



УДК 004.777(045)

[https://doi.org/10.52058/2708-7530-2022-5\(23\)-273-284](https://doi.org/10.52058/2708-7530-2022-5(23)-273-284)

Половенко Людмила Петрівна кандидат педагогічних наук, доцент кафедри економічної кібернетики та інформаційних систем Вінницького торговельно-економічного інституту ДТЕУ, 21050, м. Вінниця, Хмельницьке шосе, 25, тел.: (0432) 55-04-46, <https://orcid.org/0000-0002-9909-825X>

Мерінова Світлана Володимирівна кандидат економічних наук, доцент кафедри економічної кібернетики та інформаційних систем Вінницького торговельно-економічного інституту ДТЕУ, Хмельницьке шосе, 25, м. Вінниця, 21050, тел.: (0432) 55-04-46, <https://orcid.org/0000-0001-6563-5320>

ТЕХНОЛОГІЇ ПРОДУКУВАННЯ ЗНАНЬ НА ОСНОВІ ВЕБ-СЕРВІСІВ

Анотація. Доступність інфраструктур просторових даних та сумісних сервісів приводить до потреби видобутку знань із об'ємних даних. Це спричинено збільшенням розміру, об'ємів, різноманітності та швидкості геопросторових даних. Продуктування знань методами інтелектуального аналізу розподілених даних в інфраструктурі просторових даних залишається критичною проблемою.

В статті наведено найбільш поширені технології інтелектуального аналізу даних. Представлено процес виявлення знань у базах даних та розглянуто технології інтелектуального аналізу даних як один із кроків цього процесу.

Проаналізовано технології інтелектуального аналізу даних як основи технології web mining. Представлено категорії web mining: аналіз використання веб ресурсів; видобування веб-структур; видобування веб-контенту. Розглянуто основні механізми дослідження та добування інформації з веб-документів та сервісів. Продемонстровано застосування інструментів web-mining для аналізу інформації та її обробки за допомогою засобів інтелектуального аналізу даних.

Обґрунтовано ефективність використання веб-сервісів продуктування знань в якості прошарку над інфраструктурами просторових даних з метою надання користувачам просторовими даними та особам, які приймають рішення, можливість видобувати знання з масивів різноманітних просторових даних.

Досліджено питання видобування корисних знань із різноманітних та розподілених просторових даних за допомогою сумісних і стандартних сервісів. Підкреслено, що фреймворки Hadoop і Spark забезпечують високу продуктивність видобування шаблонів та знань з великих об'ємів просторових

даних. Веб-сервіси для видобування знань із реальних геоданих дозволяють використовувати динамічну, більш просту і значно більш швидку процедуру продукування знань із просторових даних.

Ключові слова: інтелектуальний аналіз даних, продукування знань, інтелектуальне розподілене середовище, методи web-mining, веб-сервіси видобування знань.

Polovenko Lyudmyla Petrovna Candidate of Pedagogical Sciences, Associate Professor of the Economic Cybernetics and Information Systems Department, Vinnytsia Institute of Trade and Economics, State University of Trade and Economics, Khmelnytske Shosse St., 25, Vinnytsia, 21050, tel.: (0432) 55-04-46, <https://orcid.org/0000-0002-9909-825X>

Merinova Svitlana Volodymyrivna Candidate of Economic Sciences, Associate Professor of the Economic Cybernetics and Information Systems Department, Vinnytsia Institute of Trade and Economics, State University of Trade and Economics, Khmelnytske Shosse St., 25, Vinnytsia, 21050, tel. : (0432) 55-04-46, <https://orcid.org/0000-0001-6563-53>

KNOWLEDGE PRODUCTION TECHNOLOGIES WITH THE HELP OF WEB SERVICES

Abstracts. The rapid increase in the size, volume, diversity and speed of geospatial data leads to the availability of spatial data infrastructures and compatible services, As a result we face the necessity of the knowledge extraction from bulk data. The production of knowledge by different intellectual analysis methods of distributed data in the spatial database remains a critical issue.

The article presents the most common technologies of data mining. The authors review the process of knowledge discovery in databases and consider the data mining technologies as one of the component of this process.

In the article we analyse the technologies of data mining as the basis of web mining technology. The categories of Web Mining: analysis of the use of web resources; extraction of web structures; web content extraction. The authors offer to examine the main mechanisms of research and extraction of information from web-documents and services.

In the article we sustain the using of the web services of knowledge production as a layer over spatial data infrastructures as an effective methods to provide users with spatial data and decision making, the ability to extract knowledge from arrays of heterogeneous spatial data.



