

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Інформаційних радіотехнологій та технічного захисту інформації
(повна назва)

Кафедра радіотехнологій інформаційно-комунікаційних систем
(повна назва)

АТЕСТАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти другий (магістерський)

ГЮІК.ХХХХХХ.008 ПЗ

Розробка нейромережових методів обробки даних

(тема)

Виконав: студент 2 курсу, групи ІКТм-19-1

Гулий Олександр Володимирович

(прізвище, ініціали)

спеціальності 122 Комп'ютерні науки

(код і повна назва спеціальності)

Тип програми освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Освітня програма інформаційно-комунікаційні технології

інформаційно-комунікаційні технології

(повна назва освітньої програми)

Керівник д.т.н., професор Кузьомін О. Я.

(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри

(підпис)

Цопа О. І.

(прізвище, ініціали)

2020_p

Не містить відомостей заборонених для відкритого публікування.

Керівник

д.т.н., професор Кузьомін О. Я.

Студент

Гулий О.В.

Харківський національний університет радіоелектроніки

Факультет Інформаційних радіотехнологій та технічного захисту інформації

Кафедра радіотехнологій інформаційно-комунікаційних систем

Рівень вищої освіти другий (магістерський)

Спеціальність 122 Комп'ютерні науки
(код і повна назва)

Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма інформаційно-комунікаційні технології
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

« _____ » _____ 20 ____ р.

ЗАВДАННЯ

НА АТЕСТАЦІЙНУ РОБОТУ

студентові Гулому Олександровичу Володимировичу
(прізвище, ім'я, по батькові)

1. Тема роботи: Розробка неймережових методів обробки даних

затверджена наказом по університету від 2 листопада 2020 р. № 1506 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 18 грудня 2020 р.

3. Вихідні дані до роботи Науково-технічні публікації та інтернет джерела з тематики атестаційної роботи

4. Перелік питань, що потрібно опрацювати в роботі вступ, 2 Розглянути теоретичні дані щодо вже існуючих систем знань водно-болотних угідь, 3 Провести їх порівняння для виявлення позитивних сторін і недоліків, 4 На підставі вхідних даних розробити та реалізувати алгоритм збору основних даних про водно болотні угіддя, 5 Провести порівняння обробки цих даних за допомогою людського ресурсу і розробленого алгоритму, 6 Розробити програмне забезпечення для візуалізації та модифікації зібраних даних, 7 Провести оцінку геотехнологічних можливостей розробленої платформи та оцінку можливостей для майбутнього розвитку (swot-аналіз).

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п. 5 включається до завдання за рішенням випускової кафедри)
Слайди презентації

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата
Основна частина	Проф. Кузьомін О.Я.		

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Аналіз літератури та Інтернет-джерел	10.10.20	виконано
2	Постановка задачі	25.10.20	виконано
3	Дослідження сучасного стану вирішення	08.11.20	виконано
4	Дослідження моделей та методів побудови систем	20.11.20	виконано
5	Написання пояснювальної записки	04.12.20	виконано
6	Підготовка презентації	10.12.20	виконано
7	Перевірка на плагіат	14.12.20	виконано
8	Нормоконтроль	14.12.20	виконано
9	Захист	18.12.20	виконано

Дата видачі завдання 04 10 2020 р.

Студент _____
(підпис)

Керівник роботи _____
(підпис) _____ (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка до магістерської атестаційної роботи містить:
69 с., 5 розділи, 23 рис., 2 табл., 11 джерел.

CW, geonode, postgresql, python, R-tool, automatization.

В роботі виконана розробка нейронної мережі для обробки статей та платформи знань, яка повинна стати універсальним постачальником рішень для пов'язаних з водно-болотними угіддями даних та інформації для сприяння дослідженням, розробці політики та фінансуванню, розширення прав і можливостей організацій громадянського суспільства і практиків для впровадження заснованих на природі рішень в інтересах сталого розвитку.

ABSTRACT

The explanatory note to the master's certification work contains:
69 pp., 5 sections, 23 figs., 2 table, 11 sources.

CW, geonode, postgresql, python, R-tool, automatization.

In the work, the development of a neural network for article processing and a knowledge platform that should become a universal solution provider for wetland-related data and information to promote research, policy development and funding, empower civil society organizations and practitioners to implement nature-based solutions for Sustainable Development.

ВСТУП.....	10
ПОСТАНОВКА ЗАВДАННЯ.....	11
2 ОПИС GEONODE ТА ОЦІНКА ГЕОТЕХНОЛОГІЧНИХ МОЖЛИВОСТЕЙ GEONODE.....	12
2.1 Що таке Geonode.....	12
2.2 Зберігання геопросторових даних	13
2.3 Змішування даних, створення карт	13
2.4 Типи даних	14
3 ОГЛЯД РОЗРОБЛЕНОЇ ПЛАТФОРМИ НА БАЗІ GEONODE.....	16
3.1 Опис проекту.....	16
3.2 Модель відносин сутностей бази даних да специфіка їх валідації	18
3.3 Збір даних.....	19
3.4 Візуалізація даних-карти, інформаційні листи та статистика.....	21
3.5 Постійний процес подачі потоку даних та інформації для платформи.....	26
3.6 Напівавтоматичний збір даних.....	29
4 ОЦІНКА ГЕО-ТЕХНОЛОГІЧНИХ МОЖЛИВОСТЕЙ РОЗРОБЛЕНОЇ ПЛАТФОРМИ ТА ОЦІНКА МОЖЛИВОСТЕЙ ДЛЯ МАЙБУТНЬОГО РОЗВИТКУ (SWOT-АНАЛІЗ).....	30
4.1 Сильні сторони.....	32
4.2 Можливості.....	32
4.3 Загрози.....	33
4.4 Слабкі сторони.....	34
4.5 Переоцінка можливостей для майбутнього розвитку.....	35
4.6 Опис можливостей технологій GeoWeb для подальшого розвитку.....	36
4.7 Порівняння можливостей технологій GeoWeb для подальшого розвитку.....	38
5 МЕХАНІЗМ ВИЛУЧЕННЯ АТРИБУТІВ З ПОБУДОВАНОЇ ІНФОРМАЦІЇ ПРО ВОДНО-БОЛОТНІ УГІДДЯ.....	44
5.1 Процедура розробки.....	45
5.2 Скринінг та попередня обробка рецензованих статей.....	47
5.3 Очищення та розділення даних	47

5.4 Витяг атрибутів	49
5.5 Збіг ключових слів.....	51
5.6 Експорт інформації в PostgreSQL	56
5.7 Перевірка ефективності.....	58
ВИСНОВОК	60
ПЕРЕЛІК ПОСИЛАНЬ.....	61
ДОДАТОК А.....	62
ДОДАТОК Б.....	69

**ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ,
ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ**

CW – constructed wetlands.

MDG - Millennium Development Goals.

SDGs - Sustainable Development Goals.

SDG - Sustainable Development Goal.

NADB - National Archeological Database.

CSOs - Civil Society Organizations.

NbS - Nature-based Solutions.

ВСТУП

Інтерес до штучних нейронних мереж (НМ) швидко зростає за останні півтора десятка років. Це відродження інтересу викликано як теоретичними, так і прикладними досягненнями, оскільки відкрилися можливості використання нейрообчислень у сферах, які до цього часу відносились лише до області людського інтелекту. З'явилися і широко використовуються можливості створення машин, здатність яких навчатись і запам'ятовувати нагадує розумові процеси людини [11].

Сьогодні при розробці та використанні обчислювальної техніки та систем керування велика увага приділяється підвищенню швидкості обчислень через розпаралелення процесів обробки інформації за допомогою нейромережових технологій. Адже на протигагу звичайній фон-Неймановській архітектурі, яка є послідовною за своєю природою, штучні нейронні мережі виграють за рахунок обробки з масовим паралелізмом. Отже, задачі обробки, аналізу зображень та сигналів, ідентифікації та керування динамічними об'єктами, розпізнавання образів, керування фінансовою діяльністю та інші з великою продуктивністю розв'язуються системами, розробленими на базі нейромережової технології.

Тому було прийнято рішення використовувати нейронні мережі для обробки статей з тематики водно-болотних угідь для побудови бази знань про них. На даний момент існує більше десяти тисяч наукових праць про водно-болотні угіддя і обробка навіть однієї такої статті вимагає великих витрат людиногодина.

Робляться деякі зусилля по регіональному збору інформації в базах даних, наприклад в Північній Америці (NADB) і в ряді європейських країн (Великобританія, Ірландія, Франція, Німеччина, Данія, Італія і т.д.). Однак на сьогоднішній день, мабуть, немає глобального центру для збору та аналізу кількості стічних вод, очищених побудованими водно-болотними угіддями, а також того, як вони працюють.

1 ПОСТАНОВКА ЗАВДАННЯ

Вхідними даними для проекту є тринадцять досліджень про водно-болотні угіддя.

В даній атестаційній роботі потрібно розглянути наступні положення:

- Розглянути теоретичні дані щодо вже існуючих систем знань водно-болотних угідь;
- Провести їх порівняння для виявлення позитивних сторін і недоліків;
- На підставі вхідних даних розробити та реалізувати алгоритм збору основних даних про водно болотні угіддя;

Провести порівняння обробки цих даних за допомогою людського ресурсу і розробленого алгоритму;

- Розробити програмне забезпечення для візуалізації та модифікації зібраних даних;
- Провести оцінку геотехнологічних можливостей розробленої платформи та оцінку можливостей для майбутнього розвитку (swot-аналіз).

2 ОПИС GEONODE ТА ОЦІНКА ГЕОТЕХНОЛОГІЧНИХ МОЖЛИВОСТЕЙ GEONODE

2.1 Що таке Geonode

GeoNode - це система управління геопросторовим контентом, платформа для управління та публікації геопросторових даних. Він об'єднує стабільні проекти програмного забезпечення з відкритим кодом під послідовним та простим у використанні інтерфейсом, що дозволяє неспеціалістам обмінюватися даними та створювати інтерактивні карти.

Інструменти управління даними, вбудовані в GeoNode, дозволяють інтегровано створювати дані, метадані і візуалізацію карт. Кожен набір даних в системі може бути загальнодоступним або обмеженим, щоб дозволити доступ тільки певним користувачам. Соціальні функції, такі як профілі користувачів, системи коментування та рейтингування, дозволяють створювати спільноти навколо кожної платформи для полегшення використання, управління та контролю якості даних, що містяться в екземплярі GeoNode.

Він також розроблений як гнучка платформа, яку розробники програмного забезпечення можуть розширювати, змінювати або інтегрувати відповідно до вимог своїх власних додатків. На рис. 2.1 представлена спрощена архітектура GeoNode.

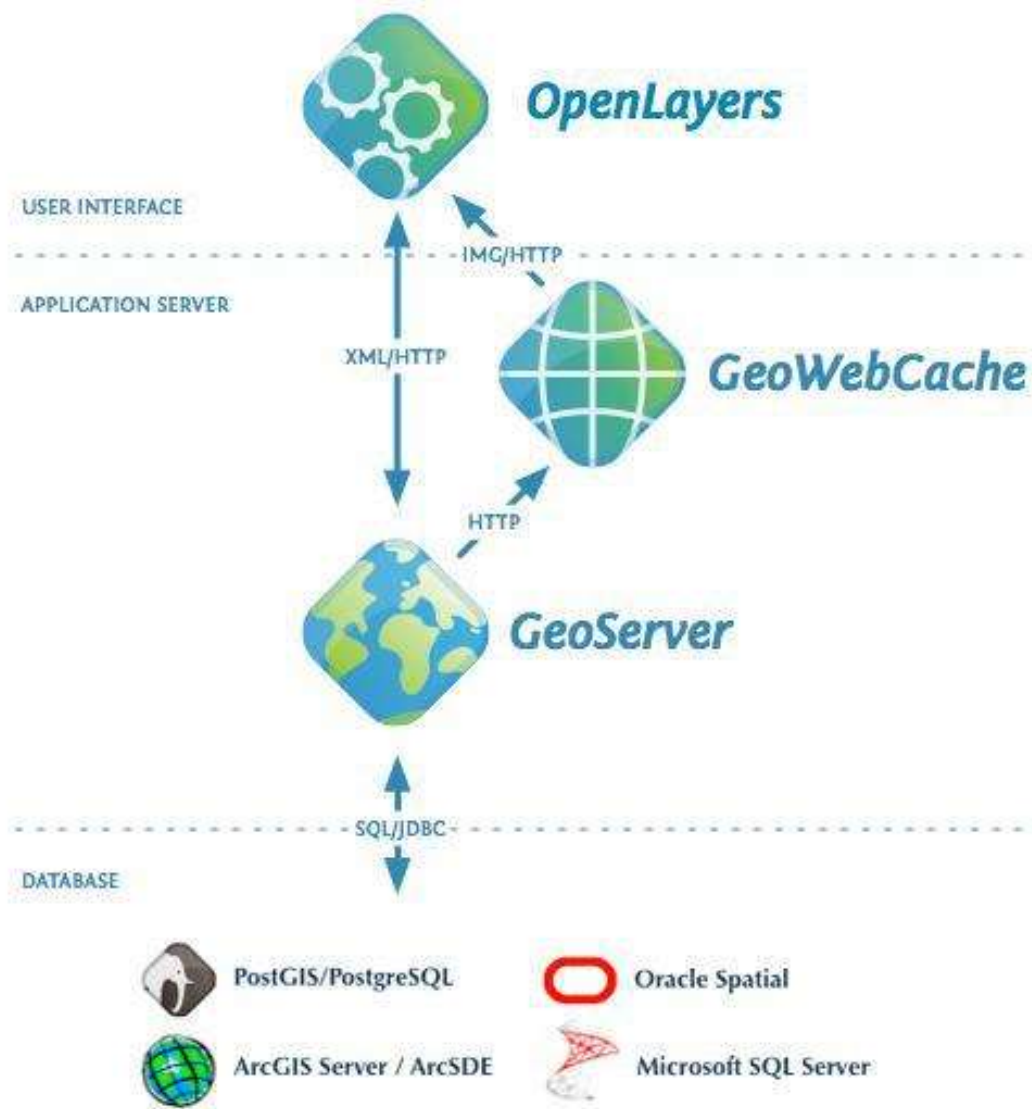


Рисунок 2.1- Спрощена архітектура GeoNode [1]

2.2 Зберігання геопросторових даних

GeoNode дозволяє користувачам завантажувати векторні дані (в даний час SHP-файли, json, csv, kml і kmz) і растрові дані в їх вихідних проекціях за допомогою веб-форми. [1]

Векторні дані перетворюються в геопросторові таблиці в БД, супутникові знімки та інші види растрових даних зберігаються у вигляді геотипів.

