

**В.Б. Мокін**  
**Є.М. Крижановський**  
**М.А. Гораш**  
**Д.В. Пасічнюк**  
**О.В. Радецький**  
**С.М. Хмарук**

## **ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ЗБИРАННЯ ТА СИСТЕМАТИЗУВАННЯ ДАНИХ ПРО СТАН ТА ПРОБЛЕМИ МАСИВІВ ВОД НА ПРИКЛАДІ БАСЕЙНУ РІЧКИ ПІВДЕННИЙ БУГ**

Вінницький національний технічний університет, Україна

### **Анотація**

*Запропоновано комплекс інформаційних систем та технологій збирання та систематизації даних про стан та проблеми масивів вод на прикладі басейну р. Південний Буг за допомогою технологій розпізнавання іменованих сутностей NER, технологій опрацювання природної мови NLP та інших технологій штучного інтелекту.*

**Ключові слова:** технологія розпізнавання іменованих сутностей, технологія опрацювання природної мови, NLP, машинне навчання, штучний інтелект, басейн річки, масив вод

### **Abstract**

*The complex of information technologies of collecting and systematizing data on the state and problems of the massifs of the Southern Bug river basin with the help of the technology of recognizing the named entities of NER, the technologies of natural language processing NLP and other technologies of artificial intelligence is proposed.*

**Keywords:** named entity recognition technology, natural language processing technology, NLP, machine learning, artificial intelligence, river basin, water body

### **Вступ**

В еру інформаційних технологій інформацію про той чи інший масив вод можна отримати за допомогою мережі Інтернет, але через велику кількість неточної, дубльованої і «фейкової» інформації важко швидко знайти корисні і чіткі дані про його екологічний стан, екологічні проблеми та досвід їх вирішення. Знайдена інформація не завжди є систематизованою і конкретною, що не дає чіткого уявлення про подію чи ситуацію, яку вона описує.

### **Результати дослідження**

Було концептуально розроблену інформаційну інтелектуальну технологію, яка дозволить автоматизувати збирання та систематизація даних про стан та проблеми вод басейну р. Південний Буг з використанням технологій штучного інтелекту. Дана технологія має 5 складових (рис. 1):

- Експертна веб-система;
- Верифікація текстів з веб-ресурсів, який підтверджує проблему;
- Інтелектуальна ІТ – основна складова;
- База даних класифікованих екологічних проблем із прив'язкою у часі;
- Вибірка та аналіз щодо конкретних масивів, екологічних проблем та часу.

Охарактеризуємо ці складові, які вже створені або ще створюються під керівництвом Мокіна В.Б., більш детально.

1. Експертна веб-система. За участі Лучка А.М. та Давидюк О.М. [1, 2] розроблено технологію побудови експертної інформаційної веб-системи виявлення пріоритетних екологічних проблем у масивах вод басейну річки та пілотна версія такої системи. Система дозволяє залучити велику кількість користувачів-експертів з різною кваліфікацією та досвідом для оцінювання екологічного стану масивів вод та ідентифікації суттєвих проблем. На рисунку 2 показано форму створення оцінки водойми, а на рисунку 3 – вигляд експертної веб-системи.

