

Information technologies in economics and environmental sciences



No 1 / 2024

Information Technologies in Economics and Environmental Sciences

Founder:

National University of Life and Environmental Sciences of Ukraine

Year of foundation: 2017

Published 2 times a year

Editors office address:

National University of Life and Environmental Sciences of Ukraine

03041, 15 Heroiv Oborony Str., Kyiv, Ukraine

E-mail: it-journal@nubip.edu.ua

<https://journals.nubip.edu.ua/index.php/Inf/en>

Інформаційні технології в економіці та природокористуванні

Засновник журналу:

Національний університет біоресурсів і природокористування України

Факультет інформаційних технологій

Рік заснування: 2017

Виходить 2 рази на рік

Адреса редакції:

Національний університет біоресурсів і природокористування України

03041, вул. Героїв Оборони, 15, м. Київ, Україна

E-mail: it-journal@nubip.edu.ua

<https://journals.nubip.edu.ua/index.php/Inf/uk>

EDITORIAL BOARD

Editor-in-Chief

Glazunova Olena

Doctor of Pedagogical Sciences, Professor, National University of Life and Environmental Sciences of Ukraine

Deputy Editor-in-Chief

Kravchenko Volodymyr

D.Sc. (Economics), Associate Professor, National University of Life and Environmental Sciences of Ukraine

Executive Secretary

Komar Kateryna

National University of Life and Environmental Sciences of Ukraine

Editorial Board Members

Bolbot Igor

D.Sc. (Technical Sciences), Professor, National University of Life and Environmental Sciences of Ukraine

Golub Bella

Ph.D. (Technical Sciences), Associate Professor, National University of Life and Environmental Sciences of Ukraine

Gusev Borys

Ph.D. (Technical Sciences), Associate Professor, National University of Life and Environmental Sciences of Ukraine

Holub Tetiana

Ph.D. (Technical Sciences), National University "Zaporizhzhia Polytechnic", Ukraine

Ivanchenko Yevheniia

Ph.D. (Technical Sciences), Professor, National Aviation University, Ukraine

Kovalenko Oleksiy

D.Sc. (Technical Sciences), Associate Professor, National University of Life and Environmental Sciences of Ukraine

Lobanchykova Nadiia

Ph.D. (Technical Sciences), Associate Professor, Zhytomyr Polytechnic State University, Ukraine

Mikriev Maksym

Ph.D. (Economics), Associate Professor, National University of Life and Environmental Sciences of Ukraine

Nikitenko Yevheniy

Ph.D. (Physics and Mathematics), Associate Professor, National University of Life and Environmental Sciences of Ukraine

Олійник Андрій

D.Sc. (Technical Sciences), Professor, National University "Zaporizhzhia Polytechnic", Ukraine

Sahun Andrii

Ph.D. (Technical Sciences), Associate Professor, National University of Life and Environmental Sciences of Ukraine

Svatko Vitaliy

Ph.D. (Technical Sciences), Associate Professor, National University of Life and Environmental Sciences of Ukraine

Semko Viktor

D.Sc. (Technical Sciences), Professor, National University of Life and Environmental Sciences of Ukraine

Skrupsky Stepan

Ph.D. (Technical Sciences), Associate Professor, National University "Zaporizhzhia Polytechnic", Ukraine

Smolij Viktorija	D.Sc. (Technical Sciences), Professor, National University of Life and Environmental Sciences of Ukraine
Khilenko Volodymyr	D.Sc. (Technical Sciences), Professor, National University of Life and Environmental Sciences of Ukraine
Shvydenko Mykhailo	Ph.D. (Economics), Associate Professor, National University of Life and Environmental Sciences of Ukraine
Shkarupylo Vadym	D.Sc. (Technical Sciences), Associate Professor, National University of Life and Environmental Sciences of Ukraine

International Members of the Editorial Board

Akhmetov Bakhytzhn	D.Sc. (Technical Sciences), Professor, Turan University, Kazakhstan, Almaty
Jamil Abedalrahim Jamil Alsayaydeh	Ph.D., Universiti Teknikal Malaysia Melaka (UTeM)
Zherlitsyn Dmyro	D.Sc. (Economics), Professor, Institute of Entrepreneurship, University of National and World Economy, Bulgaria, Sofia
Mikulecky Peter	Ph.D., Professor, RNDr., University of Hradec Kralove, Czech Republic, Hradec Kralove

РЕДАКЦІЙНА КОЛЕГІЯ

Головний редактор

Глазунова Олена Григорівна	доктор педагогічних наук, професор, Національний університет біоресурсів і природокористування України
-----------------------------------	--

Заступник головного редактора

Кравченко Володимир Миколайович	доктор економічних наук, доцент, Національний університет біоресурсів і природокористування України
--	---

Відповідальний секретар

Комар Катерина Вячеславівна	Національний університет біоресурсів і природокористування України
------------------------------------	--

Національні члени редколегії

Болбот Ігор Михайлович	доктор технічних наук, професор, Національний університет біоресурсів і природокористування України
Голуб Белла Львівна	кандидат технічних наук, доцент, Національний університет біоресурсів і природокористування України
Голуб Тетяна Василівна	кандидат технічних наук, НУ «Запорізька політехніка»

Гусєв Борис Семенович	кандидат технічних наук, доцент, Національний університет біоресурсів і природокористування України
Іванченко Євгенія Вікторівна	кандидат технічних наук, професор, Національний авіаційний університет
Коваленко Олексій Єпіфанович	доктор технічних наук, доцент, Національний університет біоресурсів і природокористування України
Лобанчикова Надія Миколаївна	кандидат технічних наук, доцент, ДУ «Житомирська політехніка»
Мокрієв Максим Володимирович	кандидат економічних наук, доцент, Національний університет біоресурсів і природокористування України
Нікітенко Євгеній Васильович	кандидат фізико-математичних наук, доцент, Національний університет біоресурсів і природокористування України
Олійник Андрій Олександрович	доктор технічних наук, професор, НУ «Запорізька політехніка»
Сагун Андрій Вікторович	кандидат технічних наук, доцент Національний університет біоресурсів і природокористування України
Сватко Віталій Володимирович	кандидат технічних наук, доцент, Національний університет біоресурсів і природокористування України
Семко Віктор Володимирович	доктор технічних наук, професор, Національний університет біоресурсів і природокористування України
Скрупський Степан Юрійович	кандидат технічних наук, доцент, НУ «Запорізька політехніка»
Смолій Вікторія Миколаївна	доктор технічних наук, професор, Національний університет біоресурсів і природокористування України
Хиленко Володимир Васильович	доктор технічних наук, професор, Національний університет біоресурсів і природокористування України
Швиденко Михайло Зіновійович	кандидат економічних наук, доцент, Національний університет біоресурсів і природокористування України
Шкарупило Вадим Вікторович	доктор технічних наук, доцент, Національний університет біоресурсів і природокористування України

Міжнародні члени редколегії

Ахметов Бахитжан Сражатдінович	доктор технічних наук, професор, університет Туран (м. Алмати, Казахстан)
Жаміль Абедалярахім Жаміль Альсяядех	PhD, Технічний університет Малайзії, Мелак (UTeM)
Жерліцин Дмитро Михайлович	доктор економічних наук, професор, Інститут підприємництва Університету національної та світової економіки (м. Софія, Болгарія)
Мікулецький Пітер	PhD, професор, RNDr., Університет Градець-Кралове (м. Градець-Кралове, Чехія)

CONTESTS

Лендел Марина Іванівна, Голуб Белла Львівна ВИКОРИСТАННЯ ТЕХНОЛОГІЙ OLAP ТА DATA MINING В СИСТЕМАХ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ У ТЕПЛИЧНОМУ ГОСПОДАРСТВІ.....	8
Хоменко Андрій Олегович, Кириченко Віктор Вікторович ОСОБЛИВОСТІ РОБОТИ АЛГОРИТМІВ СОРТУВАННЯ В PYTHON	15
Густера Олег Михайлович, Ніколаєнко Дмитро Володимирович ЗГЛАДЖУВАННЯ НЕПОВНИХ РЯДІВ ДАНИХ СТАНЦІЙ ЕКОЛОГІЧНОГО МОНІТОРИНГУ З ВИКОРИСТАННЯМ ПРОГНОЗНИХ МОДЕЛЕЙ	21
Сагун Андрій Вікторович, Панаско Олена Миколаївна СИСТЕМА РОЗПІЗНАВАННЯ ОБРАЗІВ НА БАЗІ НЕЙРОННОЇ МЕРЕЖІ З ГЛИБОКИМ НАВЧАННЯМ	31
Nikitenko Yevheniy, Ryndych Yevhen, Hoida Ivan DETECTION OF COMMUNITIES IN SOCIAL NETWORKS.....	38
Смолій Вікторія Миколаївна, Смолій Натан Вікторович, Янцевич Антон Олександрович, Іскоростенський Олексій Олександрович ПРОЄКТУВАННЯ КОМП'ЮТЕРНОЇ СИСТЕМИ АНАЛІЗУ СТАНУ РОСЛИН.....	44
Nazarenko Volodymyr, Kasatkin Dmytro INTERACTION AND FEEDBACK LOOPS IN USER INTERFACE FOR DIGITAL ENGINEERING DESIGN.....	53
Ляхно Валерій Анатолійович, Касаткін Дмитро Юрійович ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ РОЗПІЗНАВАННЯ КІБЕРЗАГРОЗ, ЯК СКЛАДОВА СИСТЕМ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ НА ТРАНСПОРТІ	62
Касаткін Дмитро Юрійович, Волошин Семен Михайлович, Гусев Борис Семенович, Матієвський Володимир Валерійович АЛГОРИТМИ РОЗРОБКИ БАЗИ ЗНАНЬ ДЛЯ ВДОСКОНАЛЕННЯ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ПРИ ВИРІШЕННІ ПРОБЛЕМ КІБЕРБЕЗПЕКИ.....	69

ЗМІСТ

Лендел Марина Іванівна, Голуб Белла Львівна ВИКОРИСТАННЯ ТЕХНОЛОГІЙ OLAP ТА DATA MINING В СИСТЕМАХ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ У ТЕПЛИЧНОМУ ГОСПОДАРСТВІ.....	8
Хоменко Андрій Олегович, Кириченко Віктор Вікторович ОСОБЛИВОСТІ РОБОТИ АЛГОРИТМІВ СОРТУВАННЯ В PYTHON	15
Густера Олег Михайлович, Ніколаєнко Дмитро Володимирович ЗГЛАДЖУВАННЯ НЕПОВНИХ РЯДІВ ДАНИХ СТАНЦІЙ ЕКОЛОГІЧНОГО МОНІТОРИНГУ З ВИКОРИСТАННЯМ ПРОГНОЗНИХ МОДЕЛЕЙ	21
Сагун Андрій Вікторович, Панаско Олена Миколаївна СИСТЕМА РОЗПІЗНАВАННЯ ОБРАЗІВ НА БАЗІ НЕЙРОННОЇ МЕРЕЖІ З ГЛИБОКИМ НАВЧАННЯМ	31
Nikitenko Yevheniy, Ryndych Yevhen, Hoida Ivan DETECTION OF COMMUNITIES IN SOCIAL NETWORKS.....	38
Смолій Вікторія Миколаївна, Смолій Натан Вікторович, Янцевич Антон Олександрович, Іскоростенський Олексій Олександрович ПРОЄКТУВАННЯ КОМП'ЮТЕРНОЇ СИСТЕМИ АНАЛІЗУ СТАНУ РОСЛИН.....	44
Nazarenko Volodymyr, Kasatkin Dmytro INTERACTION AND FEEDBACK LOOPS IN USER INTERFACE FOR DIGITAL ENGINEERING DESIGN.....	53
Ляхно Валерій Анатолійович, Касаткін Дмитро Юрійович ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ РОЗПІЗНАВАННЯ КІБЕРЗАГРОЗ, ЯК СКЛАДОВА СИСТЕМ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ НА ТРАНСПОРТІ	62
Касаткін Дмитро Юрійович, Волошин Семен Михайлович, Гусев Борис Семенович, Матієвський Володимир Валерійович АЛГОРИТМИ РОЗРОБКИ БАЗИ ЗНАНЬ ДЛЯ ВДОСКОНАЛЕННЯ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ПРИ ВИРІШЕННІ ПРОБЛЕМ КІБЕРБЕЗПЕКИ.....	69

УДК: 004.9:631.544.4

Лендел Марина Іванівна*асистентка кафедри комп'ютерних наук,**Національний університет біоресурсів і природокористування України*E-mail: marynalendiel@gmail.com**Голуб Белла Львівна***кандидат технічних наук, доцент, завідувач кафедри комп'ютерних наук,**Національний університет біоресурсів і природокористування України*ORCID: <https://orcid.org/0000-0002-1256-6138>E-mail: bella.golub55@gmail.com**ВИКОРИСТАННЯ ТЕХНОЛОГІЙ OLAP ТА DATA MINING В СИСТЕМАХ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ У ТЕПЛИЧНОМУ ГОСПОДАРСТВІ**

***Анотація.** У сучасному світі технології відіграють ключову роль у багатьох напрямках життя, включаючи сільське господарство. Вирощування овочів та фруктів у спорудах закритого ґрунту стає все більш популярним і вимагає контролю над умовами та ресурсами для збільшення врожайності, проте із зростанням обсягів даних, які збираються в системах моніторингу, стає важче ефективно аналізувати зібрану інформацію. Дана стаття розглядає використання технологій OLAP та Data Mining для підвищення ефективності вирощування овочів та фруктів у спорудах закритого ґрунту. Метою дослідження є визначення доцільності та ефективності застосування цих технологій для аналізу великих обсягів даних, зібраних під час вирощування овочів та фруктів. Автори дослідження аналізують, як OLAP та Data Mining можуть надати корисну інформацію для прийняття рішень щодо оптимізації процесів вирощування для підвищення врожайності.*

***Ключові слова:** система підтримки прийняття рішень, OLAP, Data Mining, сховище даних, багатовимірний куб.*

Вступ. Сучасний світ повністю залежить від технологій, які стрімко розвиваються. Вони займають ключове місце майже у всіх напрямках нашої діяльності та допомагають у повсякденному житті майже на кожному кроці.

Під час вирощування важливим етапом є саме вивчення та аналіз всіх умов, які необхідні для нормального росту й розвитку рослини. У процесі росту рослин важливо враховувати оптимальні показники мікроклімату теплиці для підвищення ефективності використання ресурсів для вирощування врожаю. Дивлячись на згадані особливості корисним є використання програмних засобів для моніторингу, збереження та аналізу показників, які важливі для підвищення ефективності вирощування.

На основі всіх наявних даних у системі виробник може проводити аналіз всіх ключових показників, їхні зміни та вплив у розрізі часу та приймати відповідні рішення для свого підприємства. Проте створені системи з часом розширюються, а відповідно інформація у них теж, тому стає незручно та неефективно аналізувати попередньо внесені дані. У такому випадку існує необхідність у створенні системи, яка буде проводити аналіз показників на основі накопичених даних.

Саме система підтримки прийняття рішень (СППР) допомагає аналізувати накопичені дані, а також приймати рішення, які вимагають оцінки, рішучості та послідовності дій. Така інформаційна система допомагає керівництву підприємств середнього і високого рівня організації, аналізуючи великі обсяги неструктурованих даних і накопичуючи дані, у вирішенні певної проблеми та допомогти у прийнятті рішення.

На господарства СППР використовується як для одного об'єкту, так і для мережі господарств, що допоможе у кращому вивченні даних та відповідно подальшому прогнозуванні необхідних явищ, інформації про наявні та можливі факти, а також для представлення інформації користувачам у зрозумілій формі. Такі системи допомагають

своєчасно прийняти керуюче рішення для покращення результатів виробництва та відповідно збільшення прибутку.

Аналіз останніх досліджень та публікацій. У процесі аналізу предметної області було проаналізовано наукові роботи, які пов'язані з вирішенням питання аналізу роботи тепличного господарства.

Автори Ченг Ванг, Лілі Венг, Пінг Донг, Ксяо Кяо у своїй роботі «Дослідження системи підтримки прийняття рішень (DSS) для теплиць на основі видобутку даних» розглянули систему підтримки прийняття рішень для теплиці, побудованої зі сховища даних та технології видобутку дат. У системі було створено сховище даних з метою диверсифікації пам'яті, використання онлайн аналітичної обробки та видобутку даних збагачує базу знань новою інформацією про сільське господарство. Дана робота показує структуру системи підтримки прийняття рішень загалом: використання технології аналітичної обробки OLAP, яка допомагає аналітично обробляти дані, які зберігаються у базі знань. Також пояснює структуру модуля видобутку даних, а саме побудову окремих модулів, кожен з яких відповідає за моделювання та прогнозування, пошуку схожості, побудови асоціативних правил тощо [7].

В іншій роботі «Експертна система управління тепличним виробництвом» автори Йонгуанг Ху, Джиганг Ванг та Пінгпінг Лі розповіли про важливість використання теплиці та правильного управління всіма внутрішніми процесами і це підштовхує до розробки експертних систем.

У роботі описано та зображено функціональну структуру системи, яка складається з декількох підсистем. Важливу частину займає база знань, яка містить словник правил. Також було описано систему збору даних у режимі реального часу.

У результаті аналізу було описано короткі функціональні характеристики системи:

- моніторинг мікроклімату теплиці у режимі реального часу;
- оновлення інформаційної бази довкілля для експертної системи;
- виведення та відображення інформації з текстом, графічною таблицею або роздруківкою;
- тривожний сигнал у разі екстремальних ситуацій або зниження врожайності.

Крім програмного у роботі було розглянуто апаратне забезпечення та впровадження системи та її результати роботи, де показано приклад діагностики захворювання на основі введених ознак, а також аналізу оптимальних умов мікроклімату та поливу [8].

Мета дослідження. Метою дослідження є визначення доцільності застосування технологій OLAP і Data Mining для підвищення ефективності вирощування овочів та фруктів у спорудах закритого ґрунту.

Матеріали і методи дослідження. У процесі дослідження були використані такі методи:

1. Технології OLAP для накопичення даних у більш зручній структурі: багатомірний куб, який ефективний для проведення аналізу у розрізі різних параметрів, формування звітної інформації тощо.
2. Технології Data Mining для знаходження зв'язків між наповненими даними, які відображають результати роботи тепличного господарства.

Результати дослідження та їх обговорення. Для робочої станції оператора теплиці було розроблено підсистему моніторингу, яка працює з даними, які зберігаються в оперативній базі даних, а також апаратне забезпечення з використанням датчиків температури й вологості, які підключені до мікроконтролера Arduino. Для вимірювання показників використано датчик температури і вологості повітря DHT11. Підсистему моніторингу реалізовано з використанням мови програмування Python, де забезпечується передача даних про отримані показники температури й вологості до оперативної бази даних. На рис. 1 зображена схема оперативної бази даних, де зберігаються показники мікроклімату у теплицях, врожайність, оптимальні показники тощо.

На основі отриманого оперативного джерела даних для проведення аналізу використовується сховище даних (СД), яке дозволить проводити аналіз у різних розрізах. Це забезпечується наявністю декількох вимірів, які являють собою сукупність довідкової

інформації про вимірювану подію. У контексті сховищ даних подіями являються факти, які описують результати певного бізнес-процесу. Структура сховища даних зображена на рис. 2.

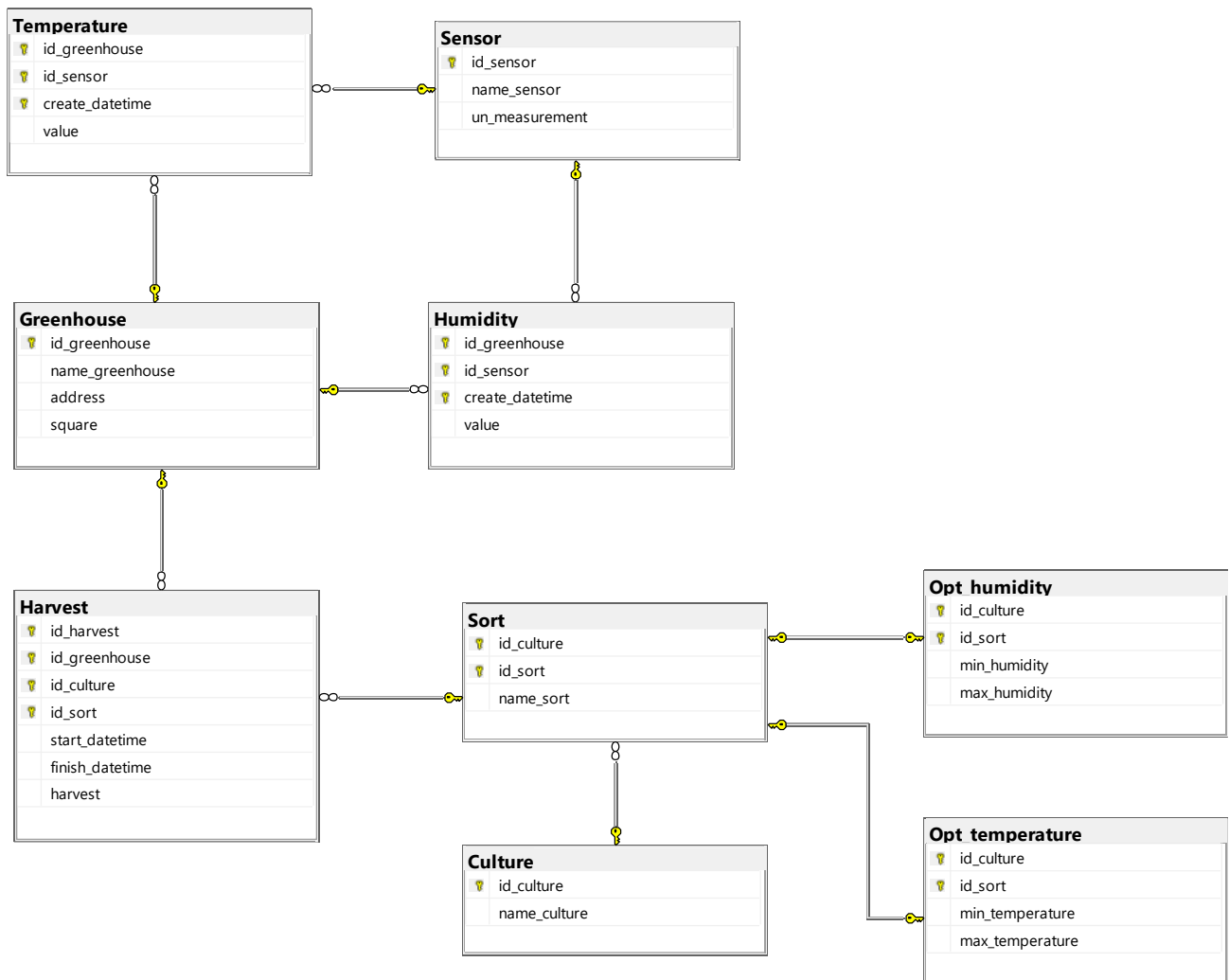


Рисунок 1 – Схема оперативної бази даних



Рисунок 2 – Вітрина тепличного господарства

Розроблене сховище даних було використано для розгортання багатовимірного кубу, використовуючи службу SQL Server Analysis Services (SSAS). Саме для аналізу даних було реалізовано наповнення сховища даних, яке відбувається на основі даних з оперативної бази даних, зображеної на рис. 1. Процес передачі даних був реалізований за допомогою служби SQL Server Integration Services.

У процесі аналізу отриманих даних у СД було реалізовано розрахунок ключового показника ефективності (KPI), який був реалізований з використанням мови запитів MDX для доступу до багатовимірних структур даних (рис. 3).

Отобразить структуру	Значение	Цель	Состояние
KPI_Sort1_Opt_Avg_Humidity	66.00%	67.5 %	
KPI_Sort1_Opt_Avg_Temp	21.11°C	22.0 °C	
KPI_Sort4_Opt_Avg_Humidity	75.22%	77.5 %	
KPI_Sort4_Opt_Avg_Temp	19.89°C	22.5 °C	

Рисунок 3 – Результати розрахунку ключового показника ефективності

Для проведення інтелектуального аналізу даних було використано SQL Server Data Tools-Business Intelligence(SSDT-BI), який містить набір технологій для бізнес аналізу: створення моделей даних Analysis Services (AS), пакетів Integration Services (IS) та звітів Reporting Services (RS).

Використовуючи розгорнутий куб було побудовано графік зміни врожайності з використанням алгоритму часових рядів, який зображено на рис. 4.