One of the main emphasis of the article is the that the Hadoop and Spark frameworks provide high productivity of extracting templates and knowledge from large amounts of spatial data. Web services for extracting knowledge from real geodata allow us to use a dynamic, simpler and much faster procedure for producing knowledge from spatial data.

Keywords: data mining, knowledge production, intelligent distributed environment, Web-mining methods, web services for knowledge extraction.

Постановка проблеми. Інтелектуалізація методів обробки й аналізу даних останнім часом дозволяє вирішувати не лише класичні завдання ухвалення рішень, виявлення причинно-наслідкових зв'язків та прихованих закономірностей в системах. Стрімкий розвиток мережевих технологій, а також новітні розробки в таких технологіях, як смартфони, геодатчики, збільшують доступність інфраструктур просторових даних та сумісних сервісів.

Існує гостра необхідність у вирішенні проблеми видобутку знань із об'ємних даних. Просторові дані – критично важливі для багатьох підприємств та установ. Сучасні технології дозволяють видобувати дані та знання з неструктурованих інформаційних ресурсів, здійснювати інтелектуальний аналіз просторових даних.

Проте, досі проблемною нішею залишається задача інтеграції даних і представлення релевантної інформації з усіх можливих джерел. Дослідники припускають [7], що відсоток отримання практично значущих знань на основі опрацювання «сирих» даних, пошук прихованих в цих даних закономірностей, поки що досить низький.

Видобування високорівневої інформації, продукування знань із просторових даних базується на методах інтелектуального аналізу просторових даних, які добре працюють в централізованих системах.

Але їх застосування до розподілених даних в інфраструктурі просторових даних з метою відкриття знань залишається критичною проблемою. В розподіленому середовищі відповідні дані часто розміщені розрізнено і для застосування методів інтелектуального аналізу даних виникає необхідність розміщення всіх необхідних даних у сховища даних. А це трудомісткий процес, який потребує надійних, масштабованих, функціонально-сумісних і розподілених фреймворків обробки великих даних [5]. Для роботи з геопросторовими даними ефективним є використання grid технологій, просторових веб-служб.

Аналіз останніх досліджень і публікацій. Вітчизняні та зарубіжні дослідники досить широко розглядають моделі та методи інтелектуальних технологій, в тому числі і технологій продукування знань [1, 3, 6, 7]. Науковці досліджують методику застосування технологій видобування даних і знань в економіці та фінансах [4], освіті [6] тощо.

Останнім часом активно розвиваються та поширюються нові методи аналізу даних і видобування знань, які ґрунтуються на інноваційних парадигмах [3]. Обґрунтовується доцільність застосування онтологій та геоонтологій для розв'язання проблем, які виникають при інтеграції геопросторових та непросторових даних [8]. Для роботи з геопросторовими даними зарубіжні дослідники пропонують застосовувати просторові веб-служби [5]. Але методам інтелектуального аналізу просторових даних з використанням стандартної сервіс-орієнтованої архітектури приділяється недостатньо уваги.

Мета статті – дослідження технології видобування знань із розподілених просторових даних на основі методів web mining та геопросторових веб-сервісів.

Виклад основного матеріалу. Основна функція інтелектуального аналізу даних (англ. Data Mining) – виявлення прихованих правил та коректних, нових, потенційно корисних і таких що інтерпретуються шаблонів в великих масивах даних (англ. Big Data). Завдяки технологіям Data Mining можна виявити неочевидні, несподівані, нетривіальні тенденції та приховані глибинні пласти знань, про існування яких навіть не здогадувались.

Нерідко Data Mining ототожнюють з виявленням знань у базах даних (англ. Knowledge Discovery in Databases), хоча більш правильно вважати Data Mining одним із кроків цього процесу.

Процес видобування знань містить послідовність дій, необхідних для побудови моделі (рис. 1). Такими знаннями можуть бути правила, що описують зв'язки між властивостями даних (дерева рішень), шаблони, які часто зустрічаються (асоціативні правила), а також результати класифікацій (нейронні мережі) та кластеризація даних (самоорганізуючі карти Кохонена) тощо.