Особливе значення надається стандартним форматам метаданих, таким як стандарти метаданих ISO 19139:2007 / ISO 19115. [1]

Як тільки завантаження завершено, користувач може заповнити метадані ресурсу, щоб зробити його раптово доступним через кінцеві точки CSW (OGC Catalog Service) і API. [1]

Користувачі також можуть завантажити XML-документ метаданих (ISO, FGDC і Dublin Core format) для автоматичного заповнення ключових елементів метаданих GeoNode. [1]

Точно так само GeoNode надає веб-стайлер, який дозволяє користувачам змінювати зображення даних і переглядати зміни в режимі реального часу. [1]

2.3 Змішування даних, створення карт

Після того, як дані були завантажені, GeoNode дозволяє користувачеві шукати їх географічно або за ключовими словами, щоб створити химерні карти.

Всі шари автоматично переробляються в web Mercator для відображення карт, що дозволяє використовувати різні популярні базові шари, такі як Open Street Map, Google Satellite або Bing layers. [1]

Після того, як карти збережені, їх можна вбудувати в будь-яку веб-сторінку або отримати PDF-версію для друку.

2.4 Типи даних

Існує три основних типи ресурсів, якими може керувати GeoNode:

Документ - GeoNode дозволяє публікувати табличні і текстові дані, а також управляти метаданими і пов'язаними з ними документами. Документи можуть бути завантажені безпосередньо з диска. Наступні типи документів допускається: .doc, .docx, .gif, .jpg, .jpeg, .ods, .odt, .odp, .pdf, .png, .ppt, .pptx, .rar, .sld, .tif, .tiff, .txt, .xls, .xlsx, .xml, .zip, .gz, .qml. [1] Карти -

Карты є основним компонентом GeoNode.

Карты складаються з різних шарів і їх стилів. Шари можуть бути як локальними шарами в GeoNode, так і віддаленими шарами, що обслуговуються

або з інших серверів WMS, або за допомогою шарів веб-сервісів, таких як Google або MapQuest. [1]

Карти геонод також містять іншу інформацію, таку як масштаб і екстент карти, порядок шарів і стилі. [1]

Ви можете створити карту на основі завантажених шарів, об'єднати їх з деякими існуючими шарами і шаром віддаленого веб-сервісу, поділитися отриманою картою для загального перегляду. Після завантаження даних GeoNode дозволяє користувачеві шукати їх географічно або за ключовими словами і створювати карти. Всі шари автоматично проєктуються в web mercator для відображення карт, що дозволяє використовувати популярні базові карти, такі як OpenStreetMap. [1]

Шари (Layers) - Шари є основним компонентом Геоноду.

Шари - це публіковані ресурси, що представляють растровий або векторний джерело просторових даних. Шари також можуть бути пов'язані з метаданими, оцінками та коментарями. [1]

GeoNode дозволяє користувачеві завантажувати векторні та растрові дані в їх вихідних проєкціях за допомогою веб-форми. [1]

Векторні дані можуть бути завантажені в багатьох різних форматах. Супутникові знімки та інші види растрових даних можуть бути завантажені у вигляді Геотипів. [1]

3 ОГЛЯД РОЗРОБЛЕНОЇ ПЛАТФОРМИ НА БАЗІ GEONODE

3.1 Структура проекту

На рис. 3.1 представлена абстрактна структура розробленого проекту.

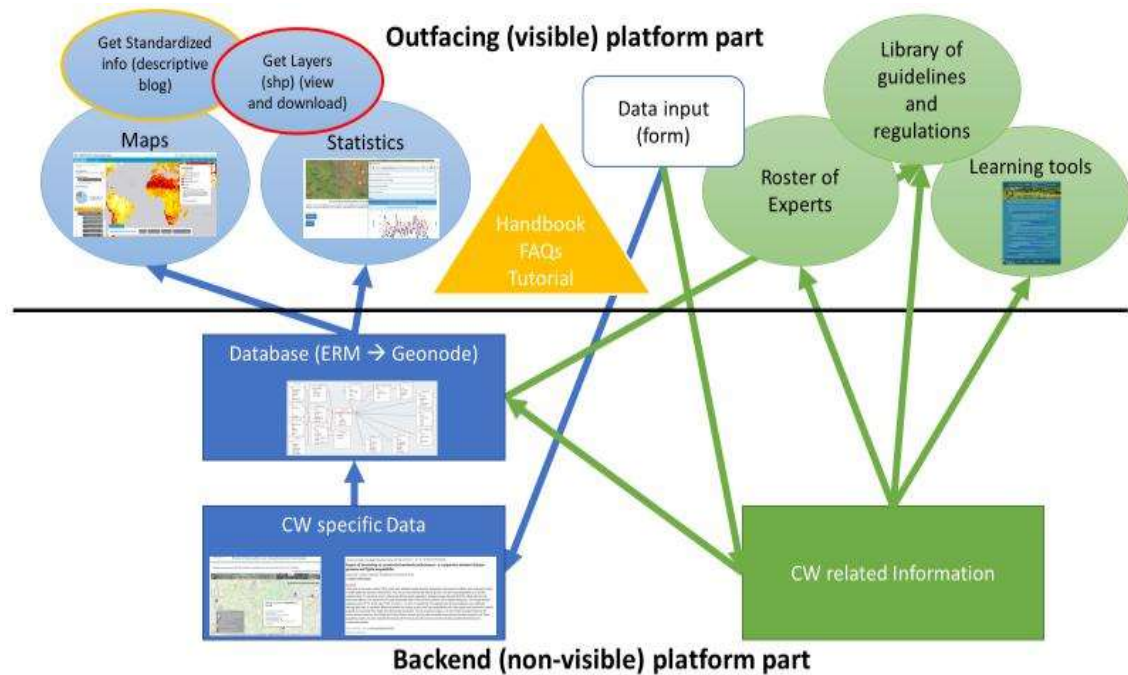


Рисунок 3.1 - Абстрактна структура розробленого проекту

В даний час обрана платформа заснована на Geonode. Об'єкт Geonode був зіставлений з елементами, описаними вище, щоб виділити, які об'єкти пов'язані з необхідними елементами. На рис. 3.2 представлена детальна структура розробленого проекту.

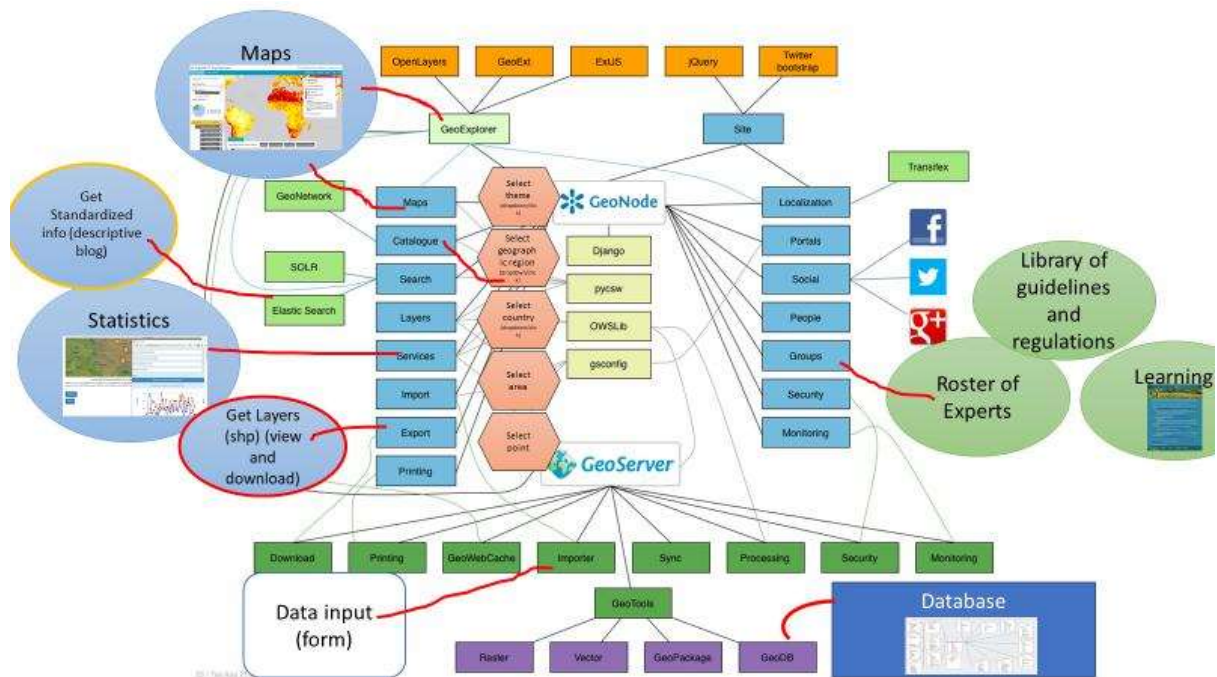


Рисунок 3.2 - Детальна структура розробленого проекту

Дані про побудовані водно-болотні угіддя збираються за допомогою різних процесів і включаються в базу даних Geonode. Інші дані, пов'язані з водно-болотними угіддями, частково зберігаються в базі даних (наприклад, керівні принципи) або пов'язані безпосередньо з веб-платформою. Побудовані дані з конкретних водно-болотних угідь надходять в основному з рецензованих журнальних статей або існуючих баз даних, але також можуть бути зібрані в конкретних (тематичних) накопичувачах даних, пов'язаних з дослідницькими пропозиціями або іншими ініціативами (наприклад, заклик до дії). Нарешті, дані можуть також надходити з інших джерел даних, таких як контрольована якістю сіра література або цивільні наукові проекти. Інформація про побудовані водно-болотні угіддя надходить в основному з інших джерел даних (наприклад, навчальні матеріали або керівні принципи). Але деяка інформація (наприклад, експертна інформація) може також надходити з журнальних статей або існуючих баз даних. На рис. 3.3 представлена схема як дані потрапляють до бази даних.

Data source and pathways of entry

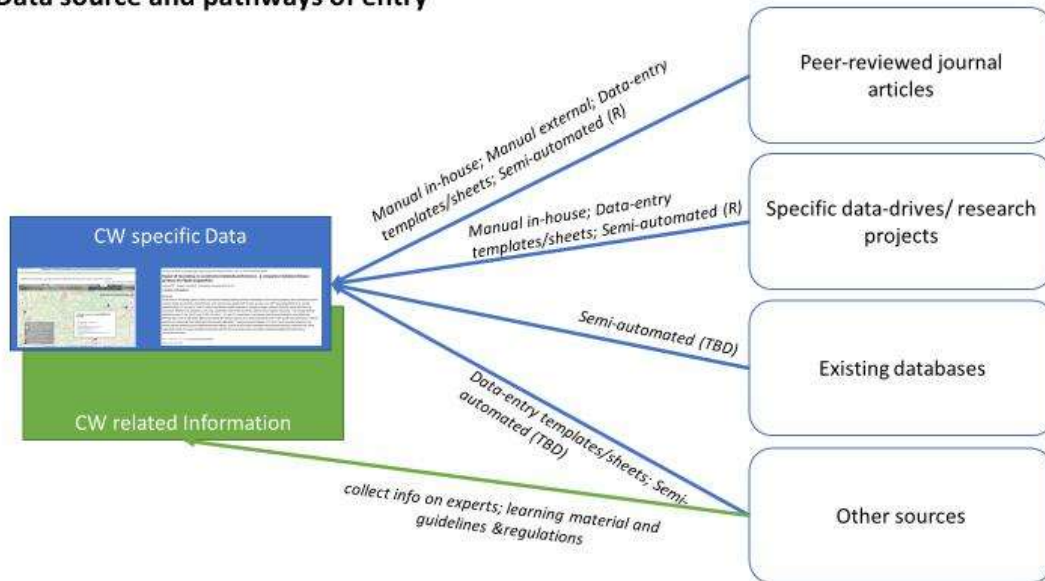


Рисунок 3.3 – схема як дані потрапляють до база даних

3.2 Модель відносин сутностей бази даних да специфіка їх валідації

Попередня база даних була підготовлена відповідно до описаних нижче кроків на рисунку 3.4, який показує етапи валідування.

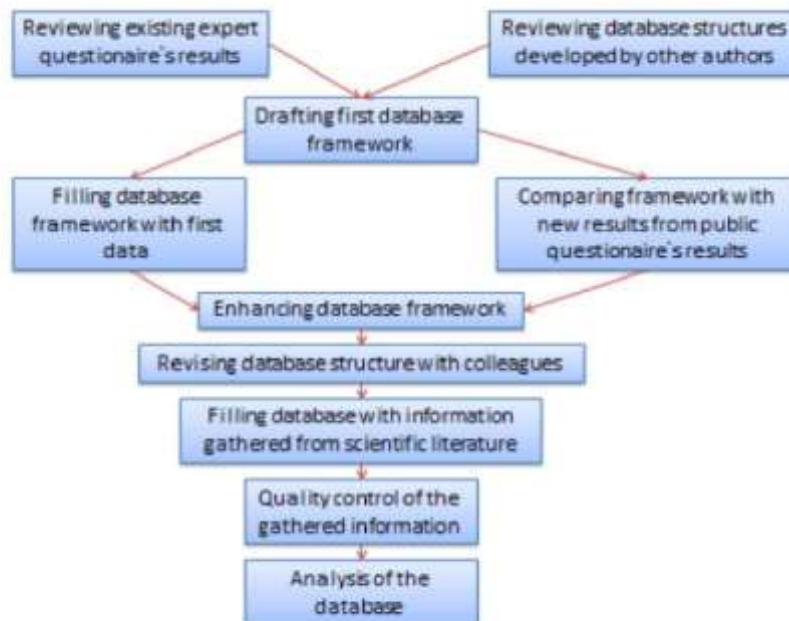


Рисунок 3.4 – Етапи валідування даних

Перше завдання в цьому розділі полягало в тому, щоб взяти до уваги інформацію, зібрану в ході першого літературного дослідження про побудовані водно-болотні угіддя, а також інформацію, взятую з експертного опитування, щоб створити структуру для побудованої бази даних водно-болотних угідь. Другим джерелом бази даних були вже існуючі бази даних про побудовані водно-болотні угіддя в Північній Америці. Перший проект, підготовлений таким чином, був потім зіставлений з результатами другого відкритого опитувальника, розісланого науковим співробітникам. Ідеї, отримані в результаті аналізу нової анкети та заповнення прототипу бази даних першими значеннями з наукової літератури, були використані для поліпшення структури, яка потім контролювалася за якістю шляхом подвійної перевірки. Система контролю якості була заповнена даними про побудовані водно-болотні угіддя з наукової літератури. Після цього якість бази даних контролювалася шляхом аналізу діапазону параметрів, що стосуються правдоподібності. За цим кроком пішов аналіз прототипу бази даних.