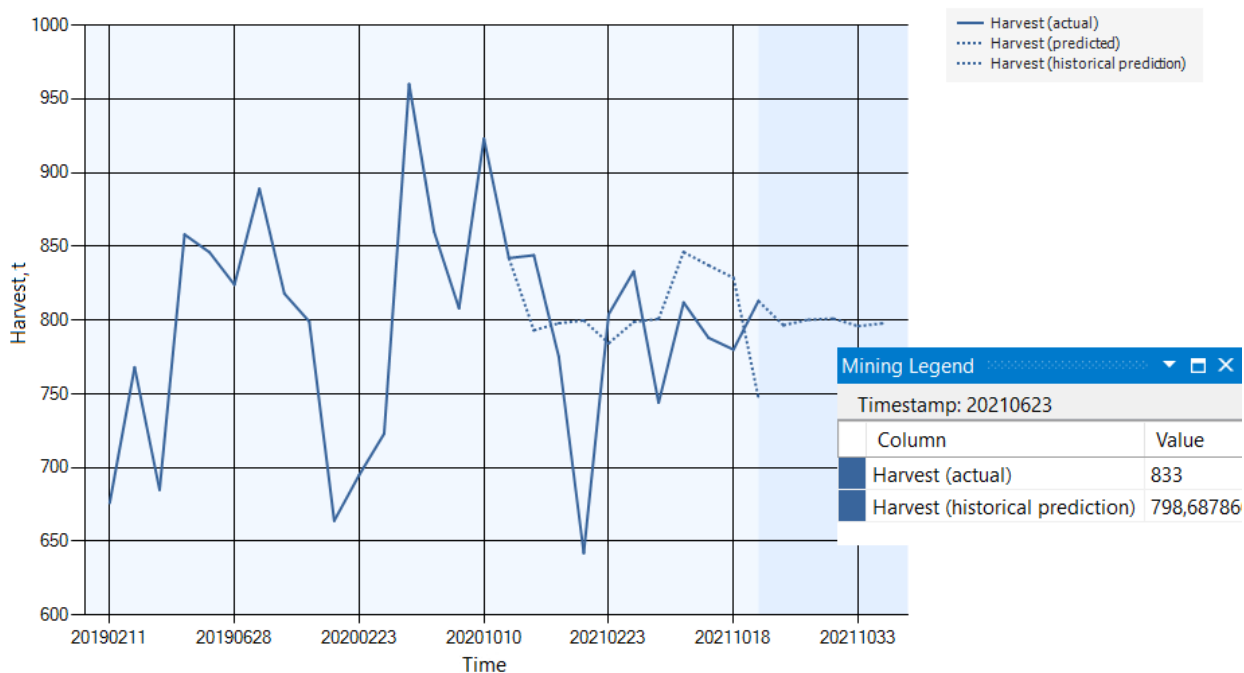


Рисунок 4 – Сформований графік з використання алгоритму часових рядів

У процесі аналізу за допомогою часових рядів було проаналізовано зміну врожайності. Часові ряди дозволяють прогнозувати як буде далі розвиватися ситуація. Тому на основі отриманого графіку можна побачити, що починаючи з початку 2021 року сумарна врожайність у теплицях зменшилась. Наприклад взимку 2021 врожайність досягла найнижчого значення, прогнозоване значення відрізняється від фактичного.

Також на основі розгорнутої структури було побудовано мережу залежностей назви культури та сорту від заданих діапазонів температури та вологості (рис. 5).

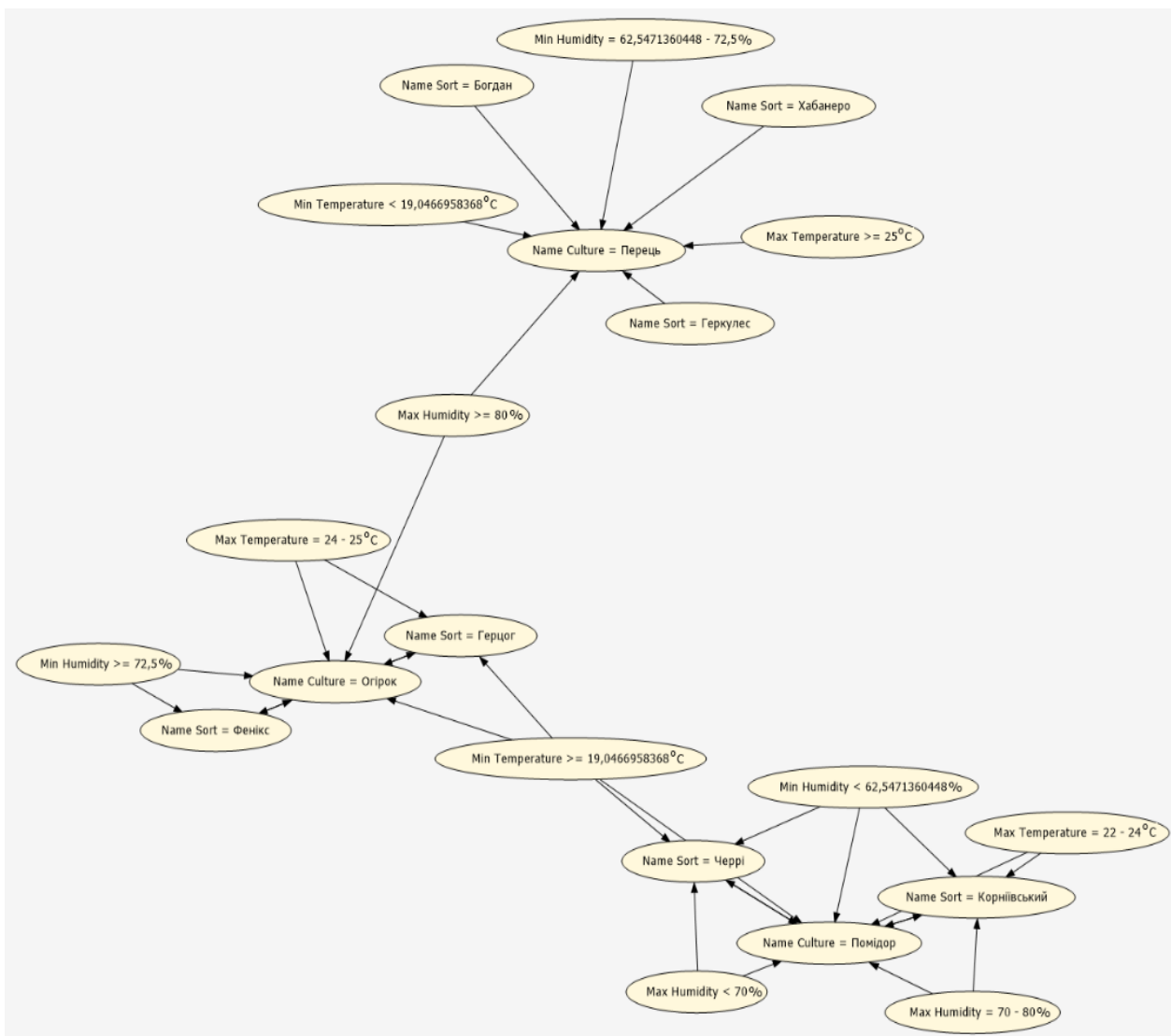


Рисунок 5 – Мережа залежностей на основі реалізованого алгоритму асоціативних правил

З наведеної схеми асоціативних правил можна визначити найбільш сприятливі умови для вирощування як окремих сортів та культур, так і про можливість комбінування сортів різних культур для вирощування в єдиному просторі. Виокремимо значущі особливості:

- сорти однієї культури мають доволі схожі вимоги до середовища, що робить процес вирощування більш зрозумілим, адже дає підстави для їх вирощування в єдиному просторі;
- перці та огірки мають однакові вимоги до максимальних та мінімальних показників вологості та температури, що дає можливість комбінувати їх сорти під час вирощування в єдиному просторі;
- щодо помідорів, то їх мінімальне значення температури співпадає із показником для огірків, але через більшу чутливість сортів цієї культури, найкращим рішенням буде вирощування їх в окремому просторі.

Розгорнувши структуру аналізу алгоритмом кластеризації було отримано модель, наведену на рис. 6.

За допомогою методу кластеризації були отримані зв'язки між різними кластерами, схожі за ознаками, культурою і умовами вирощування. У результаті розгорнутої структури можна зробити висновок, що дані є однорідними та впорядкованими. Особливо це помітно на схемі мережі зв'язків кластери майже не пов'язані один з одним.

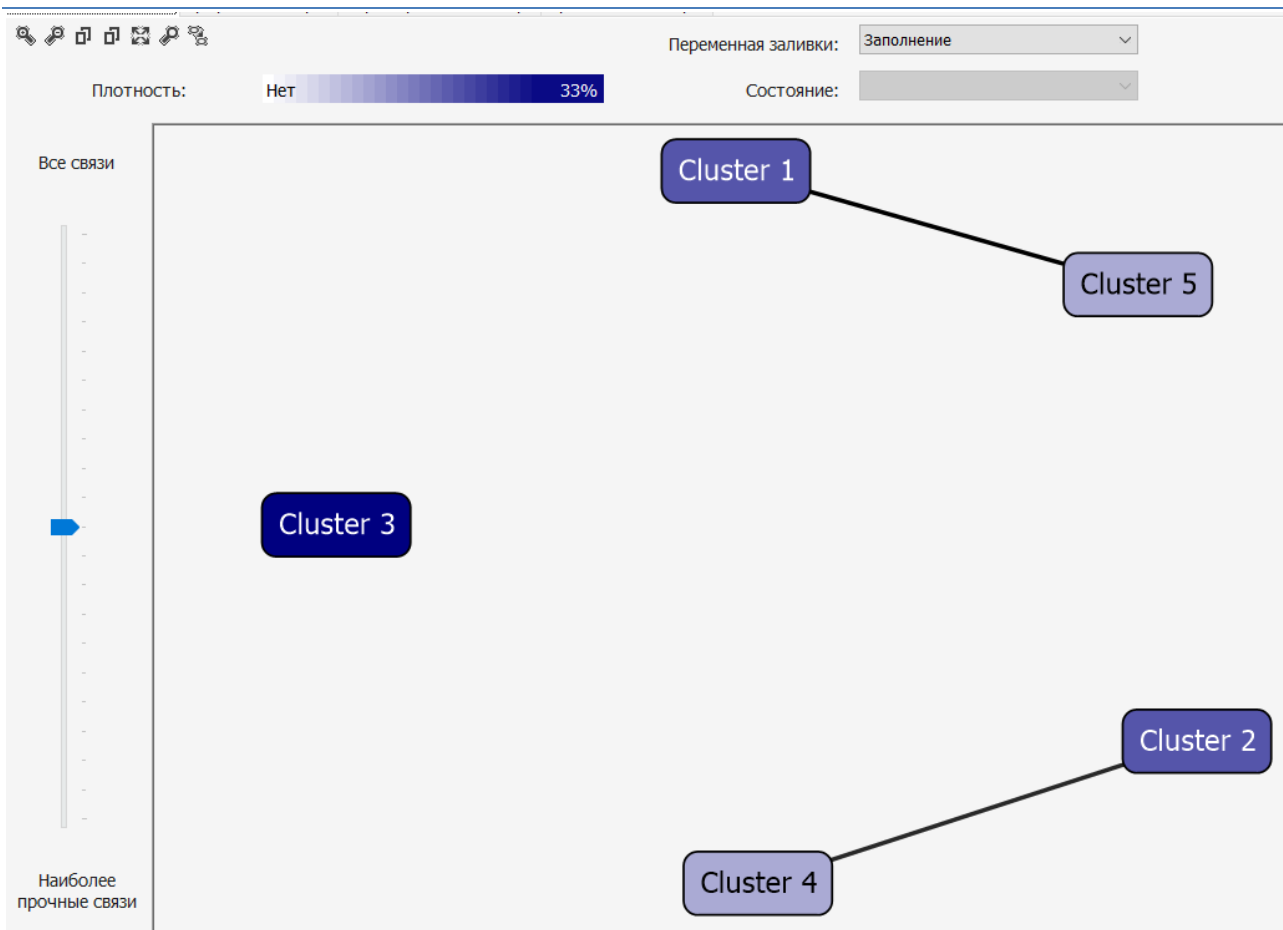


Рисунок 6 – Кластеризація наявних фактів у сховищі даних

Висновки і перспективи. Під час розробки системи підтримки прийняття рішень було представлено застосування технологій OLAP і Data Mining на прикладі тепличного господарства. Обидві технології доповнюють одна одну і в той час як DM знаходить закономірності на основі відомих знань, то OLAP аналізує дані в режимі реального часу та може стверджувати або спростовувати гіпотези, наведені у процесі інтелектуального аналізу. Отримані знання корисні у процесі формування керуючих рішень тепличного господарства, що дозволить спрямувати стратегію щодо управління окремими бізнес-процесами таким чином, щоб підвищити врожайність у теплицях та відповідно прибутковості господарства загалом. Проте застосування не обмежується лише двома параметрами, так як систему можна розширювати застосовуючи інші показники, які допоможуть розширити пошук залежностей. До прикладу це може бути освітленість, рівень CO₂ тощо. Також предметною областю може виступати не тільки тепличне господарство, дана система також може поширюватись і на інші сфери.

Список використаних джерел

1. Wang, C., Wang, L., Dong, P., Qiao, X. (2008). Research of Decision Support System (DSS) for Greenhouse Based on Data Mining. In: Li, D. (eds) Computer And Computing Technologies In Agriculture, Volume II. CCTA 2007. The International Federation for Information Processing, vol 259. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-77253-0_40.
2. Hu, Y., Wang, J., & Li, P. (2010). Expert System for Greenhouse Production Management. In Expert Systems. InTech. <https://doi.org/10.5772/7080>.
3. Microsoft. (2023). Data mining concepts. Analysis Services. <https://docs.microsoft.com/en-us/analysis-services/data-mining/data-mining-concepts?view=asallproducts-allversions>

4. TutorialsPoint. (2022). What are the basic concepts of data mining. <https://www.tutorialspoint.com/what-are-the-basic-concepts-of-data-mining>
5. Microsoft. (2024). Overview of online analytical processing (OLAP). Microsoft Support. <https://support.microsoft.com/en-us/office/overview-of-online-analytical-processing-olap-15d2cdde-f70b-4277-b009-ed732b75fdd6>
6. Microsoft. (2024). Data source views in multidimensional models. Analysis Services. <https://docs.microsoft.com/en-us/analysis-services/multidimensional-models/data-source-views-in-multidimensional-models?view=asallproducts-allversions>
7. Microsoft. (2024). Key performance indicators (KPIs) in multidimensional models. Analysis Services. <https://docs.microsoft.com/en-us/analysis-services/multidimensional-models/key-performance-indicators-kpis-in-multidimensional-models?view=asallproducts-allversions>
8. Greenhouse Grower. (2011). Monitoring the greenhouse environment. <https://www.greenhousegrower.com/production/plant-culture/special-series/monitoring-the-greenhouse-environment/>

Lendiel Maryna

*Assistant at the Department of Computer Science,
National University of Life and Environmental Sciences of Ukraine*
E-mail: marynalendel@gmail.com

Golub Bella

*Candidate of Engineering Sciences, Head of the Department of Computer Science,
National University of Life and Environmental Sciences of Ukraine*
ORCID: <https://orcid.org/0000-0002-1256-6138>
E-mail: bella.golub55@gmail.com

USE OF OLAP AND DATA MINING TECHNOLOGIES IN DECISION SUPPORT SYSTEMS IN GREENHOUSE FARMING

Abstract. *In the modern world, technologies play a key role in many areas of life, including agriculture. Growing vegetables and fruits in closed ground structures is becoming increasingly popular and requires control over conditions and resources to increase yields, however, with the increase in the amount of data collected in monitoring systems, it becomes more difficult to effectively analyze the collected information. This article considers the use of OLAP and Data Mining technologies to increase the efficiency of growing vegetables and fruits in closed ground structures. The purpose of the study is to determine the feasibility and effectiveness of using these technologies to analyze large amounts of data collected during the cultivation of vegetables and fruits. The authors of the study analyze how OLAP and Data Mining can provide useful information for making decisions on optimizing growing processes to increase yields.*

Keywords: *decision support system, OLAP, Data Mining, data warehouse, multidimensional cube.*

УДК 004.62

Хоменко Андрій Олегович*аспірант кафедри комп'ютерних наук,**Національний університет біоресурсів і природокористування України*E-mail: andmayster@gmail.com**Кириченко Віктор Вікторович***кандидат фізико-математичних наук, доцент, доцент кафедри комп'ютерних наук,**Національний університет біоресурсів і природокористування України*ORCID: <https://orcid.org/0009-0001-0575-8684>E-mail: v.kyrychenko@nubip.edu.ua**ОСОБЛИВОСТІ РОБОТИ АЛГОРИТМІВ СОРТУВАННЯ В PYTHON**

***Анотація.** Стаття присвячена аналізу різних алгоритмів сортування з використанням мови програмування Python та за допомогою алгоритмів сортування на мові програмування Cython. В дослідженні порівнюються класичні методи сортування, такі як сортування бульбашкою, вставками та швидке сортування, з метою визначення їх ефективності для великих наборів даних. Особливу увагу приділено проблемам локалізації при сортуванні рядків у неанглійських мовах, зокрема, застосуванню спеціалізованого словника для коректної обробки українського алфавіту. Представлено методики вимірювання продуктивності та візуалізації результатів у вигляді графіків, що дозволяє глибше оцінити масштабування кожного з алгоритмів в залежності від обсягу даних.*

***Ключові слова:** Python, Cython, алгоритми сортування, великі набори даних, локалізація, український алфавіт, продуктивність сортування, візуалізація даних.*

Вступ. Алгоритми сортування відіграють важливу роль у програмуванні та обробці даних, дозволяючи ефективно організувати та аналізувати інформацію. Однак, сортування даних, що містять кирилицю, може викликати певні труднощі через особливості локалізації та кодування символів. Ця стаття розглядає проблеми та особливості сортування рядкових даних на українській мові в Python, аналізує різні підходи та методи сортування, та пропонує ефективні рішення для оптимізації цього процесу.

Проблематика сортування кирилиці. Сортування текстових даних, які містять символи кирилиці, може створити виклики, пов'язані з локалізацією та упорядкуванням символів. Проблема полягає в тому, що стандартні методи сортування в Python не завжди адекватно обробляють символи кирилиці, особливо коли мова йде про українські символи, такі як 'Г', 'Є', які мають особливе місце в алфавіті. Це може призвести до неправильного порядку сортування, коли дані відсортовані за алфавітом.

Метою дослідження є оцінка продуктивності різних алгоритмів сортування під управлінням Python та Cython та надання рекомендацій щодо їх вибору для конкретних обчислювальних завдань. Стаття також спрямована на вирішення проблем, пов'язаних з локалізацією сортування рядків, з акцентом на українську мову, яка представляє певні виклики через відмінності в алфавіті порівняно з англійською мовою.

Аналіз останніх досліджень та публікацій. Сучасний стан досліджень у сфері алгоритмів сортування та їх оптимізації свідчить про значний інтерес науковців та розробників до підвищення ефективності обробки даних. Розробка мови програмування Python, зокрема її розширення Cython, забезпечує нові можливості для оптимізації вже існуючих алгоритмів та створення нових, більш ефективних методів обробки великих обсягів інформації. Значну увагу приділено також адаптації алгоритмів під особливості конкретних мов, що є важливим для міжнародних програмних продуктів.

У роботах Ван Россума Г. та Дрейка Ф. Л. [1], а також Орлова С. А. [3] наголошується на важливості забезпечення високої продуктивності програм на Python шляхом ефективного

використання існуючих алгоритмів сортування та розробки нових методів. Дослідження Бехтерева А. М. [2] підкреслює роль Cython у оптимізації Python коду, зокрема для алгоритмів сортування, демонструючи значне збільшення продуктивності.

Специфічні аспекти локалізації алгоритмів сортування, особливо для мов з нелатинською абеткою, розглянуті у працях Кормена Т. Х. та ін. [5], де аналізуються основні труднощі та пропонуються методи їх вирішення. Важливість візуалізації даних для аналізу ефективності алгоритмів підкреслена у роботах Маккінні У. [7], що вказує на необхідність застосування сучасних інструментів для графічного представлення результатів досліджень.

Загалом, аналіз останніх досліджень та публікацій показує тенденцію до пошуку нових методів оптимізації алгоритмів сортування, їх адаптації під специфіку різних мов та культур, а також розвитку інструментів для аналізу та візуалізації даних. Основними напрямками подальших досліджень у цій сфері є розвиток алгоритмів машинного навчання та штучного інтелекту для обробки великих даних, а також створення нових інструментів для ефективної взаємодії з даними в різних мовних середовищах.

Матеріали і методи дослідження. Для тестування алгоритмів сортування та демонстрації проблеми локалізації важливо мати реалістичний набір тестових даних. Один із способів генерації таких даних – використання бібліотеки Faker, яка дозволяє створювати велику кількість різноманітних даних, включаючи імена, прізвища та інше, для імітації реальних даних.

Для додавання українських 'по-батькові' до наших тестових даних, ми розширимо бібліотеку Faker власним провайдером:

```
from faker import Faker
from faker.providers import BaseProvider
import random
class ProviderUk(BaseProvider):
    def patronymic(self):
        patronymics = ['Олександрович', 'Максимович', 'Іванович', 'Петрович',
'Sергійович']
        return random.choice(patronymics)
fake = Faker('uk_UA')
fake.add_provider(ProviderUk)
```

Використовувати його можна наступним чином:

```
def generate_and_save_data_to_csv(filename, num_records):
    with open(filename, mode='w', newline='', encoding='utf-8') as file:
        writer = csv.writer(file)
        writer.writerow(['Name', 'Patronymic', 'Surname'])
        for _ in range(num_records):
            name = fake.first_name()
            patronymic = fake.patronymic()
            surname = fake.last_name()
            writer.writerow([name, patronymic, surname])
```

В цьому прикладі ми реалізуємо функцію для створення csv файлу та генерацію даних за допомогою Faker, в тому числі й 'по-батькові' яке відсутнє у стандартній бібліотеці Faker.

Методи сортування. Розглянемо декілька підходів до сортування даних в Python, кожен з яких має свої особливості та області застосування. Це допоможе зрозуміти, як можна оптимізувати процес сортування, особливо при роботі з великими обсягами даних або коли потрібно врахувати специфіку локалізації.

• **Стандартне сортування Python.** Метод `sorted()` в Python є універсальним інструментом для сортування списків. Він застосовує алгоритм сортування Timsort, який є гібридом між Merge Sort та Insertion Sort. Timsort адаптований до реальних даних, які часто містять вже відсортовані секвенції, забезпечуючи високу продуктивність.

• **Сортування з використанням NumPy.** NumPy використовує різні алгоритми сортування в залежності від конкретного випадку, включно з Quicksort, Mergesort, та Heapsort. Функція `np.sort()` за замовчуванням використовує Quicksort, який є ефективним загальним алгоритмом сортування. Його ефективність обумовлена складністю $O(n \log n)$, яка є середня і залежить від вхідних даних та глибини рекурсії. Це підтверджується в роботі [5], де докладно описується робота і ефективність алгоритму Quicksort.

• **Сортування з використанням Pandas.** Pandas надає функціонал для сортування даних у DataFrame за допомогою методу `sort_values()`. Це дозволяє сортувати дані за значеннями одного або декількох стовпців, надаючи гнучкі можливості для аналізу даних.

• **Сортування з врахуванням локалізації.** Врахування особливостей мови та локалізації може бути важливим при сортуванні рядкових даних. Використання спеціального словника для визначення порядку символів у алфавіті дозволяє точно сортувати рядки, враховуючи особливості мови, як у випадку з українською кирилицею.

У цьому розділі ми детально зупинимося на реалізації трьох ключових алгоритмів сортування за допомогою Python: сортування бульбашкою, вставками та швидке сортування (QuickSort). Ми вибрали ці алгоритми через їх різноманітність застосувань та рівнів складності, що дозволяє дослідити ефективність сортування від простих до більш складних структур даних. Кожен з них показує унікальні характеристики при роботі з наборами даних різного розміру та типу, надаючи цінне розуміння того, як оптимізації на рівні мови програмування можуть впливати на загальну продуктивність обробки даних.

Сортування бульбашкою (bubble sort) — це простий алгоритм сортування, який працює шляхом многократного проходження через список, порівнюючи кожен пару сусідніх елементів та обмінюючи їх місцями, якщо вони в неправильному порядку. Процес повторюється до тих пір, поки не буде пройдено весь список без жодного обміну, що є знаком того, що список відсортовано. Це один з найпростіших алгоритмів сортування для розуміння та кодування. Сортування бульбашкою не є найефективнішим алгоритмом для великих списків через його квадратичну складність в гіршому та середньому випадках ($O(n^2)$, де n — кількість елементів у списку). Також це стабільний алгоритм сортування, тому що він не змінює порядок однакових елементів.

Сортування вставками (insertion sort) є простим, але ефективним алгоритмом сортування, який працює набагато краще на невеликих або частково відсортованих списках. Алгоритм розділяє список на відсортовану та не відсортовану частини, і поступово вбудовує кожен елемент з не відсортованої частини в правильну позицію у відсортованій.

QuickSort є одним з найшвидших загальних алгоритмів сортування для великих наборів даних. Алгоритм використовує стратегію 'розділяй та володарюй', обираючи один елемент як опорний (pivot) і розділяючи список на дві частини: елементи менші за опорний і елементи більші за опорний. Цей процес рекурсивно застосовується до кожної частини.

Вирішення проблеми локалізації. Один з підходів до вирішення проблеми локалізації при сортуванні рядків, що містять символи не з базового латинського алфавіту, полягає в використанні словника для визначення порядку літер у специфічному алфавіті. Цей метод дозволяє визначити точний порядок сортування для символів, що виходять за рамки стандартного ASCII набору, забезпечуючи правильне впорядкування рядків відповідно до правил конкретної мови або локалі.