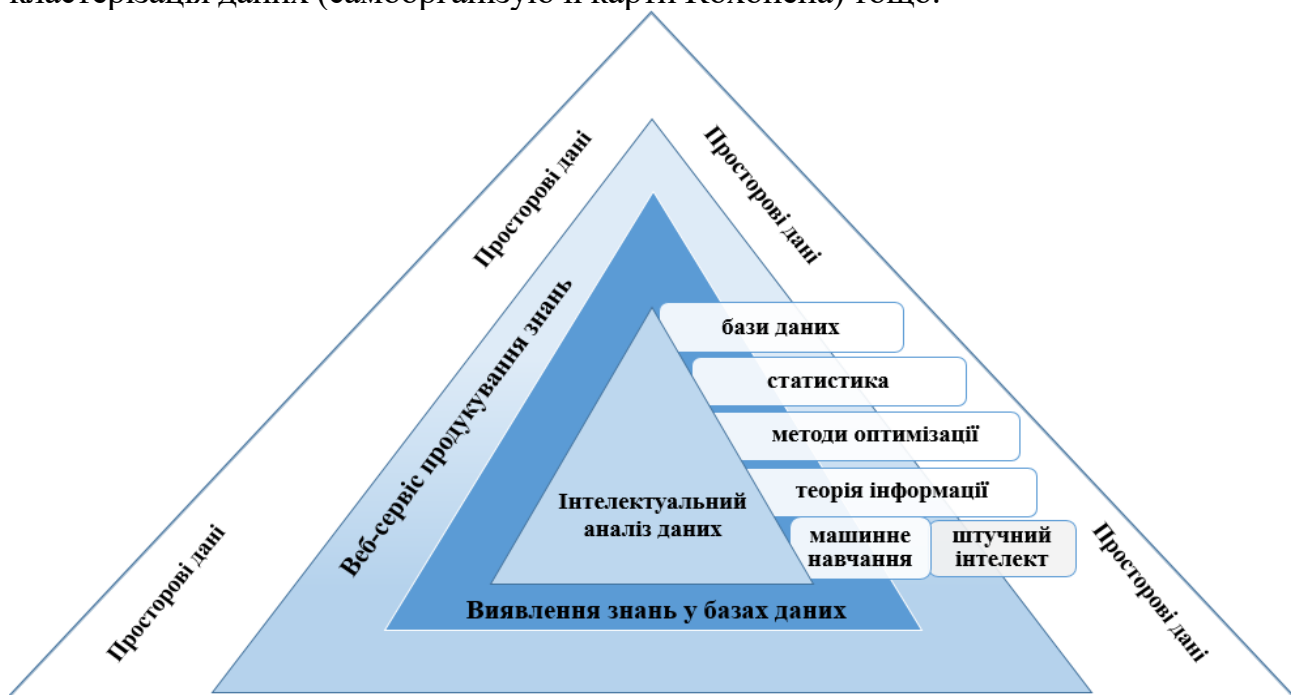


Рис. 1. Процес продукування знань



Сучасні дослідники пропонують підхід, що носить назву веб-сервісів продукування знань, який використовується в якості прошарку поверх інфраструктур просторових даних (в якості сервіс-орієнтованої структури) з метою надання користувачам просторових даних та особам, які приймають рішення, можливість видобувати знання з масивів різнорідних просторових даних [5]. Це надає можливість зосередитись на тому, що ми зазвичай хочемо від даних, замість того, щоб концентруватись на тому, як запускати алгоритми інтелектуального аналізу просторових даних.

Під поняттям технології Knowledge Mining розуміють виявлення інформації, що володіє такими ознаками знань як інтерпретованість, структурованість, зв'язність, метричність, семантичність та активність. Таким чином, технологію видобутку знань представляють у вигляді послідовного довідзначення інформації про проблемні ситуації до продукування знань шляхом розподіленого включення їх формальних ознак [4].

Видобування знань з великої кількості даних – складний ітеративний процес, що складається з таких кроків [3]:

- Очищення даних (data cleaning) – видалення завідомо неправильних даних (шумів);
- Інтеграція даних (data integration) – комбінація і об'єднання кількох витоків даних;
- Вибір даних (data selection) – відбір даних релевантних задачі аналізу;
- Трансформація даних (data transformation) – представлення даних у вигляді, прийнятному для видобування знань шляхом застосування методів аналізу даних;
- Видобування знань (data mining) – основний процес застосування інтелектуальних методів для видобування знань з даних;
- Оцінка паттернів (pattern evaluation) – процес вибору значущих і корисних для даного аналізу паттернів і шаблонів даних для видобування знань;
- Презентація знань (knowledge presentation) – процес представлення видобутих даних і знань користувачеві.

