

К. О. Бондалєтов¹В. Б. Мокін¹

АВТОМАТИЗАЦІЯ ГЕОЛОКАЦІЇ ТЕКСТОВОЇ ІНФОРМАЦІЇ ДЛЯ КОНТРОЛЮ ЕФЕКТИВНОСТІ ЗАХОДІВ З УПРАВЛІННЯ РІЧКОВИМ БАСЕЙНОМ

¹ Вінницький національний технічний університет;

Робота присвячена проблемі контролю ефективності імплементації заходів з Плану управління річковими басейном (ПУРБ) та пошуку підходів щодо їх оптимізації. Пропонується технологія автоматизованого збирання та оброблення текстової інформації з альтернативних джерел, зокрема веб-сайтів екологічних установ та соціальних мереж щодо стану водних ресурсів після застосування у них тих заходів. Використовуються методи веб-скрапінгу та спеціалізовані бібліотеки BeautifulSoup і Newspaper4k для збору текстової інформації з пошукової видачі та обраних джерел інформації. Для автоматизації геолокації зібраної інформації до конкретних масивів вод застосовується розроблений авторами метод порівняння ключових слів зі словами з опису масивів вод, а вибірка ключових слів здійснюється з використання іншого авторського методу. З метою поширення позитивного досвіду чи проблем, пов'язаний з результатами застосування заходів, з одних масивів вод та басейнів на інші застосовується метод аугментації текстової інформації, розроблений на базі бібліотеки SpaCy. Отже, запропонована технологія дозволяє підвищити ефективність контролю ефективності та аналізу аналогів щодо реалізації заходів ПУРБ у різних басейнах річок, завдяки автоматизованому збиранню та аналізу великого обсягу текстових даних із загальнодоступних альтернативних (у т.ч. неофіційних) джерел інформації. Запропонований підхід дозволяє більш ефективно співставляти та рекомендувати конкретні заходи ПУРБ до конкретних, подібних, масивів вод, що покращує можливості управління річковими басейнами. Таким чином, запропонований підхід підвищує ефективність прийняття та науково обґрунтовування рішення щодо сталого використання водних ресурсів, поліпшення екологічного стану вод, збереження та відновлення біорізноманіття та екосистем річкового басейну.

Ключові слова: Веб-скрапінг, План управління річковим басейном (ПУРБ), масив вод, аналіз даних, контроль якості вод, геолокаційні технології.

Вступ

Оптимізація та підтримка гарних значень показників якості водних ресурсів вимагає систематичних, науково обґрунтованих заходів. Тому з метою забезпечення сталого використання водних ресурсів, підвищення якості води, збереження та відновлення біорізноманіття та екосистем в усіх країнах ЄС та в Україні розробляються Плани управління річковим басейном (ПУРБ), зазвичай на шестирічні періоди [1]. Ці документи затверджуються на урядовому та міжурядовому рівнях. Вони є необхідними для інтеграції соціальних, економічних та екологічних вимог шляхом імплементації комплексів заходів, спрямованих на збалансоване використання ресурсів у межах визначеного річкового басейну.

ПУРБ є основним інструментом управління водними ресурсами [1], що містить так звану “Програму заходів”, спрямованих на забезпечення екологічного стану кожного масиву вод басейнів річок на рівні, який є не нижчим “доброго”. Цікаво, що англійським аналогом терміну “масив вод” є “water body”, але це – не дослівний переклад, це – затверджено на законодавчому рівні, в Україні Верховна рада вирішила назвати дещо по-іншому.

Для верифікації ефективності застосованих дій необхідно здійснювати оцінювання ефективності реалізації затверджених заходів ПУРБ, що, фактично, реалізують систему управління басейном річки. Важливо відслідковувати ефективність цих заходів після їх реалізації і враховувати під час оновлення ПУРБ. Однак, ці заходи та офіційні звіти про їх

запровадження можуть бути надто узагальненими, без деталізації особливостей і результатів реалізації у розрізі масивів вод. А тому з метою відслідковування змін по окремих масивах вод пропонується залучати альтернативні джерела інформації. Автори висловлюють гіпотезу про те, що, по-перше, така інформація може міститись у різноманітних джерелах текстової інформації: сайтах екологічних установ з прикладами успішної реалізації програм заходів на деяких масивах вод та в їх дописах у соціальних мережах, а також у дописах водокористувачів, що знаходяться нижче по течії, та – інших дотичних джерелах. А по-друге, що з використанням технологій штучного інтелекту можна здійснити геолокацію цих текстів на основі наведеної у них інформації та класифікацію проблем чи успіхів щодо реалізації згаданих вище заходів, що дозволить їх подальше автоматичне оброблення та узагальнення.