Спочатку реалізовується словник(dictionary), з індексами у ролі ключа та літерами у ролі значень.

```
ALPHABET_DICTIONARY = {
    char: index for index, char in
    enumerate("АабБбВвГгГгДдЕеЄєЖжЗзИиІіІіЙйКкЛлМмНнОопПпРрСсТтуУуФфХхЦцЧчШшЩщЬьЮюЯя")
}
```

Код функції `sorter`, яка використовує цей метод:

```
def sorter(compare_str: str) -> int:
    return [ALPHABET_DICTIONARY.get(char, ord(char)) for char in compare_str]
```

І тепер використовуючи стандартний метод для сортування у Python, ми можемо вказати `key`, надаючи спеціальну функцію ключа, для налаштування порядку сортування. Нижче наведено такий приклад:

```
actual_word_list = sorted(
    ["", "i", "i", "інтернаціональний", "міжнародний", "євгеній", "єлизавета", "абрикос"],
    key=sorter
)
```

Аналіз продуктивності алгоритмів сортування. Для порівняння продуктивності різних алгоритмів сортування використовується методика вимірювання часу, яка потребує фіксації часу перед початком сортування та після його завершення. Різниця між цими двома показниками вказує на тривалість виконання алгоритму. Використовуючи різні обсяги даних, можна визначити, як масштабується час виконання в залежності від розміру набору даних. Для візуалізації результатів можна створити графіки, які показують час виконання кожного алгоритму сортування в залежності від розміру набору даних. Це дозволяє візуально оцінити продуктивність алгоритмів та їхню поведінку при зміні обсягів даних. На скріншоті, представленою в дослідженні, ми бачимо, як різні алгоритми сортування масштабуються при збільшенні розмірів даних. На представленому графіку на рисунку 1 відображена динаміка часу сортування в залежності від розміру даних для різних алгоритмів сортування.

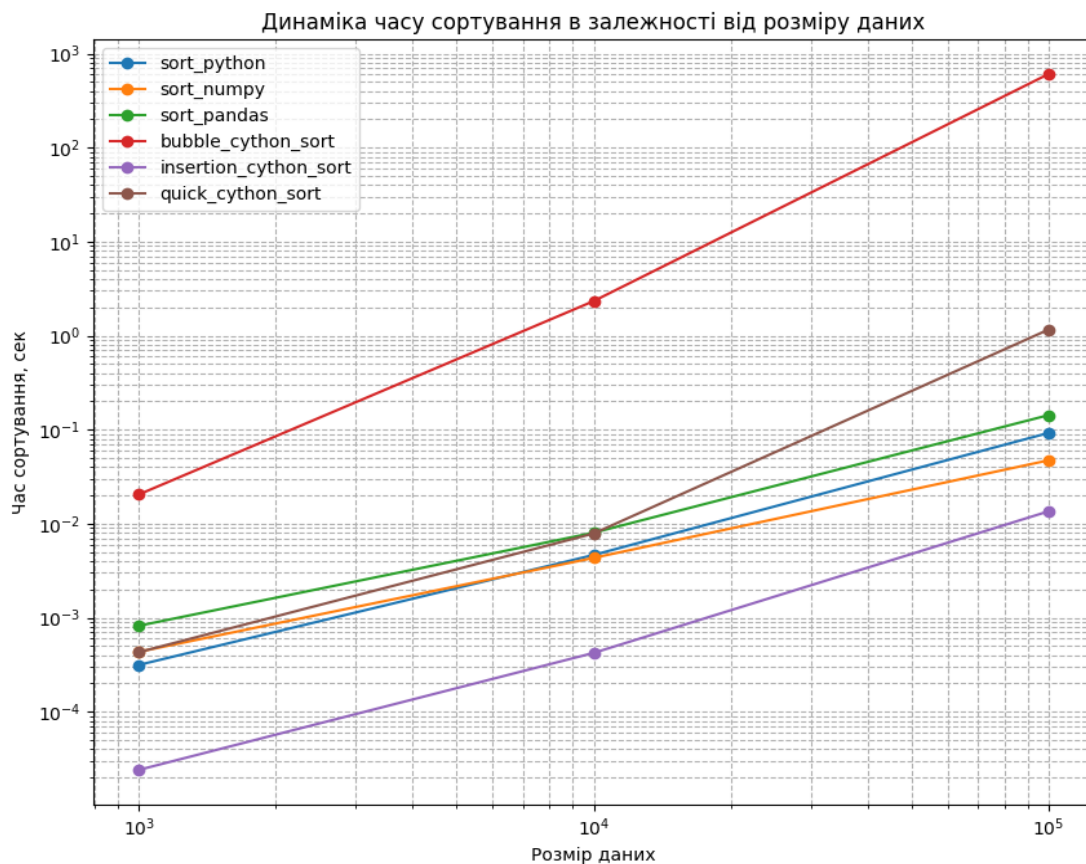


Рисунок 1 – Динаміка часу сортування в залежності від розміру даних

Видно, що *стандартне сортування в Python* (`sort_python`), яке оптимізоване для загальних випадків, має лінійно-логарифмічну складність і показує найкращі результати.

Сортування в NumPy (`sort_numpy`) та *Pandas* (`sort_pandas`) також мають схожу продуктивність, що пояснюється використанням високооптимізованих бібліотек C/C++ під капотом цих пакетів.

Алгоритми сортування, реалізовані на Cython (`bubble_cython_sort`, `insertion_cython_sort` і `quick_cython_sort`), демонструють різну продуктивність.

Сортування бульбашкою на Cython (`bubble_cython_sort`) є найменш ефективним через свою квадратичну складність і рекомендується до використання тільки на невеликих наборах даних або коли дані майже відсортовані.

Сортування вставками на Cython (`insertion_cython_sort`) показує кращу продуктивність на невеликих обсягах даних або коли дані частково відсортовані, але його ефективність різко падає зі збільшенням розміру даних.

QuickSort на Cython (`quick_cython_sort`) показує дуже добру продуктивність на великих наборах даних, маючи середню складність $O(n \log n)$. Цей алгоритм варто розглянути для більш складних випадків сортування, де необхідно оптимальне співвідношення між швидкістю та використанням пам'яті, особливо коли працюєте з великою кількістю даних.

Висновок. Проведене порівняльне дослідження алгоритмів сортування підтвердило, що ефективність сортування значно залежить від обсягу та характеру даних. Для невеликих та частково відсортованих датасетів алгоритми сортування вставками виявляються достатньо ефективними, завдяки низькій складності та відмінній продуктивності при обмеженій кількості даних. Проте, як показує дослідження, із збільшенням обсягів даних перевагу слід віддавати більш складним алгоритмам, таким як QuickSort, що, будучи оптимізованим в Cython, забезпечують значне зниження часу виконання завдяки вищій ефективності на великих масштабах. Аналіз продуктивності також розкрив важливість адаптації алгоритмів під специфіку мов та локалей, що є критичним для глобалізованих застосунків та систем. В цьому контексті введення спеціалізованого словника для сортування рядків стало ключовою оптимізацією для підтримки української мови. Враховуючи ці фактори, вибір алгоритму сортування повинен відбуватися з огляду на конкретні потреби проекту, природу даних та вимоги до швидкодії, що дозволяє досягти оптимального балансу між швидкістю обробки та використанням ресурсів.

Список використаних джерел

1. Ван Россум Г., Дрейк Ф. Л. Python. Керівництво для розробників: монографія. Київ: Діалог, 2019. 720 с.
2. Бехтерев А. М. Cython: оптимізація Python програм: монографія. Львів: Львівський технологічний, 2018. 300 с.
3. Орлов С. А. Python для комплексних задач: наука про дані та машинне навчання: монографія. Харків: Факт, 2020. 400 с.
4. Седжвік Р., Уейн К. Алгоритми: сортування, пошук, основні структури даних: підручник. Харків: Плеяди, 2017. 992 с.
5. Кормен Т. Х., Лейзерсон Ч. Е., Рівест Р. Л., Штайн К. Алгоритми: побудова та аналіз: підручник. 3-тє вид. Київ: Видавничий дім 'Вільямс', 2019. 1328 с.
6. Маккінни У. Python для аналізу даних: монографія. 2-ге вид. Київ: Основи, 2018. 540 с.
7. Cython: великий путівник для оптимізації Python коду: електрон. наук. фахове вид. 2022. URL: <https://cython.readthedocs.io> (дата звернення: 31.03.2024).
8. Гроссман С. А., МакКінни В. Ефективний Python: 59 специфічних способів написання кращого коду: монографія. Київ: Діафільм, 2016. 256 с.
9. Лутц М. Програмування на Python. Том 1: вступ до програмування: підручник. Київ: Діафільм, 2021. 640 с.

Khomenko Andrii

*Ph.D. student, Faculty of Information Technologies, Department of Computer Science,
National University of Life and Environmental Sciences of Ukraine*

E-mail: andmayster@gmail.com

Kyrychenko Viktor

*Candidate of Physical and Mathematical Sciences, Associate Professor of the Department of
Computer Science,*

National University of Life and Environmental Sciences of Ukraine

ORCID: <https://orcid.org/0009-0001-0575-8684>

E-mail: andmayster@gmail.com

CHARACTERISTICS OF SORTING ALGORITHMS OPERATION IN PYTHON

***Abstract.** This article is dedicated to the analysis of various sorting algorithms using the Python programming language and sorting algorithms in Cython. The study compares classic sorting methods, such as bubble sort, insertion sort, and quicksort, with the aim of determining their effectiveness for large datasets. Special attention is given to localization issues when sorting strings in non-English languages, particularly, the application of a specialized dictionary for the correct processing of the Ukrainian alphabet. Techniques for measuring performance and visualizing results in the form of graphs are presented, allowing a deeper assessment of the scaling of each algorithm depending on the dataset size.*

***Keywords:** Python, Cython, sorting algorithms, large datasets, localization, Ukrainian alphabet, sorting performance, data visualization.*

УДК 519.22:504.06

Густера Олег Михайлович

кандидат економічних наук, старший викладач кафедри комп'ютерних наук,
Національний університет біоресурсів і природокористування України

ORCID: <https://orcid.org/0000-0003-1010-6100>E-mail: o.gustera@nubip.edu.ua**Ніколаєнко Дмитро Володимирович**

кандидат економічних наук, старший викладач кафедри комп'ютерних наук,
Національний університет біоресурсів і природокористування України

ORCID: <https://orcid.org/0009-0008-4817-3951>E-mail: d.nikolaenko@nubip.edu.ua**ЗГЛАДЖУВАННЯ НЕПОВНИХ РЯДІВ ДАНИХ СТАНЦІЙ ЕКОЛОГІЧНОГО
МОНІТОРИНГУ З ВИКОРИСТАННЯМ ПРОГНОЗНИХ МОДЕЛЕЙ**

Анотація. У статті розглядається проблема неповних рядів даних при аналізі даних станцій екологічного моніторингу, її вплив на достовірність отриманих результатів та прогностичну придатність вхідних даних, а також методи згладжування даних, що дозволяють мінімізувати негативний вплив пропусків даних. Відсутність даних в ряді може проявлятися на практиці як хибні нульові значення, які можуть призводити до суттєвих відхилень, а також як відсутні дані, що в деяких випадках приховують тенденції зміни динаміки ряду. При цьому, аналітик може не знати про присутність пустих або нульових значень, що, в результаті, призводить до хибних висновків або прогнозів. Методи згладжування за допомогою простої ковзної середньої та екстраполяції дозволяють підвищити якість вхідних даних, та, як результат, підвищити прогностичну якість отриманих прогнозних моделей. Використання локальних прогнозів для заповнення пропущених значень дозволяє отримати найбільш точні результати замість відсутніх даних, і, як результат, підвищити прогностичну якість розроблених прогнозних моделей. Точність результатів отриманих замість відсутніх даних перевіряється розрахунком основних статистичних показників ряду з пустими значеннями та повного ряду. Розрахунок параметрів моделей прогнозування для заповнення пустих інтервалів може здійснюватися на основі попередніх даних або тенденції всього ряду. Отримані в даному дослідженні результати можуть бути використані у подальшому для заповнення неповних рядів при аналізі даних станцій екологічного моніторингу або інших рядів даних, що використовуються для прогнозування або аналітичних розрахунків.

Ключові слова: екологічний моніторинг, ряд даних, динаміка ряду даних, пропуски даних, прогноз, прогнозні моделі.

Вступ. У процесі аналізу екологічного стану шляхом отримання даних від станцій моніторингу надійність отриманих результатів напряму залежить від точності початкових даних, які отримані від датчиків та збережені у сховище даних[8]. При цьому, в сучасних умовах може виникати проблема нерівномірних інтервалів дослідження або відсутності частини спостережень, що може бути пов'язано з негативним впливом наступних факторів[6]:

- відсутність електропостачання у екологічній станції;
- відсутність зв'язку з екологічною станцією через проблеми зі зв'язком;
- некоректна робота апаратного або програмного забезпечення після відключень електроенергії;
- інші фактори, що спричиняють збої у роботі екологічної станції (DDOS-атаки, проблеми у провайдера та ін.);
- цілеспрямоване приховування або знищення даних третіми особами.

В результаті, підсумкові дані за досить тривалі періоди можуть бути відсутні, як показано на рис. 1. Ряд даних показника якості повітря, представлений на Рис. 1, у період з 12.00 10.01.2023 до 12.00 12.01.2023 є неповним. З 6.00 до 12.00 11.01.2023 дані відсутні. В таких умовах отримання достовірних та повних рядів даних від станції екологічного моніторингу ускладнюється. При цьому, неможливо заздалегідь спрогнозувати періоди, коли отримані дані будуть неповними або не коректними. Тобто, дізнатися про те, що отримані дані

не можуть бути використані для отримання достовірних результатів аналізу, як правило, можна лише у процесі аналізу[8].

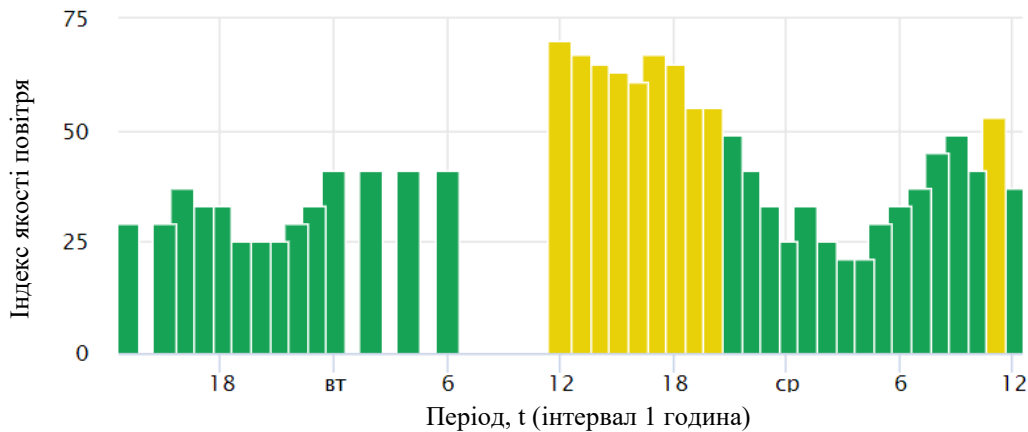


Рисунок 1 – Індекс якості повітря [18]

Аналіз останніх досліджень і публікацій. На сьогоднішній день питання екологічного моніторингу як необхідної складової частини та інформаційної бази для захисту навколишнього середовища є особливо актуальним, та висвічується в багатьох зарубіжних та вітчизняних джерелах [3,12,14,19].

В умовах постійно зростаючої кількості станцій екологічного моніторингу та інших пристроїв, що дозволяють отримувати дані про окремі показники стану навколишнього середовища зростає актуальність проблеми зберігання, обробки та аналізу накопиченої інформації для її подальшого використання [4,7,11].

Серед сучасних вітчизняних робіт у галузі екологічного моніторингу слід відзначити розробку інформаційно-аналітичної системи оцінювання стану атмосферного повітря [1,2].

Сучасні методи інтелектуального аналізу дозволяють досить ефективно вирішувати задачі інтерпретації необроблених даних за умови їх точності та достовірності [4,5]. Таким чином, першочергове питання, якому слід приділяти достатньо уваги при екологічному моніторингу – саме отримання достовірних та точних вхідних даних, які можуть бути спотворені через ряд об'єктивних чи суб'єктивних чинників.

Метою дослідження є застосування методів згладжування для заповнення неповних рядів при аналізі даних станцій екологічного моніторингу або інших рядів даних, що використовуються для прогнозування або аналітичних розрахунків.

Матеріали і методи дослідження. На практиці найбільш негативний вплив неповних рядів даних проявляється у використанні недостовірних даних для аналізу без усвідомлення їх придатності до аналізу, що може призводити до більш негативних наслідків. Тобто, аналітики або особа що приймає рішення отримує готові результати та не усвідомлює, що вони побудовані на частково хибних даних[10].

Необхідність заповнення пропусків у ряді даних підтверджується тим, що інтервали без значень можуть бути сприйняті як нульові значення, або ж не будуть враховані у загальній тенденції ряду. В деяких випадках відрізнити реальні нульові значення від пропущених спостережень досить складно. Так, наприклад, температура повітря може приймати нульове значення, яке може бути переплутане з пустим значенням.

Відновлення пропусків у часових рядах може вирішуватись за допомогою наступних груп методів:

- прості статистичні методи (екстраполяція, ковзна середня),
- ітеративне прогнозування,
- комбіновані схеми прогнозування.

Перевагою простих статистичних методів їх простота в реалізації, що особливо зручно при автоматизованій обробці вхідних даних[12].

Результати дослідження та їх обговорення. Для того щоб більш детально проаналізувати можливі наслідки використання неповних або некоректних рядів даних без їх обробки або адаптації розглянемо наступний приклад (рис. 2). Використовуємо один і той самий ряд даних показника РМ1 – дрібнодисперсні частки у повітрі, діаметром менше 1 мкм (мікрону) у трьох можливих варіантах спостереження – без пропусків, з пропусками, та з хибними значеннями замість пропусків. Хибні значення в даному прикладі це нульові значення.

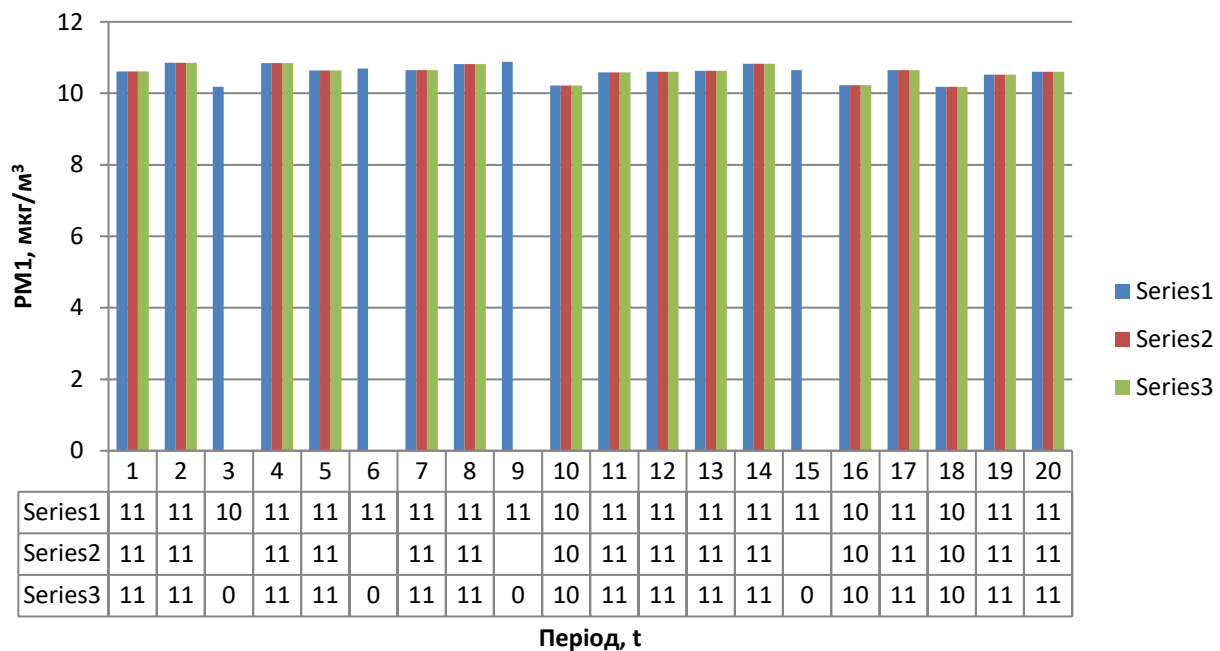


Рисунок 2 – Початковий ряд даних показника РМ1 (дрібнодисперсні частки у повітрі, діаметром менше 1 мікрону): Ряд 1 – без пропусків, Ряд 2 – з пропусками, Ряд 3 – з хибними значеннями замість пропусків (16)

У першому випадку (ряд 1) ми отримуємо повний ряд даних, і на його основі можемо побудувати прогноз на майбутній період, визначити основні статистичні характеристики ряду. В другому випадку (ряд 2) проаналізуємо аналогічний ряд, у якому будуть відсутні дані за певні періоди, тобто зімітуємо неповний ряд даних. У третьому випадку (ряд 3) замість відсутніх даних будемо використовувати некоректні дані – нулі. Для того щоб визначити, наскільки суттєво повнота ряду впливає на отримані результати, аналогічно до повного ряду визначимо його основні статистичні характеристики – середнє значення, дисперсію та середньоквадратичне відхилення (табл. 1).

Таблиця 1 – Основні статистичні характеристики досліджуваного часового ряду

Характеристика	Ряд 1	Ряд 2	Ряд 3
Середнє	10,59493	10,59324	8,474591
Дисперсія	0,048268	0,043738	17,98966
Стандартне відхилення	0,219701	0,209135	4,241422

Як видно з результатів, представлених у табл. 1, найбільш суттєво на отримані результати впливає використання для подальшого аналізу чи прогнозування використання некоректних даних. Врахування нульових значень призводить до збільшення дисперсії та стандартного відхилення у декілька разів, що, в результаті, робить побудовані прогнози моделей нераціональним через великі довірчі інтервали або низьку точність прогнозу. При цьому, врахування нульових значень залежить від абсолютних характеристик ряду даних.

Відсутність даних за певні періоди також негативно впливає на точність отриманих результатів. У випадках, коли у пропущених інтервалах відбувались суттєві зміни тенденції ряду даних, це не буде помічено та враховано при побудованні прогнози моделей або розрахунку аналітичних показників[13].

Для оцінки впливу врахування нульових та пропущених інтервалів на результати прогнозу використаємо лінійну та поліноміальну модель. Для побудованні прогнозу на основі обраних моделей використаємо MS Excel (рис. 3).

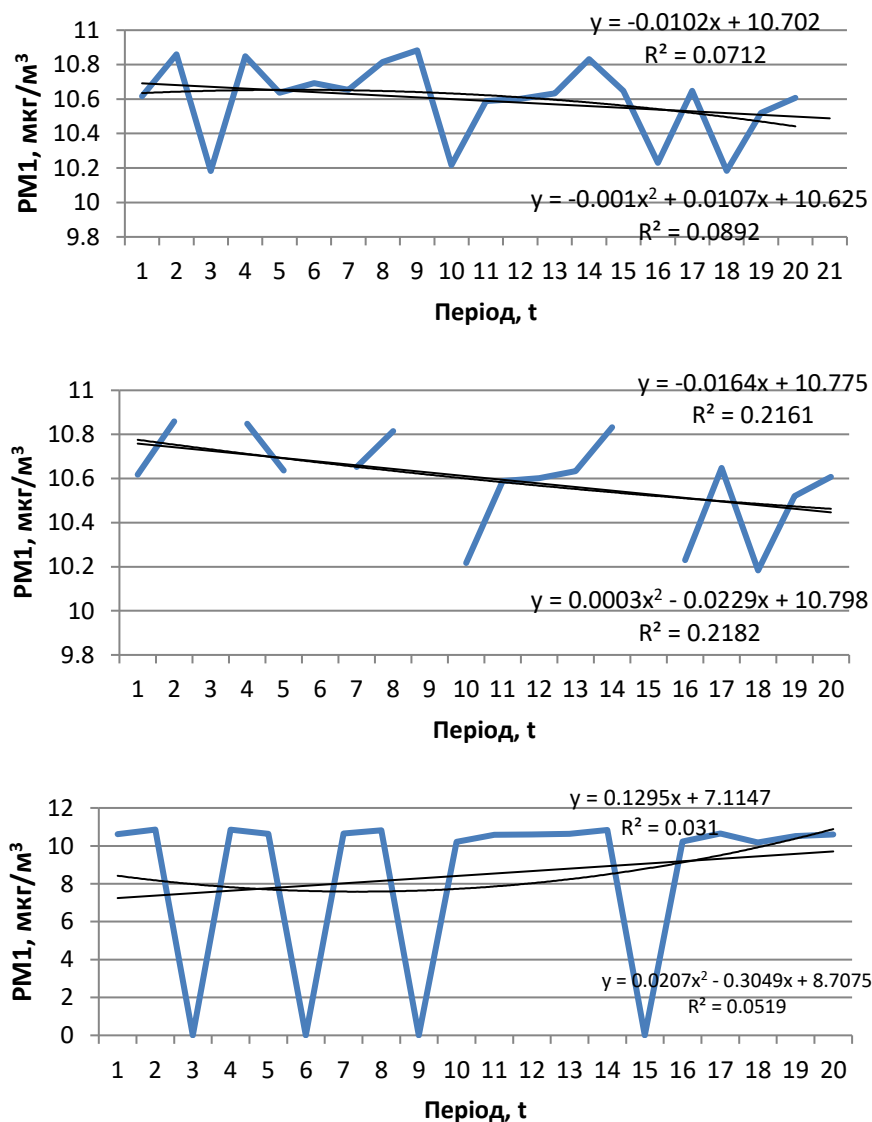


Рисунок 3 – Прогнозування на основі поліноміальної та лінійної моделі для повного ряду даних, ряду з порожніми значеннями, та ряду з нульовими значеннями

Як видно з рис. 3, використання необроблених вхідних даних для аналізу чи прогнозування може призводити до отримання результатів з низькою достовірністю або прогностичною придатністю.