При інтегруванні даних потрібно враховувати декілька аспектів, зокрема, проблеми семантичної неоднорідності. Це проблеми, пов'язані з семантично еквівалентними поняттями або семантично пов'язаними/непов'язаними поняттями. Неоднорідність може поділятися на чотири категорії: структура, синтаксис, система та семантика. Неоднорідність структури містить різні моделі даних; неоднорідність синтаксису – різні мови та відображення даних і неоднорідність системи передбачає технічні засоби та операційні системи.

Семантична неоднорідність може класифікуватись як:

– семантично еквівалентні поняття (концепції) (моделі використовують різні терміни для підтвердження подібного поняття, наприклад, синоніми; властивості, змодельовані окремими системами по-різному тощо);

– семантично непов'язані поняття (один і той самий термін може використовуватись різними системами для визначення абсолютно різних понять);

– семантично пов'язані поняття (узагальнення/деталізація, різні класифікації тощо).

Існує декілька систем, які відповідають потребам інтеграції даних. Розробники кожної системи спробували різні підходи для пошуку якнайкращого шляху для надання потрібних сервісів. Деякі з найпопулярніших систем: Garlic, TSIMMIS, ObjectGlobe, SIMS тощо. Всі вони були створені для вирішення неоднорідностей будь-яких рівнів.

Крок видобування знань може містити фактор взаємодії з користувачем, або іншою базою знань. Цікаві чи корисні шаблони (далі паттерни) знань можуть бути представлені користувачеві для його оцінки і можуть бути збережені як нові дані. Важливо наголосити, що крок видобування знань є найважливішим у процесі видобування знань і розпізнавання образів, оскільки на цьому кроці відбувається викриття прихованих паттернів і шаблонів знань.

Користувачі при різних умовах можуть переглядати дані, що знаходяться у відкритому доступі. Однак, різноманітність даних, що містяться у мережі, криє в собі проблеми, які можуть виникнути не лише на етапі аналізу даних, а й на етапі їх пошуку. Основними такими проблемами є:

– Проблема пошуку необхідної інформації зв'язана з тим, що користувач не одразу може знайти необхідні йому електронні ресурси. Лише невеликий відсоток посилань, запропонованих пошуковими системами веде до необхідних ресурсів.

– Проблема виявлення нових знань. Навіть у випадку, коли знайдена деяка множина інформації, отримання корисних даних для користувача є складною трудомісткою задачею. Сюди ж можна віднести і труднощі, пов'язані з осмисленням даних.

Оскільки більшість інформації, що розміщена на веб-ресурсах, є текстовою інформацією, то для її обробки і аналізу варто використовувати такі методи прикладного видобування даних як інтелектуальний аналіз тексту (Text Mining або Knowledge Discovery in Text).

Технологія процедур видобування просторових даних здебільшого базується на засобах Web Mining, які використовують пошукові системи на основі прихованої індексації; оптимізацію контенту веб-ресурсів; структурування шаблонів побудови веб-ресурсів.

Технології інтелектуального аналізу даних представлено на рис. 2:

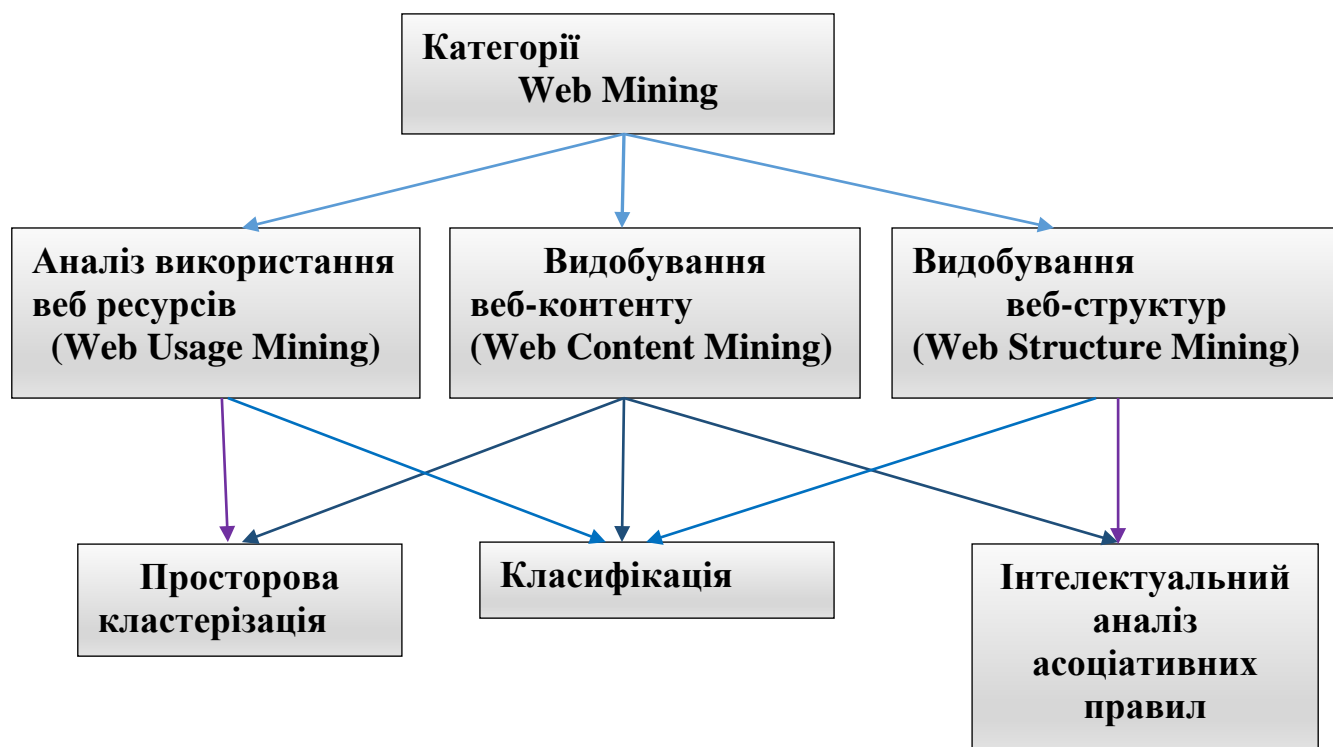


Рис. 2. Технології інтелектуального аналізу даних в процедурах Web Mining

Категорії Web Mining містять наступні складові:

- Аналіз використання веб ресурсів (Web Usage Mining);
- Видобування веб-структур (Web Structure Mining);
- Видобування веб-контенту (Web Content Mining).

Методика інтелектуального аналізу даних на основі веб-сервісів продукування знань застосовується для обробки даних, просторової кластеризації, класифікації та інтелектуального аналізу правил в сумісному середовищі.

Технологія процедур видобування тематичної інформації на засобах Web Mining дозволяє вирішувати різноманітні неформальні задачі обробки даних, прогнозування та кластеризації неструктурованих даних без попереднього формулювання гіпотез.

Отже, просторові веб-служби ми будемо визначати як набір програмних компонент, розроблених на основі стандартної сервіс-орієнтованої архітектури для підтримки міжмашинної взаємодії для управління просторовими даними в мережі.

Загальна ідея для видобування просторових знань полягає в використанні функціонально-сумісної просторової веб-служби поверх великої платформи просторових даних. Поява сучасних фреймворків великих даних, структури розподіленої та паралельної обробки дозволяють покращити обмеження

продуктивності традиційних систем, задовольнити вимоги до продуктивності обробки крупномасштабних просторових даних [5].

В цьому контексті платформа з відкритим висхідним кодом для надійних, масштабованих і розподілених обчислень Apache Hadoop [9] адаптувалась в потужну платформу для вирішення проблем з великими даними. Hadoop може ефективно зберігати і управляти великими об'ємами даних в кластерах комп'ютерів з допомогою простих моделей програмування. Основна концепція структури розділена на дві частини: розподілену файлову систему Hadoop (HDFS) для збереження даних і модель програмування MapReduce для обробки даних, які зазвичай зберігаються у HDFS [11].

Apache Spark [10], розподілені обчислення в пам'яті, – це ще одна структура, що забезпечує нову абстракцію даних, під назвою стійкі розподілені набори даних (RDD). RDD – це колекції об'єктів, розділених по кластеру машин. На сьогодні дослідниками було виявлено, що фреймворки Hadoop і Spark забезпечують високопродуктивні обчислення для видобування шаблонів і знань із великого об'єму просторових даних. Зокрема, використання методу прийому великих просторових даних з використанням паралельної техніки в кластерному середовищі; представлення структури з вбудованою підтримкою просторово-часових типів даних. Також для обробки крупномасштабних просторових даних ефективно використання середовищ Apache Sedona, які дозволяють користувачам читати різноманітні просторові об'єкти з різних форматів даних. GeoMesa забезпечує великі геопросторові запити та аналітику в розподілених обчислювальних системах. Та надає бази даних NoSQL для масового збереження точкових, лінійних і багатовимірних даних.

