басейнового управління водними ресурсами.

Мета статті – підвищити швидкість масштабованого багатокритеріального пошуку релевантної інформації в інформаційно-пошуковій системі з уніформатними підходами для парсингу вхідних даних на прикладі системи для

управління річковим басейном з можливістю пошуку даних за описом предметної області та просторовою і часовою прив'язкою цих даних.

1. Загальна постановка задачі

Як відомо, існує декілька основних підходів для створення баз даних інформаційно-пошукових систем. Найбільш поширеними підходами є, по-перше, побудова реляційних баз даних, коли усі дані чітко зв'язуються між собою за наперед заданим алгоритмом [5], але такий підхід є ефективним для формалізації лише реєстрів, наприклад КОАТУУ чи реєстру масивів вод, разом із водогосподарським та гідрографічним районуванням, оскільки гарантує однозначність віднесення заданого об'єкта до того чи іншого класу та його зв'язок з іншими. Для задач з певною невизначеністю подібних зв'язків, в першу чергу семантичною, найбільше поширення отримав підхід на основі онтологічних моделей, в Україні відомий, наприклад, за роботами О. Є. Стрижака, М. А. Попової, В. О. Подліпаєва, В. О. Шумейка, О. В. Атрасевича, В. Б. Мокіна та ін. [7–10]. Теоретично перший підхід можна вважати підвидом другого, але принцип формування у них різний (у першому – чітко формалізований, у другому – переважно евристичний).

Особливістю управління річковими басейнами, як і іншими складними розподіленими у просторі системами, що зазнають активного зовнішнього впливу є прив'язка до різних реєстрів. І така прив'язка може бути як чітко формалізованою у різних реєстрах (КОАТУУ, реєстр масивів вод України тощо), геоінформаційних системах і довідниках, так і мати лише семантичну просторову прив'язку на кшталт середня чи верхня течія (частина) річки чи її водозбірного басейну. Подібна формалізація даних і на основі довідників, і на основі семантичних онтологій, і на основі чітких відношень між даними з урахуванням їх прив'язки у просторі та у часі вже була запропонована одним із авторів у роботі [10]. Однак недоліком зазначеного підходу є те, що результатом пошуку був не набір готових для оброблення даних, а набір лише посилань на файли відкритих даних, які містять необхідну інформацію. У даній роботі ж стоїть задача отримати на виході єдиний набір даних у форматі XML, JSON чи CSV, готовий до оброблення різними математичними методами та інформаційними технологіями. Уточнимо математичну постановку задачі в термінах і позначеннях робіт [5, 7–10].

Як відомо, онтологія O_i описується трійкою [8]:

$$O_i = \langle X, R, F \rangle, \quad (1)$$

де X , R , F – кінцеві множини відповідно: X – концептів (понять, термінів), які характеризують набір даних, R – відношень між концептами цих даних, F – функцій інтерпретації X та/або R у певній інформаційній системі.

У роботі [10] пропонувалось за моделлю (1) формалізувати дані категорій «Простір» O_p (наприклад, басейн р. Південний Буг, басейн р. Соб, Хмельницька область, Вінницька область тощо з урахуванням їх співвідношень) та «Ключові слова» O_k (води суші, поверхневі води, скид зворотних вод, водозбір, гребля, ГЕС, якість вод, концентрація нітратів, розчинений у воді кисень тощо):

$$O_p = \langle X_p, R_p, F_p \rangle, \quad (2)$$

$$O_K = \langle X_K, R_K, F_K \rangle, \quad (3)$$

Формалізація прив'язки даних у часі пропонувалась у вигляді [10]

$$T_i = [T_{i11}, T_{i12}] \cup [T_{i21}, T_{i22}] \cup [T_{i31}, T_{i32}] \cup \dots \cup [T_{iww1}, T_{iww2}], \quad (4)$$

де T_{i11} та T_{i12} – початковий та кінцевий моменти j -го періоду ($j = 1, 2, \dots, w$) i -го набору даних у днях з розбиттям часового періоду на чіткі фрагменти, наприклад рік (01.01-31.12), квартал, місяць, декада, тиждень.

Для кожного файлу з набором даних здійснювалась прив'язка в часі, просторі і по ключових словах [10]:

$$D = [T, \{O_p\}, \{O_K\}], \quad (5)$$

де символ $\{\}$ означає, що можлива багатоваріантність прив'язки і в просторі (басейн р. Південний Буг, Вінницька область), і по ключових словах (води суші, поверхневі води, якість вод, концентрація нітратів) до відповідних семантичних мереж, побудованих на основі онтологічних моделей (2) та (3). За цими принципами протягом 2016-2017 рр. була створена пілотна версія інформаційно-пошукової системи ISODATA (isodata.mmss.vntu.edu.ua/). Однак її практичні випробування виявили ряд недоліків. По-перше, на виході є цінними не стільки набори даних, скільки – окремі дані, які максимально є релевантними заданому запиту. По-друге, для часової прив'язки теж можливе застосування онтологічних моделей (наприклад, тепла чи холодна пора року):