Для того щоб заповнити дані за певні періоди спочатку використаємо метод ковзної середньої, тобто використаємо середнє значення за найближчі періоди для періоду згладжування m :

$$\hat{x}_i = \frac{1}{p} \sum_{j=i-m}^{i+m} x_j \quad (1)$$

Також відсутні дані можна замінити методом екстраполяції існуючих:

$$\hat{x}_i = \begin{cases} x_{i-1}, & \text{якщо } x = 0, \\ x_i, & \text{якщо } x > 0. \end{cases} \quad (2)$$

Розглянемо використання методу простої ковзної середньої для заповнення відсутніх елементів ряду даних (рис. 4).

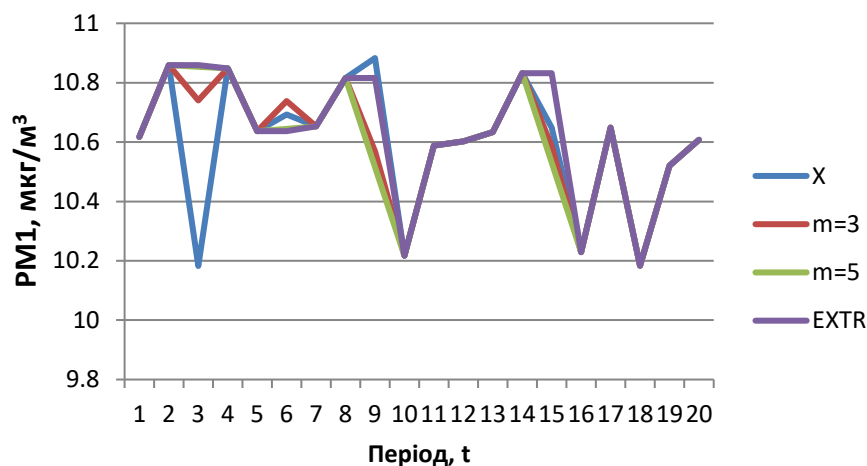


Рисунок 4 – Використання простої ковзної середньої для заповнення відсутніх даних часового ряду (період згладжування $m=3$, $m=5$) та екстраполяції

Як видно з результатів розрахунку основних статистичних характеристик досліджуваного часового ряду (X), ряду отриманого методом згладжування ($m=3$, $m=5$) та методом екстраполяції (EXTR), наведених у табл. 2, навіть прості методи згладжування дають більш достовірні результати у порівнянні з екстраполяцією. Тим не менш, методи згладжування хоча й можуть бути достатньо ефективно використані з метою усунення пропусків або некоректних нульових значень, однак не дозволяють врахувати наявність тенденції у часовому ряді і можливість її прояву саме у пропущеному інтервалі.

Таблиця 2. – Основні статистичні характеристики досліджуваного часового ряду (X) та ряду отриманого методом згладжування ($m=3$, $m=5$) та методом екстраполяції (EXTR)

Характеристика	X	m=3	m=5	EXTR
Середнє	10,59493	10,60622	10,60185	10,63172
Дисперсія	0,048268	0,036987	0,038935	0,042445
Стандартне відхилення	0,219701	0,19232	0,197319	0,206021

З метою врахування тенденції часового ряду при заповненні ряду використаємо прогнозування (3), реверсивне прогнозування (4) для локальних інтервалів на основі лінійної та поліноміальної моделі:

$$\hat{x}_i = \begin{cases} a_0 + a_1 x_{i-1} \vee a_0 + a_1 x_{i-1} + a_2 x_{i-1}^2, & \text{якщо } x = 0, \\ x_i, & \text{якщо } x > 0. \end{cases} \quad (3)$$

$$\hat{x}_i = \begin{cases} a_0 + a_1 x_{i+1} \vee a_0 + a_1 x_{i+1} + a_2 x_{i+1}^2, & \text{якщо } x = 0, \\ x_i, & \text{якщо } x > 0. \end{cases} \quad (4)$$

Для лінійної моделі, використаємо наступні параметри (рис. 5):

$$\begin{aligned} a_0 &= 10,702 \\ a_1 &= -0,0102 \end{aligned}$$

Для поліноміальної моделі, відповідно:

$$\begin{aligned} a_0 &= 10,625 \\ a_1 &= 0,0107 \\ a_2 &= -0,001 \end{aligned}$$

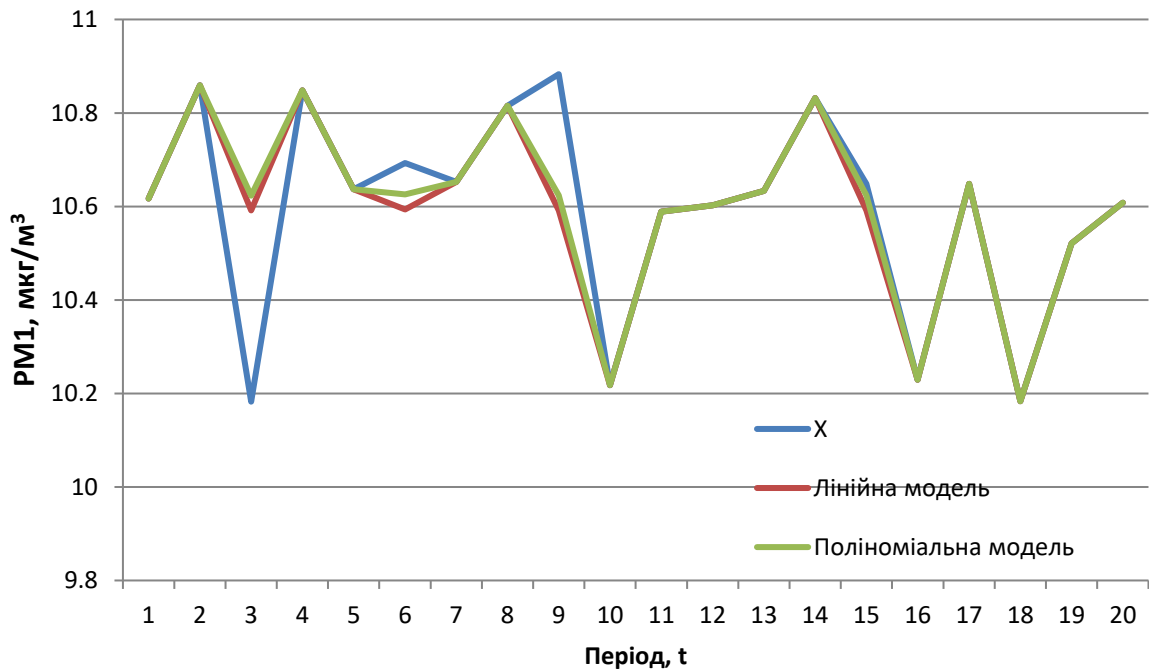


Рисунок 5 – Використання локального прогнозування для заповнення відсутніх даних часового ряду (лінійна та поліноміальна модель)

Середнє, дисперсія та стандартне відхилення часового й інтерпольованого методом локального прогнозування рядів наведені в табл. 3.

Таблиця 3 – Основні статистичні характеристики досліджуваного часового ряду (X) та ряду отриманого шляхом заповнення порожніх значень методом локального прогнозування для заповнення відсутніх даних часового ряду (лінійна та поліноміальна модель)

Характеристики	X	Лінійна модель	Поліноміальна модель
Середнє	10,59298848	10,59940492	10,59298848
Дисперсія	0,034990411	0,035142256	0,034990411
Стандартне відхилення	0,18705724	0,18746268	0,18705724

Для побудовання реверсивного прогнозу виконаємо транспонування початкового ряду даних та розрахуємо параметри моделей, що проілюстровано на рис. 6.

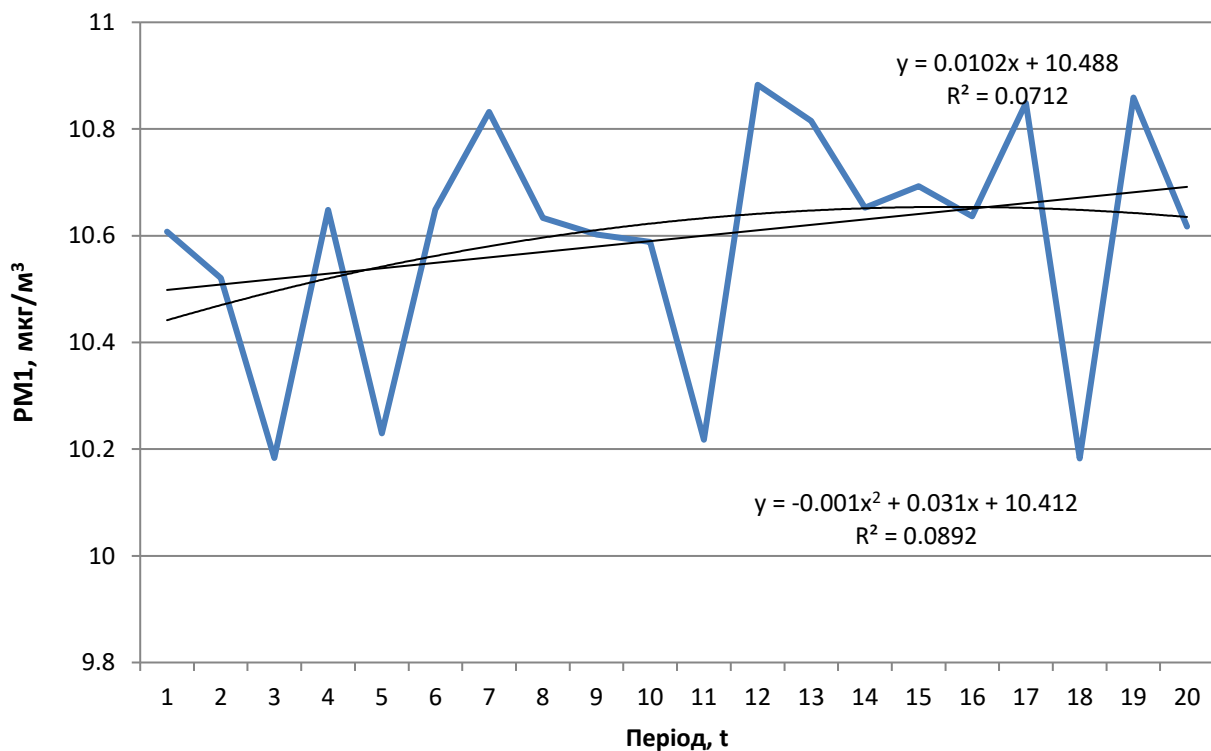


Рисунок 6 – Розрахунок параметрів лінійної та поліноміальної моделі для реверсивного прогнозування

Таким чином, для лінійної моделі, використаємо наступні параметри:

$$a_0 = 10,488$$

$$a_1 = 0,0102$$

Для поліноміальної моделі, відповідно:

$$a_0 = 10,412$$

$$a_1 = 0,031$$

$$a_2 = -0,001$$

Графіки часового ряду і прогнозних рядів за лінійної та поліноміальною моделями з використанням локального реверсивного прогнозування для заповнення відсутніх даних часового ряду наведено на рис. 7.

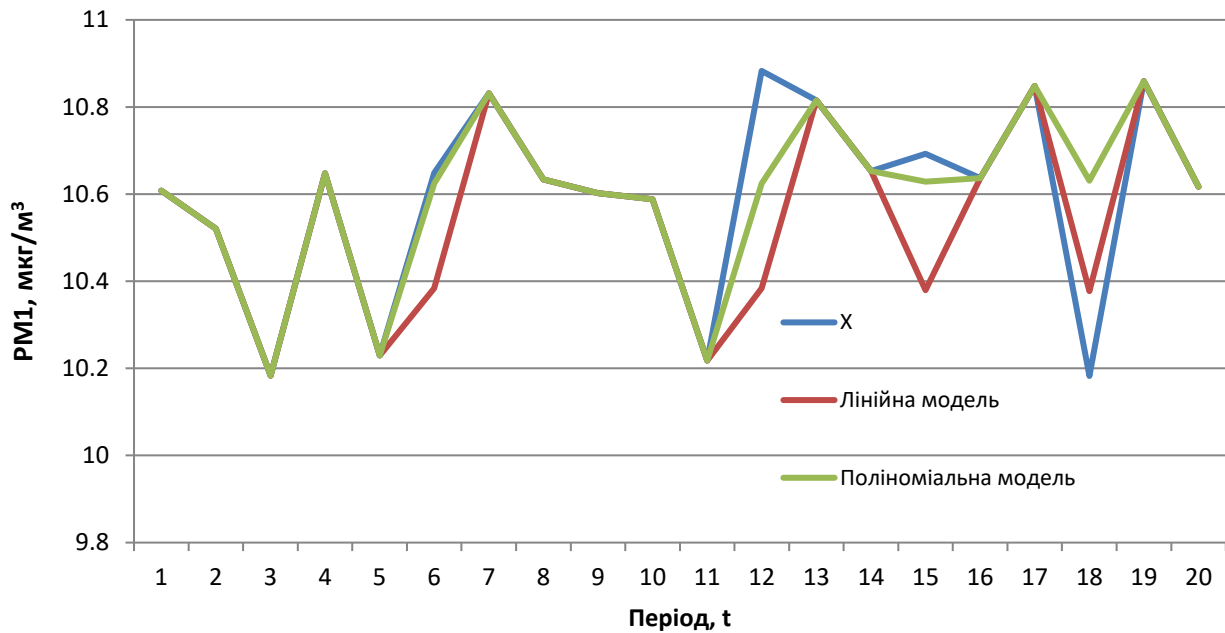


Рисунок 7 – Використання локального реверсивного прогнозування для заповнення відсутніх даних часового ряду (лінійна та поліноміальна модель)

В табл. 4 наведені статистичні характеристики часових рядів, в тому числі, отриманих за рахунок реалізації лінійної та поліноміальної моделей.

Таблиця 4 – Основні статистичні характеристики досліджуваного часового ряду (X) та ряду отриманого шляхом заповнення порожніх значень методом локального реверсивного прогнозування для заповнення відсутніх даних часового ряду (лінійна та поліноміальна модель)

Характеристики	X	Лінійна модель	Поліноміальна модель
Середнє	10,59493	10,5508	10,6
Дисперсія	0,048268	0,042197	0,035174
Стандартне відхилення	0,219701	0,205418	0,187548

Для досліджуваного прикладу, згідно з результатами отриманими з таблиці 3 та таблиці 4, більш точні результати при заповненні порожніх значень ряду дозволяє отримати метод локального прогнозування. В той же час, в деяких випадках використання реверсивного прогнозування є єдиним можливим способом отримання відсутніх значень. Наприклад, якщо порожніми є одні з перших спостережень, виявлення чіткої тенденції на основі декількох попередніх спостережень або при їх повній відсутності не дає достовірних результатів. У таких випадках, коли реверсивне прогнозування дає менш точні результати у порівнянні із прогнозуванням на основі попередніх значень, і при цьому відсутні початкові значення ряду,

можливе використання комбінованого методу. Тобто, для частини порожніх значень використовується прогнозування на основі попередніх значень, а там де це є необхідним – на основі послідуєчих значень, тобто реверсивне прогнозування.

Висновки. Використані у дослідженні методи згладжування за допомогою простої ковзної середньої та екстраполяції дозволяють підвищити якість вхідних даних, та, як результат, підвищити прогностичну якість отриманих прогнозних моделей. В той же час, застосування локальних прогнозів для заповнення пропущених значень дозволяє отримати найбільш точні результати замість відсутніх даних, і, як результат, підвищити прогностичну якість розроблених прогнозних моделей. Точність результатів отриманих замість відсутніх даних перевіряється розрахунком основних статистичних показників ряду з пустими значеннями та повного ряду. Розрахунок параметрів моделей прогнозування для заповнення пустих інтервалів може здійснюватись на основі попередніх даних або тенденції всього ряду.

Отримані в даному дослідженні результати можуть бути використані у подальшому для заповнення неповних рядів при аналізі даних станцій екологічного моніторингу або інших рядів даних, що використовуються для прогнозування або аналітичних розрахунків. Вибір моделі прогнозування або аналітичного методу, за допомогою яких здійснюється згладжування та заповнення порожніх значень даних може залежати від загальної тенденції часового ряду.

Список використаних джерел

1. Bogolyubov V.M., Golub B.L. Information-analytical system for assessing the state of atmospheric air / Sustainable development — 21st century. Discussions 2021: collective monograph / National University - Kyiv-Mohyla Academy / edited by Prof. Khlobistova E.V. — Kyiv, 2021. — 469 p. ISBN: 978-617-7668-22-9.
2. Bogoliubov V.M. and other. Optimization of the Structure of Atmospheric Air Monitoring System <https://chmnu.edu.ua/wpcontent/uploads/MONOGRAPH-2.pdf>.
3. Khan, S.; Anjum, R.; Raza, S.T.; Ahmed Bazai, N.; Ihtisham, M. Technologies for Municipal Solid Waste Management: Current Status, Challenges, and Future Perspectives. *Chemosphere* 2022, 288, 132403.
4. Andeobu, L.; Wibowo, S.; Grandhi, S. A Systematic Review of E-waste Generation and Environmental Management of Asia Pacific Countries. *Int. J. Environ. Res. Public Health* 2021, 18, 9051.
5. Ma, S.; Zhou, C.; Chi, C.; Liu, Y.; Yang, G. Estimating Physical Composition of Municipal Solid Waste in China by Applying Artificial Neural Network Method. *Environ. Sci. Technol.* 2020, 54, 9609–9617.
6. Lin, K.; Zhao, Y.; Tian, L.; Zhao, C.; Zhang, M.; Zhou, T. Estimation of Municipal Solid Waste Amount Based on One-Dimension C N Network and Long Short-Term Memory with Attention Mechanism Model: A Case Study of Shanghai. *Sci. Total Environ.* 2021, 791, 148088.
7. Sharma, M.; Joshi, S.; Kannan, D.; Govindan, K.; Singh, R.; Purohit, H. Internet of Things (IoT) Adoption Barriers of Smart cities' Waste Management: An Indian Context. *J. Clean. Prod.* 2020, 270, 122047.
8. Jassim, M.S.; Coskuner, G.; Zontul, M. Comparative Performance Analysis of Support Vector Regression and Artificial Neural Network for Prediction of Municipal Solid Waste Generation. *Waste Manage. Res.* 2022, 40, 195–204.
9. Lin, K.; Zhao, Y.; Kuo, J.-H.; Deng, H.; Cui, F.; Zhang, Z.; Zhang, M.; Zhao, C.; Gao, X.; Zhou, T.; et al. Toward smarter management and recovery of municipal solid waste: A critical review on deep learning approaches. *J. Clean. Prod.* 2022, 346, 130943.
10. Fasano, F.; Addante, A.S.; Valenzano, B.; Scannicchio, G. Variables Influencing per Capita Production, Separate Collection, and Costs of Municipal Solid Waste in the Apulia Region (Italy): An Experience of Deep Learning. *Int. J. Environ. Res. Public Health* 2021, 18, 752.

11. Hussain, A.; Draz, U.; Ali, T.; Tariq, S.; Irfan, M.; Glowacz, A.; Antonino Daviu, J.A.; Yasin, S.; Rahman, S. Waste Management and Prediction of Air Pollutants Using IoT and Machine Learning Approach. *Energies* 2020, 13, 3930.
12. Ihsanullah, I.; Alam, G.; Jamal, A.; Shaik, F. Recent Advances in Applications of Artificial Intelligence in Solid Waste Management: A Review. *Chemosphere* 2022, 309, 136631.
13. Hettiarachchi, H.; Meegoda, J.N.; Ryu, S. Organic Waste Buyback as a Viable Method to Enhance Sustainable Municipal Solid Waste Management in Developing Countries. *Int. J. Environ. Res. Public Health* 2018, 15, 2483.
14. Rahman, M.W.; Islam, R.; Hasan, A.; Bithi, N.I.; Hasan, M.M.; Rahman, M.M. Intelligent Waste Management System Using Deep Learning with IoT. *J. King Saud Univ.-Comput. Inf. Sci.* 2022, 34, 2072–2087.
15. Mookkaiah, S.S.; Thangavelu, G.; Hebbar, R.; Haldar, N.; Singh, H. Design and Development of Smart Internet of Things–based Solid Waste Management System Using Computer Vision. *Environ. Sci. Pollut. Res.* 2022, 29, 64871–64885.
16. Nowakowski, P.; Pamuła, T. Application of Deep Learning Object Classifier to Improve E-waste Collection Planning. *Waste Manag.* 2020, 109, 1–9.
17. Niu, D.; Wu, F.; Dai, S.; He, S.; Wu, B. Detection of Long-Term Effect in Forecasting Municipal Solid Waste Using a Long Short-Term Memory Neural Network. *J. Clean. Prod.* 2021, 290, 125187.
18. Сайт моніторингу рівня забруднення атмосферного повітря у місті Київ <https://www.saveecobot.com/maps/kyiv>
19. Woodward, W. A., Gray, H. L. & Elliott, A. C. (2012), *Applied Time Series Analysis*, CRC Press
20. Needleman, Saul B. & Wunsch, Christian D. (1970). "A general method applicable to the search for similarities in the amino acid sequence of two proteins". *Journal of Molecular Biology*.

Hustera Oleg

*PhD in Economics, Senior Lecturer of the Department of Computer Science,
National University of Life and Environmental Sciences of Ukraine*

ORCID: <https://orcid.org/0000-0003-1010-6100>

E-mail: o.gustera@nubip.edu.ua

Nikolaienko Dmytro

*PhD in Economics, Senior Lecturer of the Department of Computer Science,
National University of Life and Environmental Sciences of Ukraine*

ORCID: <https://orcid.org/0009-0008-4817-3951>

E-mail: d.nikolaenko@nubip.edu.ua

SMOOTHING INCOMPLETE DATA SERIES OF ENVIRONMENTAL MONITORING STATIONS USING PREDICTIVE MODELS

Abstract. *The article examines the problem of incomplete data series in the analysis of data from environmental monitoring stations, its impact on the reliability of the obtained results and the prognostic suitability of input data, as well as data smoothing methods that allow minimizing the negative impact of data gaps. The absence of data in the series can manifest itself in practice as false zero values that can lead to significant deviations, as well as missing data, which in some cases hides trends in the dynamics of the series. At the same time, the analyst may not be aware of the presence of empty or null values, which, as a result, leads to false conclusions or predictions. Smoothing methods using a simple moving average and extrapolation allow to improve the quality of input data, and, as a result, to improve the predictive quality of the obtained predictive models. Using local forecasts to fill in missing values allows you to get the most accurate results instead of missing data and, as a result, improve the predictive quality of the developed forecast models. The accuracy of the results obtained instead of missing data is checked by calculating the main statistical indicators of the series with empty values and the complete series. Calculation of the parameters of forecasting models to fill the empty intervals can be based on previous data or the trend of the entire series. The results obtained in this study can be used in the future to fill incomplete series in the analysis of data from environmental monitoring stations or other series of data used for forecasting or analytical calculations.*

Keywords: *environmental monitoring, data series, data series dynamics, data gaps, forecast, predictive models.*

УДК 004.896.8:004.451.426

Сагун Андрій Вікторович

кандидат технічних наук, доцент, доцент кафедри комп'ютерних систем, мереж та кібербезпеки,

Національний університет біоресурсів і природокористування України

ORCID: <https://orcid.org/0000-0002-5151-9203>

E-mail: a.sagun@nubip.edu.ua

Панаско Олена Миколаївна

кандидат технічних наук, доцент, доцент кафедри робототехнічних і телекомунікаційних систем та кібербезпеки,

Черкаський державний технологічний університет, м. Черкаси

ORCID: <https://orcid.org/0000-0002-0510-7742>

E-mail: lena.pa@ukr.net

СИСТЕМА РОЗПІЗНАВАННЯ ОБРАЗІВ НА БАЗІ НЕЙРОННОЇ МЕРЕЖІ З ГЛИБОКИМ НАВЧАННЯМ

Анотація. Розроблена система розпізнавання об'єктів, яка ґрунтується на нейронних мережах з глибоким навчанням. Така система розпізнавання образів здатна забезпечити точне та швидке розпізнавання відомих їй та близьких до таких зображень у відеоконтенті, отриманому з IP-камер відеонагляду. Залежно від умов зйомки відео та кута огляду IP-камери, алгоритм роботи даної системи досягає точності розпізнавання, що складає 96,38%. Такий відсоток розпізнавання образів є практично сталим відносно 11 класів об'єктів ідентифікації та розпізнавання. Такі високі результати досягаються завдяки використанню навчальної бази відеоданих CamVid в якості навчальної вибірки нейронної мережі. Така база сформована на основі 421 навчального та 280 тестових відеозображень. В моделі системи розпізнавання передбачені можливості оптимізації параметрів функції навчання та ідентифікації, а також зміни методу вимірювання відстані між векторами ознак (метрики вимірювання відстані точок).