На сьогодні для кращого управління геопросторовими даними на основі просторових веб-служб розроблено набір стандартів. Ці стандарти можна розділити на наступні групи: моделі даних, кодування, пошук, збереження, обробка, відображення та публікація геопросторових даних. В контексті гібридних ГІС-сервісів, спостерігається зрушення парадигми ГІС «від ізольованої архітектури до взаємодіючої структури, від автономного рішення до розподіленого підходу, від індивідуальних приватних форматів даних до відкритого обміну специфікаціями даних, від настольної платформи до Інтернет-середовища».

Географічно розподілена інфраструктура (Data grid) – ще одна технологія, яка об'єднує множину різних типів, доступ до яких користувач може отримати з будь-якої точки, незалежно від місця їх розміщення. Grid надає колективний розподілений режим доступу до ресурсів і до зв'язаних з ними послуг в рамках глобально-розподілених організацій (підприємств, які спільно використовують глобальні ресурси, бази даних, спеціалізоване програмне забезпечення). Розробники національної Grid інфраструктури України зв'язують майбутнє Data Mining з їх використанням в якості Grid інтелектуальних додатків, вбудованих в



віртуальні чи корпоративні сховища даних, а також в мережу Світових Центрив Даних.

Розглянемо архітектуру програмного забезпечення на основі сервіс-орієнтованої архітектури для видобування корисних і нових знань з використанням методів інтелектуального аналізу даних для отримання шаблонів, які можна використовувати в процесі прийняття рішень.

Представлена архітектура включає п'ять рівнів: рівень даних, корпоративний компонент, послуги, бізнес-процес і рівень подання.

Рівень сервісу пропонується у вигляді веб-сервісу, що робить його легко доступним із будь-якого клієнтського застосунку.

Прогрес розвитку паралельних та розподілених систем ГІС на сьогодні дозволяє вирішити питання видобування корисних знань з різнорідних і розподілених просторових даних з допомогою сумісних і стандартних веб-сервісів. Неоднорідність потребує взаємодії і стандартів серед інструментів обробки даних.

Тому питання вирішення проблеми сумісності пропонується вирішити шляхом поєднання можливостей методів сервіс-орієнтованої архітектури та методів інтелектуального аналізу просторових даних в сумісному та паралельному обчислювальному механізмі з метою полегшення процесу продукування знань з інфраструктур просторових даних.

Архітектура містить чотири основних рівні для забезпечення бажаних функцій та можливостей:

(1) рівень інфраструктури просторових даних відповідає за інтеграцію розподілених і різнорідних даних із різноманітних просторових даних з використанням сучасних технологій збереження великих даних;

(2) рівень механізму знаходження, виявлення знань - підтримує високо продуктивні методи інтелектуального аналізу просторових даних в кластерах комп'ютерів, які називаються обробниками;

(3,4) безшовна та сумісна взаємодія між клієнтами та нижніми рівнями, що забезпечуються веб-службою продукування знань.

Компонента першого рівня є основною компонентою інтеграції просторових даних. Відповідає за завантаження різних джерел даних в кластер, який підтримує дуже великі файли. Добре підходить для застосунків, які слідує семантиці «один раз запис - багато читання» і вимагають, щоб вони були «прочитані» для задоволення вимог швидкості потокової передачі. Це відповідає цілям інтелектуального аналізу даних в інфраструктурі просторових даних.

Для управління даним процесом ефективно утиліти управління сховищем [5]. Вони надають широкий спектр форматів просторових даних, включаючи XML, JSON, CSV та інші традиційні векторні формати в HDFS. На сьогодні можна ефективно обробляти великі об'єми даних з допомогою

інструментів ETL в пам'яті. Керований метаданими просторовий ETL може використовуватись для інтеграції джерел даних інфраструктури просторових даних в HDFS з допомогою веб-сервісів сервісно-орієнтованої архітектури.

Компонента другого рівня - це основна компонента обробки для видобутку знань із платформи великих просторових даних. Щоб уникнути затримки в часі на цьому рівні використовуються рішення для паралельних та кластерних обчислень. Користувачі також можуть зчитувати різноманітні просторові об'єкти з різних форматів даних і виконувати паралельні завдання просторової обробки. Тут доцільне використання інноваційного середовища паралельних обчислень з відкритим вихідним кодом.

Серед популярних методів інтелектуального аналізу просторових даних варто приділити увагу алгоритмам просторової кластеризації, головна ідея якої полягає в розподілі геопросторових даних по категоріях.