Оскільки обсяг наукової літератури, що стосується побудованих водно-болотних угідь, величезний, основна увага в літературних дослідженнях, проведених для заповнення цього першого прототипу, була приділена видаленню азоту з стічних вод через побудовані водно-болотні угіддя.

3.3 Збір даних

Сучасні методи отримання специфічних даних про водно-болотні угіддя обмежуються даними з рецензованих журнальних статей. В основному це пов'язано з тим, що за останні 20 років було опубліковано близько 10 000 статей,

і ці дані вважаються якісними. Були досліджені три основні шляхи:

- Ручний - внутрішній введення в Excel/Access, який може бути розширений до зовнішнього, якщо це необхідно або визнано необхідним;
- Розробка напівавтоматичної системи пошуку обмеженого підмножини елементів даних через R;
- Був розроблений шаблон для збору основної інформації, щоб полегшити збір даних від третіх сторін. Були зроблені обмежені спроби вивчити

- органічна речовина: карта органічної речовини, BOD і COD, оброблених системою водно-болотних угідь. Середня системна концентрація припливу і відтоку BOD і COD, в мг/л;

- поживні речовини: карта N, P, NO₃N, NH₄N, оброблених системою водно-болотних угідь. Середня системна концентрація припливу і відтоку N, P, NO₃N, NH₄N, в мг / л;

- біологічні індикатори: карта E-coli і фекальних колиформ, оброблених системою водно-болотних угідь. Ефективність видалення кишкової палички і фекальних колиформ в відсотках.

Повинно бути можливо створити вищезгадані 8 таблиць в БД для експорту сланцевих файлів і роботи з конкретними статистичними запитами, а також графіки для створення динамічних карт.

3.4 Візуалізація даних-карти, інформаційні листи та статистика

Конкретні дані по водно-болотних угіддях повинні візуалізуватися за допомогою двох основних механізмів: інтерактивного картографічного об'єкта та інструменту статистичних обчислень.

Карти повинні бути здатні відображати різні географічні області і теми. Географічні області включають глобальні, регіональні (як на континентах), національні, точкові і випадково обрані області. Теми є специфічними для теми дослідження і можуть відображатися як конкретні і незалежні шари. Те ж саме вірно і для довільно вибраної області. Кожен конкретний тип карти повинен відображати певну описову інформацію, а також статистичні елементи. На рис. 3.6 представлені режими візуалізації.

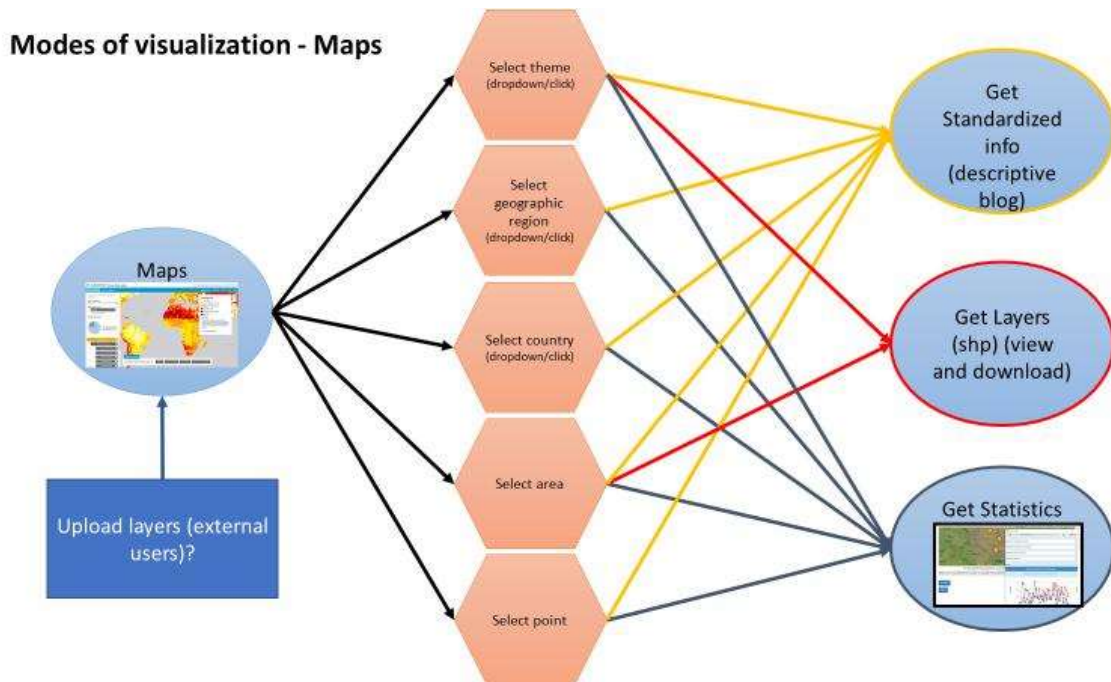


Рисунок 3.6 - Режими візуалізації – карти

Існує безліч механізмів, які можуть бути застосовані для візуалізації як бази геоданих, так і тематичної інформації. Такий дизайн візуалізацій залежить від ресурсу даних / інформації, що надаються партнерами.

Нижче наведені приклади поточних візуалізацій, які реалізовані або були реалізовані раніше.

Тематична і площадна інформація повинна відображатися в зручному для пошуку вигляді. Ці, а також інші характеристики повинні бути пов'язані зі стандартним описом інформації. На рис. 3.7 представлена візуалізація каталогу даних.

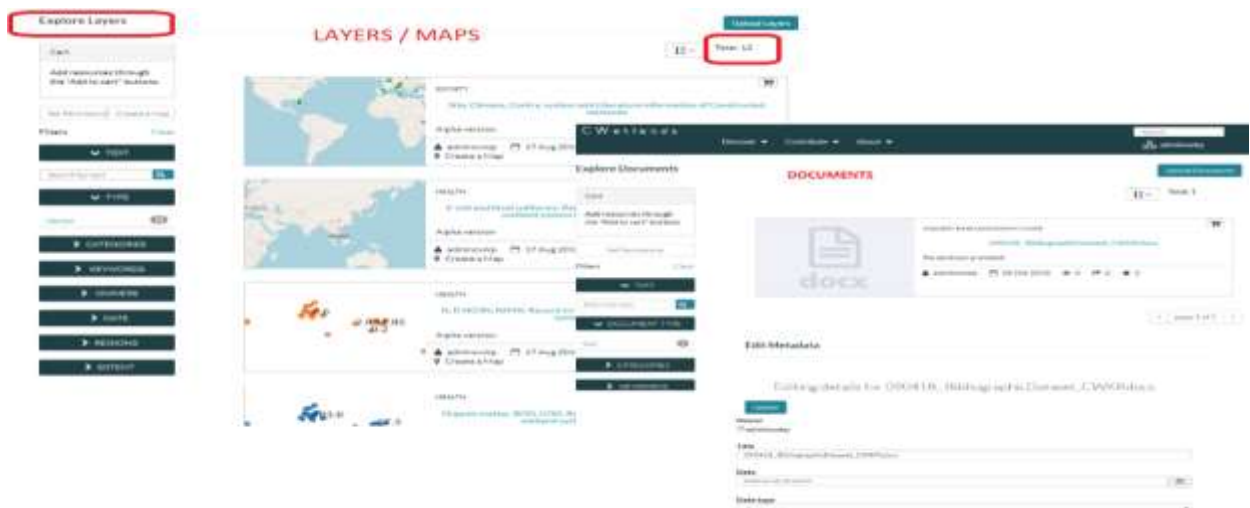


Рисунок 3.7 - Візуалізація каталогу даних

болотне угіддя. Популярним варіантом є використання сюжетних карт, де можна зібрати інформацію в декількох форматах. На рис. 3.9 представлена візуалізація тематичного контенту.

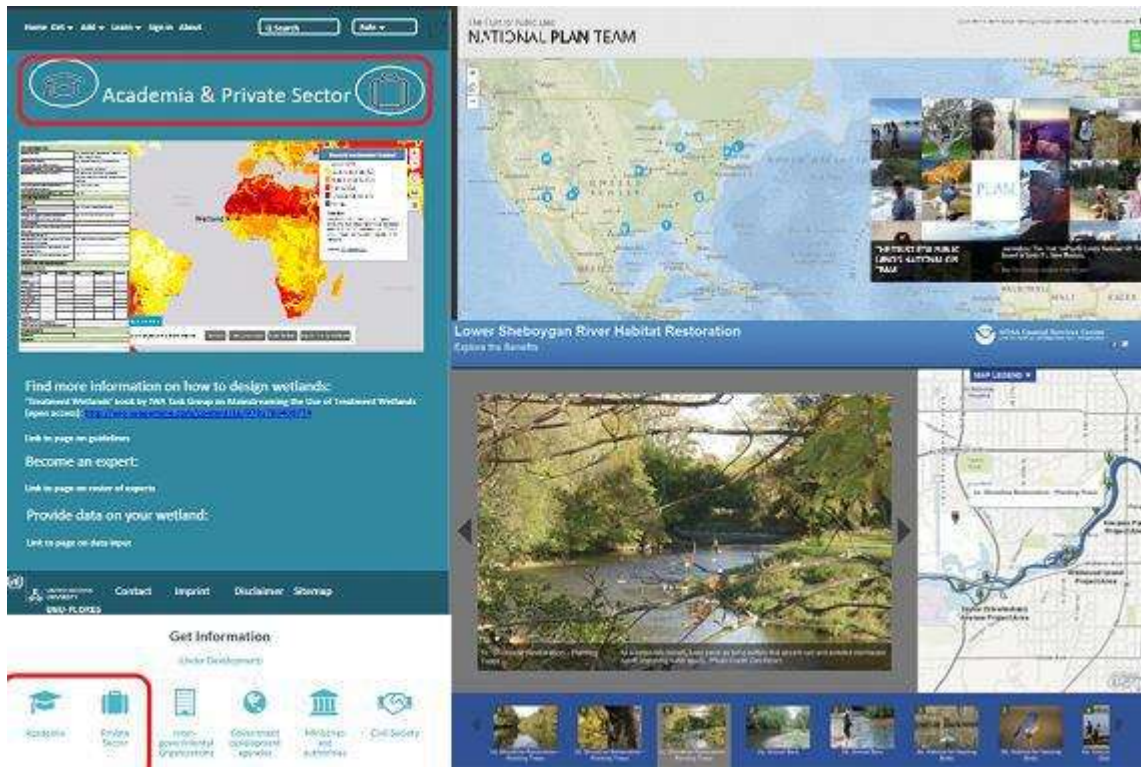


Рисунок 3.9 - Візуалізація тематичного контенту: сюжетні карти для водно-болотного угіддя.

На рис. 3.10 представлені всі можливі режими візуалізації.

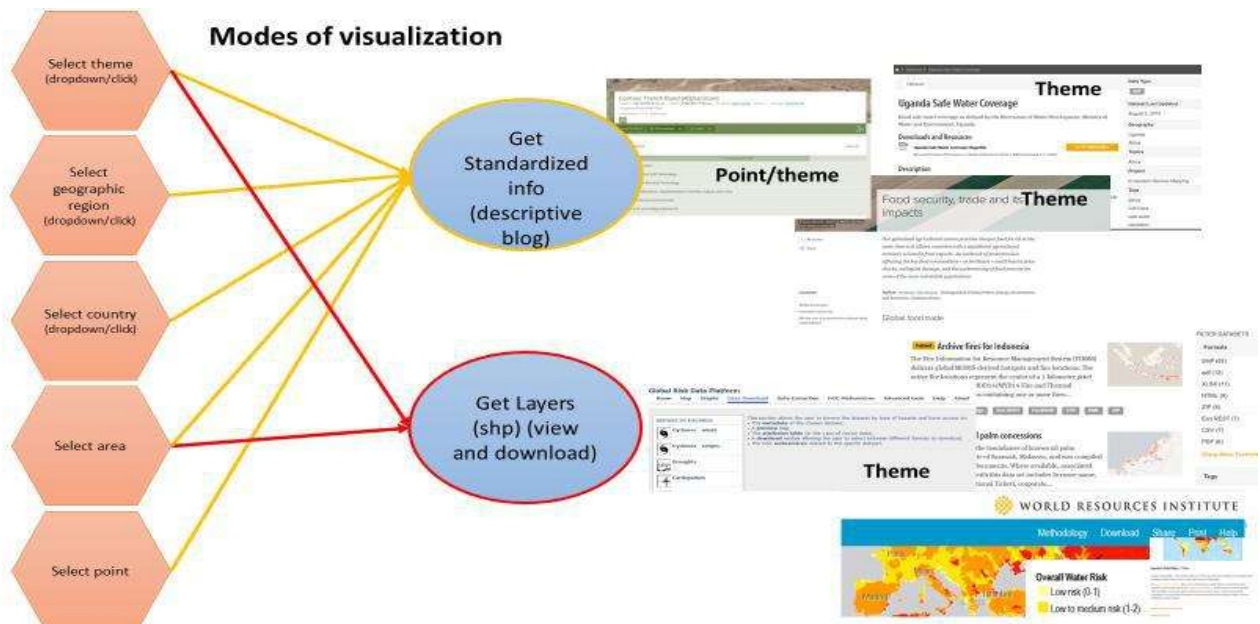
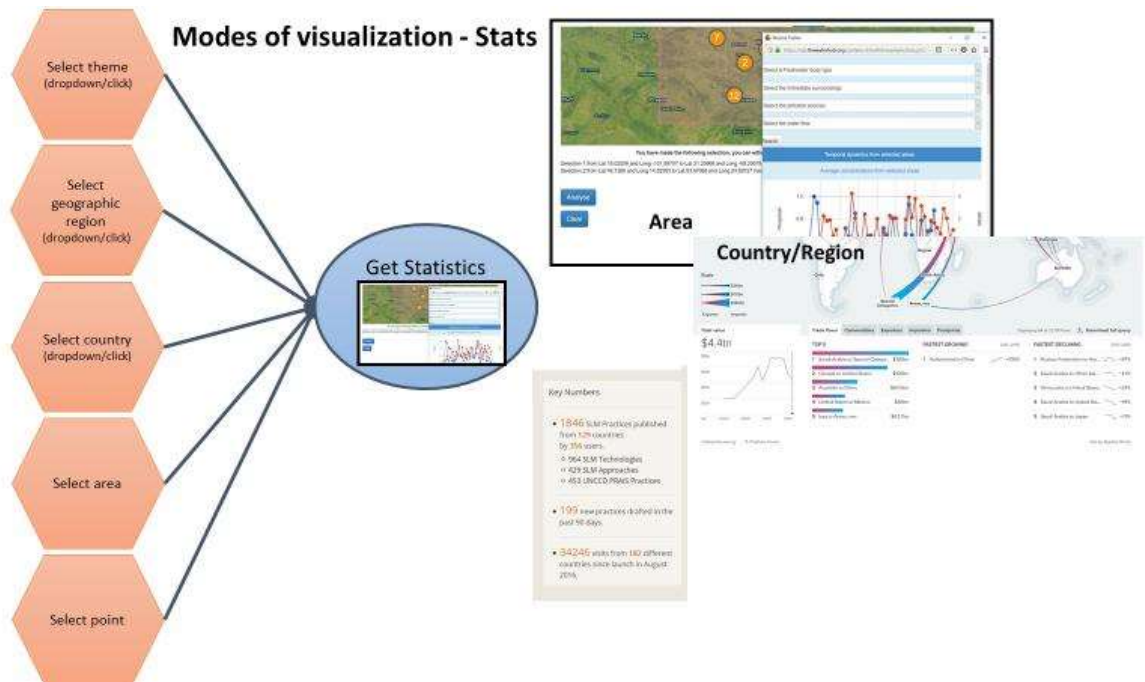


Рисунок 3.10 - Режими візуалізації – обидва варіанти

Статистична інформація повинна містити ключову інформацію і бути візуально привабливою. Ключові цифри деяких параметрів повинні бути представлені у вигляді інформаційного боксу. На рис. 3.11 представлені всі режими візуалізації у виді статистики.