Ключові слова: нейронна мережа, глибоке навчання мережі, класифікатор ознак, опорний інформаційний вектор, графічний фрейм, машинний зір.

Мета дослідження полягає в розробці ефективного алгоритму розпізнавання графічних об'єктів з відеоконтенту, отриманого з камер IP-камер системи відеонагляду. Точність розпізнавання образів в разі практичної реалізації алгоритма має складати не менше 95% для відеоданих, максимально наближених до реальних умов.

Аналіз останніх досліджень та публікацій. На підставі проведеного аналізу наукових досліджень, присвячених алгоритмам та методам розпізнавання об'єктів типу «графічний образ», які наведені в численних джерелах, зокрема в [1-9], можна зробити певні висновки стосовно моделей та математичного апарату прикладних систем розпізнавання. В такій ролі найбільш доцільно застосовувати нейронні мережі (НМ). Моделі на базі НМ є найбільш ефективними для розробки систем розпізнавання образів, які оброблюють графічний або відеоконтент. Для створення системи розпізнавання образів важливо сформулювати адекватну математичну модель, на основі якої можливо буде реалізувати підсистему ідентифікації та класифікації графічних об'єктів. Цілком логічно використати модель НМ, що ґрунтується на алгоритмах нейронних мереж з навчанням або без такого. Перевагою подібних моделей є можливість навчання. Навчання моделі НМ дає змогу актуалізувати і значно підвищувати якість процесу розпізнавання і наступної класифікації, що вказано в джерелах [6,7]. В комп'ютерних інтелектуальних системах для розпізнавання об'єктів у відеоконтенті машинне навчання часто є фактором, який суттєво підвищує адекватність та точність алгоритму визнання та ідентифікації, як зазначено у джерелах [6-8]. Тому вибір моделі системи на базі НМ є цілком виправданий.

Результати дослідження та їх обговорення. Система розпізнавання образів з відеотрафіку IP-камер є актуальною. Подібні системи умовно складаються з декількох

підсистем: 1) технічна частина (підсистема відеонагляду); 2) підсистема аналізу та класифікації елементів зображення; 3) підсистема реєстрації та узагальнення даних розпізнавання. Структурна схема системи розпізнавання образів в загальному вигляді показана на рис. 1.

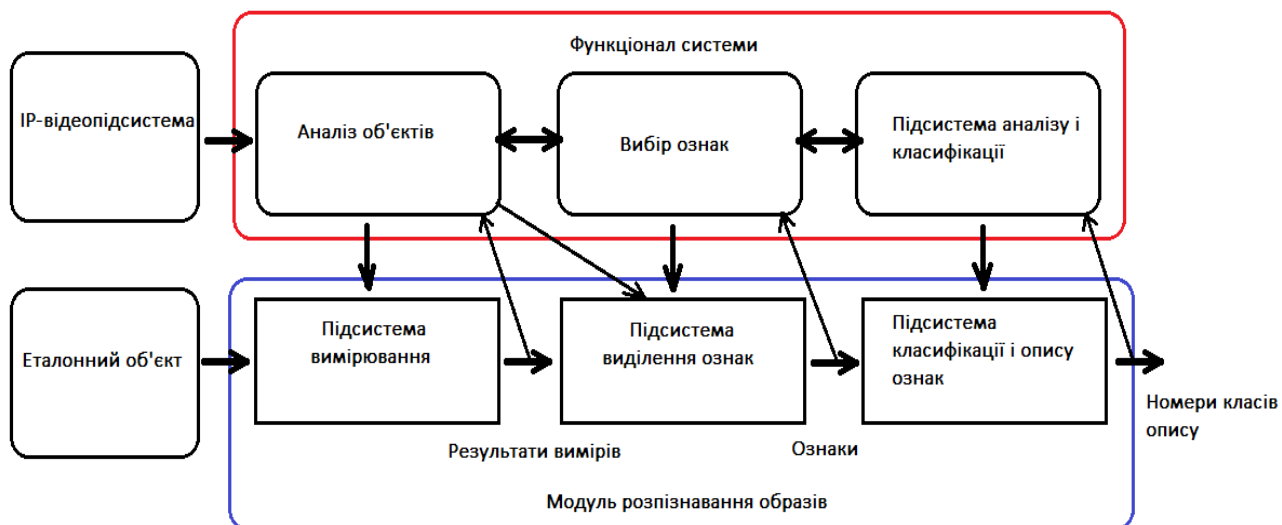


Рисунок 1 – Структурна схема системи розпізнавання образів

Важливим етапом у формуванні алгоритму є вибір моделі і типу навчання НМ. На сьогодні існують декілька типів моделей систем розпізнавання та ідентифікації за видами навчання [10]. З огляду на використання в алгоритмі математичного апарату НМ можна обрати наглядне навчання (Supervised Learning). Цей тип навчання використовується для навчання моделі на основі пар вхідних даних та відповідних міток (позначок), що узгоджується з використанням існуючих у вільному доступі тренувальних виборок. Модель НМ з глибоким навчанням вчиться передбачати відповідні мітки для нових вхідних даних на основі попередньо навчених прикладів.

У побудованій НМ, яка обрана в якості основи для підсистеми ідентифікації графічних об'єктів, передбачено кілька шарів нейронів (рис.2). З огляду на наведену на рис. 2 схему нейронів НМ математичної моделі системи розпізнавання образів з ознакою Хаара, найбільш ефективною для випадку розроблюваної системи буде НМ з глибоким навчанням. Обрана модель НМ дозволяє проводити навчання НМ для передбачення результатів на основі вхідних даних.

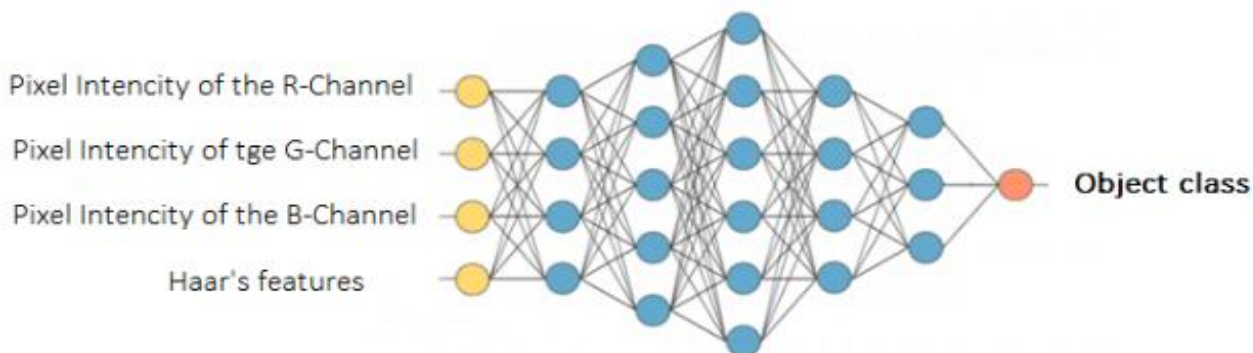


Рисунок 2 – Схема нейронних шарів прикладу нейромережі системи розпізнавання з ознакою Хаара

Вхідний шар НМ отримує початкові дані, включаючи інтенсивність кожного пікселя та ознаки Хаара для різних об'єктів (дерева, автомобілі, дороги, небо, елементи тротуару або огорожі, пішоходи тощо). Даний метод базується на використанні фільтрів Хаара для виявлення різниць у значеннях пікселів у певних областях зображення. В якості основної ідеї використання ознак Хаара для ідентифікації графічних об'єктів є те, що на зображеннях шукають певні характеристики об'єктів, такі як границі, лінії або текстури, які можуть бути присутні у зображенні. Наприклад, для задачі розпізнавання автомобілів можуть використовуватися ознаки Хаара для виявлення коліс, фар, або інших характерних деталей автомобіля. Використовуючи такі ознаки алгоритм на основі глибокої нейронної мережі відрізняємо на графічному фоні інформаційний вміст об'єкту ідентифікації (на рис.3 цей об'єкт – автомобіль) [10].

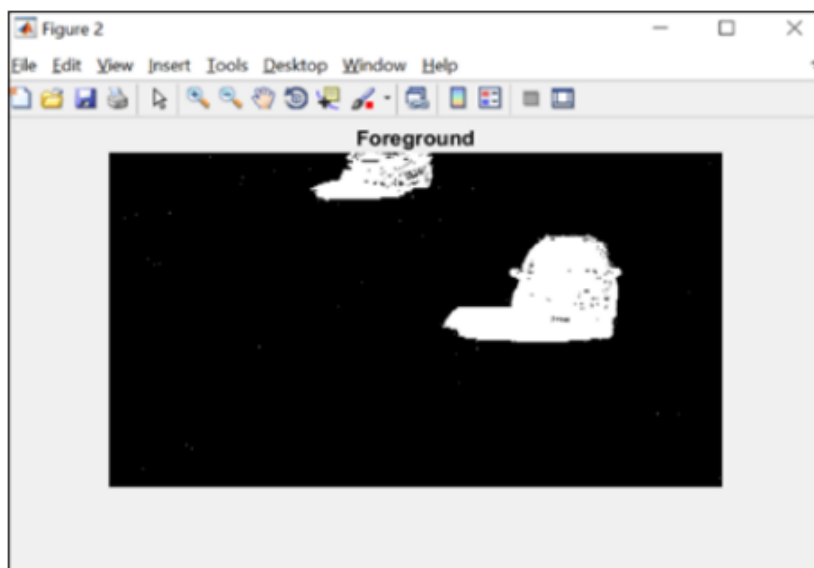


Рисунок 3 – Приклад виокремлення на графічному фоні інформаційний вміст об'єкту ідентифікації розробленим алгоритмом

Далі, ці дані передаються до першого прихованого шару НМ, де відбуваються математичні обчислення з використанням вхідних даних. Нарешті, вихідний шар формує фінальний результат – ідентифікацію типу об'єкта, присутнього на конкретному зображенні.

Для практичної реалізації розробленої математичної моделі системи розпізнавання об'єктів було вибрано середовище розробки MATLAB. У цьому середовищі доступна вбудована функція `vgg16()`, що реалізує в собі архітектуру глибокої нейронної мережі з 16 згорткових і повністю зв'язаних шарів нейронів, включаючи 13 згорткових шарів та 3 повністю зв'язаних шарів. Дану функцію можна використати для класифікації зображень в процесі розпізнавання. Згадана функція `vgg16`, що присутня в MATLAB, повертає об'єкт нейронної мережі, але вона не містить конкретного методу для обчислення відстані (метрики) між векторами ознак оброблених зображень. Через це для розв'язку подібних задач використовуються спеціальні методики обчислень [9]:

- евклідова відстань;
- квадрат евклідової відстані;
- відстань Чебишева;
- хемінгова відстань.
- мангеттенська відстань;

Для моделі системи ідентифікації ми будемо використовувати класичну відстань Евкліда в якості методу визначення відстані між точками тестових та еталонних зображень. Обчислення відстані за допомогою метрики Евкліда є найпростішим та ефективним способом отримання значення прямої відстані між двома точками у просторі. При ідентифікації

графічних зображень відстань Евкліда використовується, як міра порівняння виводу моделі з очікуваними значеннями, які були отримані під час навчання нейронної мережі, а також для подальшої класифікації зображень на основі їхніх властивостей.

Для розрахунку метрик відстаней точок застосовуємо вираз для визначення Евклідової відстані, наведений нижче:

$$\rho(x, x') = \sqrt{\sum_{i=1}^n (x_i - x'_i)^2}.$$

Відповідно, у наведеному вище виразі: x_i – перший n -вимірний вектор точок зображення; x'_i – другий n -вимірний вектор точок зображення. Дані інформаційні n -вимірні вектори в створеній моделі формуються за принципом, які можуть бути описані наступними виразами:

$$x = (x_1; x_2; \dots x_n). \quad x' = (x'_1; x'_2; \dots x'_n).$$

Тренування та тестування системи образів.

Високі показники системи розпізнавання образів можливо отримати лише тоді, коли її алгоритм моделі, побудований на базі глибокого навчання, пройде тренування на високоякісній відеовибірці. Для тренування нейронної мережі використовувалися еталонні дані відеобазы CamVid, що включає спеціальні мітки. Дана тренувальна база даних складається з колекції відеофрагментів з семантичними мітками відповідних класів об'єктів, доповнених метаданими. Основний набір міток еталонної бази CamVid асоціює кожен піксель з одним з 32 семантичних класів об'єктів, що ідентифікує система.

Однозначно визначені класи об'єктів ідентифікації моделі системи розпізнавання образів за типом:

- "Небо" ("Sky");
- "Будівля" ("Building");
- "Стовп" ("Pole");
- "Дорога" ("Road");
- "Тротуар" ("Pavement");
- "Дерево" ("Tree");
- "ЗнакСимволу" ("SignSymbol");
- "Паркан" ("Fence");
- "Автомобіль" ("Car");
- "Пішохід" ("Pedestrian");
- "Велосипедист" ("Bicyclist").

Для режиму тренування роздільна здатність графічних фремів була обмежена параметром 360x480 точок у відеовибірці.

Для того, щоб отримати середні значення для впорядкованого набору тестових даних можна використати функції `median()`, яка присутня в середовищі моделювання та тестування створеного алгоритму НМ Matlab. Значення для параметрів ініціалізації, функції втрат та тренування нейронної мережі наведені в табл. 1.

Таблиця 1 – Параметри ініціалізації функції тренування нейронної мережі

Назва аргумента функції <code>trainingOptions()</code>	Значення параметра
Momentum	0.9
InitialLearnRate	$1 \cdot 10^{-3}$
L2Regularisation	0,0005

З метою виконання кластерного аналізу та наступної фінальної класифікації розпізнаних та ідентифікованих графічних об'єктів в підсистемі ідентифікації слід визначати шкалу оцінок для класифікатора отриманих з відеоконтенту IP-камер графічних зображень. Для основи формування шкали оцінок в роботі використано інформаційну складову існуючих в комп'ютерній моделі RGB каналів, що визначають графічний фрейм у відповідному графічному форматі: $\{R \in (0; 255), G \in (0; 255), B \in (0; 255)\}$.

Еталонні вектори, з якими порівнюються отримані графічні кадри з IP-камер, сформовані відповідно до RGB-моделі кольору. Еталонні вектори в спеціалізованих джерелах часто згадаються, як опорні (назва-аналог). Крім того, такі вектори для створюваної системи розділені на 11 класів розпізнавання:

- будівлі, інфраструктурні об'єкти;
- небо, горизонт;
- дорожня інфраструктури та її елементи;
- дорожнє покриття;
- пішохідні доріжки, пандуси, пішоходи тощо;
- дерева, кущі, рослинність;
- дорожні, інформаційні знаки, світлофори;
- паркани та огорожі;
- автівки, великогабаритні засоби пересування та потяги;
- дорослі пішоходи та діти, тварини, візки;
- мотоцикли та мопеди, скутери.

Графічна карта RGB-каналів кольорів для відокремлення класів ідентифікованих графічних об'єктів містить певний набір значень каналу кольору для всіх існуючих а системі розпізнавання 11 основних класів об'єктів. В таблиці 2 наведені значення параметрів для розпізнавання базових об'єктів у основних класах ідентифікації на базі RGB-моделі.

Таблиця 2 – Базові об'єкти основних класів ідентифікації та їх характеристики

Назва класу розпізнавання	Характеристики класів розпізнавання (RGB-ознаки)
Sky	128 128 128
Building	128 0 0
Pole	192 192 192
Road	128 64 128
Pavement	60 40 222
Tree	128 128 0
SignSymbol	192 128 128
Fence	64 64 128
Car	64 0 128
Pedestrian	64 64 0
Bicyclist	0 128 192

Тренування одержаного алгоритму на навчальному наборі відеоданих базується на глибокому навчанні нейронної мережі. Такий підхід дає змогу ефективно відрізнити фон зоряюення від інформаційного вмісту об'єктів ідентифікації та розпізнавання. Математичну модель навчають на вибірці, що складається з 421 еталонних зображень, з яких 280 використовуються для подальшого тестування. Після тренування визначаються ваги об'єктів для кожного з відомих класів (таблиця 2). Наприклад, вага об'єктів класу «Sky» на графічних фреймах відеопотоку з IP-камери становить 0.318184709354742 при частоті 0,9266.

Висновки. Під час серії експериментів з розпізнаванням графічних образів, проведених в середовищі Matlab, алгоритм системи продемонстрував високу точність у роботі з графічною інформацією. В якості відеоданих для отримання результатів тестування був використаний

відеоряд, що містив базові об'єкти основних класів ідентифікації, перераховані в таблиці 2. Точність розпізнавання ідентифікованих об'єктів склала 96,38% для відеопотоку з IP-камери. Для моделі НМ використовувалися тестові дані, що включали 12 фрагментів, кожен з яких мав тривалість 68,35 секунди. Порівнюючи результати точності розпізнавання графічних об'єктів з максимально наближеними за математичним апаратом та доменною зоною використання можна зробити певні висновки за результати, показаними у табл. 3.

Таблиця 3 – Порівнювана точність розпізнавання графічних образів

Тип наборів тестових даних	Точність, %
Casting*	77,62
Defect*	97,88
Magnetic*	92,67
Tech*	94,23
Bridge*	99,02
Solar*	76,78
Базові об'єкти основних класів ідентифікації створеного алгоритму (таблиця 2)	96,38

Для розпізнавання графічних виборок на наборах тестових даних, позначених знаком «*» при навчанні використовувалися спеціалізовані для кожного набору навчальні вибірки [12]. В якості математичної основи для алгоритмів розпізнавання наборів тестових даних таблиці 3 використано нейронну мережу з глибоким навчанням VGG19. Така модель має на 3 шари нейронів більше, порівняно з тією НМ, яка використана в даній роботі. Відомим фактом є те, що зі збільшенням кількості шарів та нейронів у нейронній мережі, як правило, збільшуються обчислювальні витрати на розпізнавання образів. Отже, збіжаться і вимоги до апаратного забезпечення цільової платформи ідентифікації і розпізнавання графічних об'єктів.

Можна стверджувати, що комбінація НМ типу VGG16 з методом для обчислення відстані (метрики) між векторами ознак оброблюваних зображень на базі евклідової відстані дають прийнятну точність розпізнавання, достатню для застосування в ряді систем машинного зору. Отримані показники співставні з більш продуктивними, але складними алгоритмами на базі НМ VGG-16 при більшій обчислювальній складності останніх.

Перспективи вдосконалення. Для подальшого вдосконалення точності розпізнавання та ідентифікації графічних об'єктів можна використати інший метод обчислення метрики між векторами ознак зображень, наприклад, на основі чотирьох альтернатив: квадрату евклідової відстані; мангеттенська відстань; відстань Чебишева; хемінгова відстань. Хоча ці методи є обчислювально більш складними для потенційної апаратної бази застосування, але їх використання може підняти точність розпізнавання графічних образів в прикладному застосуванні.

Список використаних джерел

1. Moskalenko, V.V., & Korobov, A.H. (2019). Optimization parameters of intellectual identification system of objects on the terrain. *Radioelectronic and Computer Systems*, (2), 32–39. <https://doi.org/10.32620/reks.2016.2.05>.
2. Lakhno V., Akhmetov B., Chubaievskiy V., Desiatko A., Palaguta K., Blozva A., Chasnovskiy Y. (2021). Information Security Audit Method Based on the Use of a Neuro-Fuzzy System. *Lecture Notes in Networks and Systems*, 232 LNNS, pp. 171–184.
3. Herasina, O.V., & Korniienko, V.I. (2010). Alhorytmy hlobalnoi i lokalnoi optymizatsii v zadachi identyfikatsii skladnykh dynamichnykh system. *Systemy obrobky informatsii*, (6), 73–77.

4. Lakhno V., Kryvoruchko O., Desiatko A., Blozva A., Semidotska V. Development strategy model of the informational management logistic system of a commercial enterprise by neural network apparatus (2020). CEUR Workshop Proceedings, 2746, pp. 87–98.
5. Schuster H.G. (1992). Deterministic Chaos: Introduction and Recent Results. In: Thomas H. (eds) Nonlinear Dynamics in Solids. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-95650-8_2.
6. Ljung L., Andersson C., Tiels K., & Schön T.B. (2020). Deep Learning and System Identification. IFAC-PapersOnLine, 53(2), 1175–1181. <https://doi.org/10.1016/j.ifacol.2020.12.1329>
7. Nelles O. (2001). Nonlinear System Identification: From Classical Approaches to Neural and Fuzzy Models. Springer.
8. Goodfellow I., Bengio Y., & Courville A. (2016). Deep Learning. The MIT Press, London, England, 777 pages.
9. Lee, J.-H. (2019). Minimum Euclidean distance evaluation using deep neural networks. AEU – International Journal of Electronics and Communications, 112, 152964. <https://doi.org/10.1016/j.aeue.2019.152964>.
10. Sahun, A., Khaidurov, V., & Bobkov, V. (2024). Model of Graphic Object Identification in a Video Surveillance System based on a Neural Network. Y CPITS-2024: Cybersecurity Providing in Information and Telecommunication Systems. <https://ceur-ws.org/Vol-3654/short2.pdf>
11. Sturges, Paul. (2016). Oxford Brookes University. CamVid Database.
12. Ioannis D. Apostolopoulos, Mpesiana Tzani (2022). Industrial object, machine part and defect recognition towards fully automated industrial monitoring employing deep learning. The case of multilevel VGG19. Journal of Ambient Intelligence and Humanized Computing. <https://doi.org/10.1007/s12652-021-03688-7>.

Sahun Andii

PhD, Associated Professor of the Department of Computer Systems, Networks and Cybersecurity, National University of Life and Environmental Sciences of Ukraine

ORCID: <https://orcid.org/0000-0002-5151-9203>

E-mail: a.sagun@nubip.edu.ua

Panasko Olena

PhD, Associated Professor of the Department of Telecommunication and Robotics Systems and Cybersecurity,

Cherkasy State Technological University

ORCID: <https://orcid.org/0000-0002-0510-7742>

E-mail: [lena.pa@ukr.net](mailto:lana.pa@ukr.net)

IMAGE RECOGNITION SYSTEM BASED ON A NEURAL NETWORK WITH DEEP LEARNING

Abstract. An object recognition system based on deep learning neural networks has been developed. Such an image recognition system is capable of accurately and quickly recognizing images known to it and similar to them in video content obtained from IP surveillance cameras. Depending on the video shooting conditions and the viewing angle of the IP camera, the algorithm of this system achieves a recognition accuracy of 96.38%. This image recognition rate is practically constant for 11 classes of identification and recognition objects. Such high results are achieved through the use of the CamVid video database as a training sample for the neural network. This database is based on 421 training and 280 test video images. The recognition system model provides for the optimization of training and identification function parameters, as well as changes in the method of measuring the distance between feature vectors (point distance measurement metrics).

Keywords: neural network, deep learning network, feature classifier, support vector, graphic frame, machine vision.

УДК 378:004

Nikitenko Yevheniy

Ph.D., Associate Professor of the Department of Computer Systems, Networks and Cybersecurity, National University of Life and Environmental Sciences of Ukraine

ORCID: <https://orcid.org/0000-0002-9222-644X>

E-mail: ev.nikitenko@nubip.edu.ua

Ryndych Yevhen

Ph.D, Associate Professor,

National University "Chernihiv Polytechnic"

ORCID: <https://orcid.org/0000-0002-2723-4144>

E-mail: zkasterwork@gmail.com

Hoida Ivan

Student of the Department of Computer Systems, Networks and Cybersecurity Department, National University of Life and Environmental Sciences of Ukraine

E-mail: vanyahoyda@gmail.com

DETECTION OF COMMUNITIES IN SOCIAL NETWORKS

Abstract. *One of the main threats is malicious programs (bots), fake accounts capable of imitating human behavior. At the moment, bots create a lot of problems, both for ordinary users and for those who use social networks to conduct a marketing campaign or conduct social research. Using bot profiles in social networks greatly distorts information about the real benefits and interests of portal users. Therefore, it is necessary to determine which users of the social network are programmed, and to be able to divide the flow of data into that generated by bots and by humans. The threatening scale of the use of social bots requires the creation of effective algorithms for their detection. Internet platforms and social services themselves are not too concerned about this problem. As a result, both ordinary users who organize various communities and companies that promote goods, brands, and services through social networks suffer. Thus, the task of recognizing malicious accounts in social networks and combating them remains relevant in the issue of cyber security. The solution to the problem will be the development of network analysis methods that are designed to identify and classify communities in social networks, assess their connectivity, degree of trust, as well as develop effective algorithms for detecting malicious accounts. The purpose of the work is to develop an improved algorithm for detecting malicious accounts in social networks, which is based on the study of modern methods.*

Keywords: *social bots, detection, malicious accounts, social networks, cluster analysis.*

Introduction. Currently, social networks are an integral part of most spheres of human life, integrating almost all existing Internet sources. They effectively structure users from political or religious views, interests and passions, affecting almost all segments of the population, and are a powerful tool for self-organization of both individual groups and society as a whole. Social networks, which unite 40% of the planet's population every day, have become not only a means of communication, but also a great source of information, entertainment hosting, a commercial platform with a set of effective tools for distributing services and goods. It is natural that interested persons seek to use such limitless potential for profit and to achieve their far from noble goals.