$$O_T = \langle X_T, R_T, F_T \rangle, \quad (6)$$

крім того, часто дата невідома точно – є лише семантична прив'язка, наприклад «1-й квартал» без дати, яку формально замінити на 01.01-31.03 чи на 15.02 не зовсім коректно. По-третє, онтологічні моделі містять певну невизначеність та багатоваріантність, що ускладнює використання цих моделей для формалізації реєстрів, де часто забороняється неоднозначність трактування – для них частіше використовуються моделі, подібні до тих, які наведені у роботі [5]:

$$R_1 \subset R_2 \subset R_3 \subset R_4 \subset R_5, \quad (7)$$

де індекс біля R означає рівень агрегації утворення (1 – найменший). Іноді на якомусь рівні допускаються й варіанти (наприклад, 2-й рівень – обласні ради, 3-й – районні або міські ради). Як правило, реєстри мають чітке кодування об'єктів у них, яке й відтворює модель (6). Наприклад, водогосподарська ділянка № М5.1.5.59 означає басейн р. Остер («.59») суббасейну річки Десна («.5»), району басейну річки Дніпро («.1»), який впадає у Чорне море («М5») [2].

Далі пропонуємо модель вигляду (6) позначати як $\{\{R_{1-5}\}\}$, що означає 5-рівневий реєстр об'єктів. У загальному вигляді для просторових об'єктів реєстри пропонуємо позначати як $\{\{R_p\}\}$, а для інших семантичних зв'язків (наприклад, реєстр чи довідник показників якості води) – $\{\{R_K\}\}$. Позначенням

$\{\{\{R_p\}\}\}$ пропонуємо позначати множину реєстрів просторових об'єктів, яких, в загальному випадку, може бути багато за різними критеріями (реєстр адміністративних утворень КОАТУУ, реєстр масивів вод, реєстр рибогосподарських водойм, реєстр об'єктів природно-заповідного фонду тощо). Аналогічно, позначенням $\{\{\{R_r\}\}\}$ пропонуємо позначати множину інших реєстрів або довідників можливих видів об'єктів (не за просторовими ознаками), яких теж може бути багато (довідник показників якості води, довідник видів очисних споруд, довідник видів гідротехнічних споруд тощо). А позначенням $X_p \subset \{\{\{R_p\}\}\}$ будемо позначати належність елемента X_p кожному з цих реєстрів.

Реєстри можуть задаватись у різний спосіб. Наприклад, в країнах ЄС існує директива INSPIRE (Infrastructure for Spatial Information in the European Community), яка, за даними однойменного порталу inspire.ec.europa.eu, спрямована на створення інфраструктури просторових даних країн-членів ЄС для цілей екологічної політики і політики та діяльності країн ЄС, які можуть вплинути на навколишнє середовище [11], згідно з якою усі можливі просторові особливості просторових об'єктів формалізуються, як правило, у форматі XML.

Отже, задача, яка має бути розв'язана в даній статті, це – побудова моделі даних інформаційної системи, яка забезпечить результат пошуку:

$$D = \{D\} : \left((X_T = [T_1^*, T_2^*]) \cup (X_T = X_T^*) \cup (X_p = \{\{R_p\}\}) \cup (\{\{R_p\}\} = X_p^*) \cup (X_p = X_p^*) \cup (\{\{R_K\}\} = X_K^*) \cup (X_K = X_K^*) \right), \quad (8)$$

тобто пошук множини даних $\{D\}$ такої, що, по-перше, прив'язка в часі X_T кожного з цих даних буде охоплюватись заданим періодом, по-друге, семантично цей час буде відповідати онтології X_T^* , по-третє, просторова прив'язка усіх даних X_p^* буде стосуватись елемента відповідного реєстру $\{\{R_p\}\}$, вчетверте, семантично ця прив'язка буде відповідати онтології X_p^* , вп'яте, по інших (непросторових) ключових словах прив'язка усіх даних X_K^* буде стосуватись елемента відповідного реєстру $\{\{R_K\}\}$, вшосте, семантично ця прив'язка буде відповідати онтології X_K^* .

Крім цього, така система повинна забезпечувати евристичний пошук, коли додатковими умовами будуть ще й такі:

$$\left((X_p = X_p^*) \subset \{\{\{R_p\}\}\} \right) \cup \left((X_K = X_K^*) \subset \{\{\{R_K\}\}\} \right), \quad (9)$$

тобто просторова прив'язка усіх даних X_p (наприклад, «Волинська область», «Рівненська область», «район басейну річки Дніпро») семантично буде відповідати онтології X_p^* («басейн річки Прип'ять»), але ця прив'язка, у свою

чергу, буде відповідати певним реєстрам $\{\{R_p\}\}$ (в даному випадку – КОАТУУ та гідрографічного районування), аналогічно – для X_R .