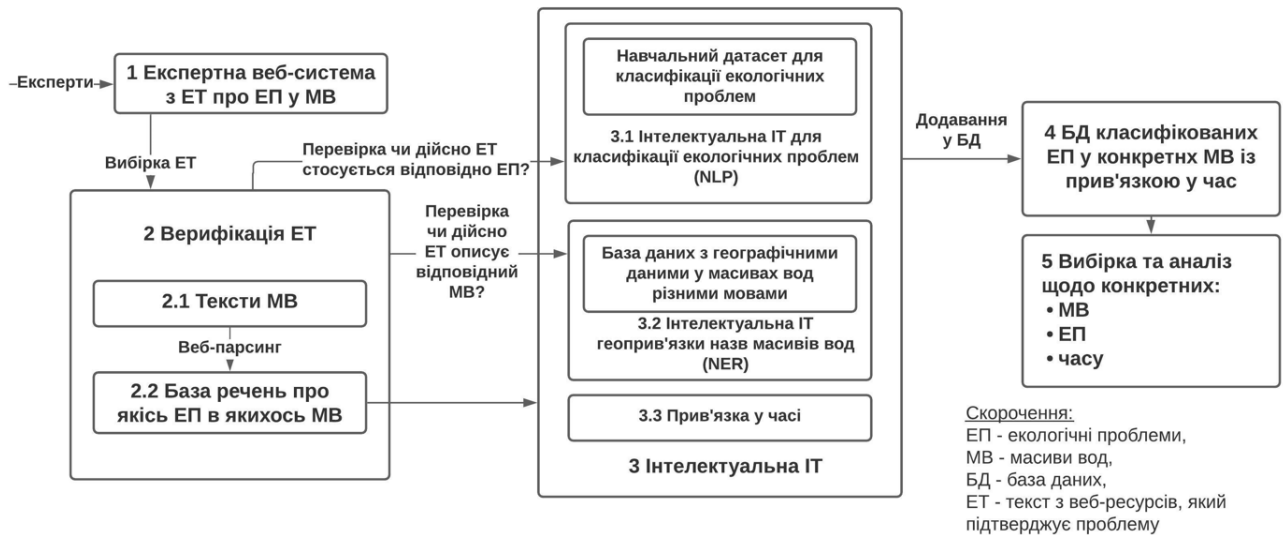


Рисунок 1. Складові інформаційної технології

## Форма (пілотна версія)

Форма для заповнення основних проблем масиву вод, які негативно впливають на його екологічний стан

@gmail.com (без совместного доступа)

[Сменить аккаунт](#)

\* **Обязательно**

Код масиву вод: \*

Наприклад, UA\_M5.4\_0013. Взяти з <http://geoportal.davr.gov.ua:81/#arraysSurfaceWatersSidebar>



Рисунок 2. Форма створення оцінки водойми

Створити оцінку Тяжилівка
✕

**Код об'єкту:**

**Опис:**

**Автор: \***

**Зробити оцінку анонімною:**

**Проблема: \***

**Ступінь впливу проблеми на екологічний стан води: \***

Високий
▾

**Тип джерела інформації: \***

Із соцмереж
▾

**Веб-посилання: \***

**Коментар:**

Закрити
Відправити

Рисунок 3. Експертна веб-система (пілотна версія)

2. Верифікація ЕТ. Верифікація текстів з веб-ресурсів складається з таких частин, як веб-парсинг, верифікація виду екологічних проблем і верифікації щодо геоприв'язки масивів вод.

Веб-парсинг. Верифікаційна технологія, створена з використанням штучного інтелекту, перевіряє чи дійсно вказаний експертом текстовий доказ кожної зазначеної ним проблеми підтверджує її. Для цього проводиться парсинг веб-ресурсу (рис. 4) та формується база речень для проведення такої перевірки. Верифікація виду екологічних проблем – це інтелектуальна технологія, підготовлена за участі Пасічнюка Д. В. та Радецького О. В. [3], створена з використанням навчального датасету, перевіряє чи дійсно вказаний експертом текстовий доказ стосується зазначеної ним екологічної проблеми.

Верифікація до геоприв'язки масивів вод – це інтелектуальна технологія, створена з використанням інтелектуальних технологій NER та NLP, перевіряє чи дійсно вказаний експертом текстовий доказ стосується зазначеного ним МВ. Випробування цієї технології охарактеризовано в роботі [4, 5], підготовленої за участі Крижановського Є. М. та Хмарука С. М.

База даних класифікованих екологічних проблем із прив'язкою у часі – формується систематизована БД класифікованих ЕП у конкретних МВ із прив'язкою у часі, в якій буде зручно робити пошук (рис. 5) та аналіз за різними критеріями.

Вибір та аналіз щодо конкретних масивів, екологічних проблем та часу – у спеціальній формі користувач вибирає конкретну МВ, ЕП та діапазон часу і отримує список текстових доказів таких ЕП. Також, можна буде робити ранжування ЕП в МВ і визначення пріоритетних заходів для усунення найбільш проблемних ЕП. Також, буде можливість автоматичного синтезу опис масиву (рис. 6) вод та його ЕП з використанням сучасних технологій штучного інтелекту (приклад наведено у [6]).

```
<li class="item-123">
  <a href="/vodni-resursy/protsedura-ukladannia-dohovoru-na-orendu-vodnykh-objektiv">
    Процедура укладання договору на оренду водних об'єктів
  </a>
</li>
<li class="item-124 parent">
  <a href="/vodni-resursy/vodohospodarska-obstanovka">
    Водогосподарська обстановка
  </a>
</li>
```

```
import numpy as np
import pandas as pd
import requests
from bs4 import BeautifulSoup as bs
r = requests.get('https://buvrpb.davr.gov.ua/vodni-resursy/hidrografichna-merezha')
soup = bs(r.content)
print(soup.prettify())
ad = soup.find('div', itemprop = "articleBody")
text_all = ad.getText()
text_all
```