Over the past decade, social networks have grown significantly along with the speed of Web2.0 application development. Millions of people are registered on social networks such as Facebook, Twitter, LinkedIn, etc. For example, Facebook has about 2,94 billion users as of March 2023. As the number of users only increases, the complexity of detecting social communities also increases. Detecting and clustering data and users in social media communities is an important and challenging problem in creating effective marketing models in the changing and evolving social systems. Such marketing models are based on individual product purchase decisions influenced by friends and acquaintances. This is leading to new models that treat users as part of an online social network rather than traditionally as marketing individuals.

Social connections play an important role in determining user behaviour. For example, a user may buy a product that a friend of theirs has recently purchased. This phenomenon is called social

influence and is used to study how strongly the action of one user can provoke others to take certain actions.

Literature review. Network science, which began its journey in the 1700s with Euler's Seven Bridges of Königsberg, has gone through many stages, including the emergence of graph theory, sociograms (graphical representations of social connections), and the emergence of social network analysis, culminating in a boom and establishment as a discipline. Only after some of the most recent important discoveries, such as the development of scale-free networks, did the first social networking sites appear within a decade, with Facebook accounting for 51% of the world's active online users [1].

Social network analysis focuses on connections rather than actors. As a rule, a social network is described by a graph or matrix of relationships (Figure 1) [2].

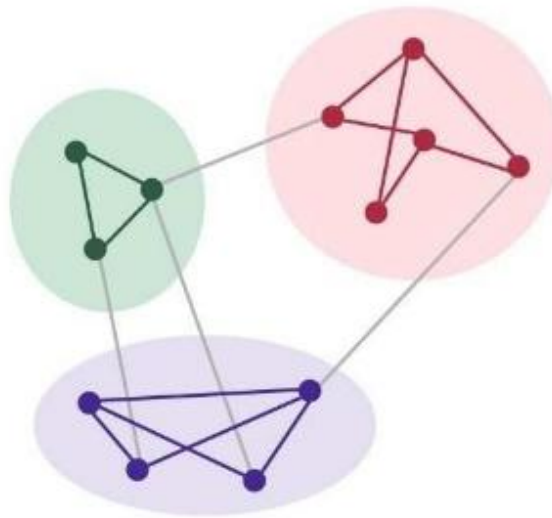


Figure 1 – Example of a community structure

The main areas of social media research are as follows:

1. Identification of communities in online social networks.
2. Search for key nodes in any society.
3. Building a set of nodes that are used to spread influence on an online social network.

The significance of social networks is based on the fact that, on the one hand, they are the subject of socialisation of people, and on the other hand, they are the most powerful and accessible political, ideological and economic tool [3]. The methods of social media analysis include the following:

- methods of graph theory, in particular, the study of oriented graphs and matrices representing them, used to study the structural relationships of a network participant;
- methods of finding local properties of participants, for example, centrality, influence, position, belonging to certain subgroups;
- methods for determining the equivalence of participants, including their structural equivalence;
- probabilistic models, including Markov process models.

Works [4] and [5] study the problem of spam bot detection using a popular social network as an example. The authors propose to distinguish ordinary users from malware using machine learning classification. Traditional classification algorithms are used for this purpose: decision trees, neural networks, and a naive Bayesian classifier. The number of subscribed and read users, as well as graph-oriented user relationships, were taken as features.

A new methodology for identifying communities was put forward by the authors of the article [6]. It consists in the construction and analysis of scalograms that reflect the interaction of users in a

social network. Scalograms reveal a lot of hidden information about the nature of non-stationary processes. They are used in various fields: predicting the consequences of an earthquake, analysing earth movements, analysing the resistance of buildings to hurricanes, analysing the stability of bridges, etc.

Purpose. This article focuses on community detection as one of the most promising tools for world-class marketing research. The task of researchers is to find groups of users who have similar interests or a high level of connections between them. Several existing community detection methods have been studied. The research extended the CNM (Clauaset-Newman-Moore) [9] algorithm, to use the Jaccard similarity measure. First, a graph from the original network is output, and as a result, it is labelled as a similarity social network or a virtual social network. The method based on this algorithm shows that by pre-processing the original network, it is possible to obtain higher quality community structures. Another algorithm for building strong communities online is the ECD-Jaccard algorithm (ECD – «Enhanced Community Detection»), which enriches a virtual social network with edge weights, then applies a quality-optimised version of the CNM algorithm to detect communities. This algorithm has shown that by pre-processing the original network, it is possible to obtain higher quality community structures.

Results and Discussion. A study was conducted to summarise the state of the environment affecting social influence in online social networks, and different approaches were evaluated, which are combined into an original categorisation system to understand commonalities and distinguish differences.

Clustering was used to separate groups from the total population of users. The advantages of cluster analysis are that it allows you to split objects not by one parameter, but by a whole set of features. In addition, cluster analysis, unlike most mathematical and statistical methods, does not impose any restrictions on the type of objects under consideration and allows you to consider a variety of source data of an almost arbitrary nature. Cluster analysis allows you to consider a fairly large amount of information and dramatically reduce and compress large amounts of information, making them compact and visual. Moreover, clustering can be used cyclically. In this case, the research is carried out until the required results are achieved. In this case, each cycle can provide information that can greatly change the focus and approaches to further application of cluster analysis.

The task of cluster analysis is to divide the set of objects X into m (m – integer) clusters (subsets) Q_1, Q_2, \dots, Q_m based on the data contained in the set G so that each object Q_j belongs to one and only one subset of the division. And objects belonging to the same cluster were similar, while objects belonging to different clusters were heterogeneous. The solution to the cluster analysis problem is a partition that satisfies some optimality criterion.

It is possible to define clustering in the context of a real social network by grouping people with high friendships internally and scattered friendships externally. With clustering, you can identify interest groups or communities that share common properties that can be used to learn about these groups and understand their behaviour. For example, Amazon provides users with recommendations based on their purchase history. Twitter also recommends new friends to follow members based on several factors, such as being a friend of the same user.

It is difficult to imagine dividing social media user accounts with a large number of unequal criteria into groups with two degrees of membership of 0 or 1. It is more natural to use a partial membership in the range from 0 to 1, which will allow users whose characteristics are on the boundaries between several clusters to belong to them with different degrees. Therefore, the fuzzy clustering method was chosen as the main method for dividing user accounts into groups. The steps of this algorithm are as follows:

1. The initial information for clustering is an observation matrix U of size $n \times k$ (1):

$$U = \begin{bmatrix} u_{11} & \cdots & u_{1k} \\ \cdots & \cdots & \cdots \\ u_{n1} & \cdots & u_{nk} \end{bmatrix}, \quad (1)$$

where n – is the number of users, k – is the number of features.

2. Using the matrix U , we can find the value of the fuzzy error criterion (2):

$$E^2(X, U) = \sum_{i=1}^N \sum_{k=1}^K U_{ik} \|x_i^{(k)} - c_k\|^2, \tag{2}$$

where c_k – «centre of mass» of the fuzzy cluster k (3):

$$c_k = \sum_{i=1}^N U_{ki} x_i. \tag{3}$$

3. After regrouping the objects to reduce this value of the fuzzy error criterion, it is necessary to return to step 2 until the changes in the matrix U become insignificant.

The minimal spanning tree algorithm [8] first builds a minimal spanning tree on the graph and then sequentially removes the edges with the highest weight. The figure shows the minimum spanning tree obtained for nine objects.

By deleting the link labelled CD with a length of 6 units (the edge with the maximum distance), we obtain two clusters: $\{A, B, C\}$ and $\{D, E, F, G, H, I\}$. The second cluster can be further divided into two more clusters by removing the edge EF , which has a length of 4.5 units (Fig. 2).

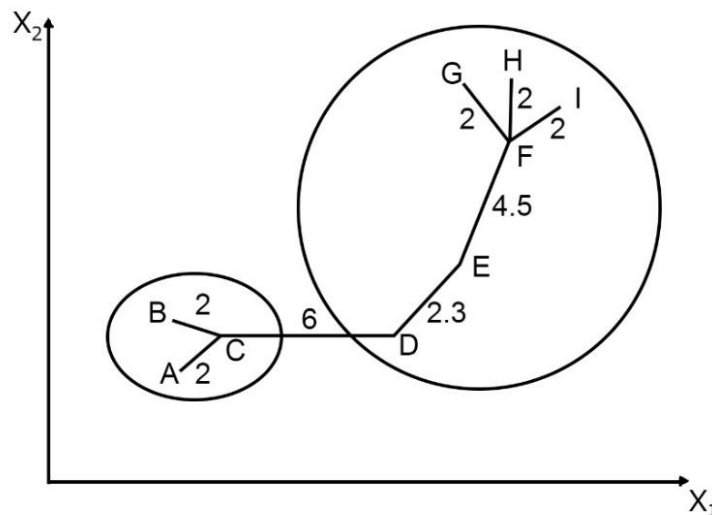


Figure 2 – Illustration of the minimum spanning tree algorithm

In the n -dimensional metric feature space, the distance between two objects is considered to be a measure of the «similarity» of two objects.

The Mankowski distance family is a very common class of distance functions and can be represented as follows (4):

$$D(p_i, p_j) = w \sum (p_i - p_j)w, \tag{4}$$

where w – is a parameter with a value greater than or equal to 1. Based on the value of w , different distance functions can be represented, such as the Hamming distance ($w = 1$), the Euclidean distance ($w = 2$) and the Chebyshev distance ($w = \infty$). Other similarity measures are the cosine correlation measure and the Jaccard measure.

To investigate the proposed approaches, experimental information was used for both synthetic and popular real-world datasets. The modularity of Q and the normalised mutual information used previously in other experiments were used as evaluation metrics to show the performance and accuracy of the proposed algorithms. The maximum modularity Q_{max} was estimated using the CNM

algorithm provided by Clauset et al. [9]. The normalisation of mutual information was also calculated using the formula of Danon et al. [7].

Based on the data in the developed software application, a network was generated and the results were studied.

For the Facebook network, the maximum modularity value for similarity-CNM outperforms the original CNM algorithm by more than 67%. And for the American football network, the maximum value of modularity for similarity-CNM exceeds the value obtained by the original CNM algorithm by more than 23%. The number of steps to achieve Q_{max} is lower for CNM than for similarity-CNM.

For Facebook, the number of steps when running the original CNM is 9879 steps, while in the similarity-CNM it is reduced to 8843 steps. For the American football network, the number of steps is also reduced from 108 steps for the original CNM to 100 steps for the similarity-CNM algorithm.

In the course of the study, software was developed that implements existing and proposed similarity-CNM and ECD-Jaccard algorithms to provide a better community structure. The similarity-CNM algorithm detects similarities between nodes and creates a corresponding virtual social network. Similarly, the ECD-Jaccard algorithm also calculates the similarity of nodes as a preprocessing step, but these values are then assigned as weights to the edge of the network, resulting in a weighted virtual social network, unlike the similarity-CNM approach, which has no weights. The CNM algorithm is then used to detect communities in both approaches. Experimental analysis shows that these preprocessing methods have an advantage over the original CNM algorithm in terms of community modularity.

Conclusions. The article proposes a solution relevant to the cybersecurity of social networks, which involves the development of a new method for detecting social communities. The following analytical and practical results were obtained in the course of solving the task:

1. An analysis of modern detection methods using various pattern recognition approaches and mathematical models. The results of testing existing methods were compared, which showed that most of them work with a certain amount of data. In this regard, there is a need to develop a new algorithm for detecting social communities, which allows to reduce the number of unnecessary user checks.

2. An analysis of clustering methods, taking into account their advantages and disadvantages, is carried out in order to select the most effective ones for solving the tasks set in the paper.

3. A method for improving the quality of account sorting with further analysis of the results and decision-making for individual user groups is proposed.

4. A combined method for tracking the behaviour of suspicious accounts in the network is proposed, which can significantly reduce the number of additional checks.

References

1. C. K. Leung, F. Jiang, T. W. Poon, P.-É. Crevier. Big Data Analytics of Social Network Data: Who Cares Most About You on Facebook? // Springer, Studies in Big Data, – 2018. – vol. 27, pp. 1-17.
2. Stanford Large Network Dataset Collection [Online resource]. – URL: <http://snap.stanford.edu/data/>
3. Alaa Alden Al Mohamed, Sobhi Al Mohamed and Moustafa Zino. Application of fuzzy multicriteria decision-making model in selecting pandemic hospital site// Future Business Journal, Springer. – 2023, vol. 9, №14. – pp. 1–22.
4. Vaishali Chawla, Yatin Kapoor. A hybrid framework for bot detection on twitter: Fusing digital DNA with BERT // Multimedia Tools and Applications, – 2023. – vol.82, №20. – pp. 30831-30854.
5. Amit Pratap Singh, Maitreyee Dutta. An Efficient Classifier for Spam Detection in Social Network // International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2019. – vol.9, №1. – pp. 2323–2328.

6. Manzoor Hussain. Digital divide in the use of social networking sites: a study of P.G. students (gender-wise) through scalogram analysis// International Journal of Research in Economics and Social Sciences (IJRESS). – Vol. 7. – Issue 9, September – 2017, pp. 527–536.
7. Andreas Kanavos, Nikos Antonopoulos, Ioannis Karamitsos and Phivos Mylonas. A Comparative Analysis of Tweet Analysis Algorithms Using Natural Language Processing and Machine Learning Models// 2023 18th International Workshop on Semantic and Social Media Adaptation & Personalization (SMAP), 2023.
8. Christopher Braker, Stavros Shiaeles, Gueltoum Bendiab, Nick Savage, Konstantinos Limniotis. BOTSPOT: DEEP LEARNING CLASSIFICATION OF BOT ACCOUNTS WITHIN TWITTER // Internet of Things, Smart Spaces and Next Generation Networks and Systems, – 2021. – pp. 165–175.
9. Martijn Gösgens, Remco van der Hofstad and Nelly Litvak. The projection method: a unified formalism for community detection// Frontiers in Complex Systems, – 2024. – pp. 1-18.

Нікітенко Євгеній Васильович

кандидат технічних наук, доцент кафедри комп'ютерних систем, мереж та кібербезпеки, Національний університет біоресурсів та природокористування України

ORCID: <https://orcid.org/0000-0002-9222-644X>

E-mail: ev.nikitenko@nubip.edu.ua

Риндич Євген Володимирович

кандидат технічних наук, доцент, Національний університет «Чернігівська політехніка»,

ORCID: <https://orcid.org/0000-0002-2723-4144>

E-mail: zkasterwork@gmail.com

Гойда Іван Сергійович

студент кафедри комп'ютерних систем, мереж та кібербезпеки, Національний університет біоресурсів та природокористування України

ORCID: <https://orcid.org/0000-0002-2723-4144>

E-mail: vanyahoyda@gmail.com

ВИЯВЛЕННЯ СПІЛЬНОТ У СОЦІАЛЬНИХ МЕРЕЖАХ

Анотація. Однією з головних загроз є шкідливі програми (боти) – фейкові акаунти, здатні імітувати поведінку людини. На даний момент боти створюють багато проблем як для звичайних користувачів, так і для тих, хто використовує соціальні мережі для проведення маркетингових кампаній або соціальних досліджень. Використання профілів-ботів у соціальних мережах значно спотворює інформацію про реальні переваги та інтереси користувачів порталу. Тому необхідно визначити, які користувачі соціальної мережі є запрограмованими, і мати можливість розділити потік даних на той, що генерується ботами, і той, що генерується людьми. Загрозливий масштаб використання соціальних ботів вимагає створення ефективних алгоритмів для їх виявлення. Інтернет-платформи та соціальні сервіси самі по собі не надто переймаються цією проблемою. В результаті страждають як звичайні користувачі, які організують різні спільноти, так і компанії, які просувають товари, бренди та послуги через соціальні мережі. Таким чином, завдання розпізнавання зловмисних акаунтів у соціальних мережах та боротьби з ними залишається актуальним у питанні кібербезпеки. Рішенням проблеми стане розробка методів мережевого аналізу, які призначені для ідентифікації та класифікації спільнот у соціальних мережах, оцінки їх зв'язності, ступеня довіри, а також розробка ефективних алгоритмів виявлення зловмисних акаунтів. Метою роботи є розробка вдосконаленого алгоритму виявлення зловмисних акаунтів у соціальних мережах, який базується на вивченні сучасних методів.

Ключові слова: соціальні боти, виявлення, зловмисні акаунти, соціальні мережі, кластерний аналіз.

УДК 378:004

Смолій Вікторія Миколаївна

Доктор технічних наук, професор, професор кафедри інформаційних систем та технологій,
Національний університет біоресурсів і природокористування України

ORCID: <https://orcid.org/0000-0002-1268-7837>E-mail: vmsmolij@nubip.edu.ua**Смолій Натан Вікторович**

Студент кафедри інформаційних систем та технологій,
НТУУ «КПІ імені Ігоря Сікорського», м. Київ

ORCID: <https://orcid.org/0009-0002-3763-6726>E-mail: hoibbitizukrainy@gmail.com**Янцевич Антон Олександрович**

Студент кафедри інформаційних систем та технологій,
Національний університет біоресурсів і природокористування України

E-mail: ist21-a.yantsevych@nubip.edu.ua**Іскоростенський Олексій Олександрович**

Студент кафедри інформаційних систем та технологій,
Національний університет біоресурсів і природокористування України

E-mail: ist21-o.iskorostenskyi@nubip.edu.ua

ПРОЄКТУВАННЯ КОМП'ЮТЕРНОЇ СИСТЕМИ АНАЛІЗУ СТАНУ РОСЛИН

Анотація. Робота присвячена актуальній тематиці відновлення аграрного господарства на звільнені території, що пов'язано з ризиками замінування, забруднення, захламлення, подолання наслідків екологічних катастроф, псування ґрунтів, тощо територій, необхідності проведення зазначених робіт та проведення досліджень щодо можливості вирощувати рослини в поствоєнних умовах та перспективи ефективного вирощування рослин в зоні ризикованого землеробства. Мета статті полягає в висвітленні створення інформаційної системи аналізу стану рослин, призначеної для здійснення агропромислової діяльності щодо вирощування біоенергетичних та злакових культур, яка забезпечує збір, обробку та рекомендації щодо стану рослин, що в свою чергу дозволяє визначити доцільність (раціональність) вирощування певного сорту рослин в певному регіоні, запланувати заходи польових робіт, обробки та зрошування рослин.

В матеріалі статті обґрунтовано необхідність проведення численних експериментів щодо виявлення тенденцій розвитку рослин в умовах досліджуваних регіонів. Пропонується за зібраними статистичними даними визначити доречну (мінімально необхідну) кількість біоенергетичних та злакових культур, необхідну статистику щодо опорних показників нормального розвитку рослин, індикаторів плодоносіння. Запропоновано спосіб отримання інформації про поточний стан рослин з детальною фіксацією умов розвитку та результатів врожайності, спираючись на які засобами абстрагування слід розробити підґрунтя дорадницької інформаційної системи, побудованої відповідними методами аналізу і синтезу. Наведено математичне описання визначення ефекту середовища та умов вирощування (вплив на рослини) за будь-яким біопараметром: за масою рослини, довжиною кореневої або стеблової системи, кількістю ушкоджених рослин або кількістю сходів тощо, призначеного для оцінки придатності рослин для аграрного бізнесу у визначеному територіальному регіоні. Запропоновані Use-case-, BPMN- та Dataflow- діаграми, які ілюструють з певним ступенем деталізації послідовність дій і обробку інформації для процесу дослідження аналізу стану рослин, що становить основу для побудови інформаційної системи аналізу станів рослин. До перспектив подальших розвідок слід віднести необхідність більш детальної проробки кожного з компонентів запропонованої інформаційної системи зі з'ясуванням механізмів генерації висновків щодо доцільності вирощування рослин у певному регіоні та компонентів дорадництва щодо організації діяльності у визначеному аграрному секторі економіки.

Ключові слова: післявоєнне відновлення, агропромислова галузь, сільськогосподарські культури, визначення стану рослин, розрахунок параметрів розвитку, агродрони, задача ідентифікації параметрів рослин, інформаційна система, Use-case, Business Process Model and Notation та Dataflow діаграми.

Вступ. Післявоєнна відбудова України потребуватиме повернення економіки від військових потреб (залишаючи повне з надлишком! для сталого розвитку забезпечення потреб Армії у мирний час) до потреб бізнесів і економіки в цілому. Безповоротним і вхідним впливом

у цій ситуації є повне руйнування цивільного, індустріального, аграрного, інфраструктурного, кадрового, фінансового, тощо загалом будь-якого сегменту резерву Східних та Південних областей України і Республіки Крим. Тобто є в наявності повне знищення країною – агресоркою росією всіх елементів життя, промисловості, сільського господарства, житлового і соціального фонду на окупованих територіях, яка залишила після себе вижену і заміновану багатьма шарами землю. Єдиним виходом в такій ситуації є створення умов для малого бізнесу, як важеля відновлення територіальних громад і стимулювання відновлення демографічної ситуації на звільнених територіях.

Майже всі громади на звільнених територіях будуть перебувати в однаково поганих і ризикованих початкових точках розвитку бізнесів, що буде потребувати величезних обсягів залучення капіталів, що було неможливим і в довоєнний час, особливо для депресивних регіонів. Особливу увагу, як рушійній силі, слід приділити саме малому багатопрофільному рентабельному бізнесу, який би можливо було тиражувати від громади до громади та який би забезпечив виживання-відновлення прошарку населення з низьким і нижче середнього рівнем статків.

Основним акцентом розвитку слід виділи саме аграрний бізнес, спрямований на забезпечення населення продуктами харчування, джерелом палива та підґрунтям для розвитку громад зі знищеними промисловими, соціальними та цивільними об'єктами. Самі звільнені території характеризуються надвисоким ступенем пошкодженості через мінування, бомбардування та спричинені окупантами екологічні катастрофи надвеликих масштабів. Окрім того кліматичні умови звільнених територій Сходу та Півдня належать до зон ризикованого землеробства [1, 11]

Аналіз останніх досліджень та публікацій. При дослідженні визначеної предметної галузі значну увагу слід приділити визначенню який саме агробізнес слід (доцільно) розвивати в зазначених регіонах, але спільним є необхідність використання надсучасних технологій розмінування, повернення родючості ґрунтів і організації бізнесу, вибору безпосередньо рослин для здійснення господарювання. В основу організації зазначеного малого багатопрофільного рентабельного бізнесу слід покласти створення продукту. В якості продукту, враховуючи розташування зазначених регіонів у зонах ризикованого землеробства [2], слід розглянути об'єкт – біоенергетичні та злакові культури, що в свою чергу дозволить забезпечити продуктами харчування людей і тварин та організувати виробництва біопалива, що забезпечить потреби зазначених регіонів. Площа зазначених територій буде збільшуватись, що зручно для подальшого масштабування бізнесу, за умови узгодження динамік окупності і необхідного масштабування. Слід звернути увагу на «інтелектуалізацію» вибору груп і сортів рослин, оскільки продуктивність вирощування залежить від великої об'єктивної низки факторів припускаючи менший вплив суб'єктивної складової, обумовленої невеликою кількістю залучених фахівців через війну [2, 3].

Згідно документів [4, 5] є регламентований план дій з відновлення деокупованих територій, який системно враховує військові, адміністративні та інші аспекти. З точки зору системності процесу відновлення економіки на зазначених територіях, виділяють наступні задачі в організації агробізнесу: розмінування, дослідження ґрунтів, що обумовлено подоланням наслідків екологічних катастроф штучно створених ворогом, здійснення (за потреби) заходів щодо відновлення та/або підвищення родючості ґрунтів, збір, обробка та узагальнення даних щодо контролю етапів розвитку рослин у певному регіоні (області, районі, ділянки), систематизації для них відповідних заходів обробки, живлення та зрошення. Здійснення планування робіт та інтегроване проведення потрібних операцій, потребу в яких визначатиме оброблені стохастичні дані з проведення діяльності у певний період розгортання зазначеної технології відновлення агробізнесу. Великий обсяг первинних даних, зібраний у зазначений період, дозволить екстрапольовати отримані судження та результати на певні господарські області, здійснивши планування та проведення польових робіт при забезпеченні оптимальності витрат для досягнення ефективності господарської діяльності [12]. При цьому актуальною є оптимізація зрошень та обробок рослин, при безперервній динаміці

спостереження та аналізу розвитку рослин (стан розвитку, наявність хвороб та шкідників, необхідність підживлення та її поєднання зі зрошенням, потреба у захисті рослин) [10].

Мета дослідження полягає в створенні інформаційної системи аналізу стану рослин, призначеної для здійснення агропромислової діяльності щодо вирощування біоенергетичних та злакових культур, яка забезпечує збір, обробку та рекомендації щодо стану рослин, що в свою чергу дозволяє визначити доцільність (раціональність) вирощування певного сорту рослин в певному регіоні, запланувати заходи польових робіт, обробки та зрошення рослин.

До основних завдань досліджень слід віднести: визначення піддослідних рослин, узгодження методики оцінки параметрів розвитку, з'ясування способу отримання даних про стан рослин, визначення стану розвитку рослин, фіксація наявності та величини умов зовнішнього середовища, які вплинули на стан рослини, встановлення факту доцільності промислового вирощування піддослідної рослини, обґрунтування способу отримання інформації про рослину для оперативного і достатнього аналізу з метою визначення придатності для промислового вирощування, генерація інформаційної системи для реалізації визначених задач.

Матеріали і методи дослідження. Слід зауважити про необхідність проведення численних експериментів щодо виявлення тенденцій розвитку рослин в умовах досліджуваних регіонів. За зібраними статистичними даними визначити доречну (мінімально необхідну) кількість біоенергетичних та злакових культур, необхідну статистику щодо опорних показників нормального розвитку рослин, індикаторів плодоносіння. Обов'язково слід обґрунтувати спосіб отримання інформації про поточний стан рослин з детальною фіксацією умов розвитку та результатів врожайності, спираючись на які засобами абстрагування треба розробити підґрунтя дорадницької інформаційної системи, побудованої відповідними методами аналізу і синтезу.