Результати дослідження

В сучасних дослідженнях соціальних мереж та порталів новин важливу роль відіграють методи збирання та аналізу даних, зокрема веб-скрапінг та постоброблення текстових даних. Веб-скрапінгом є автоматизований процес вилучення даних із веб-сайтів, який дозволяє ефективно видобувати усі доступні актуальні дані з відомих чи очікуваних джерел інформації. У рамках задачі автори використовують спеціалізовані бібліотеки [2]: BeautifulSoup та Newspaper4k, які забезпечують автоматизоване збирання даних (рис. 1).

Notebook Input Output Logs Comments (1) Settings				
Out[6]:				
	domain	keyword	title	link
0	fakty.com.ua	Південний Буг	Південний Буг - Факти ICTV	https://fakty.com.ua/ua/tag/pivdennyj-bug/
1	fakty.com.ua	Південний Буг	Південний Буг стрімко всихає: як рятують річку	https://fakty.com.ua/ua/videos/pivdennyj-bug-s...
2	fakty.com.ua	Південний Буг	Одна з найбільших річок України перетворюється...	https://fakty.com.ua/ua/ukraine/suspilstvo/201...
3	fakty.com.ua	Південний Буг	Битва за Південний Буг: чому обміліла річка та...	https://fakty.com.ua/ua/videos/bytva-za-pivden...
4	fakty.com.ua	Південний Буг	На березі Південного Бугу побудували незаконні...	https://fakty.com.ua/ua/videos/_trashed-3/
5	fakty.com.ua	Південний Буг	Була прив'язана мотузка з каменем: на Миколаїв...	https://fakty.com.ua/ua/proisshestvija/2019080...
6	fakty.com.ua	Південний Буг	топ-5 знахідок археологів у Меджибзькій фортеці	https://fakty.com.ua/ua/ukraine/20210815-kubok...
7	fakty.com.ua	Південний Буг	Мер Миколаєва розповів, коли у місті відновлять...	https://fakty.com.ua/ua/ukraine/20220505-mer-m...
8	fakty.com.ua	Південний Буг	Радонове озеро та рафтинг: скільки коштує відп...	https://fakty.com.ua/ua/videos/radonove-ozero-...

Рис. 1 – Приклад результату первинного автоматизованого веб-скрапінгу проблем про стан масивів вод басейну р. Південний Буг з обраних джерел інформації [2]

Розроблена у статті [3] за участю авторів цього дослідження методологія дозволяє здійснити аналіз текстової інформації про проблеми чи успіхи реалізації заходів та прив'язати її до

Similarity of separate word
Південний Буг 0.7775
ж. Вінниця 0.3543
с. Стрижавка 0.2833
с. Гушниця 0.1487
с. Мезівка 0.1124
с. Медвівка 0.2695
с. Лаврівка 0.2273
водозабір Вінницького каналу 0.4462
Київський міст 0.7352
Центральний міст 0.7881
Старомиський міст 0.7684
"Вінницька реберна" 0.3897
пляж "Хімік" 0.468
пляж "Кумбар" 0.2882
пляж "Спартак" 0.4835
пляж "Гонти" 0.2933
мікрорайон "Старе місто" 0.535
набережна "Ромен" 0.4885
місце впадіння р. Тажилівка 0.2195
Десна 0.2957
Периорка 0.2149
Вини 0.3466
Згар 0.4248
Постолова 0.2387
Вінницька область 0.3992
Вінниччина 0.2857
Хмельницький район 0.6878
Вінницький район 0.7491
мостовий шляхпровід автодороги "Вінниця – Калинівка" 0.5817

конкретних масивів вод, тобто забезпечити геолокацію цієї текстової інформації (рис. 2). Прив'язка забезпечується по ключових словах, які можна видобути у різний спосіб, але автори пропонують використовувати їх метод, викладений у роботі [4]. Для поширення інформації про одні масиви вод на інші, подібні за параметрами (фактично, для текстової "аугментації" даних) було використано бібліотеку SpaCy, що дозволило збільшити обсяг інформації для аналізу та узагальнення [3]. Узагальнену по багатьох таких текстах інформацію можна використати для контролю ефективності реалізації програми заходів ПУРБ та для пошуку напрямків їх більш ефективної реалізації, з урахуванням досвіду, отриманого щодо інших масивів вод та басейнів.