3.11 - Режими візуалізації – статистика

3.5 Постійний процес подачі потоку даних та інформації для платформи

На рис. 3.12 представлений процес подачі даних для платформи.

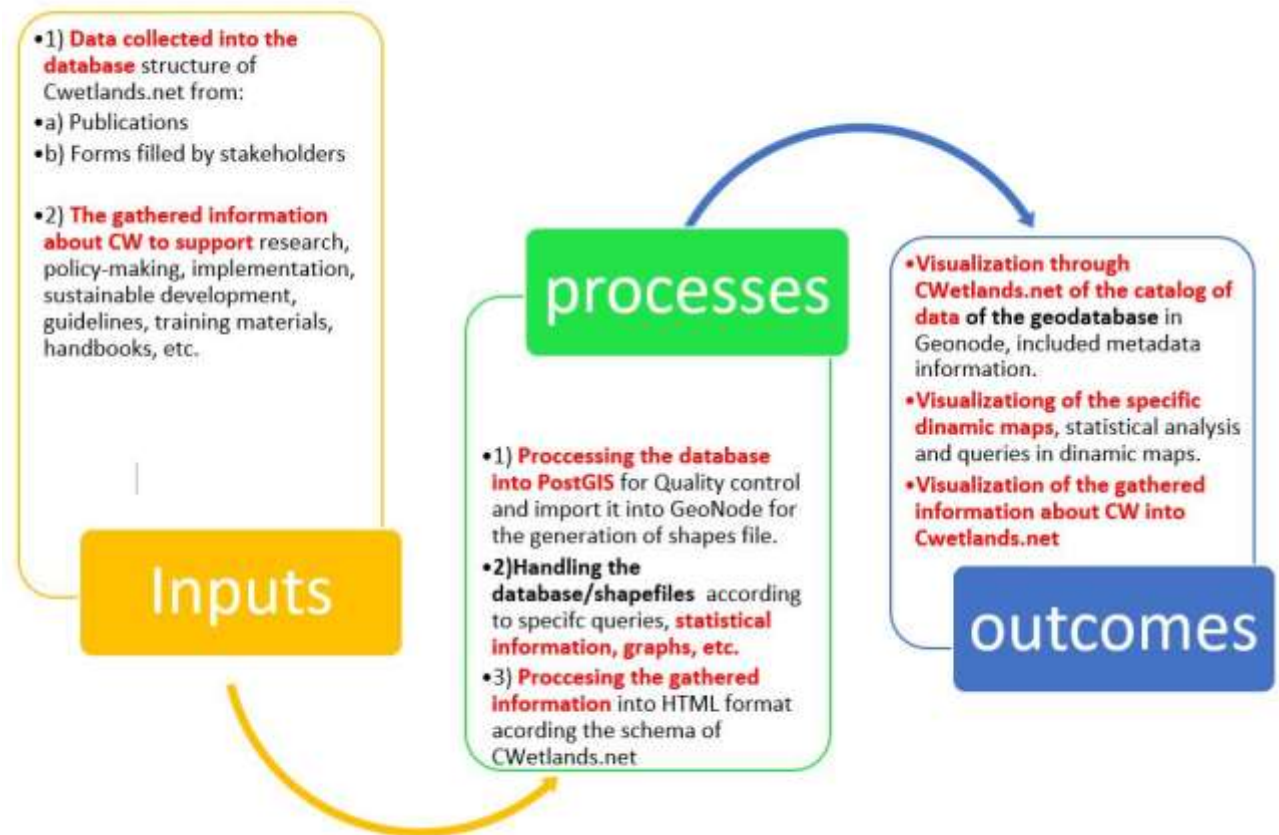


Рисунок 3.12 - Триступневий безперервний процес подачі даних для платформи

Етап обробки необхідний для того, щоб обробляти дані та інформацію структурованим чином, щоб привести їх у відповідність з масштабом мети платформи.

Структура тематичного контенту залежить від організації тематичного контенту і рішення зацікавлених сторін про те, як вони очікують використовувати і знаходити інформацію через платформу.

До цих пір існують деякі процеси і міркування, які повинні бути застосовані, пов'язані з БД, перш ніж інформація може бути завантажена в БД.

Нижче перераховані поточні процеси і міркування.

Такі мінімальні значення визначають, чи варто згадувати конкретне водно-болотне угіддя, надане: зацікавленою особою (з використанням заповненої форми) або журнальною публікацією.

Дуже важливо розуміти, що постійний потік інформації буде постійно оновлюватися платформою весь час, така здатність повинна бути відображена в платформі. У загальних рисах зацікавлені сторони завжди будуть постачати платформу даними та інформацією. Вони являють собою два основних змісту, що живлять базу даних і тематичний зміст платформи. Такі дані / інформація повинні бути належним чином оброблені, проаналізовані, синтезовані та підготовлені для навколишнього середовища на геовузлі, після того як візуалізація інформації може бути підготовлена для платформи за допомогою широкого спектру технологій GeoWeb.

Ідея полягає в тому, що зацікавлені сторони вносять свій внесок у платформу, щоб регулярно постачати її даними та інформацією за допомогою наступних варіантів, які повинні бути надані (дані, керівні принципи, довідники тощо.). На рис. 3.13 представлений зразок інтерфейсу.

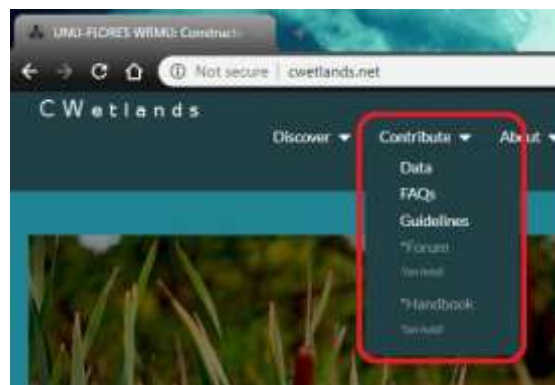


Рисунок 3.13 - Введення інформації через меню "внести свій вклад"

На рис. 3.14 представлений приклад форми для подачі даних.

SITE INFORMATION					
Name of the site	e.g. Dresden Kaditz Wastewater Treatment Plant or BOKU Vienna Campus				
Municipality/Country (Country after UNISO MMS standard)	e.g. Dresden/Germany or Vienna/Austria				
Latitude/longitude (in decimal degrees)	e.g. 51.070544, 13.6786233				
Wastewater type (based on NADREV2)	e.g. agricultural, municipal, stormwater, industrial, other and unknown (other should be specified)				
Long-term annual mean temperature (in °C, including reference period)	e.g. 8.9°C (1961-1990)				
Long term annual precipitation (in mm)	e.g. 567				
SYSTEM PROPERTIES					
System name					
Scale	e.g. full scale or experimental scale				
Pre-treatment					
Scheme incl. type of constructed wetland cells (HF, VF, FWS), after pre-treatment	e.g. HF, VF-HF or (VF/VF/F)-(VF/VF)				
Post-treatment					
System area (sum of the surface area of all CW cells, in m ²)					
Influent flow rate (m ³ /d)					
Plant species (Latin name, separated by comma if more than one species)	e.g. Typha latifolia or Cyperus papyrus				
Intensification (aeration, recirculation, use of reactive media etc.)					
Capital cost (in USD), with and without land acquisition	...				
OPERATION AND MAINTENANCE					
Performance indicators					
Parameter	Inflow	Outflow	Data	Method	Standard
Total solids (mg/L)					
BOD ₅ (mg/L)					
COD (mg/L)					
TN (mg/L)					
NH ₄ -N (mg/L)					
NO ₃ -N (mg/L)					
TP (mg/L)					
E. coli					
Faecal coliforms					
O&M Activities and Frequencies					
Annual cost for Operation & Maintenance					
SOURCES					

Рисунок 3.14 - Приклад можливої форми, що надається зацікавленим сторонам для подачі бази даних

На рис. 3.15 представлено як виглядає база даних всередині динамічної карти.

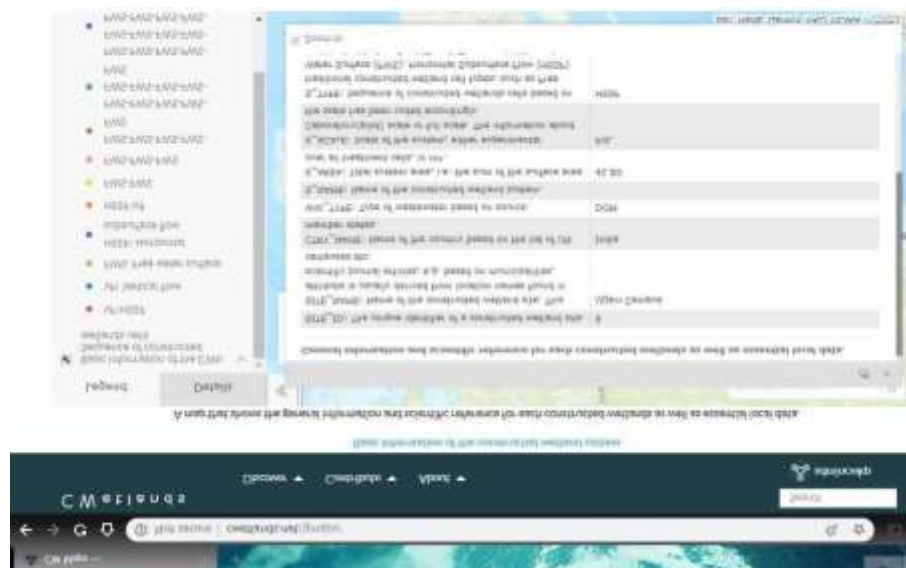


Рисунок 3.15 - Як виглядає база даних всередині динамічної карти.

3.6 Напівавтоматичний збір даних

Розробляється інструмент для вилучення даних з журналів за допомогою мови R.

Інструмент R буде вибирати і витягувати параметри бази даних з журналів. Після вилучення вони повинні бути включені в структуру бази даних (БД). База даних повинна бути розроблена в PostGIS відповідно до наступної діаграми відносин сутностей. Крім того, цей же процес повинен бути розглянутий, щоб об'єднати обидва введення інформації в БД, один через журнали, а інший через форми, заповнені зацікавленими сторонами. На рис. 3.16 представлена діаграма відносин сутностей платформи.



Рисунок 3.16 - Діаграма відносин сутностей для платформи

Geonode, як технологічна платформа, дозволяє користувачам завантажувати та оновлювати нову інформацію у вигляді SHP-файлів. Кожен SHP-файл являє собою таблицю БД в базі даних. SHP-файли, завантажені в Geonode, можна знайти в каталозі БД за допомогою інструменту пошуку.

4 ОЦІНКА ГЕО-ТЕХНОЛОГІЧНИХ МОЖЛИВОСТЕЙ РОЗРОБЛЕННОЇ ПЛАТФОРМИ ТА ОЦІНКА МОЖЛИВОСТЕЙ ДЛЯ МАЙБУТНЬОГО РОЗВИТКУ (SWOT-АНАЛІЗ)

Наступний розділ спрямований на те, щоб запропонувати деякі шляхи, які можуть бути прийняті для подальшого розвитку платформи. На таблиці 4.1 представлений огляд технологічних можливостей.

Таблиця 4.1 – Огляд технологічних можливостей

Частина	Пріоритет	
	Обов'язкові	Приємно-мати
Загалом	<ul style="list-style-type: none"> - Платформа надає функціональні можливості для візуалізації агрегованих даних, наприклад, по географічному регіону, країні або субнаціональному рівню; - Платформа повинна візуалізувати дані, наприклад, у вигляді часових рядів або карт; - Платформа повинна забезпечувати функціональні можливості для включення інформації метаданих; - Дані про продуктивність повинні бути фільтрованими, наприклад, за роками або за призначенням; - Платформа повинна бути здатна зберігати зображення, бази даних, текстові та ГІС-формати; - Платформа повинна дозволяти завантажувати, запитувати і завантажувати дані, генерувати статистичну інформацію і візуалізувати 	<ul style="list-style-type: none"> - Інформація повинна бути здатна інтегруватися в екологічні інформаційні системи органів влади; Платформа повинна забезпечувати: - функція завантаження вибраних записів у машиночитуваних форматах; - посилання на Google maps або супутниковий знімок побудованих водно-болотних угідь; - можливість пожертвувати гроші на подальший розвиток і обслуговування платформи; - Платформа повинна бути пов'язана з іншими технологіями децентралізованого очищення стічних вод;

Частина	Пріоритет	
	Обов'язкові	Приємно-мати
	<p>дані, а також завантажувати карти;</p> <ul style="list-style-type: none"> - Платформа повинна дозволяти фільтрувати інформацію за часом, місцем розташування та типом джерела інформації; - Платформа повинна враховувати авторські права; - Платформа повинна бути доступна безкоштовно. 	
База даних	<ul style="list-style-type: none"> - База даних платформи повинна мати можливість зберігати інформацію про: <ul style="list-style-type: none"> - о розташування побудованого водно-болотного ділянки, наприклад топографічна інформація, така як висота над рівнем моря; - клімат на об'єкті, наприклад, температура і опади, а також, якщо це можливо, випаровування, вологість і швидкість вітру. Представлені значення повинні бути, принаймні, річним або краще щомісячним дозволом; - форму і геометричних розмірів кожної клітини побудовані водно-болотних угідь; - розмір площі кожного збудованих водно-болотних угідь; - платформа повинна дозволяти користувачеві експортувати дані, наприклад, для цілей проектування; 	

Частина	Пріоритет	
	Обов'язкові	Приємно-мати
	- Платформа повинна забезпечувати середовище для зберігання літератури про побудовані водно-болотні угіддя, наприклад, для цілей проектування.	
WebGIS	- Платформа повинна забезпечувати засоби представлення існуючих побудованих водно-болотних угідь в області; - Платформа повинна дозволяти експортувати карти.	