А найбільш цінною особливістю такої інформаційно-пошукової системи має стати те, що модель даних повинна враховувати, що ряд ознак об'єктів вищого рівня можуть поширюватись і на об'єкти нижнього рівня, наприклад статус «рівнинна річка», віднесений до усієї річки, може автоматично бути поширений і на кожну її ділянку.

2. Ідеї розв'язання задач

Для розв'язання поставлених задач пропонуються такий ряд принципів та підходів:

1. Для прив'язки кожного даного до усього розмаїття характеристик, реєстрів, показників тощо пропонується відмовитись від намагання оптимізації комп'ютерного місця (пам'яті), яке на це виділяється, і зберігати та індексувати окремо кожне дане, а не набір даних (таблицю). Це є обґрунтованим саме для задачі управління річковими басейнами, де інформації не так вже й багато, але уся вона є дуже цінною в соціальному, екологічному та економічному плані. Крім того, в наш час хмарних технологій, датацентрів та постійного здешевлення вартості комп'ютерної пам'яті для зберігання інформації такий підхід є економічно обґрунтованим. Саме такий принцип забезпечив успіх відомого агрегатора версій програмного забезпечення GitHub, який на відміну від аналогів, що зберігають лише відмінності однієї версії від іншої і що спричиняє багато ускладнень подібних зв'язків, зберігає весь програмний код усіх версій кожної програми.

2. Для автоматизованого формування семантичних мереж на основі онтологічних моделей пропонується, за аналогією з роботою [10], використати принцип mesh-мереж, коли у разі формування нового елемента такої мережі (до якого прив'язується певне дане D) вказується його відповідність лише об'єктам вищого рівня.

3. Для введення кожного нового набору даних в систему слід створити автоматизований конструктор парсеру, який дозволить один раз налаштувати, яким чином кожне дане цього набору буде максимально ув'язуватись з усіма наявними реєстрами та онтологічними моделями в системі.

4. Для формалізації реєстрів створити типові конструктори парсерів, які дозволять для кожного реєстру створити його інформаційний аналог з можливістю оновлення, який буде зручно інтегруватись з іншими видами даних, передусім з онтологічними моделями, що дасть змогу досить швидко наповнити систему необхідними довідковими даними.

5. Забезпечити можливість порівняння альтернативних варіантів назв елементів семантичних мереж, які будуть обробляться як однакові (наприклад, «Південний Буг» = «П. Буг» = «Southern Bug» = «Southern Booh» = «S. Bug» = «S. Booh» або «БСК5» = «Біохімічне споживання кисню» = «Біохімічна потреба кисню» = «Biochemical Oxygen Demand» = «Biological Oxygen Demand» = «BOD»), що суттєво розширить можливості системи.

3. Алгоритм роботи парсеру системи

На даний момент існує певна кількість баз відкритих даних про водні ресурси України. Проте важливою їх особливістю є те, що кожна з них зберігає дані у власному форматі, вони – не уніфіковані. Тому тут постає проблема синхронізації даних між цими системами. Більшість систем дозволяють завантажити дані у вигляді файлів форматів csv, xls,xlsx, dbf та ін. Деякі з них підтримують і мають хоча б якийсь API, проте воно майже в усіх системах не є відкритим і про нього взагалі ніякої інформації не надають. Тому формально, якщо конкретизувати задачу, потрібно реалізувати парсери файлів зі специфічною структурою даних, які надають сервіси.

Розглянемо принципи роботи запропонованої технології на прикладі двох сервісів відкритих даних про стан водних ресурсів: системи «Чиста вода» (texty.org.ua/water) і системи «Моніторинг та екологічна оцінка водних ресурсів України» Державного агентства водних ресурсів України (monitoring.davr.gov.ua). Обидва з них мають власне API, яке на виході повертає csv-файл з певним форматом даних.

Дані в цих файлах мають спільні риси в структурі: вони складаються з інформації про водні об'єкти (назва, географічне положення та інша інформація), характеристик, по яких здійснювались вимірювання на цих об'єктах і, власне, значень цих вимірювань. Відповідно до цього пропонуємо структуру уніфікованої бази даних (БД) для зберігання даних цих сервісів, подану на рис. 1.

