

РОЗРОБЛЕННЯ ЕКСПЕРТНОЇ ІНФОРМАЦІЙНОЇ ВЕБ-СИСТЕМИ ВИЯВЛЕННЯ ТА ВЕРИФІКАЦІЇ ПРІОРИТЕТНИХ ЕКОЛОГІЧНИХ ПРОБЛЕМ У МАСИВАХ ВОД БАСЕЙНУ РІЧКИ

Вінницький національний технічний університет

Анотація

У даній роботі розглянуто питання збирання, верифікації та узагальнення великої кількості експертних оцінок про наявний стан вод та наявні екологічні проблеми. Задача ускладнюється великою кількістю таких масивів вод, оскільки зібрати достовірну інформацію про об'єкти, що розташовані у кожному з них, вкрай важко. Розв'язати таку проблему дозволить створення веб-системи з картою масивів вод і залученням великої кількості експертів, у т.ч. місцевих жителів, вимагаючи від них не просто експертні оцінки на основі єдиних довідників варіантів можливих відповідей, а й – посилання на текстові веб-ресурси, які підтверджують їх оцінки.

Ключові слова: інформаційна веб-система, масиви вод, технології захисту водних ресурсів, експертні оцінки, екологічні проблеми, машинне навчання, класифікація текстових даних, технології опрацювання природної мови, NLP.

Abstract

This paper considers the collection, verification and generalization of a large number of expert assessments of the current state of waters and existing environmental problems. The task is complicated by a large number of such water bodies, as it is extremely difficult to gather reliable information about the objects located in each of them. To solve this question will create a web system with a map of water bodies and the involvement of a large number of experts from the locals, requiring them not only expert assessments based on single directories of possible answers but also - links to text web resources that confirm their reviews.

Keywords: information web system, water protection technologies, expert assessments, environmental problems, machine learning, text data classification, natural language processing, NLP.

Вступ

Завше етапу розроблення програм заходів планів управління річкового басейну (ПУРБ) в Європі, спрямованих на досягнення чи стабілізацію доброго екологічного стану води у масивах вод цього басейну, передусє етап збирання великої кількості інформації про наявний стан вод та екологічні проблеми [1-4]. Задача ускладнюється великою кількістю таких масивів вод, тому зібрати достовірну інформацію про об'єкти, розташовані конкретно у кожному з таких, вкрай важко. Цьому може зарадити створення веб-системи з картою масивів вод і залучення великої кількості експертів, у т.ч. із числа місцевих жителів, небайдужих до проблем свого довкілля. Але виникає питання: як перевірити достовірність оцінок таких експертів, враховуючи різну їх кваліфікацію та досвід роботи, різне бачення цілей ПУРБ, та як їх порівнювати, щоб визначити найбільш вразливі регіони? Тому додатково пропонується вимагати від них не просто експертні оцінки на основі єдиних довідників варіантів можливих відповідей, а й – посилання на джерела (якісь текстові веб-ресурси), які підтверджують їх слова. А потім оцінювати наскільки ці джерела дійсно підтверджують експертну оцінку щодо даного регіону.

Отже, метою даного дослідження є розробка експертної інформаційної веб-системи виявлення та верифікації пріоритетних екологічних проблем у масивах вод басейну річки за експертними оцінками та підтверджуючими їх текстовими матеріалами, що дозволить одночасно зібрати максимально детальну і достовірну інформацію про об'єкти масивів вод та прискорити її опрацювання і ранжування.

Результати дослідження

Для досягнення поставленої задачі розроблена інформаційна веб-система, робота якої складається з наступних етапів: етап збирання та верифікації інформації, етап оброблення даних та етап візуалізації.

На етапі збирання та верифікації інформації користувач веб-системи (експерт) описує проблему та вказує посилання на текстові веб-ресурси, які містять підтвердження наявності зазначеної проблеми у заданому масиві вод. Після цього здійснюється верифікація достовірності опису.

Верифікація здійснюється за допомогою використання методів NLP (Natural Language Processing), які спочатку перевіряють чи текст містить назви відповідних географічних об'єктів. Якщо містить, то далі здійснюється перевірка наскільки кожне речення цього тексту відповідає сформульованій проблемі. Для цього слід мати наперед треновані моделі для кожного виду проблем, за якими застосовувати для кожного речення тексту метод `predict_proba`, який оцінює ймовірність кожного класу моделі [5, 6]. Для кожного речення тексту обчислюється така ймовірність p_r , а потім серед них визначається максимальне значення P і воно присвоюється усьому тексту. У разі, якщо експерт надає m текстів на підтвердження своєї оцінки, тоді загальну для них достовірність k пропонується обчислювати за таким виразом:

$$k = 1 - \prod_{q=1}^m (1 - P_q) \quad (1)$$

Але, важливо враховувати випадок, коли ймовірність P_q жодного з текстів не подолає певний мінімальний поріг P^* , наприклад у 0,6, що означатиме негативний результат етапу верифікації оцінки.

В результаті верифікації, за формулою (1), для оцінки експерта отримуємо достовірність $k_{i,j}$ де i – індекс проблеми, j – індекс водного масиву, яка вказує на ступінь достовірності опису.

На етапі оброблення даних беруться усі достовірні оцінки проблеми i та їх значення достовірності $k_{i,j}$ і обчислюється $K_{i,j}$ – вага важливості проблеми i для водного масиву j . Для водного масиву ваги важливості проблем $K_{i,j}$ сортуються і, відповідно, проблеми, для яких значенням $K_{i,j}$ є найбільшим у даному водному масиві, вважаються найбільш значущими для водного масиву.