Вимоги до побудованої платформи знань про водно-болотні угіддя, отримані на основі аналізу бажаних функцій інших платформ.

4.1 Сильні сторони

Це 28 з 36 пунктів, які Geonode має в якості сильних сторін. Таким чином, можна стверджувати, що Geonode забезпечує більшість вимог користувача і бажаних функцій платформи. Слід мати на увазі, що навіть якщо Geonode має функціональні можливості, є ще багато роботи, як ці функції будуть сприйматися зацікавленими сторонами. Використання процесу зворотного зв'язку із зацікавленими сторонами може дати уявлення про те, що можна поліпшити і що необхідно для подальшого розвитку.

4.2 Можливості

Це 12 з 36 пунктів, які Geonode не надає прямо зараз, але їх можна розглядати як можливості. Можливості-це поточні функції, яких у Geonode немає, але їх можна розвивати. Такого роду функціональні можливості в

основному пов'язані з наданням специфічних геовізуалізацій, таких як графіки, статистичні уявлення, поліпшення обробки даних в динамічних картах і т. д.

Виходячи з цих результатів, припускається, що переоцінка геотехнологічних можливостей повинна починатися з цих 12 пунктів.

4.3 Загрози

Навіть якщо видно, що не було виявлено загроз з вищенаведеного списку, насправді існують загрози, які є ключовими моментами для розгляду при подальшому розвитку платформи. Наступні моменти пов'язані з процесом управління платформою. Той факт, що під час і після розробки платформи "звичайні користувачі" (а не просунуті фахівці в області ГІС або розробки веб-сайтів) повинні мати можливість управляти постійним потоком контенту платформи.

В цілому, виділеною метою платформи, яка робить її і видатним рішенням між іншими аналогічними платформами, повинна бути її конструкція для безперервного потоку подачі інформації і даних через активну співпрацю зацікавлених сторін. Цей пункт вимагав роботи і вдосконалення, оскільки з урахуванням нинішнього досвіду Geonode виникне ще більше питань для досягнення такої мети. Geonode сам по собі може бути поліпшений, але через можливу складність проектування продукту, який успішно націлений на цю мету, потрібен досвідчений експерт з розробки платформ.

4.4 Слабкі сторони

Geonode був задуманий як платформа з відкритим вихідним кодом, яка дозволяє подальший розвиток з одним єдиним обмеженням, пов'язаним з реальною технологічною здійсненністю, заснованої на практиці, як такої, всі вимоги і функціональні можливості можуть бути розроблені поверх цієї технології. В даний час кілька пунктів згадуються як обмеження для будь-якої

розробки Платформи (якщо метою є збір даних / інформації), вони можуть існувати і інші, з якими я не знайомий:

- Економічний, людський і часовий ресурс, який повинен бути інвестований для створення придатної для використання технології GeoWeb, в цілому, для створення мережі партнерства для генерування безперервного потоку даних та інформації.

- Будь-яка платформа залежить від даних, її структури даних та зацікавлених сторін, які її годують. Якими б хорошими не були дані, такими ж хорошими будуть і результати. Крім того, цей момент впливає на можливості розвитку візуалізацій і складного розвитку.

4.5 Переоцінка можливостей для майбутнього розвитку

12 можливостей, виявлених в ході SWOT-аналізу, згадуються в якості орієнтира для повторних оцінок платформи.

- візуалізуйте агреговані дані, наприклад, по географічному регіону, країні або субнаціональному рівню;

- візуалізуйте дані, наприклад, у вигляді часових рядів або карт;

- візуалізуйте статистичну інформацію та дані;

- здатні бути інтегрованими в екологічні інформаційні системи органів влади;

- можливість пожертвувати гроші на подальший розвиток і обслуговування платформи;

- платформа повинна забезпечувати засоби представлення існуючих побудованих водно-болотних угідь в області, що цікавить, наприклад в країні або річковому басейні;

- складіть карту, яку легко зрозуміти, щоб її можна було використовувати інтуїтивно;

- включіть різні види візуалізації даних, наприклад діаграми, графіки, інфографіку тощо;

- забезпечте функцію фільтра (з не надто великою кількістю кнопок; вкл. визначень; наприклад, регіон, країна, годинниковою стрілкою тип і т.д.) і легенда на інтерактивній карті;

- забезпечте функціональні можливості для візуалізації статистичної інформації під картою;

- виділіть різні результати, засновані на інформації та візуалізації інструменту, наприклад, за допомогою історій;

- надати документи про те, як була створена система, наприклад керівництво/довідник.


У мережі існує безліч технологічних рішень, які можна використовувати для оновлення поточного статусу Geonode. Я згадаю пару з них, ті, з якими я знайомий, і які я тестував раніше, і з цієї причини я впевнений, що вони виконують деякі з 12 пунктів зверху, також це популярні варіанти, і їх можна комбінувати з Geonode. Також я переважно згадаю варіанти, пов'язані з гео-візуалізацією.


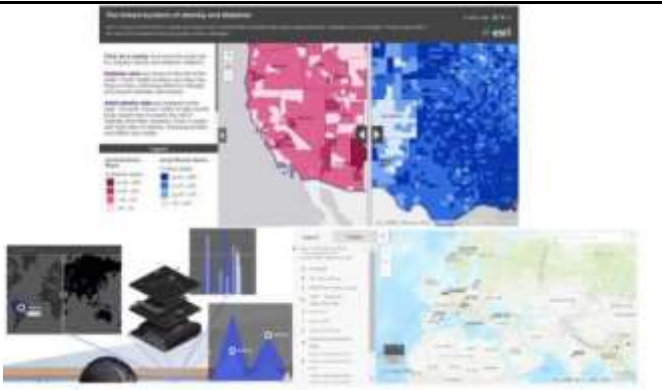
Згадані варіанти можна розділити на дві частини. Безкоштовні GeoWeb-інструменти з відкритим вихідним кодом і комерційні GeoWeb-інструменти. У конкретному контексті платформи, де це повинно бути передбачено 12 пунктами вище, ці дві категорії вимагають спеціальних знань про те, як розробляти платформи GeoWeb. Далі наводиться короткий опис і порівняння деяких інструментів GeoWeb.

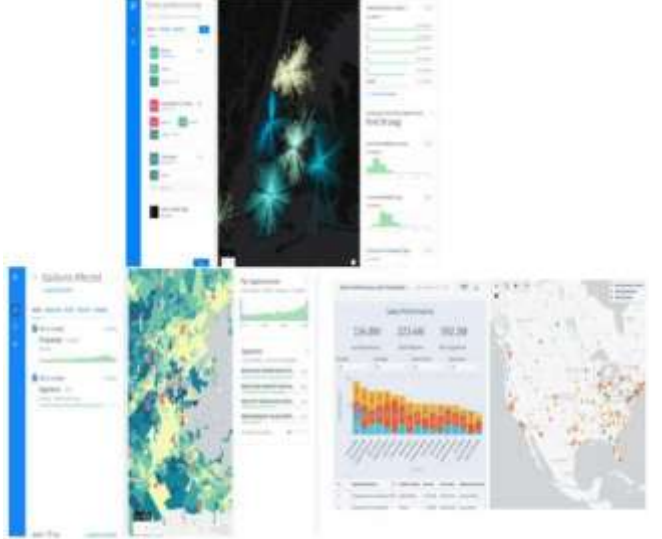
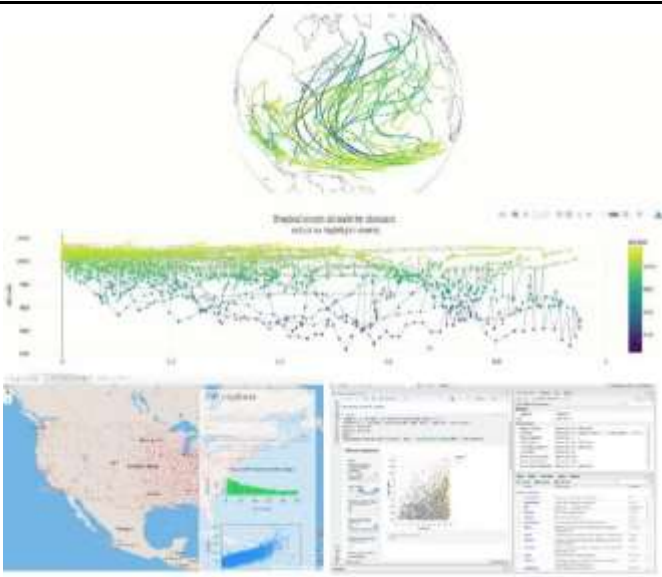
4.6 Опис можливостей технологій GeoWeb для подальшого розвитку

На таблиці 4.2 представлений опис можливостей технологій

GeoWeb. Таблиця 4.2 - Опис можливостей технологій GeoWeb

Гео-технологія	Опис	Захоплення
Leaflet (open source)	<p>Leaflet - це провідна бібліотека JavaScript з відкритим кодом для мобільних інтерактивних карт. Важить всього близько 38 КБ JS, він володіє всіма функціями відображення, які коли-небудь потрібні більшості розробників.</p> <p>Leaflet розроблена з урахуванням простоти, продуктивності і зручності використання. Він ефективно працює на всіх основних настільних і мобільних платформах, може бути розширений за допомогою безлічі плагінів, має гарний, простий у використанні і добре документований API і простий, читається вихідний код, в який приємно вносити свій внесок.</p>	

<p>D3 (open source)</p>	<p>D3 (Data-Driven Documents or D3.js) - це бібліотека JavaScript для візуалізації даних з використанням веб-стандартів. D3 допомагає вам оживити дані за допомогою SVG, Canvas і HTML.</p> <p>D3 поєднує в собі потужні методи візуалізації та взаємодії з керованим даними підходом до маніпулювання DOM, надаючи вам всі можливості сучасних браузерів і свободу проектування правильного візуального інтерфейсу для даних.</p>	
<p>ArcOnline (commercial)</p>	<p>ArcOnline: з'єднуйте людей, місця розташування та дані за допомогою інтерактивних карт, створених за допомогою ArcGIS Online.</p>	

<p>CARTO (commercial)</p>	<p>CARTO (CartoDB)-це платформа хмарних обчислень Software as a Service (SaaS), що надає інструменти ГІС і веб-картографії для відображення у веб-браузері. Компанія позиціонується як платформа геолокаційної розвідки завдяки інструментам зі здатністю до аналізу і візуалізації даних, які не вимагають попереднього досвіду роботи з ГІС або розробки.</p>	
<p>RStudio Shiny (commercial)</p>	<p>Shiny-це пакет R, який дозволяє легко створювати інтерактивні веб-додатки прямо з R. ви можете розміщувати автономні програми на веб-сторінці або будувати їх у документи R Markdown або створювати панелі моніторингу. Ви також можете розширити свої блискучі програми за допомогою CSS-тем, htmlwidgets і дій JavaScript.</p>	

4.7 Порівняння можливостей технологій GeoWeb для подальшого розвитку

На таблиці 4.3 представлений опис можливостей технологій GeoWeb для подальшого розвитку.

Таблиця 4.3 - Опис можливостей технологій GeoWeb для подальшого розвитку.

Гео-технологія	Коментарі в контексті платформи	Споживання людських ресурсів з метою розвитку	Вартість	Кількість можливостей, виявлених в SWOT-аналізі, які можуть бути вирішені
Leaflet (open source) + D3 (open source)	<ul style="list-style-type: none"> - Безкоштовна технологія; - Поєднання цих технологій поверх платформи легко налаштовується, і вони можуть бути повністю сумісні з Geonode; - Ви фізично володієте всіма послугами, так само як ви володієте програмним і апаратним забезпеченням; 	<ul style="list-style-type: none"> - Для роботи з цими технологіями ІТ-розробникам потрібна висока кваліфікація, оскільки для створення очікуваних результатів задіяно кілька мов коду і технологій; - Існує значна потреба в часі, який може знадобитися розробникам. Отже, 	Вільний	1,2,3,5,6, 7,8,9,10,11

	- Цей варіант забезпечує довгострокове і остаточне рішення.	високий попит на розробників.		
ArcOnline (comertial)	<p>- Цей продукт має високі функціональні можливості для фахівців з ГІС;</p> <p>- Платний;</p> <p>Як тільки ви перестаете платити, ви втрачаєте доступ до послуг;</p> <p>- Дуже можливо, що ви платите занадто багато грошей за послуги;</p> <p>- Це "легко", щоб створити різноманітність карт динаміки, як історія карт і т.д. але деякі шаблони за замовчуванням вимагають роботи з кодом;</p> <p>- Неможливо володіти обладнанням, на якому будуть</p>	<p>- Попит на високу кваліфікацію і досвід невеликий, так як багато функціональні можливо можуть бути легко вкористані і реалізовані в платформі (поточні візуалізації);</p> <p>- Налаштування карт вимагає роботи з декількома мовами і кодом;</p> <p>- Можливість встановлення на сервері програми Arc не здійсненна, тому що це означає, що ви повинні вибрати, Geonode або Arc.</p>	2.300\$/рік	1,2,3,6,7, 8,9,10,11