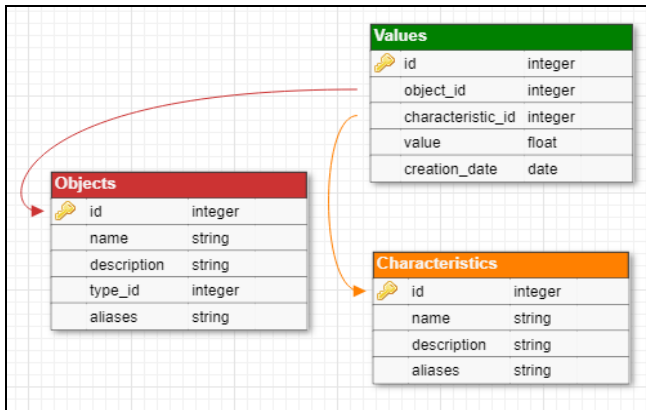


Рисунок 1 – Загальна структура БД розроблювальної системи

Таблиця «Objects» містить дані про об'єкти (водні, водогосподарські тощо), «Characteristics» – опис характеристик цих об'єктів (показники якості води, параметри гідротехнічних споруд, джерело даних (власник) тощо), «Values» – значення характеристик об'єктів.

Усі дані у файлах відкритих даних csv, xls, dbf, як правило, згруповані по стовпцях, тому конструктору парсерів просто потрібно вказувати, який стовпець є, наприклад, назвою об'єкта, який – назвою характеристики і т.д. В парсері ми вказуємо перелік стовпців файлу і зв'язки з полями відповідної таблиці (рис. 2) – це типовий підхід для такого роду сервісів.

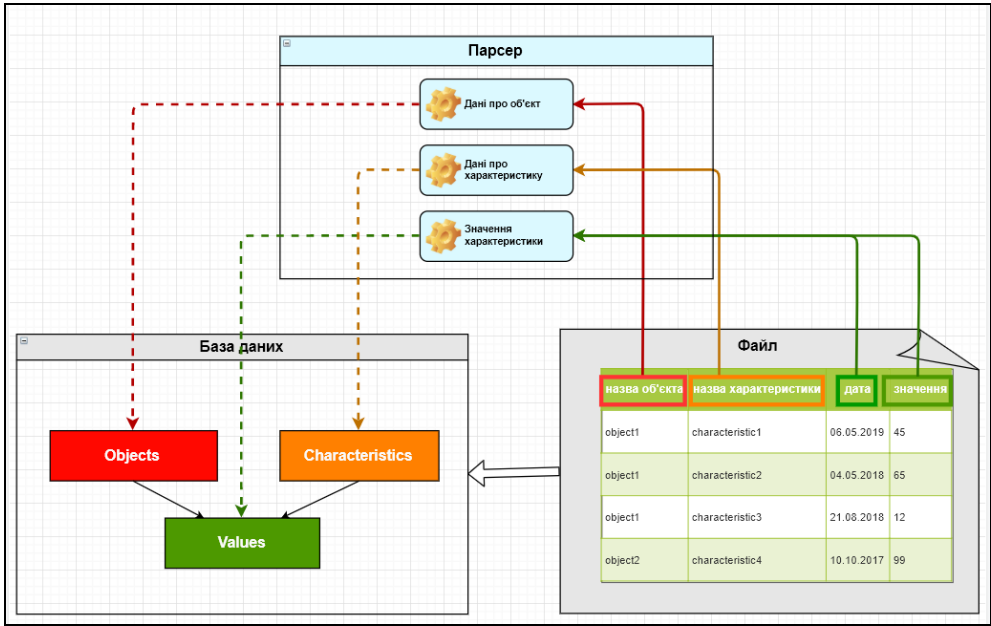


Рисунок 2 – Схема парсеру сервісів даних, необхідних для управління водними ресурсами

Далі після того, як визначені зв'язки між стовпцями файлу і відповідними полями сутностей БД, парсер буде проходити по усіх записах у файлі і конвертувати (уніфікувати) дані у БД. Якщо даних в БД для даного запису немає, то записуємо його, якщо є – пропускаємо. Слід зазначити, що об'єкти і характеристики в БД мають набір «псевдонімів» (поле «aliases»), оскільки ці сутності можуть мати альтернативні назви, яку було зазначено вище щодо назви р. Південний Буг та показника якості води «БСК5».