На етапі візуалізації на мапі відображаються масиви вод водного басейну та їх найбільш значущі проблеми (рис. 1). Тому, можна легко побачити, які проблеми є важливими для водного масиву.

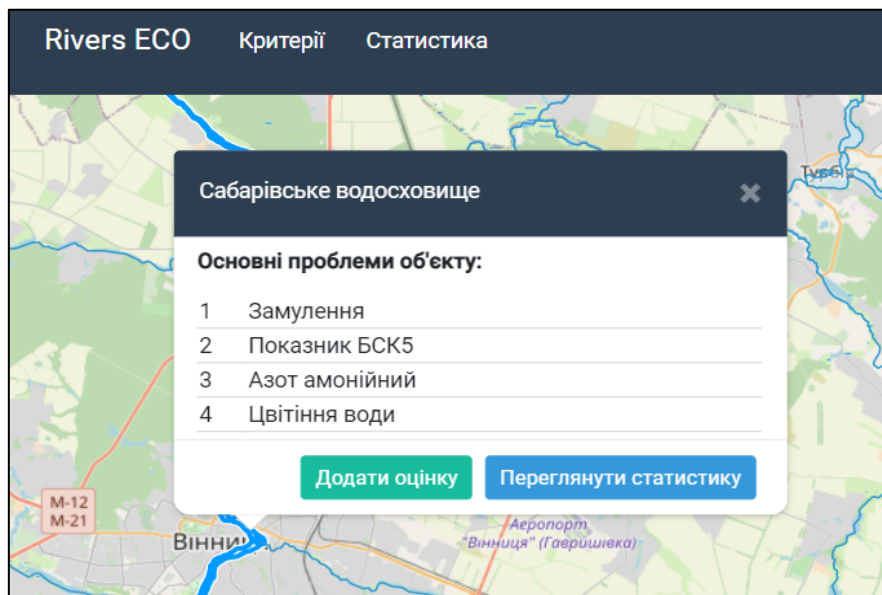


Рис. 1. Інтерфейс прототипу створеної експертної веб-системи: масив вод із списком значущих проблем

Висновки

У роботі розглянуто питання збирання великої кількості експертних оцінок про наявний стан вод та наявні екологічні проблеми. Для розв'язання поставленої задачі запропоновано створити веб-систему з картою масивів вод, яка дозволить залучати велику кількість експертів. Також запропоновано вимагати від експертів не просто оцінки на основі єдиних довідників варіантів можливих відповідей, а

й – посилання на джерела, які підтверджують їх слова – це дозволить усунути проблему перевірки оцінок на достовірність.

Розроблено прототип експертної інформаційної веб-системи виявлення пріоритетних екологічних проблем у масивах вод басейну річки. Відзначено, що в цій технології можна виділити три етапи: етап збирання та верифікації інформації про проблеми, етап оброблення зібраних даних та етап візуалізації результатів. Розглянуто метод опрацювання експертних оцінок, який може бути застосований не тільки для аналізу екологічного стану масивів вод водних об'єктів України, а й для інших подібних об'єктів, де є доцільним залучення великої кількості експертів з різною кваліфікацією та досвідом.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Постанова Кабінету Міністрів України № 336 від 18.05.2017 року «Про затвердження Порядку розроблення плану управління річковим басейном». Режим доступу до ресурсу: <https://www.davr.gov.ua/postanova-kabinetu-ministriv-ukraini-vid-18-travnua-2017-roku--336-pro-zatverdzhennya-poryadku-rozroblennya-planu-upravlinnya-richkovim-basejnom>
2. Директива 2000/60/ЄС Європейського Парламенту і Ради "Про встановлення рамок діяльності Співтовариства в галузі водної політики" від 23 жовтня 2000 року.
3. Водний Кодекс України (Відомості Верховної Ради (ВВР), 1995, № 24, ст. 189) (введений в дію Постановою ВР № 214/95-ВР від 06.06.95).
4. The EU Water Framework Directive - integrated river basin management for Europe Directive 2000/60/EC establishing a framework for the Community action of water policy (Water Framework Directive).
5. Scikit-learn 0.24.1. Machine Learning in Python. User Guide. – Режим доступу: https://scikit-learn.org/stable/user_guide.html
6. Mokin V. «NLP for EN : BERT Predict_proba in Water Report» / V. Mokin, A. Luchko, O. Davidiyuk // Kaggle. – Режим доступу: <https://www.kaggle.com/vbmokin/nlp-for-en-bert-predict-proba-in-water-report?scriptVersionId=53779986>

Мокін Віталій Борисович — д-р техн. наук, професор, завідувач кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: vbmokin@gmail.com

Лучко Андрій Михайлович — аспірант кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: andriyluchko@gmail.com

Давидюк Оксана Миколаївна — аспірант кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, Вінниця, e-mail: davidyuk-ok@ukr.net

Mokin Vitalii B. — Dr. Sc. (Eng.), Professor, Head of the Chair of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: vbmokin@gmail.com

Luchko Andrii M. — Postgraduate student of the Chair of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: andriyluchko@gmail.com

Oksana Davydyuk M. — Postgraduate student of the Chair of System Analysis and Information Technologies, Vinnytsia National Technical University, Vinnytsia, e-mail: davidyuk-ok@ukr.net