	розміщуватися дані; - Взаємодія з Geopode не є повним, так як ця опція є зовнішнім сервісом.			
CARTO (comercial)	- Цей продукт орієнтований на візуалізацію і надає високорозвинених інструментів, як Arc-Online; Те ж, що і Arc-Online: - Платний; тільки перестанете платити, втрачаєте доступ до послуг. - 'Легко' створювати різноманітні карти динаміки (сюжетні карти можуть не включатися) і т.д. але деякі шаблони за замовчуванням ви-	- Попит на високу кваліфікацію і досвід невеликий, так як багато функціональні можливо можуть бути легко використані і реалізовані в платформі (поточні візуалізації); - Налаштування карт вимагає роботи з декількома мовами і кодом.	3.400\$/рік	1,2,3,7, 8,9,10,11

	<p>магають роботи з кодом;</p> <p>- Неможливо володіти обладнанням, на якому будуть розміщуватися дані;</p> <p>- Взаємодія з Geonode не є повним, так як ця опція є зовнішнім сервісом.</p>			
<p>RStudio Shiny (commercial)</p>	<p>- Цей продукт орієнтований на статистику, він не надає таких складних інструментів, як ArcOnline, хоча і надає розширені функції для статистичного аналізу;</p> <p>-Платний.</p> <p>Як тільки ви перестаєте платити, ви втрачаєте доступ до послуг;</p> <p>- "Легко" створювати різноманітні карти динаміки (сюжетні карти</p>	<p>- Це вимагає досвіду роботи в R. Додатки створюються за допомогою R studio;</p> <p>- Попит на високу кваліфікацію і досвід невеликий, так як багато функціональні можливості можуть бути легко використані і реалізовані в розробленій платфо-</p>	9,995\$/рік	1,2,3,7, 8,9,10,11

	<p>можуть не включатися) і т.д. але деякі шаблони за замовчуванням вимагають роботи з кодом;</p> <p>- Неможливо володіти обладнанням, на якому будуть розміщуватися дані;</p> <p>- Взаємодія з Geonode не є повним, так як ця опція є зовнішнім сервісом.</p>	<p>рмі. (поточні візуалізації);</p> <p>- Налаштування карт вимагає роботи з декількома мовами і кодом.</p>		
--	---	--	--	--

Загалом, причина, по якій буде краще розвивати 12 пунктів можливостей

з використанням відкритого вихідного коду (D3 & Leaflet) для оновлення платформи - це той факт, що вам не потрібно інвестувати в покупку технологічного продукту на відміну від комерційних продуктів. Крім того, партнери матимуть право присвоювати нову розробку кодування. Це ключовий момент, якщо партнери вибирають комерційний продукт GeoWeb, то як тільки партнери перестають платити, їх обслуговування припиняється, це може включати розробку кодування, яке повинно бути розроблено для налаштування послуг.

5 МЕХАНІЗМ ВИЛУЧЕННЯ АТРИБУТІВ З ПОБУДОВАНОЇ ІНФОРМАЦІЇ ПРО ВОДНО-БОЛОТНІ УГІДДЯ

За результатами роботи нейронної мережі була створена структура для платформи більш ніж на 100 атрибутах, включаючи критерії для критеріїв проектування, експлуатації, ефективності, клімату та інших географічних факторів. Це дослідження спрямоване на розробку інструменту механізму вилучення значень атрибутів в R для вилучення значущої інформації з рецензованих журнальних статей надійним і швидким способом.

Інструмент фокусується на витяганні вісімнадцяти різних атрибутів, зібраних в 4 класи, які описують основні характеристики систем водно-болотних угідь.

Процес містить 3 підпроцеси:

- здійснюється доступ до документів та їх попередня обробка;
- атрибути витягуються за допомогою двох методів інтелектуального аналізу даних: зіставлення ключових слів і веб-брухту;
- значення експортуються в базу даних.

Для розробки і тестування інструменту було використано 13 статей. Середній показник успіху цього інструменту склав 79% за 30 хвилин. Це менше в порівнянні з 480 хвилинами, необхідними при ручному підході. В наступних версіях очікується, що інструмент отримає більш високий відсоток успіху за всіма атрибутами.

5.1 Процедура розробки

Інструмент був розроблений в R studio версії 3.5.0 на 64-бітній операційній системі Windows 10. Необхідні програми установки-PostgreSQL 9.5.13 64-біт,

PhantomJS 2.1.1 64-біт і Java 64-біт. Відповідні пакети R були встановлені та завантажені відповідно до вимог R.

Код був перевірений шляхом зіставлення його результатів з даними раніше існуючої бази даних, яка була розроблена вручну і містила дані приблизно з 100 англomовних рецензованих журнальних статей. Для розробки кодексу було розроблено тринадцять статей маючи більше трьох атрибутів з класу технічних специфікацій і опубліковані в журналах з найбільшою кількістю публікацій про побудовані водно-болотні угіддя (наприклад, екологічна інженерія, водна наука і техніка, технологія біоресурсів), були обрані з вже існуючої база даних. Отримані результати коду порівнювалися з даними з бази даних для перевірки надійності коду.

Вхідні дані інструменту для водно-болотних угідь з використанням R формуються 4 підпапками, як показано на рисинку 5.1.

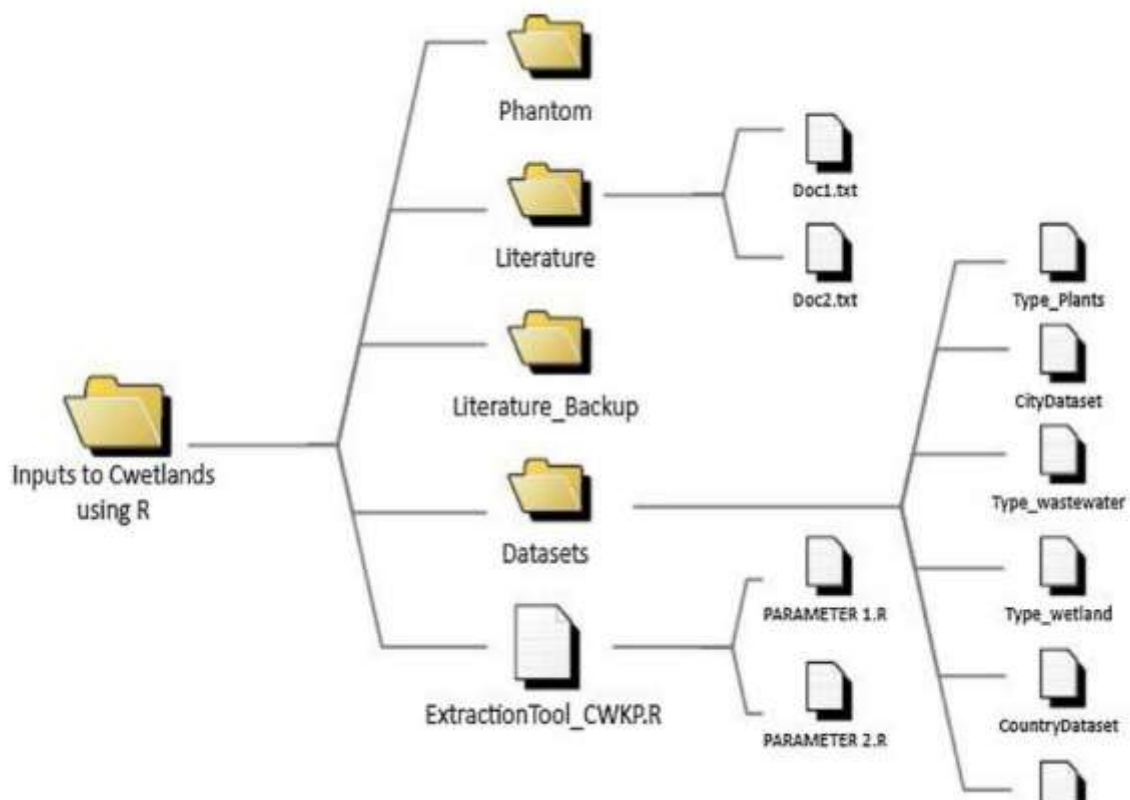


Рисунок 5.1 - Огляд структури папок для інструменту

Релевантність кожної підпапки в різних процесах представлена в графічному резюме, а то, як користувачі повинні редагувати / вводити

інформацію для отримання адекватних результатів від інструменту, пояснюється наступним чином.

Phantom-містить програмні файли PhantomJS. Інструмент використовує цю програму для доступу до компонентів Java HTML-сторінок. Файл Excel HTML_Links повинен бути заповнений посиланнями, з яких були завантажені документи, використовувані на екрані процесу і попередньо оброблені рецензовані статті. Інформація в цій папці є попередньою вимогою для виконання підпроцесу Web Scrap.

Literature -файли .txt, створені на екрані процесу і попередньо оброблені рецензовані статті, повинні бути збережені в цій папці. Інструмент зчитує файли з цієї папки для виконання подальших процесів.

Literature_backup - після останнього процесу експорту інформації в PostgreSQL інструмент копіює файли в папку Literature для створення резервної копії. Пізніше інструмент усуває файли в папці. Якщо користувач хоче запустити інструмент знову наступного разу, то папка Literature вже порожня, тому немає ніякого подвійного аналізу файлів з попереднього запуску інструменту.

Datasets - доступні набори даних, наприклад CountryDataset, CityDataset, Type_wastewater, зберігаються в цій папці. Інструмент зчитує файли з цієї папки в якості вимоги для виконання зіставлення ключових слів підпроцесу. Користувачі можуть змінювати файли наборів даних у форматі excel. Файли у форматі .txt (CountryDataset, CityDataset) повинен залишатися незмінним. ExtractionTool_CWKP.R - основний вихідний код у форматі R. На рис. 5.2 представлений список витягуваних атрибутів з рецензованих журнальних статей.

Extractable Attribute Description	Extractable Attribute name	Entity	Classes/Groups used by CWetlands-Extraction tool
Title of the article	TITLE	LITERATURE	metadata
Year of Publication	YEAR	LITERATURE	metadata
DOI of the article ^a	-	-	metadata
Journal name	JOURN_NAME	LITERATURE	metadata
Publisher	PUBLISHER	LITERATURE	metadata
Name of the author/s	AT_FIRST_N	AUTHORS	metadata
Last name of the author/s	AT_LAST_N	AUTHORS	metadata
Country name	CTRY_NAME	COUNTRY	location
Name of the municipality ^b	MUNI	SITES	location
Latitude	LAT	SITES	location
Longitude	LONG	SITES	location
Wastewater type	WW_TYPE	SITES	technical specifications
System area	S_AREA	SYSTEMS	technical specifications
Plant species	PLANT_SPEC	C_PLANTS	technical specifications
Plant genus	PLANT_GENUS	C_PLANTS	technical specifications
Type of Constructed Wetland	C_TYPE	CELLS	technical specifications
Average BOD ₅ inflow concentration	C_BOD_IN	C_ORG	treatment performance
Average BOD ₅ outflow concentration	C_BOD_OUT	C_ORG	treatment performance

Рисунок 5.2 - Список витягваних атрибутів з рецензованих журнальних статей.

Наведена нижче процедура відповідає крокам, описаним в графічній анотації.

5.2 Скринінг та попередня обробка рецензованих статей

Доступ до рецензованих журнальних статей здійснювався і завантажувався зі служби індексації цитування Web of Science. Теоретично цей інструмент може також працювати для інших онлайн-документів, які можна завантажити, таких як звіти і книги після декількох коригувань. Рецензовані журнальні статті були перетворені з файлів .pdf в .txt за допомогою MS Word і збережені у форматі doc[#].txt, наприклад doc1.txt. Для перетворення файлів в текстовий формат можна використовувати і інші платформи.

5.3 Очищення та розділення даних

Отримані файли в форматі .txt були додатково оброблені, щоб А) видалити спеціальні символи, які в іншому випадку заважають інтелектуальному аналізу тексту, і Б) розділити його на підрозділи, щоб забезпечити більш цілеспрямований пошук слів. Оригінальні рецензовані журнальні статті містять набір неструктурованого тексту, такого як таблиці, рівняння і малюнки, які при

перетворенні в файл .txt виглядають як комбінація спеціальних символів без лінгвістичного значення, наприклад *><\ -. Ці невпорядковані рядки, а також розділові знаки, наприклад ;?!, багатозначні прогалини (коли між послідовними словами більше одного пробілу) і стоп-слова (словосполучення типу "деякі", "кожен", "коли") ускладнюють обробку регулярних виразів, тому їх необхідно видалити.

Наступна функція коду піклується про видалення спеціальних символів:
`removeSpecialChars<-function(x) gsub("[^a-zA-z0-9.]","",x)`. Вираз в дужках зберігає всі цифри, букви і точки в тексті, а також усуває інші символи, наприклад спеціальні символи та інші розділові знаки. Точки зберігаються, тому що інформація в числовому десятковому форматі, наприклад 4.5, має в якості десяткового роздільника точку. Усунення множини пробіли виконуються за допомогою рядка коду: `mydocs<-tm_map(mydocs, stripWhitespace)`, яка гарантує, що між кожним словом є тільки один пробіл. Усунення стоп-слів здійснюється за допомогою рядка: `mydocs<-tm_map(mydocs,removeWords,stopwords('en'))`. На рис. 5.3 представлений приклад документа до і після очищення.