4. Приклад реалізації системи

Першим етапом розв'язання поставленої задачі є наповнення інформаційно-пошукової системи даними, за допомогою сконструйованих за запропонованими вище підходами парсерів. На основі наведеного вище алгоритму було створено парсер для сервісу «Чиста вода». Цей сервіс дозволяє отримати дані про стан водних ресурсів, які доступні за API у вигляді csv-файлу, що має структуру, наведену на рис. 3.


```

1 "id","name","Відомчий.код.ПС","river","Назва.лабораторії","date","key","value","norm","doc","mean","dev","size","last_date","lat","lon"
2 27222,"р. Дністер, 936 км. м. Залізняк, в/з, прав. берег, 50 м вище мосту через річку по дорозі Черніці - Тернопіль","ЧЕР ДНЕСТР 0000 0000 0000
3 27223,"р. Дністер, 940 км. с. Митків, правий берег, 200 м вище станції першого підйому в/з м. Черніці","ЧЕР ДНЕСТР 0000 0000 0000 0000 0000
4 27224,"р. Дністер, 926 км. м. Хотин, в/з, лівий берег, 600 м вище мосту дороги Черніці-Камен.-Под.", "ЧЕР ДНЕСТР 0000 0000 0000 0000 0000 0000 0000
5 27226,"р. Дністер, 708 км. с. Кормань, Дністрівське вдкш., пр. берег, безпосередньо в районі насосної станції водозабору","ЧЕР ДНЕСТР 0000 0000 0000
6 27227,"р. Дністер, 675 км. Дністрівське вдкш. в/б'єр, пр. берег, 500 м вище греблі ГЕС - 1","ЧЕР ДНЕСТР 0000 0000 0000 0000 0000 0675 13 77","Дніс
7 27228,"р. Дністер, 674 км. Дністрівське вдкш. в/б'єр, пр. берег, 500 м вище греблі ГЕС - 1","ЧЕР ДНЕСТР 0000 0000 0000 0000 0000 0674 14 77","Дні
8 27229,"р. Дністер, 658 км. с. Наславка, кордон з Молдовою, н/б'єр Нижньодністрівської ГЕС","ЧЕР ДНЕСТР 0000 0000 0000 0000 0000 0658 15 77","Дніс
9 27230,"р. Дністер, 631 км. м. Могилів - Подільський, (міст), Вінницька обл., митний перехід з Республікою Молдова","ЧЕР ДНЕСТР 0000 0000 0000 0000
10 27231,"р. Дністер, 550 км. с. Шиньківка, Ямпільський р-н, Вінницької обл. кордон з Молдовою, лівий берег, після складу з очисних споруд м. Сорочин
11 27303,"р. Прут, 418 км. с. Черепівці, міст, прикордонна зона з Румунією","ЧЕР ДУНАЙ 0181 0000 0000 0000 0000 0418 02 77","Дунай","Василькова німе
12 27304,"р. Прут, 448 км. м. Стуржоньча, в/з","ЧЕР ДУНАЙ 0181 0000 0000 0000 0000 0448 03 77","Дунай","Василькова лабораторія моніторингу вод та зг
13 27307,"р. Прут, 790 км. см. Неполоківці, кордон Ів.-Франківської і Чернівецької обл.", "ЧЕР ДУНАЙ 0174 0000 0000 0000 0000 0790 03 77","Дунай","Ге
14 27308,"р. Прут, 772 км. с.Ленківці, в/з м. Чернівці, лівий берег, 500 м вище мосту по дорозі на Чернівці","ЧЕР ДУНАЙ 0174 0000 0000 0000 0000 0772
15 27309,"р. Прут, 759 км. с. Магала, лівий берег, 3 км вище мосту по дорозі Чернівці-Хотин, 600 м вище складу стічних вод м. Чернівці","ЧЕР ДУНАЙ 01
16 27310,"р. Прут, 712 км. с. Тарасівці, (міст) - Орофтяни, прикордонна зона з Румунією","ЧЕР ДУНАЙ 0174 0000 0000 0000 0000 0712 06 77","Дунай","Вас
17 27311,"р. Прут, 697 км. с. Костичани, кордон з Румунією та Молдовою, 200 м вище впадіння в р. Прут з Червана","ЧЕР ДУНАЙ 0174 0000 0000 0000 0000
18 27312,"р. Прут, 678 км. с. Мамалига, прикордонна зона з Румунією та Республікою Молдова.", "ЧЕР ДУНАЙ 0174 0000 0000 0000 0000 0678 08 77","Дунай"
19 26960,"р. Уж, 172 км. м. Коростень, лівий в/з","ЧЕР ДНІПР 0981 0028 0000 0000 0000 0172 01 18","Дніпро","Лабораторія моніторингу вод та зруччя з
20 26951,"р. Ужгород, 120 км. с. Рудня Хоминська, кордон з Білоруссю","ЧЕР ДНІПР 0981 0280 0000 0000 0000 0122 01 18","Дніпро","Лабораторія моніторин
21 26952,"р. Случ, 155 км. с. Устя, кордон Рівненської і Житомирської обл-й","ЧЕР ДНІПР 0981 0412 0105 0000 0000 0155 03 56","Дніпро","Лабораторія м
22 26956,"р. Льва, 100 км. с. Переброди, кордон з Білоруссю","ЧЕР ДНІПР 0981 0339 0043 0000 0000 0100 01 56","Дніпро","Лабораторія моніторингу вод та
23 26953,"р. Случ, 4 км. с.Копи", "ЧЕР ДНІПР 0981 0412 0105 0000 0000 0038 04 56","Дніпро","Лабораторія моніторингу вод та зруччя Рівненської обл",
24 26955,"р. Сивачка, 95 км. с. Полонна, кордон з Білоруссю","ЧЕР ДНІПР 0981 0339 0000 0000 0000 0095 01 56","Дніпро","Лабораторія моніторингу вод та
25 26954,"р. Жомора, 52 км. м. Полонне, лівий в/з міста","ЧЕР ДНІПР 0981 0412 0105 0268 0000 0052 01 68","Дніпро","Лабораторія моніторингу вод та зг
26 26951,"р. Случ, 203 км. м. Ковторгад-Волницький, лівий в/з","ЧЕР ДНІПР 0981 0412 0105 0000 0000 0203 02 18","Дніпро","Лабораторія моніторингу вод
27 26950,"р. Случ, 406 км. с.Червоніака, Красільського р-ну, лівий в/з м. Хмельницький","ЧЕР ДНІПР 0981 0412 0105 0000 0000 0406 01 68","Дніпро","Г
28 26949,"р. Устя, 24 км. м. Ізясє, 2 км вище міста, вплив бурж", "ЧЕР ДНІПР 0981 0412 0302 0000 0000 0024 01 56","Дніпро","Лабораторія моніторингу
29 26948,"р. Горинь, 478 км. м. Славутич, зона впливу ХАЕС","ЧЕР ДНІПР 0981 0412 0000 0000 0000 0478 05 68","Дніпро","Лабораторія моніторингу вод та з
30 26446,"р. Гопинь, 67 км. с. Висоцьк, кордон з Білоруссю","ЧЕР ДНІПР 0981 0412 0000 0000 0000 0647 06 66","Дніпро","Лабораторія моніторингу вод та