Результати дослідження та їх обговорення. Згідно [5,6] для зернових культур оцінка зростання здійснюється за системою ВВСН, де базовою є шкала Задокса. Відповідно до ВВСН До основних етапів росту рослин відносять: проростання, розвиток листя, кушіння, подовження стебла, розвиток вегетативних органів рослин, поява суцвіть, цвітіння, утворення плодів, дозрівання плодів, відмирання.

До визначення необхідності зрошення та живлення, слід зібрати статистику щодо впливу найнесприятливіших умов на стан рослини, її розвиток та в наступному на її врожайність, зокрема дослідити розвиток квітів, ділення клітин, розростання та ендоредуплікацію клітин плоду та дозрівання.

Визначення стану посівів здійснюється на підставі використання об'єктивних чисельних та якісних показників, регламентованих нормативними документами на законодавчому рівні [7]. Передбачено системне дослідження точності та рівномірності висівання, аналіз розподілу густини рослин по площі поля, а також кількість, видовий стан, фази розвитку, розподіл по площі бур'янів, тощо.

Безпосередньо моніторинг посівів здійснюється способами, наведеними у [8], до яких відносять: супутниковий, з використанням дронів та листову діагностику рослин і аналіз ґрунтів. Переваги і недоліки кожного зрозумілі, але перспективними є перші два, які є базою для обчислення NDVI індексу.

Потрібно зібрати статистичні вибірки, для яких визначити відхилення від еталонних значень. В даному контексті після проведення вимірювань для кожного з досліджуваних варіантів обчислюють середню довжину надземної і кореневої частин $\bar{x} \pm m$, де m – помилка середнього арифметичного, яку визначають так:

$$m = \sqrt{\frac{\sigma^2}{N}}, \quad (1)$$

де N – кількість результатів; σ^2 – дисперсія, яку визначають за виразом:

$$\sigma^2 = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}. \quad (2)$$

Достовірність різниці середніх арифметичних t розраховується за критерієм Стьюдента-Фішера:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{m_1^2 + m_2^2}}, \quad (3)$$

де \bar{x}_1 – середнє арифметичне значення показника в контрольному досліді; \bar{x}_2 – середнє арифметичне значення показника у досліджуваному варіанті; m_1 – помилка середнього арифметичного в контрольному досліді; m_2 – те ж у досліджуваному варіанті.

Якщо фактично встановлена величина t більше або дорівнює критичному (стандартному) значенню t_{st} роблять висновок про існування статистично достовірної різниці між середніми арифметичними у досліджуваному та контрольному варіанті. Якщо ж фактична величина t менша за t_{st} , різницю між середніми вважають статистично недостовірною.

Відсутність статистично достовірної різниці між середніми значеннями біопараметра у контрольному та досліджуваному варіанті свідчить про відсутність значних змін ростових процесів у біоіндикаторів, в порівнянні з контрольним варіантом.

Тобто умови та фактори у досліджуваному варіанті майже такі самі, як і в контрольному досліді та не мають пригнічуючих (стимулюючих) впливів. І навпаки, статистично достовірна різниця між варіантом та контрольним дослідом вказує на те, що для дослідного зразка є наявності пригнічуючі (стимулюючі) впливи.

Ефект середовища та умов вирощування (вплив на рослини) визначається у відсотках за будь-яким біопараметром: за масою рослини, довжиною кореневої або стеблової системи, кількістю ушкоджених рослин або кількістю сходів тощо. Розраховується за формулою:

$$\Phi E = \frac{M_o - M_x}{M_o} \cdot 100, \% \quad (4)$$

де M_o – значення біопараметра (маса рослин, висота паростків, довжина корінців та ін.) у «еталонної» рослини; M_x – значення аналогічного біопараметра у дослідної рослини.

Виходячи з наявності обладнання, першочергово слід визначитись з методом отримання інформації про стан рослин, а далі для кожної рослини слід визначити параметри розвитку за зображенням рослини. Важливо також узагальнити отримані по рослинах результати у статистичні виборки. Тобто актуальною є задача детекції рослин, вирішення якої дозволить визначити необхідність проведення процедур зрошення, живлення, тощо та в подальшому обумовить рішення щодо раціональності впровадження дослідної культури в асортимент ведення господарчої діяльності.

Виходячи з викладеного, на Use-case діаграмі інформаційної системи аналізу стану рослин (рис. 1) передбачено наявність чотирьох акторів: замовник, експерт, рослина та обладнання. В якості замовника виступає зацікавлена особа, мається на увазі суб'єкт господарської діяльності, зацікавлений в налагодженні аграрної діяльності в умовах деокупації територій. Експертом можуть бути споріднені фахівці з більш тривалим досвідом роботи в аналогічних ситуаціях та з накопиченими статистичними даними щодо процесу діяльності взагалі і окремо по суб'єкту господарювання зокрема. В якості актора-обладнання може виступати як агродрон, так і людина, виконуючи методику ідентифікації стану рослини, поєднані іншими акторами прецедентами оцінки та надання рекомендацій щодо живлення та зрошення, спостереження за навколишнім середовищем та впливу на параметри навколишнього середовища, зокрема температуру, тиск, склад повітря та ґрунту, через використання інтелектуальних теплиць [9].

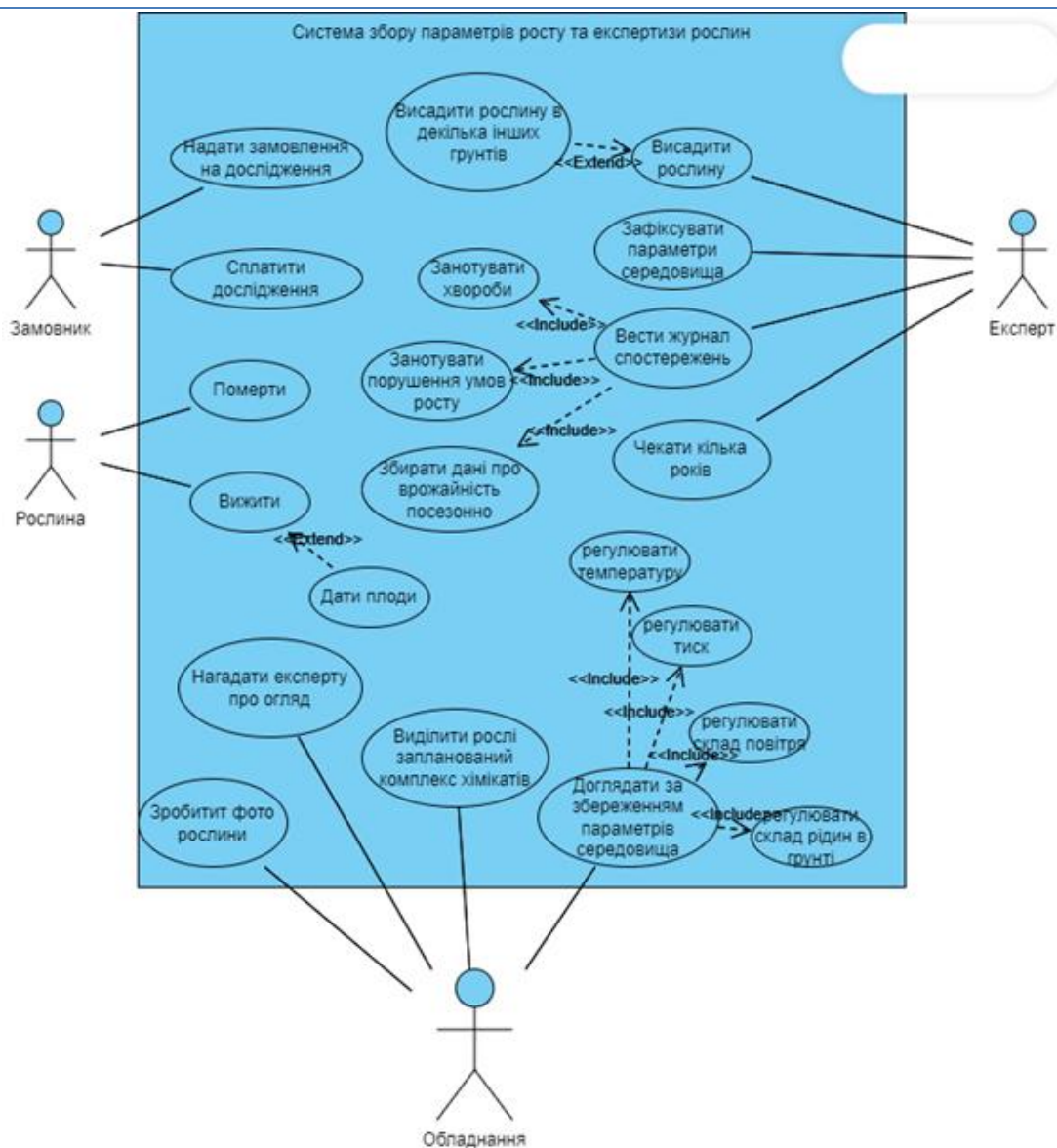


Рисунок 1 – Use-case діаграма системи збору параметрів росту та експертизи рослин

Окрему увагу при розгляді прецедентів (права частина рис.1), пов'язаних з умовним актором «експерт», приділено процесу збору статистичних даних про об'єкт господарювання щодо отримання, обробки, зберігання статистичних даних про параметри розвитку та врожайність рослин, та про фіксацію залежностей від умов навколишнього середовища (клімату), особливостей ґрунтів та певних методик операцій зрошення та обробок, що в свою чергу дозволяє отримати математичне підґрунтя для побудови продукційних залежностей для експертної системи у межах створюваної інформаційної системи аналізу стану рослин.

Слід звернути увагу на отримання репрезентативних результатів для обробки статистичних даних, що досягається горизонтом планування експериментів щодо накопичення статистичних даних у декілька сезонів плодоношення.

Згідно [4] акторами в Use-case діаграмі виступають військові адміністрації населених пунктів, органи місцевого самоврядування, сільськогосподарські підприємства, Держекоінспекція, підприємства, установи та організації незалежно від форми власності, зокрема будуть задіяні держслужбовці, підприємці, аграрії, вчені, експерти, тощо. Це в свою чергу зумовить потребу в обробці різнопланової неструктурованої інформації у великих обсягах, що підтвердить ефективність запропонованого підходу в організації інформаційної системи аналізу стану рослин з чіткою диференціацією функцій обладнання та людини і необхідності визначення регламентів їх взаємодії на кожному з етапів роботи.

На рис. 2 наведено BPMN (Business Process Model and Notation) діаграми деяких бізнес-процесів створюваної системи, які містять послідовність робочих дій і рухи потоків інформації, спрямовані на розв'язання конкретних нагальних потреб інформаційної системи аналізу стану рослин.

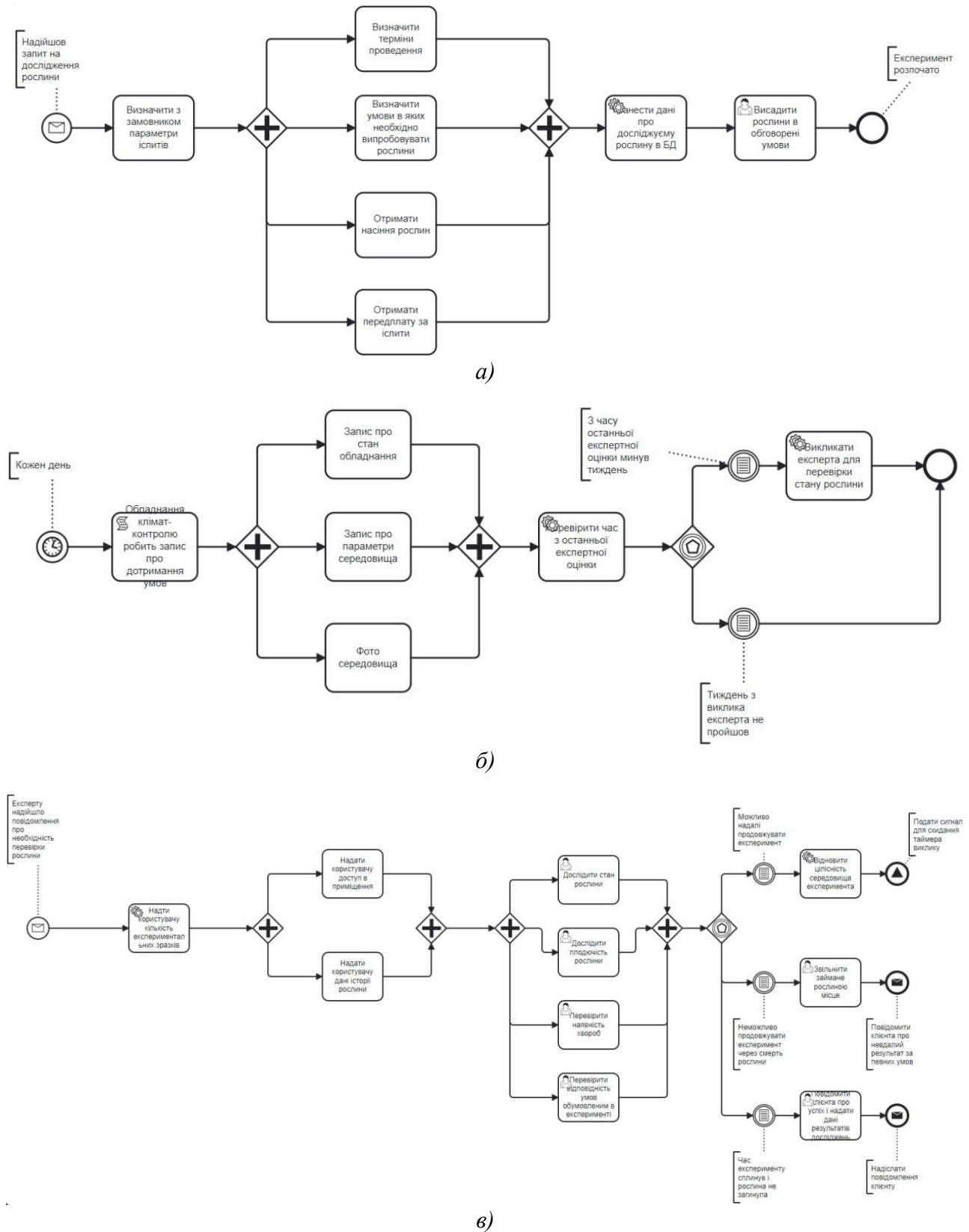


Рисунок 2 – BPMN діаграма бізнес-процесів системи збору параметрів росту та експертизи рослин

На рис. 2.а наведено обробку запиту на дослідження певної рослини, процес якого включає з'ясування умов замовлення та оплати, передачі піддослідного матеріалу, створення інформаційного забезпечення БД, генерацію певного сценарію дій щодо дослідницької діяльності стосовно впровадження рослини в аграрне виробництво. Рис. 2.б ілюструє обробку інформації щодо стану і регламентів роботи використаного обладнання для фіксації зовнішніх умов, можливого підналаштування їх під потреби рослин та залучення фахівців-експертів для визначення стану рослин. Фіналізація досліджень з можливими результатами нормального розвитку рослини або загибелі проілюстрована на рис. 2.в. Детальну увагу приділено інформації щодо продуктивності рослин, визначення наявності певних хвороб, детальному відслідковуванню факторів зовнішнього середовища для подальшої фіксації і обробки інформаційних потоків у відповідних розділах БД.

Для візуалізації потоків даних при проектуванні інформаційної системи аналізу стану рослин була запропонована Dataflow діаграма, яку зображено на рис. 3.

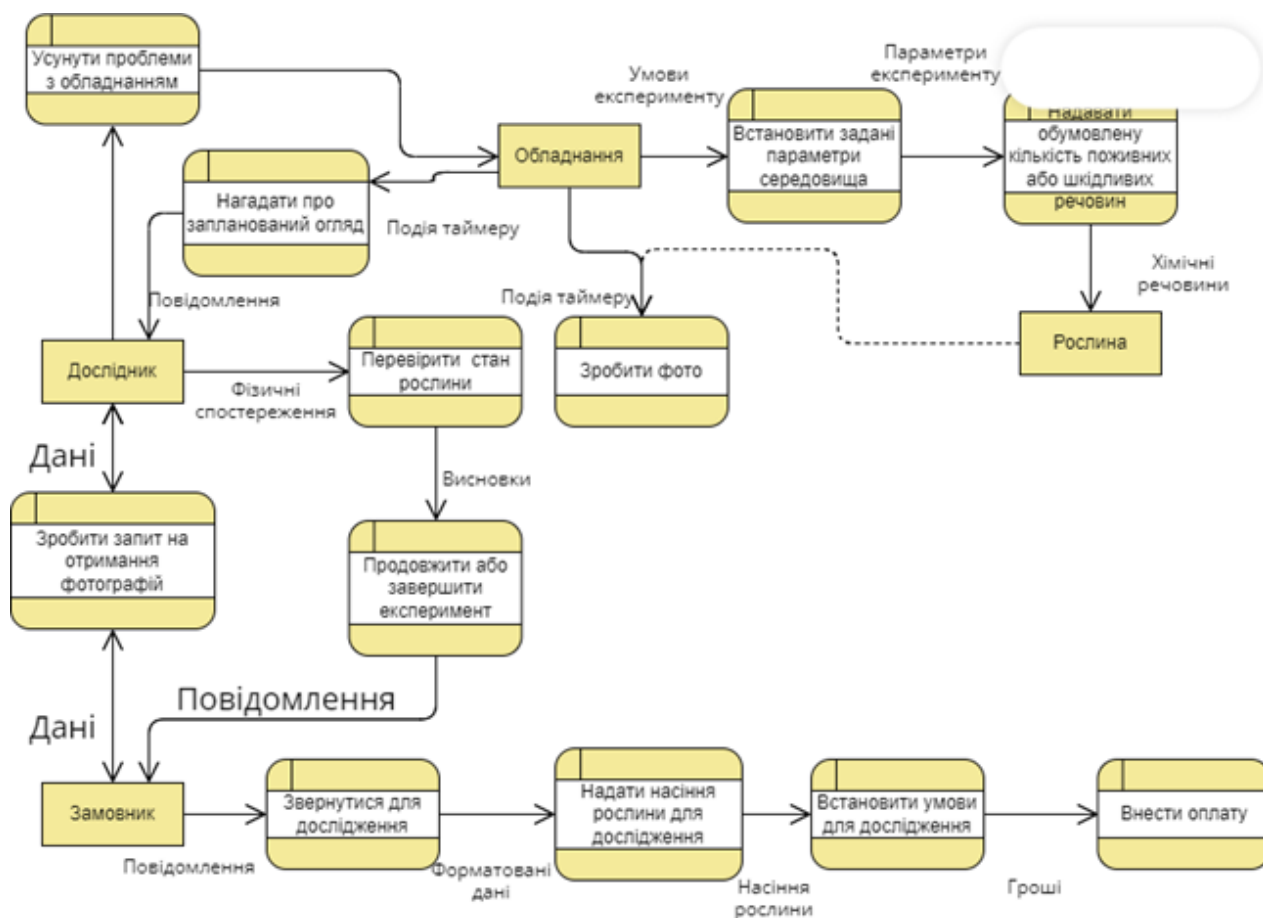


Рисунок 3 – Dataflow діаграма системи збору параметрів росту та експертизи рослин

Діаграма потоків даних відображає наявні потоки даних в ІС та їх використання різними сутностями, зокрема детальна увага приділена спостереженням за станом рослин та фіксації отриманих результатів з допомогою обладнання та фіксації певних поточних результатів з передачею даних аналізу замовнику досліджень.

Висновки і перспективи. В матеріалі статті обґрунтовано необхідність проведення численних експериментів щодо виявлення тенденцій розвитку рослин в умовах досліджуваних регіонів. Пропонується за зібраними статистичними даними визначити доречну (мінімально необхідну) кількість біоенергетичних та злакових культур, необхідну статистику щодо опорних показників нормального розвитку рослин, індикаторів плодоносіння. Запропоновано спосіб отримання інформації про поточний стан рослин з детальною фіксацією умов розвитку

та результатів врожайності, спираючись на які засобами абстрагування слід розробити підґрунтя дорадницької інформаційної системи, побудованої відповідними методами аналізу і синтезу.

Наведено математичне описання визначення ефекту середовища та умов вирощування (вплив на рослини) за будь-яким біопараметром: за масою рослини, довжиною кореневої або стеблової системи, кількістю ушкоджених рослин або кількістю сходів тощо, призначеного для оцінки придатності рослини для аграрного бізнесу у визначеному територіальному регіоні.

Запропоновані Use-case-, BPMN- та Dataflow- діаграми, які ілюструють з певним ступенем деталізації послідовність дій і обробку інформації для процесу дослідження аналізу стану рослин, що становить основу для побудови інформаційної системи аналізу станів рослин.

До перспектив подальших розвідок слід віднести необхідність більш детальної проробки кожного з компонентів запропонованої інформаційної системи зі з'ясуванням механізмів генерації висновків щодо доцільності вирощування рослин у певному регіоні та компонентів дорадництва щодо організації діяльності у визначеному аграрному секторі економіки.

Список використаних джерел

1. Україна переходить у зону ризикованого землеробства? URL: <https://agro-business.com.ua/agro/ahronomiia-sohodni/item/18236-ukraina-perekhodyt-u-zonu-ryzykovanoho-zemlerobstva.html> (дата звернення: 24.03.2024).
2. BBCH working group. (2001). Growth stages of mono-and dicotyledonous plants. (U. Meier, Ed.) (2nd ed.). Federal Biological Research Centre for Agriculture and Forestry.
3. Kimball, J. W. (2021) Biology. Tufts University & Harvard University.
4. Уряд розробив план дій з відновлення деокупованих територій. URL: <https://www.kmu.gov.ua/news/uryad-rozrobiv-plan-dij-z-vidnovlennya-deokupovanih-teritorij> (дата звернення: 24.03.2024).
5. Про затвердження плану дій органів виконавчої влади з відновлення деокупованих територій територіальних громад. URL: <https://zakon.rada.gov.ua/laws/show/1219-2022-r#Text> (дата звернення: 24.03.2024).
6. Zadoks, J.C.; T.T. Chang; C.F. Konzak (1974). "A decimal code for the growth stages of cereals". Weed Research. 14 (6): 415–421. doi:10.1111/j.1365-3180.1974.tb01084.x
7. Про затвердження Методики польового оцінювання насінневих посівів соняшнику та ріпаку Наказ, Методика від 21.11.2018 № 558. URL: https://ips.ligazakon.net/document/view/re32965?an=2&ed=2018_11_21 (дата звернення: 23.03.2024).
8. Моніторинг полів в сільському господарстві: як застосовувати отримані дані. URL: <https://blog.agrokebety.com/monitorynh-poliv-v-silskomu-hospodarstvi>. (дата звернення: 23.03.2024).
9. Інтелектуальне управління тепличним господарством. URL: <https://www.proxis.ua/uk/solution/automation-of-the-agricultural-sector/> (дата звернення: 23.03.2024).
10. Kalpana P., Anandan R., Hussien A.G., Migdady H., Abualigah L. Plant disease recognition using residual convolutional enlightened Swin transformer networks (2024) Scientific Reports, 14 (1), art. no. 8660. DOI: 10.1038/s41598-024-56393-8 <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85190363094&doi=10.1038%2fs41598-024-56393-8&partnerID=40&md5=541f6be704177217ac6584da3a9c231a>
11. Voothuluru P., Wu Y., Sharp R.E. Not so hidden anymore: Advances and challenges in understanding root growth under water deficits (2024) The Plant cell, 36 (5), pp. 1377 – 1409. DOI: 10.1093/plcell/koae055 <https://www.scopus.com/inward/record.uri?eid=2-s2.0-85192027980&doi=10.1093%2fplcell%2fkoae055&partnerID=40&md5=bf1414b6b3a24cc50a4a7f485814643f>

12. Zhang Y., Lu Y., Guan H., Yang J., Zhang C., Yu S., Li Y., Guo W., Yu L. A Phenotypic Extraction and Deep Learning-Based Method for Grading the Seedling Quality of Maize in a Cold Region (2024) *Agronomy*, 14 (4), art. no. 674. DOI: 10.3390/agronomy14040674 [https://www.scopus.com/inward/record.uri?eid=2-s2.0-85191484052&doi=10.3390 %2fagronomy 14040674&partnerID= 40&md5=8a38bc5896f2a9385352ee5389b12d49](https://www.scopus.com/inward/record.uri?eid=2-s2.0-85191484052&doi=10.3390%2fagronomy14040674&partnerID=40&md5=8a38bc5896f2a9385352ee5389b12d49).