Рис. 2 Приклад геолокаційної прив'язки текстових даних до масивів басейну по ключових словах повідомлення

Висновки

Таким чином, запропоновано технологію автоматизованого збирання, у т.ч. у соцмережах, та оброблення текстової інформації із забезпеченням її геолокації щодо конкретних масивів вод заданого басейну вод для контролю ефективності реалізації програм заходів Планів управління річковими басейнами та пошуку шляхів удосконалення цих програм, з урахуванням позитивного досвіду, отриманого щодо інших масивів вод чи щодо інших басейнів річок. Ця технологія дозволить забезпечити більш швидке та масштабне збирання різноманітної текстової інформації, її автоматизоване оброблення та, в цілому, сприятиме прийняттю більш оптимальних рішень щодо поліпшення екологічного стану вод басейнів річок.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] Кабінет Міністрів України, Постанова від 18.05.2017 р. No 336, Про затвердження Порядку розроблення плану управління річковим басейном [Електронний ресурс]. Режим доступу: <https://www.kmu.gov.ua/npas/249999756>. Дата звернення 07.08.2024.
- [2] Bondaletov K. Collecting Google Search Data with BeautifulSoup. Kaggle Notebook, 2024. Режим доступу <https://www.kaggle.com/code/bondaletov/collecting-google-search-data-with-beautifulsoup>. Дата звернення 07.08.2024.
- [3] Мокін В. Б., Бондалетов К. О., Крижановський Є. М., і Каравасв В. О. Метод аугментації текстів про стан масивів вод на основі інтелектуальної прив'язки до багатозв'язних геоінформаційних систем іменованих сутностей, Вісник Вінницького політехнічного інституту, вип. 3, с. 55–65, Черв. 2023. URL: <https://doi.org/10.31649/1997-9266-2023-168-3-55-65>.
- [4] В. Б. Мокін, і К. О. Бондалетов, Інтелектуальні методи видобування ключових словосполучень із тексту для побудови онтологічних моделей інформаційно-пошукових систем. Інформаційно-комунікаційні технології та сталий розвиток, колективна моногр. за матеріалами XXI Міжнародної науково-практичної конференції, Київ, 14-16 листопада 2022 р., С. О. Довгий, Заг. ред. К., Україна: ТОВ «Видавництво «Юстон», 2022, С. 242.

Бондалетов Костянтин Олегович — аспірант кафедри системного аналізу та інформаційних технологій Вінницького національного технічного університету; e-mail: bondaletov.k@gmail.com;

Мокін Віталій Борисович — д-р техн. наук, професор, завідувач кафедри системного аналізу та інформаційних технологій Вінницького національного технічного університету; e-mail: vbmokin@vntu.edu.ua.

K. O. Bondaletov¹

V. B. Mokin¹

Automation of Geolocation of Textual Information for Controlling the Effectiveness of River Basin Management Measures

¹Vinnitsia National Technical University

The work is devoted to the problem of controlling the effectiveness of the implementation of measures from the River Basin Management Plan (RBMP) and finding approaches to their optimization. The technology of automated collection and processing of textual information from alternative sources, in particular the websites of environmental institutions and social networks regarding the state of water resources after the application of those measures, is proposed. Web scraping methods and specialized libraries BeautifulSoup and Newspaper4k are used to collect text information from search results and selected information sources. To automate the geolocation of collected information to specific bodies of water, the method developed by the authors of comparing keywords with words from the description of bodies of water is used, and the selection of keywords is carried out using another author's method. In order to spread positive experience or problems related to the results of the application of measures, the text information augmentation method developed on the basis of the SpaCy library is used from one body of water and basins to others. Therefore, the proposed technology allows to increase the effectiveness of controlling the effectiveness and analysis of analogues regarding the implementation of RBMP measures in various river basins, thanks to the automated collection and analysis of a large amount of text data from publicly available alternative (including unofficial) sources of information. The proposed approach makes it possible to more effectively compare and recommend specific PURB measures to specific, similar, bodies of water, which improves the management capabilities of river basins. Thus, the proposed approach increases the effectiveness of decision-making and scientific justification for the sustainable use of water resources, improvement of the ecological state of waters, preservation and restoration of biodiversity and ecosystems of the river basin..

Keywords: components Web scraping, River Basin Management Plan (RBMP), water body, data analysis, water quality control, geolocation technologies.

Bondaletov Kostyantyn Olehovych — Post-graduate student of the Chair of System Analysis and Information Technology, email: bondaletov.k@gmail.com;

Mokin Vitalii Borysovych — Dr.Eng.Sc., Prof., head of the department of System Analysis and Information Technologies of the Vinnitsia National Technical University; e-mail: vbmokin@vntu.edu.ua.