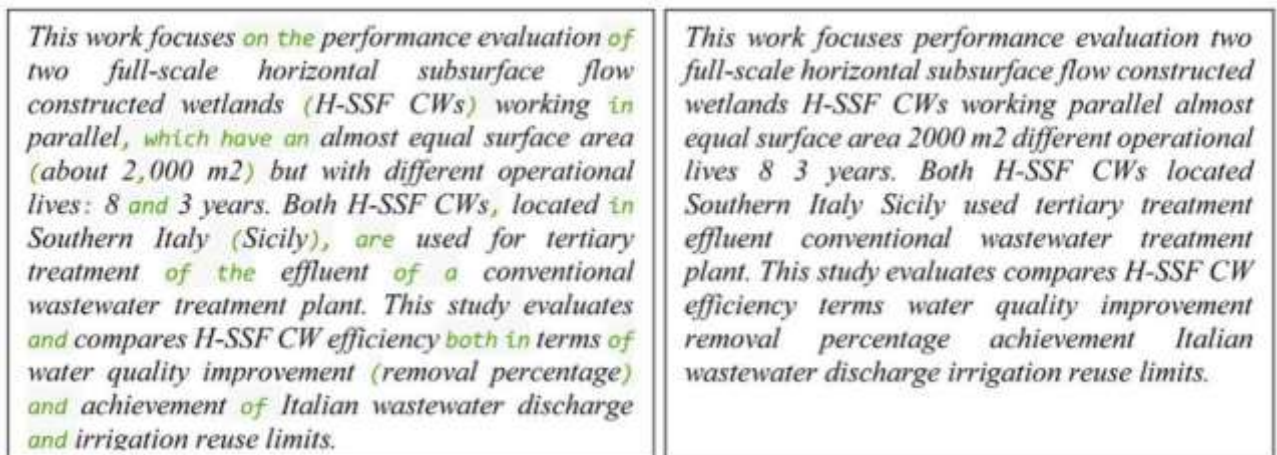


Рисунок 5.3 - Приклад документа до очищення (ліворуч) і після очищення (праворуч)

Ці функції шукають в тексті конкретні слова / пропозиції, які характеризують кожну з частин, наприклад введення, методи/матеріали і методи, а потім виключають всю інформацію до слова/пропозиції, наприклад введення, і всю інформацію після наступного слова / пропозиції, наприклад методи і методи,

щоб розділити текст на чотири основні частини. Частина висновків не розглядалася, так як необхідна інформація вже витягнута з інших частин. Результатом цього процесу є очищений і розділений текст в структурі даних R text, званої VCorpus, яка сумісна з функціями, використовуваними для наступних кроків. На рис. 5.4 представлені ключові функції.

```
## Step 6: To clean information before the word "Abstract"
removeAbstract<-function(x) gsub("(.*?)A(?:)bstract(?:-i)", "", x)
mydocs<-tm_map(mydocs,removeAbstract)
## Step 7: To clean information after the word "Conclusions"
removeConclusions<-function(x) sub("C(?:)onclusion(?:-i){.*}", "", x)
mydocs<-tm_map(mydocs,removeConclusions)
## Step 8: To divide the text into four main parts: Abstract, Introduction, Materials and Methods and Results
## Step 8.1: Abstract
removeIntroduction_down<-function(x) sub("I(?:)ntroduction(?:-i){.*}", "", x)
str1<-tm_map(mydocs,removeIntroduction_down)
## Step 8.2: Introduction
removeIntroduction_up<-function(x) sub("(.*?)I(?:)ntroduction(?:-i)", "", x)
str2_fs1.1<-tm_map(mydocs,removeIntroduction_up)
removeMethods_down<-function(x) sub("M(?:)ethods(?:-i){.*}", "", x)
str2_fs1.2<-tm_map(str2_fs1.1,removeMethods_down)
removeMethods_down<-function(x) sub("M(?:)aterial(?:-i){.*}", "", x)
str2<-tm_map(str2_fs1.2,removeMethods_down)
## Step 8.3: Materials and Methods
removeMethods_up<-function(x) sub("(.*?)M(?:)aterial(?:-i)", "", x)
str3_fs1.1<-tm_map(mydocs,removeMethods_up)
removeMethods_up<-function(x) sub("(.*?)M(?:)ethods(?:-i)", "", x)
str3_fs1.2<-tm_map(str3_fs1.1,removeMethods_up)
removeResults_down<-function(x) sub("R(?:)esults(?:-i){.*}", "", x)
str3<-tm_map(str3_fs1.2,removeResults_down)
## Step 8.4: Results
removeResults_up<-function(x) sub("(.*?)R(?:)esults(?:-i)", "", x)
str4<-tm_map(mydocs,removeResults_up)
```

Рисунок 5.4 – Ключові функції

5.4 Витяг атрибутів

Атрибути, пов'язані з водно-болотними угіддями, згадані в журнальних статтях, описують різні аспекти побудованих водно-болотних угідь. Ці аспекти можуть бути згруповані в різні класи, такі як технічні характеристики (наприклад, вид рослин, рід рослин, тип побудованих водно-болотних угідь, тип стічних вод, Площа системи), ефективність очищення (наприклад, середня концентрація припливу BOD5, середня концентрація відтоку BOD5), метадані (наприклад, назва статті, ім'я автора, назва журналу, Рік публікації) і місце

розташування (наприклад, назва країни, назва муніципалітету, широта, довгота).

У своїй першій версії інструмент фокусується на цих класах.

Інші класи як експлуатаційні (наприклад, технологія попередньої обробки, технологія подальшої обробки, частота технічного обслуговування), так і економічні аспекти (наприклад, інвестиційні витрати, витрати на технічне обслуговування) можуть бути включені в подальшу версію інструменту. Для розробки інструменту витягу даних було вибрано невелику підмножину атрибутів (див. таблицю 5.1). Ці атрибути були обрані тому, що вони є найбільш бажаними атрибутами для потенційних користувачів платформи, а також мають високу частоту появи в рецензованих журнальних статтях і, як очікується, будуть витягуватися надійним способом з використанням принципів textmining. Деякі інші атрибути, такі як загальна кількість зважених твердих частинок (TSS), також задовольняли критеріям відбору, які повинні бути вилучені інструментом. Вони не були включені в інструмент в першу чергу тому, що основною метою цієї першої версії була розробка і перевірка методу вилучення. Тому для більш легкого досягнення цієї мети було використано меншу кількість атрибутів.

У наступних версіях інструменту планується розширення списку витягуваних атрибутів.

Процес вилучення атрибутів здійснюється двома різними способами залежно від атрибута:

- за збігом ключових слів;
- по Web Scrap.

Для першого шляху інструмент шукає відповідні значення матриці текстового документа і набору виразів в папці Datasets, наприклад Type Plants. У другому випадку інструмент шукає інформацію безпосередньо на вихідній веб-сторінці, де публікується рецензована стаття. В обох випадках результати згодом уточнюються регулярними виразами для отримання фактичного значення параметра.

З практичної точки зору витяг інформації з веб-сторінки, тобто Web Scrap, простіше і надійніше, ніж витяг інформації з PDF-файлів, тобто збігу ключових слів, оскільки перетворення з .pdf в .txt займає більше часу.

Перетворення також може призвести до невпорядкованих виразів / символів, що ускладнює реалізацію надійних регулярних виразів для пошуку необхідної інформації. Структура таблиць також втрачається при переході з одного формату

в інший, що ускладнює Витяг змістовної інформації всередині них. З іншого боку, HTML-файли, крім того, що зрозумілі будь-якій людині, організовані в структури, які легко доступні для мови програмування, такої як R, за допомогою вихідного HTML-коду, що дозволяє уникнути більшості недоліків підходу PDF.

Причина, по якій він не використовується повсюдно, полягає в тому, що, на жаль, не всі журнали пропонують свої рецензовані статті в HTML-версіях. У більшості випадків анотація і метадані-це єдина інформація, яка відображається в HTML-версії. У цьому сенсі ключове слово match доповнює інформацію, яка не може бути витягнута за допомогою Web Scrap. Отже, найбільш ефективним способом вилучення інформації з якомога більшої кількості джерел є інтеграція обох підходів.

5.5 Збіг ключових слів

У цьому процесі інструмент зіставляє матрицю текстового документа, створену інструментом на основі двох заздалегідь встановлених критеріїв:

- кількість слів;
- текстовий розділ.

А також набір ключових слів, який був створений авторами. Атрибути, витягнуті таким чином, є частиною технічних специфікацій групи та розташування групи.

Цей процес ділиться на наступні три підпроцеси:

- Визначення критеріїв: мета полягає в тому, щоб визначити критерії для створення матриць текстових документів. Ці матриці генеруються процесом токенізації, який ділить текст, збережений в структурі VCorpus, на послідовні рядки з n слів. кількість слів: цей критерій являє собою діапазон для змінної N для кожного атрибута. Визначення N залежить від атрибутів, можливими значеннями яких є рядки з декількох слів. Наприклад, атрибутивна назва країни може бути назвою країни, підтвердженим двома словами 'Південна Африка' або тільки одним словом "Колумбія". Цей критерій був визначений для кожного атрибута шляхом підрахунку кількості слів кожного з значень, витягнутих з вибірки з 13 документів, а потім визначення мінімальної і максимальної кількості слів. На рис. 5.4 представлені атрибути ключових слів, а на рис. 5.5. результат зіставлення ключових слів.

Attribute Description	Keywords	OUTPUT column	File ^a
<i>Location</i>			
Country name	Dataset with the name of the countries in the world.	-	CountryDataset.txt
Name of the Municipality	Dataset with the name of the municipalities/cities in the world.	-	CityDataset.txt
<i>Technical Specifications</i>			
Plant species/Plant Genus	<i>Typha latifolia, Typha, latifolia, Phragmites australis, Phragmites, australis, Cyperus alternifolius, Cyperus, alternifolius, Arundo donax, Arundo, donax, etc.</i>	-	Type_plants.xlsx
Type of Constructed Wetland	HFCW, Horizontal Flow, Horizontal Flow Constructed Wetland, horizontal flow, surface flow constructed wetlands.	HORIZONTAL SURFACE FLOW	Type_wetland.xlsx
	Vertical Flow System, VFCW, Vertical Flow Constructed Wetland, vertical flow, vertical flow wetland	VERTICAL FLOW	
	Horizontal Subsurface Flow system, horizontal subsurface flow, Horizontal subsurface flow, subsurface flow wetland	HORIZONTAL SUBSURFACE FLOW	
Type of wastewater	Municipal raw wastewater, municipal wastewater, domestic wastewater	MUNICIPAL	Type_wastewater.xlsx
	Food processing industry, Dairy milking parlor, Potato starch processing, eutrophic lake water, industrial.	INDUSTRIAL	
	urban runoff agriculture, agricultural, swine urine.	STORMWATER AGRICULTURAL	

Рисунок 5.4 - Атрибути ключових слів

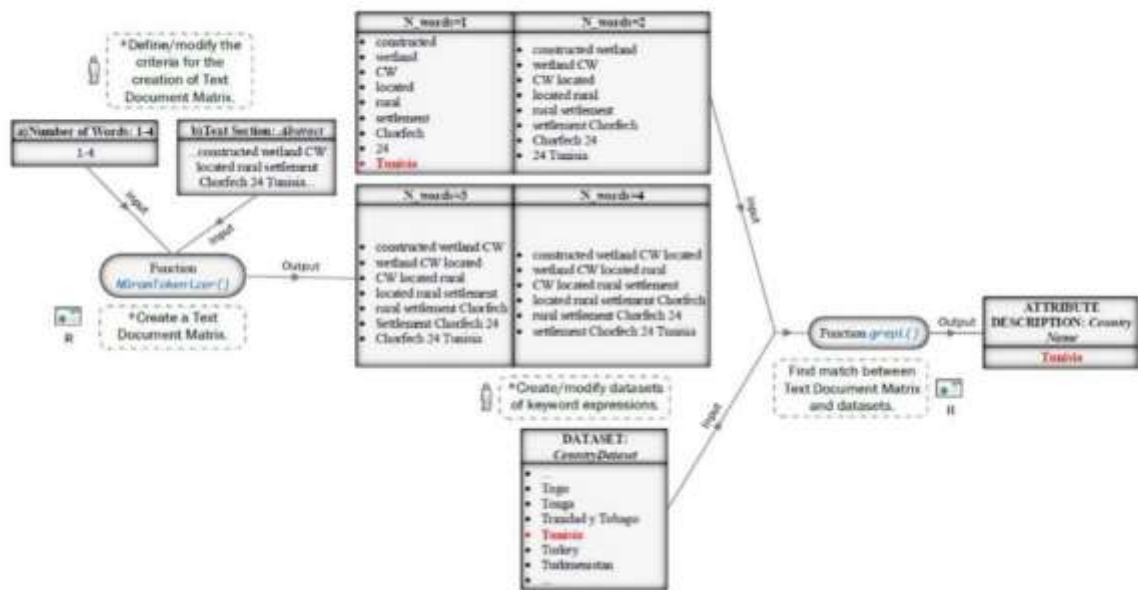


Рисунок 5.5 - Результат зіставлення ключових слів підпроцесу

Ім'я, діапазон був налаштований в 1-4, що означає, що цей атрибут може приймати значення, підтвержені 1, 2, 3 або 4 словами.

База даних виразів ключових слів: був розроблений набір даних виразів ключових слів для кожного з атрибутів. Ці набори даних являють собою список можливих значень, які може приймати атрибут. Вони засновані на аналізі відібраних 13 рецензованих статей. Наприклад, у випадку атрибутивного типу рослин вирази, які посилалися на назву рослини, були виділені в кожній статті і потім вставлені в файл Type_plants.xlsx. Різні набори даних, знайдені в папці Datasets, можуть бути легко розширені за допомогою додаткових виразів, витягнутих з аналізу більшої кількості peerreviewed статей в майбутньому. Ці вирази повинні бути додані з урахуванням великої літери, щоб отримати збіг в ключовому слові підпроцесу, оскільки інструмент розглядає великі і малі символи як різні, наприклад Турпа відрізняється від turpa. Тільки набори даних CountryDataset і CityDataset повинні залишатися незмінними, оскільки назви міст і країн світу, як очікується, не зміняться в найближчому майбутньому;

У фіналі суб-процес, програма використовує функцію NGramTokenizer (), щоб створити текстовий документ, таблицю, засновану на критерії і функції

grepI() виконує пошук збігаються значень текстовий документ, створений матриці і наборі сайту вирази зберігаються в папці набору даних.

У прикладі назва країни, перший критерій був встановлений в 1-4. Weka_control(хв = 1, Макс = 4), і другий критерій був встановлений як «автореферат», «матеріали і методи», тобто VCorpus_docs1 відноситься до абстрактних і VCorpus_docs3 відноситься до «матеріали і методи». Огляд ключових слів, витягнутих з літератури і використовуваних в процесі розробки, можна побачити нижче.

Різні автори рецензованих статей можуть виражати одне і те ж значення атрибута, наприклад HSSFHorizontal Subsurface Flow, використовуючи різні вирази, наприклад subsurface flow wetland або horizontal subsurface flow. Це характерна риса природної мови. Щоб отримати узгоджене і однорідне значення атрибута, необхідно згрупувати вирази, що представляють одне і те ж значення, в кластерну групу.

Набори ключових слів, що відповідають атрибутам тип побудованого водно-болотного угіддя і тип стічних вод, містять додатковий стовпець під назвою OUTPUT, який містить назви кластерних груп, визначених авторами. Інструмент спочатку знаходить відповідний вираз ключового слова після відповідності ключового слова підпроцесу, потім шукає відповідне ім'я групи кластерів і зберігає його в якості значення атрибута.

Web scrap цьому процесі інформація витягується безпосередньо з веб-сторінок. У HTML інформація на веб-сторінці, наприклад таблиці, графіки, зображення, основний текст, заголовки, ідентифікується за допомогою мови запитів XPath через вирази XPath. Знаючи ці вирази, кілька функцій використовуються в R для отримання даних. Параметри, витягнуті за допомогою цього підходу, є частиною метаданих групи, наприклад назва журналу, назва статті, ім'я автора і продуктивність групи, наприклад середня концентрація припливу BOD5, середня концентрація відтоку BOD5.