Компонента механізму видобування знань підтримує декілька алгоритмів розподілу, ієрархії та кластеризації на основі щільності. Оскільки такі алгоритми мають справу лише з подібністю атрибутів, пропонується метод явної просторової кластеризації [5]. Даний метод дозволяє, крім групування геопросторових об'єктів, ідентифікувати гарячі й холодні точки, а також просторові викиди.

Найбільш важливими концепціями, які використовуються в веб-службі видобування знань є сервіс, інтерфейс і робота.

Висновки. Запропонована структура продукування знань із розподілених просторових даних базується на ідеї поєднання методів сервісно-орієнтованої архітектури та методів інтелектуального аналізу просторових даних для забезпечення процесів видобування знань.

Видобування знань із розподілених геопросторових даних на основі просторових веб-сервісів забезпечує основу для розподіленого інтелектуального аналізу просторових даних в середовищі інфраструктур просторових даних.

Просторові веб-сервіси визначаються як стиль, концепція або парадигма програмної архітектури, яка містить наступні принципи: слабкий зв'язок, можливість повторного використання, функціональна сумісність, масштабованість, гнучкість, незалежність від технологій за рахунок використання стандартизованих та модульних компонент, які називають сервісом.

Перспективою розвитку просторових веб-сервісів є дослідження різних аспектів управління просторовими даними та використання парадигми сервісно-орієнтованої архітектури в продукуванні знань на основі геопросторових даних.

Література:

1. Глуменко А.О., Стельмах Д.Е., Маттиас Колонко, Арсирий Е.А. Использование многовариантного хранения и мультимодельных СУБД для обработки разнообразных данных. *Сучасні інформаційні технології* : мат. X Міжнар. наук. конф. Одеса : Наука і техніка. 2020. С. 118–119.



2. Буслов П.В. Розробка класифікації систем консолідації комерційної інформації. *Вчені записки Таврійського національного університету ім. В.І. Вернадського, серія: Технічні науки*, 2018. Т.29 (68). №3. С. 105-108.
3. Катеринич Л.О., Петелько Ю.Ю. Видобування та аналіз інформації, що знаходиться у вільному доступі. *WORLD SCIENCE*. № 3(31). Vol.7. March 2018. С. 10-13. Режим доступу: <https://journals.indexcopernicus.com/api/file/viewByFileId/220091.pdf>
4. Ковальчук К. Ф., Никитенко О. К. Специфіка прогнозування фінансових ринків на основі технології Knowledge mining. *Економічний вісник*. 2013. №4. С. 139–146. Режим доступу: https://ev.nmu.org.ua/docs/2013/4/EV20134_139-146.pdf
5. Омидипур М., Тооманиан А., Нейсани Самани Н., Мансуриан А. Веб-сервис обнаружения знаний для инфраструктур пространственных данных. *Международный журнал геоинформации ISPRS*. 2021. 10 (1). Режим доступу: <https://doi.org/10.3390/ijgi10010012>
6. Паласиос Калифорния, Рейес-Суарес Дж. А., Беарзотти Л. А., Лейва В., Марчант К. Открытие знаний для удержания студентов высших учебных заведений на основе интеллектуального анализа данных: алгоритмы машинного обучения и тематическое исследование в Чили. *Энтропия*. 2021. 23 (4). Режим доступу: <https://doi.org/10.3390/e23040485>
7. Топорівський Б.П., Гагарін О.О. Технології Web mining для систем персоналізації контенту. *Сучасні аспекти розробки програмного забезпечення: зб. наук. праць міжнародної ІV наук.-практ. дистанційної конференції молодих вчених і фахівців з розробки програмного забезпечення*. Черкаси: видавець Чабаненко Ю.А.. 2017. Режим доступу: <http://sargo.kiev.ua/materials/2017/2017.pdf>
8. Філатов В.О., Костіна В.Л. Мультиагентний підхід до видобування знань в розподілених інформаційних системах. *Використання інноваційних технологій у попередженні злочинів : матеріали наук.-практ. семінару. МВС України, Харк. нац. ун-т внутр. справ. X. : ХНУВС, 2012. С. 172-174.*
9. Apache Hadoop. Режим доступу: (<https://hadoop.apache.org/>).
10. Apache Spark. Режим доступу: <https://spark.apache.org/>
11. Alarabi, L.; Mokbel, M.F.; Musleh, M. St-hadoop: A mapreduce framework for spatio-temporal data. *GeoInformatica* 2018, 22, 785–813.