```

Рисунок 3 – Приклад csv-файлу з даними сервісу «Чиста вода»

Файл містить наступні інформативні поля (стовпці): "id" – ідентифікатор запису; "name" – назва посту; "Відомчий.код.ПС" – відомчий код; "river" – назва річки, на якій здійснили дане вимірювання (просторова прив'язка); "Назва.лабораторії" – назва лабораторії; "date" – дата вимірювання (часова прив'язка); "key" – назва характеристики (назва забруднюючої речовини); "value" – значення характеристики; "lat" – широта; "lon" – довгота GPS-координат місця спостереження.

Виділимо з цих полів ті, що відносяться до опису об'єктів. Ними є такі поля: "name", "Відомчий.код.ПС", "river", "Назва.лабораторії". Виділимо поля, що відносяться до опису характеристики. Це – лише поле "key". Виділимо поля, які містять значення характеристик: "date" і "value". Також до них належать поля "lat" і "lon", які містять в собі зарезервовані в системі характеристики для об'єкта «широту» і «довготу», відповідно. Отже, ми визначили для даного парсеру необхідні дані з файлу. Далі відбувається зчитування даних з файлу, конвертація і їх запис в БД розроблювальної системи. Такий парсер реалізовано в пілотній версії системи (рис. 4).

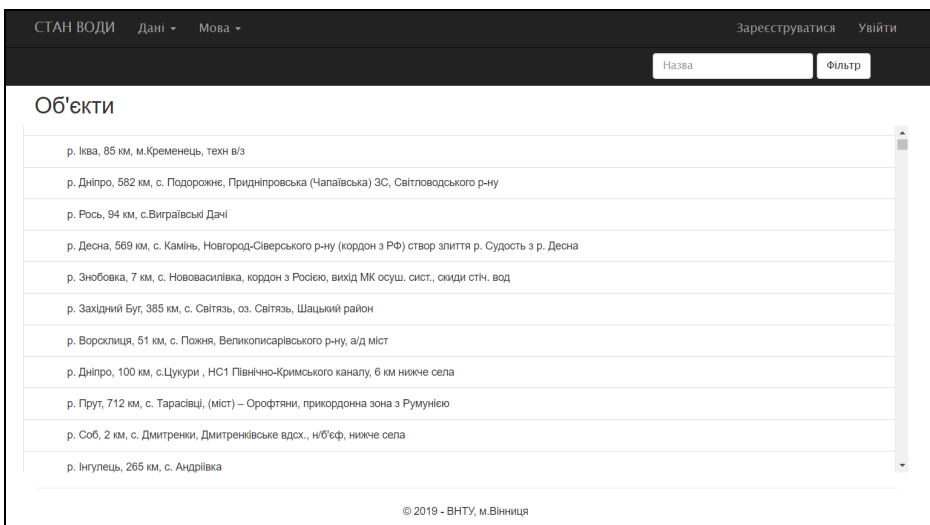


Рисунок 4 – Синхронізовані об'єкти

Додатково були створені передумови для формалізації і реалізації парсерів для завантаження в систему довідкових даних та для формування на основі даних реєстрів та результатів роботи парсерів інших вхідних даних онтологічних моделей та семантичних мереж системи.