Цей процес ділиться на наступні 3 підпроцеси:

- користувач повинен вставити в файл посилання на сторінки, з яких були завантажені рецензовані статті, а також назву відповідних документів, наприклад Doc1, Doc2 HTML_links.xlsx з папки Фантом;
- були знайдені конкретні вирази XPath, які ідентифікують інформацію, яку ми хотіли витягти. Наприклад, для сторінки "Ельзевір", в XPath="період.заголовок-текст" визначає назву назва рецензованої статті. Кожна веб-сторінка має різні вирази XPath для отримання тієї ж інформації. В даному випадку ми використовували вирази чотирьох сайтів, а саме: Elsevier,NCBI, IWA ResearchGate;
- функція R `html_nodes()` використовувалася для доступу до інформації, ідентифікованої певним виразом XPath, за певним посиланням. Функція `html_text()` витягує інформацію, до якої раніше зверталися, коли це текст, А функція `html_tables()` - коли це таблиця. У випадку останньої функції іноді веб-сторінки містять інформацію, закодовану компонентами Java, тому необхідно викликати програму під назвою PhantomJS, щоб отримати доступ до табличної інформації. Програма не потребує установки, і її файли знаходяться в папці Phantom. Варто відзначити, що код, який використовується для виклику цієї програми в R, налаштований для роботи тільки з операційною системою Windows і що файли в папці були завантажені для 64-бітної версії. На рис. 5.6 представлені XPath і регулярні вирази.

Attribute Descriptions	XPath Expression	Regular Expression/s
Web Scrap		
Title of the article	<code>span.title-text</code>	-
Year of Publication	<code>div.text-xs</code>	<code>grep(pattern="(?)volume(?:-i)")</code> <code>sub("d u0080 u0093","-")</code> <code>sub("(.)? . ","")</code> <code>sub("[. .)","")</code> <code>sub("(.)?doi.org","doi:")</code>
DOI of the article	<code>a.doi</code>	-
Journal name	<code>h2#publication-title.publication-title</code>	-
Publisher	-	<code>grep("sciencedirect")</code>
Name of the author/s	<code>span.text.given-name</code>	<code>sub("Ä", "ä")</code> <code>sub("Åµ", "ö")</code> <code>sub("Ä u009c", "Ü")</code> <code>sub("Ä u0096Ä¶", "Öö")</code>
Last name of the author/s	<code>span.text.surname</code>	
Keyword Match	-	
Latitude	-	<code>sub("^(d+{2}).*\$", " 1")</code> <code>sub("^{0-9}{2}.*", " 1")</code> <code>sub("^(^0-9)*(d+)(^0-9 .)*", " 2")</code> <code>grep(pattern="^[0-9]{5}W")</code> <code>grep(pattern="^[0-9]{5}E")</code>
Longitude	-	<code>sub("^(d+{2}).*\$", " 1")</code> <code>sub("^{0-9}{2}.*", " 1")</code> <code>sub("^(^0-9)*(d+)(^0-9 .)*", " 2")</code> <code>grep(pattern="^[0-9]{5}W")</code> <code>grep(pattern="^[0-9]{5}E")</code>
Wastewater type	-	-
System area	-	<code>grep(pattern="[0-9]m2\$")</code> <code>grep(pattern="[0-9] m2.\$")</code> <code>grep(pattern="[0-9] m2\$")</code> <code>grep(pattern="[0-9]m2.\$")</code> <code>sub("[0-9].*", "")</code> <code>sub("m(.*)", "")</code> <code>sub("[[: space:]]+", "")</code>
Average BOD ₅ inflow concentration	-	<code>grep("(?)influent(?:-i)") grep("(?)inflow(?:-i)")</code> <code>sub("(?)Ä(.*)?(?-i)", "")</code> <code>grep("BOD")</code> <code>grep("(?)deviation(?:-i)")</code> <code>grep("(?)in(?:-i)")</code>
Average BOD ₅ outflow concentration	-	<code>grep("(?)effluent(?:-i)") grep("(?)outflow(?:-i)")</code> <code>sub("(?)Ä(.*)?(?-i)", "")</code> <code>grep("BOD")</code> <code>grep("(?)deviation(?:-i)")</code> <code>grep("(?)out(?:-i)")</code>

Рисунок 5.6 - XPath і регулярні вирази.

Інформація, витягнута за допомогою зіставлення ключових слів або веб-фрагмента, може потребувати подальшої обробки, щоб бути готовою до експорту в базу даних PostgreSQL. Деякі непотрібні слова або вирази повинні бути видалені. Для цього використовуються регулярні вирази через функцію `Rgsub()`. Ця функція шукає небажані слова / Символи за допомогою шаблонів пошуку, визначених регулярними виразами, а потім усуває їх.

Наприклад, у випадку системи атрибутів Area процес зіставлення ключових слів виробляє наступний набір виразів: "area 5000 m2."i" single 806 m2", з яких повинні бути вилучені тільки числові символи. Для цього інструмент виконує наступні кроки. Якщо будь-яке з виразів містить тільки букви в якості символів, воно відкидається за допомогою регулярного виразу "[0-9].*" тобто у цьому прикладі обидва вирази зберігаються, оскільки вони включають числа 5000 і 806. (2) залишилися вирази уточнюються шляхом виключення одиниці виміру, наприклад m2, прогалін та інших слів, наприклад "single", за допомогою регулярних виразів "m(.*) "i" [[: space:]]+".

Результатом цього процесу є: "5000–806".

Регулярні вирази були визначені шляхом аналізу результатів процесу вилучення атрибутів у вибірці з 13 документів. Регулярні вирази можуть бути додатково покращені в залежності від атрибутів і характеристик, знайдених в подальших статтях.

5.6 Експорт інформації в PostgreSQL

Метою цього процесу є експорт значень атрибутів, витягнутих в базу даних в PostgreSQL: системі управління об'єктно-реляційними базами даних з відкритим вихідним кодом, підключеної до платформи. Структура бази даних складається з 34 таблиць, які корелюються зовнішніми ключами, тобто інформацією стовпця, що розділяється різними таблицями. Було вирішено використовувати цю структуру реляційної бази даних, оскільки вона розбиває інформацію на різні сутності, що полегшує навігацію / зміну бази даних в PostgreSQL.

Інструмент генерує фрейм даних, тобто табличну структуру в R, де інформація, уточнена на етапі використання регулярного виразу, зберігається разом з похідними атрибутами і первинними ключами. Останні і- це стовпці в таблицях бази даних PostgreSQL, які однозначно ідентифікують кожен рядок у таблиці. Необхідно дотримуватися таких особливостей: кожен рядок повинен містити унікальне значення і не може містити нульових значень. Наприклад, в табличній літературі первинним ключем є стовпець LIT_ID, який може

приймати значення LIT_1, LITE_2, LIT_3. Значення первинних ключів для нової інформації, яка буде експортована, повинні бути визначені в наступній послідовності:

інформація вже є в базі даних PostgreSQL. Наприклад, якщо останнім первинним ключем в таблиці літератури PostgreSQL є LIT_8, то нові інформаційні первинні ключі повинні починатися з LIT_9. Для цього використовується функція `r dbGetQuery ()`, яка зчитує таблиці PostgreSQL в R, відстежуючи поточну послідовність первинних ключів. Крім того, похідні атрибути-це атрибути, які не були отримані в процесі вилучення атрибутів, але які створюються шляхом об'єднання. На рис. 5.7 представлений список похідних атрибутів.

Derived Attribute Descriptions	Extractable Attributes	Derived Attribute names	Entities	Classes/Groups used by CWetlands-Extraction tool
Citation of the peer-review article	TITLE	LIT_CITE	LITERATURE	metadata
	YEAR			
	DOI			
	JOURN_NAME			
	PUBLISHER			
	AT_FIRST_N			
Country code	AT_LAST_N			
	CTRY_NAME	CTRY_NAME	SITES	location
Latitude	MUNI	LAT	SITES	location
Longitude	MUNI	LONG	SITES	location

Рисунок 5.7 - Список похідних атрибутів

Наприклад, атрибут CITATION виходить шляхом об'єднання атрибутів в сутності література і автори за допомогою функції `paste0`. Крім того, похідні атрибути можуть бути отримані шляхом вставки значень витягнутого атрибута в функцію. Наприклад, атрибути LAT і LONG виводяться за допомогою функції

getGeoData(), яка дає значення координат, вставляючи витягнутий атрибут MUNI. Інструмент використовує останній підхід, коли в рецензованих статтях відсутня інформація про координати.

5.7 Перевірка ефективності

Інструмент використовує ключові слова і дані, витягнуті з 13 статей, і був перевірений на наявність тих же статей. Середній час, витрачений автоматичним підходом (інструментом), порівнювався з часом, витраченим ручним підходом, шляхом зупинки спостереження за часом, використовуваним обома підходами для обробки одних і тих же рецензованих журнальних статей. Автоматичний процес зайняв 30 хв для всіх 13 виробів (час обробки), тоді як ручний процес екстракції зайняв близько 480 хв.

Ефективність кожного атрибута, що відповідає кожному виробу у вибірці, наведена в таблиці 5: значення 0 означає, що інструмент не зміг витягти атрибут або що витягнуте значення не було таким же, як записане в ручному процесі. Значення 1 означає, що значення атрибутів рівні в обох випадках процесу. Нарешті, значення н / Д означає, що не було ніякої інформації про атрибут в рецензованих статті. Показники успіху (значення 1, де значення атрибутів були рівні) варіюються від 57,14%, як у випадку атрибута MUNI, до 100%, як у випадку атрибутів JOURNAL_NAME або PUBLISHER. Загальний показник успіху за всіма параметрами і всіма статтями склав 87%. Атрибути в entity LITERATURE, COUNTRY I C_PLANTS показують високі показники успіху (90-100%), що означає, що інструмент добре працює при витяганні інформації з цих сутностей. Атрибути в сайтах сутностей і S_ORG показують низькі показники успіху (75-22,2%), а атрибути в системах сутностей і осередках - середні показники успіху (90-76, 9%). На рис. 5.8 представлений список ефективності по атрибутам

Entity	Attributes	Article													Efficiency
		1	2	3	4	5	6	7	8	9	10	11	12	13	
COUNTRY	Country name	1	1	1	N/A	1	1	1	1	1	1	1	1	1	100
SITES	Name of the Municipality	1	0	N/A	N/A	N/A	0	N/A	N/A	N/A	1	1	0	1	57.1
SITES	Latitude & Longitude	1	1	N/A	N/A	N/A	0	1	N/A	N/A	1	1	0	1	75
SITES	Wastewater type	1	1	0	1	1	0	0	1	0	N/A	1	1	1	66.7
SYSTEMS	System area	0	1	1	0	1	1	1	1	1	1	1	0	1	76.9
CELLS	Plant species/genus	1	1	0	1	1	N/A	1	1	1	1	N/A	N/A	1	90
CELLS	Type of Constructed Wetland	1	1	1	0	1	0	0	1	1	1	1	1	1	76.9
AUTHORS	Name/Last name of the authors	1	1	1	1	1	1	1	1	1	1	1	1	1	100
LITERATURE	Title of the article	1	1	1	1	1	1	1	1	1	1	1	1	1	100
LITERATURE	Publisher	1	1	1	1	1	1	1	1	1	1	1	1	1	100
LITERATURE	Year of Publication	1	1	1	1	1	1	1	1	1	1	1	1	1	100
LITERATURE	Journal Name	1	1	1	1	1	1	1	1	1	1	1	1	1	100
LITERATURE	DOI	1	1	1	1	1	1	1	1	1	1	1	1	1	100
S_ORG	BOD inflow	0	N/A	N/A	N/A	0	N/A	1	0	1	0	0	0	0	22.2
S_ORG	BOD outflow	0	N/A	N/A	N/A	0	N/A	1	0	1	0	0	0	N/A	25

Рисунок 5.8 - Ефективність по атрибутам

ВИСНОВКИ

В ході виконання випускної кваліфікаційної роботи бакалавра були отримані наступні результати:

- 1) Розглянуто теоретичні дані щодо вже існуючих систем знань водно-болотних угідь;
- 2) Проведено їх порівняння для виявлення позитивних сторін і недоліків;
- 3) На підставі вхідних даних за допомогою нейронних мереж реалізовано алгоритм збору основних даних про водно болотні угіддя;
- 4) Проведено порівняння обробки цих даних за допомогою людського ресурсу і розробленого алгоритму;
- 5) Розроблено програмне забезпечення для візуалізації та модифікації зібраних даних;
- 6) Проведено оцінку геотехнологічних можливостей розробленої платформи та оцінку можливостей для майбутнього розвитку.

Розроблена платформа може стати універсальним постачальником рішень для пов'язаних з водно-болотними угіддями даних та інформації для сприяння дослідженням, розробці політики та фінансуванню, розширення прав і можливостей організацій громадянського суспільства і практиків для реалізації заснованих на природі рішень в інтересах сталого розвитку.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Офіційний сайт з документацією по технології Geonode
Режим доступу: <http://geonode.org/> (дата звернення 28.11.2020)
2. Офіційний сайт з документацією по Python
Режим доступу: <https://www.python.org/> (дата звернення 27.11.2020)
3. Офіційний сайт з документацією по postgresql
Режим доступу: <https://www.postgresql.org/> (дата звернення 29.11.2020)
5. Jim Thatcher, Josef Eckert, Andrew Shears. Thinking Big Data in Geography, Nebraska Press, 2018, - 324с.
6. Ilya Grigorik. High Performance Browser Networking. Massachusetts: O'Reilly Media, 2013. – 582с.
7. I. Feinerer, K. Hornik, D. Meyer, Text mining infrastructure in R, J. Stat. Softw., 2008, 1–54с.
8. A. Ghrabi, L. Bousselmi, F. Masi, M. Regelsberger, Constructed wetland as a low cost and sustainable solution for wastewater treatment adapted to rural settlements: the Chorfech wastewater treatment pilot plant, Water Sci. Technol., 2011, 3012с.
9. K. Welbers, W. Van Atteveldt, K. Benoit, Text analysis in R, Commun. Methods Meas., 2017. 265с.
10. S. Prost-Boucle, P. Molle, Recirculation on a single stage of vertical flow constructed wetland: treatment limits and operation modes, Ecol. Eng., 2012, 267с
11. Уоссермен Ф. Нейрокомпьютерная техника: Теория и практика / Ф. Уоссермен; [пер. с англ.]. – М. : Мир, 1992. – 240 с.