'\nГідрографічна мережа\ха0басейну річки\ха0Південний\ха0Буг\ха0\nРічки\nУ басейні Південного Бугу налічується 6582 малих річок, загальною довжиною близько 20 тис. км, 11 середніх річок загальною довшиною понад 1,6 тис. км та 1 велика річка Південний Буг.\nPівденний Буг бере початок на Волино-Подільській височині поблизу с.Холодець Хмельницької області і впадає в Дніпро-Бузький лиман Чорного моря. Довжина річки становить 806 км, загальне падіння\ха0321 м, середній схил\ха0 водної поверхні 0,4 ‰.\n\ха0зера\nУ Кіровоградській, Одеській та Миколаївській областях зустрічаються невеликі озера. Їх загальна площа складає\ха0190 га.\n\ха0\nМоря\nПівденний Буг належить до басейну Чорного моря.\n\ха0\nлимани\nДніпро-Бузький лиман є мілководною затокою Чорного моря, що відокремлений від нього Кінбурнською косою. Сполучається лиман з морем Кінбурнською протокою, через яку і здійснюється водообмін. На сході лиман межує з гирлом Дніпра, на півночі поступово переходить у р. Південний Буг.\nЗа своїми розмірами Дніпро-Бузький лиман є найбільшим на Чорному морі. Його площа складає 800 км2, об'єм води – 3 км3. Лиман є мілководним, його середня глибина становить 3,5-4 м.\nДніпро-Бузький ли

Рисунок 4. Приклад роботи веб-парсинг (фрагмент вхідного html-тексту, фрагмент коду на Python та фрагмент отриманого результату)

|                |  |  |
|----------------|--|--|
| Об'єкт         | <input type="text" value="Південний Буг"/> | <input type="button" value="Оберіть об'єкт"/> <input type="button" value="✕"/>         |
| Характеристика | <input type="text" value="БСК5..МгО.дм3"/> | <input type="button" value="Оберіть характеристику"/> <input type="button" value="✕"/> |
| Тип            | <input type="text" value="Тип"/>           | <input type="button" value="Оберіть тип об'єкту"/> <input type="button" value="✕"/>    |
| Дата           |  |  |
| Від            | <input type="text" value="01.04.2018"/>    |  |
| До             | <input type="text" value="30.04.2018"/>    |  |
|                | <input type="button" value="Створити"/>    |  |

Рисунок 5. Форма пошуку

### Висновки

Розроблено інформаційну систему для синтезу та систематизування даних про стан та проблеми масивів вод р. Південний Буг з використанням сучасних технологій NLP та NER, та мов програмування C# та Python. Дана технологія дозволить сформувати чітку та систематизовану базу даних про стан водного басейну річки Південний Буг за допомогою інформації від місцевих експертів та інформаційних джерел із мережі Інтернет.



**ORIGINAL TEXT:**

Despite the similar volumes of discharged wastewater major part of pollutants comes with communal WWTPs. They bring 84% of organic pollution 86% of phosphate ions and 84% of mineral nitrogen 91% of ammonia nitrogen 87% nitrate nitrogen and 79% nitrite nitrogen. The input of the industry is between 7–21% and agriculture has the lowest impact on water bodies – 0–6%. Of the 92 urban areas only 51 localities (55%) have centralized collection of communal waste waters and their monitoring. Among the 2878 villages 6 of them (0.2%) have such a monitoring.

```
gpt_summary = []
for i in range(len(bodies)):
    GPT2_model = TransformerSummarizer(transformer_type="GPT2", transformer_model_key="gpt2-medium")
    gpt_summary.append(' '.join(GPT2_model(bodies[i], min_length=min_length_text)))
```

**GPT-2 Summarizing Result:**

Despite the similar volumes of discharged wastewater major part of pollutants comes with communal WWTPs. Of the 92 urban areas only 51 localities (55%) have centralized collection of communal waste waters and their monitoring.