Наступним етапом роботи інформаційно-пошукової системи є конструювання зручної форми для вибору критеріїв пошуку даних у системі, які враховували б структуру даних та їх взаємозв'язки, у т.ч. семантичні, з іншими даними за запропонованими у даній роботі підходами. На рис. 5 наведено приклад запиту, який відповідає наведеній на рис. 1 структурі даних, а на рис. 6 – приклад його виконання.

The screenshot shows a search query form with the following fields and values:

- Об'єкт:** р. Південний Буг, 582 км, м. Вінниця, Сабарівське вдсх, питний в/з міста, Вінницяоблводоканал
- Характеристика:** БСК5..мгО.дм3
- Тип:** якість вод
- Дата:**
 - Від:** 01.03.2018
 - До:** 31.05.2018

Each field has a corresponding 'Оберіть...' button with a red 'X' icon.

Рисунок 5 – Приклад запиту в пілотній інформаційно-пошуковій системі

	A	B	C	D	E
1	Об'єкт	Суб'єкт моніторингу	Характеристика	Дата	Значення
2	р. Південний Буг, 582 км, м. Вінниця, Сабарівське вдсх, питний в/з міста, Вінницяоблводоканал		Мінімальне за місяць БСК5, мгО2/дм3	03.2018	5,68
3	р. Південний Буг, 582 км, м. Вінниця, Сабарівське вдсх, питний в/з міста, Вінницяоблводоканал		Максимальне за місяць БСК5, мгО2/дм3	03.2018	5,92
4	р. Південний Буг, 582 км, м. Вінниця, Сабарівське вдсх, питний в/з міста, Вінницяоблводоканал		Мінімальне за місяць БСК5, мгО2/дм3	04.2018	5,68
5	р. Південний Буг, 582 км, м. Вінниця, Сабарівське вдсх, питний в/з міста, Вінницяоблводоканал		Максимальне за місяць БСК5, мгО2/дм3	04.2018	5,92
6	р. Південний Буг, 582 км, м. Вінниця, Сабарівське вдсх, питний в/з міста, Держводагентство		БСК5, мгО/дм3	17.04.2018	4,3
7	р. Південний Буг, 582 км, м. Вінниця, Сабарівське вдсх, питний в/з міста, Вінницяоблводоканал		Мінімальне за місяць БСК5, мгО2/дм3	05.2018	5,04
8	р. Південний Буг, 582 км, м. Вінниця, Сабарівське вдсх, питний в/з міста, Вінницяоблводоканал		Максимальне за місяць БСК5, мгО2/дм3	05.2018	10,56

Рисунок 6 – Результат виконання прикладу запиту

Подальший розвиток системи планується у напрямку наповнення її даними, необхідними для розроблення та верифікації основних розділів Плану управління річковим басейном на прикладі басейну р. Південний Буг.

Висновки

Розглянуто проблему побудови масштабованої інформаційно-пошукової системи для управління річковим басейном на основі реєстрів та онтологічних моделей. Обґрунтовано, що для пошуку максимально повної та релевантної інформації про задані об'єкти та їх характеристики необхідно враховувати особливості прив'язки інформації у просторі, часі та до певних сутностей (ключових слів), які можуть формалізуватись як через офіційні реєстри, так і через певні семантичні описи. Крім того, слід налагодити створення автоматичних програм-парсерів для завантаження формалізованих і збережених у певний спосіб даних із відомих форматів відкритих даних та офіційних реєстрів, у т.ч. з урахуванням моделей просторових даних як вітчизняного (КОАТУУ), так і європейського досвіду (директива INSPIRE та ін.).

Запропоновано удосконалення онтологічних моделей та принципів формування таких систем шляхом збереження не таблиць даних, а окремих даних, ув'язаних з усіма можливими реєстрами, семантичними мережами та описом предметної області та їх просторової і часової прив'язки. Для формування семантичних мереж запропоновано розвиток принципу формування mesh-мереж, раніше запропонований авторами для збереження таблиць даних у подібних інформаційно-пошукових системах.

Продемонстровано працездатність розробленої моделі та підходів до формалізації побудови масштабованої інформаційно-пошукової системи для управління річковим басейном р. Південний Буг на прикладі імпорту даних про якість поверхневих вод із веб-сервісів Держводагентства та «Чиста вода». Охарактеризовано основні етапи роботи, інтерфейс та особливості функціонування створеної системи.

Широке запровадження таких інформаційно-пошукових систем у сфері управління водними ресурсами України дозволить максимально повно опрацьовувати усю наявну інформацію про водні та водогосподарські об'єкти, пришвидшити її пошук та отримання у зручному для подальшого оброблення вигляді. Нескладне удосконалення моделі дозволить її застосування і в інших галузях та сферах управління природними ресурсами країни.