References:

1. Glumenko, A.O. & Stelmah, D.E. & Mattias Kolonko & Arsiriy, E.A. (2020). Ispolzovanie mnogovariantnogo hraneniya i multimodelnyih SUBD dlya obrabotki raznoobraznyih dannyih.[Using multivariant storage and multi-model DBMS to process a variety of data]. *Suchasni informatsiyni tehnologiyi : mat. X MIzhnar. nauk. konf. – Modern information technologies: mat. X International. Science. conf.* Odesa : Nauka I tehnika [in Russian].
2. Buslov, P.V. (2018). Rozrobka klasifikatsiyi sistem konsolidatsiyi komertsyynoyi Informatsiyi. [Development of classification systems for the consolidation of commercial information]. *TavrIyskiy natsionalnogo universitetu im. V.I. Vernatskogo. – Scientific notes of Tavriya National University. V.I. Vernatsky, series: Technical Sciences*. Vol. 29 (68), 3, 105- 108 [in Ukrainian].
3. Katerinich, L.O. & Petelko Yu.Yu. Vidobuvannya ta analiz informatsiyi, scho znahoditsya u vilnomu dostupi. [Extraction and analysis of freely available information]. *WORLD SCIENCE*. (Vol.7), 3(31), 10-13. Retrieved from <https://journals.indexcopernicus.com/api/file/viewByFileId/220091.pdf>
4. Kovalchuk, K. F. & Nikitenko, O. K. (2013). Spetsifika prognozuvannya finansovih rinkiv na osnovi tehnologiyi Knowledge mining.[Specifics of forecasting financial markets based on Knowledge mining technology]. *Ekonomichniy visnik – Economic Bulletin*, 4, 139–146. Retrieved from https://ev.nmu.org.ua/docs/2013/4/EV20134_139-146.pdf [in Ukrainian].

5. Omidipur, M. & Toomanian, A. & Neysani Samani, N. & Mansurian, A. (2021). Veb-servis obnaruzheniya znaniy dlya infrastruktur prostranstvennykh daniy. [Knowledge discovery web service for spatial data infrastructures]. *Mezhdunarodnyy zhurnal geoinformatsii ISPRS – International Journal of Geoinformation ISPRS*, 10 (1). Retrieved from: <https://doi.org/10.3390/ijgi10010012> [in Russian].
6. Palasios Kaliforniya & Reyes-Suares, Dzh. A. & Bearzotti, L. A. & Leyva, V. & Marchant, K. (2021). Otkrytie znaniy dlya uderzhaniya studentov vysshikh uchebnykh zavedeniy na osnove intellektualnogo analiza daniy: algoritmy mashinnogo obucheniya i tematicheskoe issledovanie v Chili. [Discovering knowledge to retain university students based on data mining: machine learning algorithms and case studies in Chile]. *Entropiya – Entropy*, 23 (4) Retrieved from: URL: <https://doi.org/10.3390/e23040485> [in Russian].
7. Toporivskiy, B.P. & Gagarin, O.O. (2017). Tehnologiyi Web mining dlya sistem personalizatsiyi kontentu. [Web mining technologies for content personalization systems]. *Suchasni aspekti rozrobki programnogo zabezpechennya: zb. nauk.-prakt. Konf. molodih vchenih i fahilvtziv z rozrobki programnogo zabezpechennya – Modern aspects of software development: coll. Science. prot. scientific-practical. remote conference of young scientists and software developers*. Cherkasi: vidavets Chabanenko Yu.A. Retrieved from: <http://sarpo.kiev.ua/materials/2017/2017.pdf> [in Ukrainian].
8. Filatov, V.O. & Kostina, V.L. (2012). Multiagentniy pidhid do vidobuvannya znan v rozpodilenykh Informatsiynih sistemah. [Multi-agent approach to knowledge extraction in distributed information systems]. *Vikoristannya Innovatsiynih tehnologiy u poperedzhenni zlochiniv : materialy nauk.-prakt. Seminaru MVS UkraYini, Hark. nats. un-t vnutr. sprav – The use of innovative technologies in crime prevention: materials of scientific practice. Seminar. Ministry of Internal Affairs of Ukraine, Kharkiv. nat. University of Internal Affairs affairs*. H. : HNUVS, (pp 172-174) [in Ukrainian].
9. Apache Hadoop. Retrieved from: (<https://hadoop.apache.org/>).
10. Apache Spark. Retrieved from: <https://spark.apache.org/>
11. Alarabi, L. & Mokbel, M.F. & Musleh, M. (2018). St-hadoop: A mapreduce framework for spatio-temporal data. *Geoinformatica*, 22, 785–813.