Рисунок 6. Синтез опису масиву вод

**СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ**

1. Новий підхід до виявлення найбільш проблемних ділянок басейну річки за різними критеріями з використанням інформаційної аналітичної веб-системи [Електронний ресурс] / В. Б. Мокін А. М. Лучко О. М. Давидюк // Матеріали Всеукраїнської науково-практичної інтернет-конференції «Молодь в науці: дослідження, проблеми, перспективи», ВНТУ, Вінниця, 18-29 травня 2020 р. – Електрон. текст. дані. – 2020. – Режим доступу: <https://conferences.vntu.edu.ua/index.php/mn/mn2020/paper/viewFile/10502/8855>
2. Технологія побудови експертної інформаційної веб-системи виявлення та верифікації пріоритетних екологічних проблем у масивах вод басейну річки [Електронний ресурс] / В. Б. Мокін, А. М. Лучко, О. М. Давидюк, Т. Є. Вуж. – 2021. – Режим доступу до ресурсу: <https://visnyk.vntu.edu.ua/index.php/visnyk/article/view/2585>.
3. В.Б. Мокін, М.А. Гораш, Д.В. Пасічний, О.В. Радецький. Концепція інтелектуальної NLP технології для геоприв'язки та класифікації відкритої текстової інформації про масиви вод // Матеріали XV міжнародної конференції "Контроль і управління в складних системах (КУСС-2020)", м. Вінниця, 8-10 жовтня 2020 р. - Вінниця, 2020. [Електронний ресурс]. Режим доступу: <http://ir.lib.vntu.edu.ua/bitstream/handle/123456789/30607/KUSS%202020%20MHPR%20-%20NLP.pdf?sequence=1>
4. Інформаційна інтелектуальна технологія автоматизованої геоприв'язки екологічної текстової природно-мовної інформації [Електронний ресурс] / В. Б. Мокін, М. А. Гораш, Є. М. Крижановський, Т. Є. Вуж. – 2021. – Режим доступу до ресурсу: <https://praci.vntu.edu.ua/index.php/praci/article/view/624/587>.
5. Формування базової множини сутностей для автоматичної геоприв'язки текстової інформації про масиви вод [Електронний ресурс] / В. Б. Мокін, Є. М. Крижановський, С. М. Хмарук. – 2021. – Режим доступу до ресурсу: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2021/paper/view/11907/10420>.
6. Mokin V. B. Kaggle Notebook «NLP for WR : Summarizing using BERT, GPT2, XLNET» [Електронний ресурс]. – Режим доступу: <https://www.kaggle.com/code/vbmokin/nlp-for-wr-summarizing-using-bert-gpt2-xlnet>

**Мокін Віталій Борисович** — доктор технічних наук, професор кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, м. Вінниця, [vbmokin@vntu.edu.ua](mailto:vbmokin@vntu.edu.ua).

**Крижановський Євгеній Миколайович** — кандидат технічних наук, доцент кафедри системного аналізу та інформаційних технологій, Вінницький національний технічний університет, м. Вінниця, [kruzhan@gmail.com](mailto:kruzhan@gmail.com).

**Гораш Микола Анатолійович** — аспірант кафедри системного аналізу та інформаційних технологій Вінницький національний технічний університет, м. Вінниця, [kolia28011994@gmail.com](mailto:kolia28011994@gmail.com).

**Пасічний Д.В.** – студент кафедри кафедри системного аналізу та інформаційних технологій Вінницький національний технічний університет, м. Вінниця, [dimapasichniuk@gmail.com](mailto:dimapasichniuk@gmail.com).

**Радецький О.В.** – студент кафедри кафедри системного аналізу та інформаційних технологій Вінницький національний технічний університет, м. Вінниця, [s.radt02@gmail.com](mailto:s.radt02@gmail.com).

**Хмарук С.М.** – студент кафедри кафедри системного аналізу та інформаційних технологій Вінницький національний технічний університет, м. Вінниця, [khmarukserhii0@gmail.com](mailto:khmarukserhii0@gmail.com).

**Mokin V. B.** - Doctor of Technical Sciences, Professor, Head of the Chair of System Analysis and Information Technology, Vinnytsia National Technical University, Vinnytsia, [vbmokin@vntu.edu.ua](mailto:vbmokin@vntu.edu.ua).

**Kryzhanovsky Y. M.** - Candidate of Technical Sciences, Associate Professor of the Chair of System Analysis and Information Technology Vinnytsia National Technical University, Vinnytsia.

**Horash M. A.** - postgraduate student the Chair of System Analysis and Information Technology, Vinnytsia National Technical University, Vinnytsia, [kolia28011994@gmail.com](mailto:kolia28011994@gmail.com).

**Pashichniuk D.V.** - student of the Chair of System Analysis and Information Technology, Vinnytsia National Technical University, Vinnytsia [dimapasichniuk@gmail.com](mailto:dimapasichniuk@gmail.com).

**Radetskyi O.V.** - student the Chair of System Analysis and Information Technology, Vinnytsia National Technical University, Vinnytsia, [s.radt02@gmail.com](mailto:s.radt02@gmail.com).

**Khmaruk S.M** - student of the Chair of System Analysis and Information Technology, Vinnytsia National Technical University, Vinnytsia, [khmarukserhii0@gmail.com](mailto:khmarukserhii0@gmail.com).