СПИСОК ЛІТЕРАТУРИ

1. Постанова Кабінету Міністрів України № 336 від 18.05.2017 року «Про затвердження Порядку розроблення плану управління річковим басейном».
2. Томільцева А. І., Яцик А. В., Мокін В. Б., Овчаренко І. І. та ін. Екологічні основи управління водними ресурсами : підручник. Інститут екологічного управління та збалансованого природокористування, 2017. – 200 с.
3. Постанова Верховної Ради України № 24, 1995 р. «Водний Кодекс України».
4. Дрозд Н. И. Материалы по типизации рек Украинской ССР. Академия наук Украинской ССР. 1953. Вып. № 452. С. 350.
5. Мокін В.Б. Комп'ютеризовані регіональні системи державного моніторингу поверхневих вод: моделі, алгоритми, програми: монографія. УНІВЕРСУМ-Вінниця. 2005. 310 с.
6. Клименко М. О., Мокін В. Б., Овчаренко І. І., Крижановський Є. М., Яцолт А. Р. та ін. [15 співавторів]. Рациональне використання та відновлення водних ресурсів : монографія. Житомир: Вид-во ЖДТУ ім. І. Франка, 2016. – 250 с.
7. Стрижак О. Є., Попова М. А. Формування таксономій шарів карт в ГІС-середовищах на основі онтологій натуральних систем. Радіоелектронні і комп'ютерні системи. 2013. Вип. № 4 (63). С. 46-54.
8. Стрижак О. Є. Онтологічні інформаційно-аналітичні системи. Радіоелектронні і комп'ютерні системи. 2014. Вип. № 3. С. 71–76.
9. В. О. Подліпаєв, В. О. Шумейко, О. В. Атрасевич. Атрибутивний пошук геопросторових даних та інформації про геопросторові об'єкти у трансдисциплінарному інформаційному середовищі. Математичне моделювання в економіці. 2018. Вип. № 4. С. 21-26.
10. Мокін В. Б., Довгополук С. О., Боцула М. П., Коханський М. В. Розробка комплексної моделі інформаційно-пошукової веб-системи відкритих числових даних. Вісник Вінницького політехнічного інституту. 2017. Вип. № 1. С. 62-69.
11. Van Loenen B. INSPIRE Empowers Re-Use of Public Sector Information / B. van Loenen, M. Grothe // International Journal of Spatial Data Infrastructures Research. – 2014. – # 9. – Pages 86–106. – doi:10.2902/1725-0463.2014.09.art4.

REFERENCES

1. Resolution of the Cabinet of Ministers of Ukraine No. 336 dated May 18, 2017 "On Approval of the Procedure for the Development of the River Basin Management Plan".
2. Tomiltseva A.I., Yatsyk A.V., Mokin V.B., Ovcharenko I.I. and others. Environmental Basics of Water Management: A Textbook. Institute for Environmental Management and Sustainable Use, 2017. – 200 p.
3. Resolution of the Verkhovna Rada of Ukraine No. 24 of 1995 "The Water Code of Ukraine".
4. Drozd N. I. Materials on typification of the rivers of the Ukrainian SSR. Academy of Sciences of the Ukrainian SSR. 1953. Issue No. 452. S. 350.
5. Mokin B. B. Computerized regional systems of state monitoring of surface waters: models, algorithms, programs: monograph. UNIVERSUM-Vinnitsia. 2005. 310 p.
6. Klimenko M. O., Mokin V. B., Ovcharenko I. I., Kryzhanovsky E. M., Yaschalt A. R. and others. [15 collaborators]. Rational Use and Recovery of Water Resources: Monograph. Zhytomyr: View at ZHTTU them. Ivan Franko, 2016. – 250 p.
7. Stryzhak O. Ye., Popova M. A. Formation of taxonomies of card layers in GIS-environments on the basis of ontologies of natural systems. Radio electronic and computer systems. 2013. Issue No. 4 (63). Pp. 46-54.
8. Strizhak O. E. Ontological information-analytical systems. Radio electronic and computer systems. 2014. Issue No. 3. P. 71-76.
9. V.O. Podlipaev, V.A. Shumeiko, O.V. Atrasevich. Attributive search of geospatial data and information on geospatial objects in a transdisciplinary informational environment. Mathematical Modeling in Economics. 2018. Vip. No. 4. p. 21-26.
10. Mokin V. B., Dovgopolyuk S. O., Botsula M. P., Kokhansky M. V. Development of a comprehensive model of information retrieval web system of open numerical data. Bulletin of the Vinnitsa Polytechnic Institute. 2017. Issue No. 1 P. 62-69.
11. Van Loenen B. INSPIRE Empowers Re-Use of Public Sector Information / B. van Loenen, M. Grothe // International Journal of Spatial Data Infrastructures Research. – 2014. – # 9. – Pages 86–106. – doi:10.2902/1725-0463.2014.09.art4.

Стаття надійшла до редакції 18.04.2019.