Smolij Viktorija

Doctor of Technical Sciences, Professor, Professor of the Department of Information Systems and Technologies,

National University of Life and Environmental Sciences of Ukraine

ORCID: <https://orcid.org/0000-0002-1268-7837>

E-mail: vmsmolij@nubip.edu.ua

Smolij Natan

Postgraduate Student of the specialty "Information Systems and Technologies",

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

ORCID: <https://orcid.org/0009-0002-3763-6726>

E-mail: hoibbitizukrainy@gmail.com

Yantsevych Anton

Student of the specialty "Information Systems and Technologies",

National University of Life and Environmental Sciences of Ukraine

E-mail: ist21-a.yantsevych@nubip.edu.ua

Iskorostenskyi Olexiy

Student of the specialty "Information Systems and Technologies",

National University of Life and Environmental Sciences of Ukraine

E-mail: ist21-o.iskorostenskyi@nubip.edu.ua

DESIGNING A COMPUTER SYSTEM FOR ANALYSING THE CONDITION OF PLANTS

Abstract. This research is devoted to the topical issue of restoring agriculture in liberated territories, which is associated with the risks of mining, pollution, littering, overcoming the consequences of environmental disasters, soil degradation, etc., the need to carry out the specified work and conduct research on the possibility of growing plants in post-war conditions and the prospects for effective plant cultivation in areas of risky agriculture. The purpose of the article is to highlight the creation of an information system for analyzing the condition of plants, designed for agro-industrial activities related to the cultivation of bioenergy and cereal crops, which provides for the collection, processing, and recommendations on the condition of plants, which in turn allows determining the feasibility (rationality) of growing a particular plant variety in a particular region, plan field work, processing, and irrigation of plants.

The article substantiates the need to conduct numerous experiments to identify plant development trends in the regions under study. It is proposed to use the collected statistical data to determine the appropriate (minimum necessary) amount of bioenergy and cereal crops, the necessary statistics on the reference indicators of normal plant development, and indicators of fruiting. A method is proposed for obtaining information about the current state of plants with detailed recording of development conditions and yield results, based on which means of abstraction should be used to develop the basis for an advisory information system built using appropriate methods of analysis and synthesis. A mathematical description of the determination of the effect of the environment and growing conditions (impact on plants) for any bioparameter is given: by plant weight, root or stem system length, number of damaged plants or number of seedlings, etc., designed to assess the suitability of a plant for agribusiness in a specific geographical region. The proposed Use-case, BPMN, and Dataflow diagrams illustrate, with a certain degree of detail, the sequence of actions and information processing for the plant condition analysis research process, which forms the basis for building an information system for plant condition analysis. Prospects for further research include the need for a more detailed study of each component of the proposed information system, clarifying the mechanisms for generating conclusions about the feasibility of growing plants in a particular region and components of advisory services for organizing activities in a specific agricultural sector of the economy.

Keywords: post-war recovery, agro-industrial sector, agricultural crops, plant condition assessment, development parameter calculation, agrodrones, plant parameter identification task, information system, Use-case, Business Process Model and Notation, and Dataflow diagrams.

UDC 004.2:004.5

Nazarenko Volodymyr

Ph.D., Associate Professor of the Department of Computer Systems, Networks and Cybersecurity, National University of Life and Environmental Sciences of Ukraine,

ORCID: <https://orcid.org/0000-0002-7433-2484>

E-mail: volodnz@nubip.edu.ua

Kasatkin Dmytro

Candidate of Pedagogical Sciences, Associate Professor, Head of the Department of Computer Systems, Networks and Cybersecurity,

National University of Life and Environmental Sciences of Ukraine,

ORCID: <https://orcid.org/0000-0002-2642-8908>

E-mail: ksim@nubip.edu.ua

INTERACTION AND FEEDBACK LOOPS IN USER INTERFACE FOR DIGITAL ENGINEERING DESIGN

Abstract The article aims to review existing issues and look for novel ways of human-computer digital systems design and interaction. Such interaction takes place, as a rule, through an appropriate interface, which can be mechanical, digital, analog, with a touch screen, etc. For the end user of the relevant application or system, it is important how this interaction process will be implemented, presented, and programmed. That is, what controls, actions for interaction, and feedback mechanisms from the system to the user will be available. The research methods are based on the main methodologies of design and analysis of digital products with an emphasis on user experience, research of user requirements, the context of application use, and the presentation of the menu of interaction with a digital service. The article presents the results of a detailed description and modeling of the main process of user interaction with the feedback system (cyclic process). The context of using the application, the main desires and goals of the user when interacting, as well as the types of available options for interacting with the interactive system are taken into account. The seven-step process of user interaction presented in the paper includes functions, data, software systems and modules, appropriate software and hardware architecture, and methods of internal communication between cycle functions. A process of interactive interaction has been developed, which describes the full cycle from the beginning to the end of the user's work with any digital, analog, computing device, application, or service. This process can be used by designers, software developers, and scientists/researchers at the stage of planning a new interactive system or to improve an existing one.

Keywords: HCI, user interface, computer engineering, design engineering, design.

Introduction. Software and digital information applications are used in every area of life. Hardware and technological advancements led to a great number of digital interactive devices and screens available to users around the world. Digital and analog devices are used not only by the general populace but by many professionals in all industries, from construction to aerospace engineers to mechanical engineering to military and health care, among many other industries. What users see and interact with, first of all, is the user interface or mechanical interface. The Human-Computer Interaction field is responsible for researching new ways and improving existing user-computer (user-devices) mechanics. The process of user-computer interaction happens in loops, and for every user action or series of activities, systems provide appropriate feedback. To take a step further interaction design is not just centered on, limited to input and output operations cycles, it affects the whole software/hardware/system. Digital human-computer interaction systems design can be classified as a type of engineering systems architecture and should be studied appropriately.

Analysis of research and publications. There are many prominent researchers and research work in the fields of HCI, UI research, and interaction engineering. The most common work in the area of HCI is user interface and interaction feedback loops [1-5]. Besides, fundamental studies have been carried out in the area of user interface design principles and design systems [6-9]. Emerging Technologies, such as VR, AR, and new touch interfaces require a new view of how the interaction loop operates [10]. Even though the area of HCI and interaction research is not new, many systems and possibilities were left unstudied or haven't received enough focus. Among them, the field of

digital design is focused on comprehensive system design that includes multiple simultaneous use complex interaction actions, and multi-device feedback.

Purpose. The research work aims to describe the HCI interaction loop, break it down into functional components, and present a robust model for interaction design context, interaction process, as well as the view on User Interaction – Feedback Systems, and States.

Methods. List of main HCI research methods, components, and interaction systems are listed in Table 1. Human-computer interaction is conducted using various mechanical, screen, and sound processing devices. Major interaction systems include digital electronic and electrical devices, used for input and output operation, data and signal processing modules. Each type of interaction has a corresponding sample system type, ranging from LCDs to touchscreen interfaces to found speakers and spatial sensors.

*Table 1 – User-Interface Interaction Classification**

Interaction Type	Components	Data	Example systems
Mechanical	Input Device Input Sensor Output Device Group of devices Processing Machine Singal Processing	Analog/Digital Signal Spatial coordinates Input/Output Type Input/output signal End-point comms. message Device control data	Electrical device Singal processors Transistors Programmable components Camera and motion recognition, tracker Touch screen device Mechanical control device
Text	Direct physical input/output Virtual Input/output	Screen size and type Language model Input/output text	BW screen LCD screen Paper/Magnetic Ink
Voice	Physical microphone/speaker Audio processing Audio recognition software	Signal data (wave) Recognition and sensory data Timer and buffer data	Sound wave signal converters Sound processing engine
Visual (CGI)	GPU Rendering Software	2D/3D Coordinates Geometry VFX and Animation sequences	Game Engine Modeling Software End-user application

** prepared based on author work and public research data [1-10]*

Data and programming functions are included in user, software, and hardware system design. Data types are broken down into categories based on interaction type, and listed in Table 1, column 3. List of the most common user-interface interaction functions (based on Table 1 Interaction Types), each of which can take input data and/or return result data:

- ProcessingEngine();
- ScreenRender();
- ViewMode();
- InteractionMode();
- MenuSystem();
- GetInput();
- SetOutput();
- StoreData();
- ConnectToDevice();
- SingalProcess();
- and others.

When developing and designing new interaction systems both scientists and industry specialists have a wide range of tools at their disposal. The most commonly used hardware and software solutions are listed in columns 1 and 2 of Table 2. HCI and Interaction design rely on many methodologies and patterns, as the product/applications design process starts with the Business and

product planning, followed by design and technology development (column 3 in Table 2). What the end-user gets is not some random number of signals or misc. messages, but specific text, and visual elements and has access to several possible menus and navigation panels (column 4 in Table 2).

Table 2 – Hardware, Software, Methodology Classification in HCI*

Hardware	Software	Methodology	Presentation
Glasses: - VR - AR - MR	Software development suits and stacks: - native platform - cross-platform - use case specific	Business and Product development: - Stakeholders - Requirements - Planning and Resources allocation	Physical dimension: - presentation - size - interaction type - context of use
Sensor Screen: - phone - tablet - general screen	Interface design: - physical - digital graphic design - digital UX, UI	Design and Interaction: - Form - Interface - Presentation	Input and Output: - direct - loop - contextual
Mechanical: - joystick/ gamepad/controller - Virtual Mouse and Keyboard - Mechanical Mouse and Keyboard	2D and 3D graphics programs: - modeling - rigging - animation - industrial	Technology and Design: - Analytics - Development - Materials - Electronics - Aesthetics	UX: - Objective/Goal - Functionality - Interaction - Information - Sensory
Sensors and trackers: - detection and recognition - motion	Embedded and hardware development software: - firmware - application	Manufacturing and Production: - Packaging - Supply and Delivery	UI: - Layout - Navigation - Menu - Screen Elements
Misc. Input&Output devices: - digital - analog	Testing and Evaluation: - UX - product and market - software	User and Behavior: - Goals and desires - Information Architecture - Psychology and behaviors - social and group interaction	Special: - notifications - sound - music - 3D effects - spatial effects - misc.

* prepared based on author work and public research data [1-10]

Human-computer interaction is a three-phase process, as presented in Figure 1.

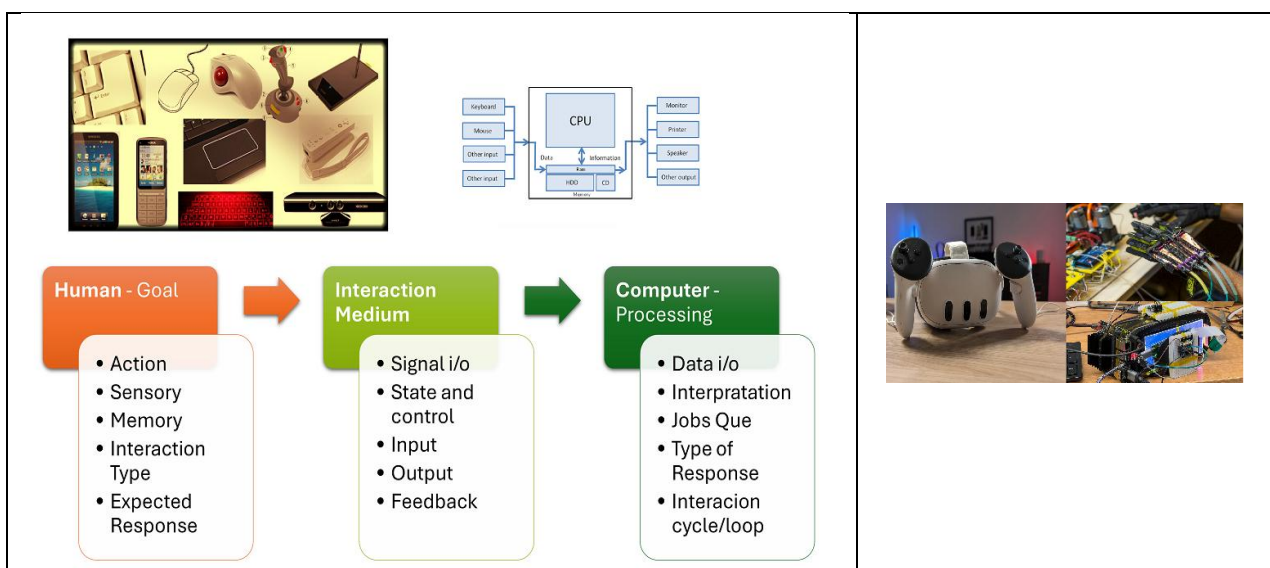


Figure 1 – General Human-Computer Interaction Components [6-9]

The essential first step is the user (human) defined goal that he/she needs to accomplish with the help (usage) of digital/analog/mechanical/misc. type of interactive device. User input is processed depending on the interaction medium and input type by the software system (computer, microcomputer, etc.). A list of underlying components, actions, and conceptual functions is presented in the bottom section of Figure 1 under the respective Interaction phase.

Results. At the center of applications, software, and systems interaction is human – the user. The reason the interaction process is initiated and instanced is due to the user having a single or several goals that they wish to accomplish. The interaction systems in the process are just a tool used by the user to achieve the desired goal. Figure 2 highlights the user-centered interaction loop with a focus on the interaction processing feedback loop. The interaction activity consists of the user interface, input, and output processing functions, that send appropriate commands towards the main System Processor unit.

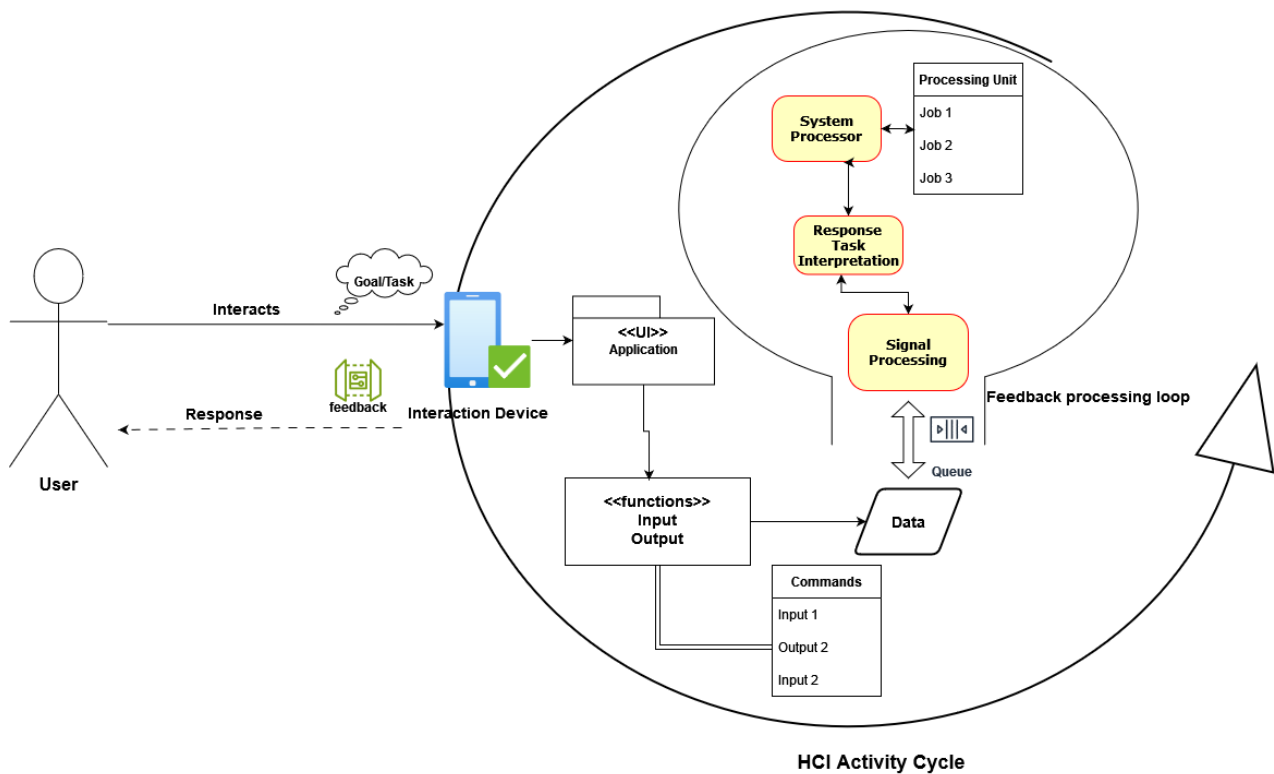


Figure 2 – Feedback loop in UI and Digital Engineering Design Systems

Figure 3 illustrates the Human-Computer Interaction Process with a focus on major users, functions, and underlying systems. The process is divided into three layers – conceptual programming functions, applications modules and components, main systems, and data processing units. There are several types of users who directly or indirectly affect the interaction cycle, user or player directly influences and interacts with the systems, while the rest of the stakeholder's team is responsible for the production and maintenance of the deployed application/system. Systems' internal communication and processing logic can be broken down into two parts – user-centered services and system internal processing services modules.

Interaction processes happen not in the vacuum space, but rather in a real-world environment. This is especially true for industrial, medical, and mobile applications. Figure 4 visualizes the context of use and design impacting factors. Each of the interactive systems should consider usage environment, time and location, where the device and application can be and will be used. Certain tasks and interaction activities influence how a user will interact with the given system, which might add additional constraints and limitations. Differentiating user type and purpose of the system usage is vital in early system design stages, as certain industries and professional environments place a large

weight on the end-user goal and step-by-step user-system interaction flow. All of these factors and more are highlighted in Figures 3 and 4.

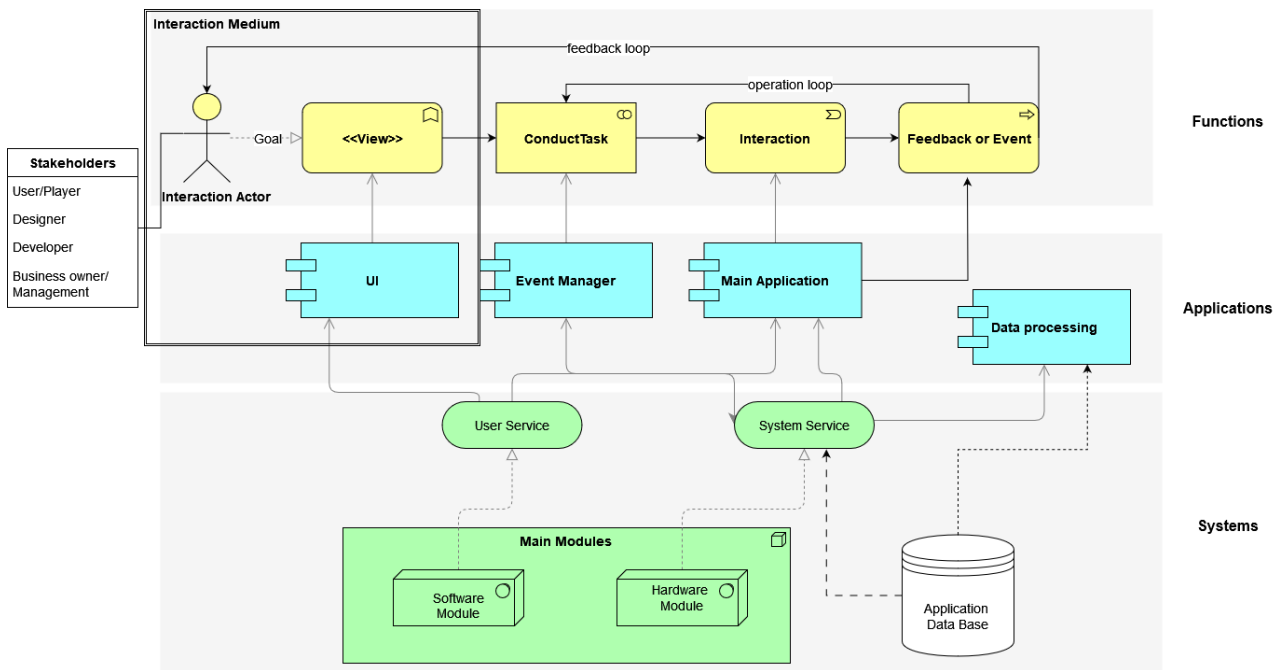


Figure 3 – Human-Computer Interaction Process – Users, Functions and Systems

Interaction processes happen not in the vacuum space, but rather in a real-world environment. This is especially true for industrial, medical, and mobile applications. Figure 4 visualizes the context of use and design impacting factors. Each of the interactive systems should consider usage environment, time and location, where the device and application can be and will be used. Certain tasks and interaction activities influence how a user will interact with the given system, which might add additional constraints and limitations. Differentiating user type and purpose of the system usage is vital in early system design stages, as certain industries and professional environments place a large weight on the end-user goal and step-by-step user-system interaction flow. All of these factors and more are highlighted in Figures 3 and 4.

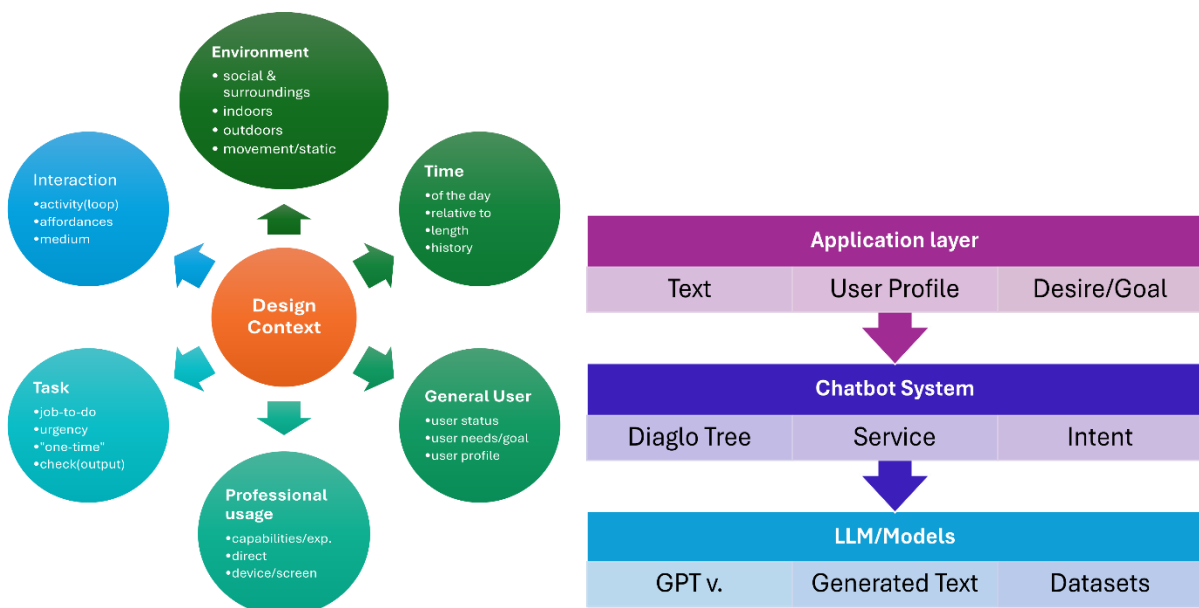


Figure 4 – Interaction Design – Context of Use

Table 3 presents six core interaction process phases (stages, steps). Each new interaction loop starts with the system boot (loading) and concludes with the system shutdown. After the system is loaded and operational, it loads the appropriate user interface (or similar input-output system) and enables a user-system interaction interface. The next phase is the user-system interaction loop and feedback loop, both of them are operating, and loading in parallel. Feedback and interaction can consist of “n” number steps, which are determined by the user or/and by the system. At each interaction cycle of the loop, the decision can be made to either move to the next step or get to the initial system step, or completely shut down the system. The interaction loop stops when the user decides to stop the interaction and the system saves its state and powers off. With this phase the interaction loop is complete, and system is disabled or is operating in the background listening to new input or internal event messages.

Table 3 – Interaction Process

№	Interaction phase	Systems	Components Actors	Result Feedback
1	System boot	Power controls Direct Interaction Authentication	Electricity Device User	Load the previous state Load new session Status indicator
2	Load UI Enable Interaction	GUI Input/Output functions Menu/Controls	Electronic device PC Hardware Application SDK/Libraries	Enable HCI process Controls -> On Interface -> Render Direct/Indirect Ctrl's
3...n	Interaction Loop	Tasks General Functionality Enhanced Input	User Interface system UI Menu and Navigation	System Input Interpretation User Task(s) Execution
3...n	Feedback Loop	Enhanced Output Notification Event manager	System Event Manager UI Manager Data Manager	User Feedback System Task(s) Execution, Scheduling
n	Decision making	Action Evaluation System Planning	Application User Task Manager	Internal Application Data save/load Job Status Check
n+1	Shutdown	System save state Power Controls	Electricity Device User Data Manager System Event Manager	Save State Check Special Status Status Notification

The more detailed presentation of phase number 3 can be viewed in Figure 4 – Interaction and Feedback loops. Two actors interact with the systems during the loop – the user and the computational (general) system. In the center of the diagram denoted by the yellow rectangles are major system states. The interaction starts with the user action and the input information/task threads. Each new action is recorded and currently being processed by the event queue and Event Manager systems. While the user interacts and waits for the response, computational systems also engage in the interaction process by firing appropriate response and communications messages. The interaction process is part of larger server/client applications and database services (listed at the bottom section of Figure 4).

Algorithm step-by-step for feedback loop design (Figure 4):

1. The user has a goal – and needs to complete the action;
2. System loads and provides appropriate interaction interface;
3. A user interacts with the UI via micro-interactions, input action can be single or multiple;
4. Event Manager processes actions and makes a queue (in case of multiple user activities or based on system requirements/architecture);

5. System loads output – provides feedback based on User Action and Event Manger System via Communication interface system depends on software and hardware architecture, as well as specific application and embedded device software;

6. A user reviews feedback or receives notification from the list of output types, he chooses whether to continue the micro-interaction or stop (or a system event can force a user to stop or continue);

7. The system stops or moves into hibernation/standby modes.

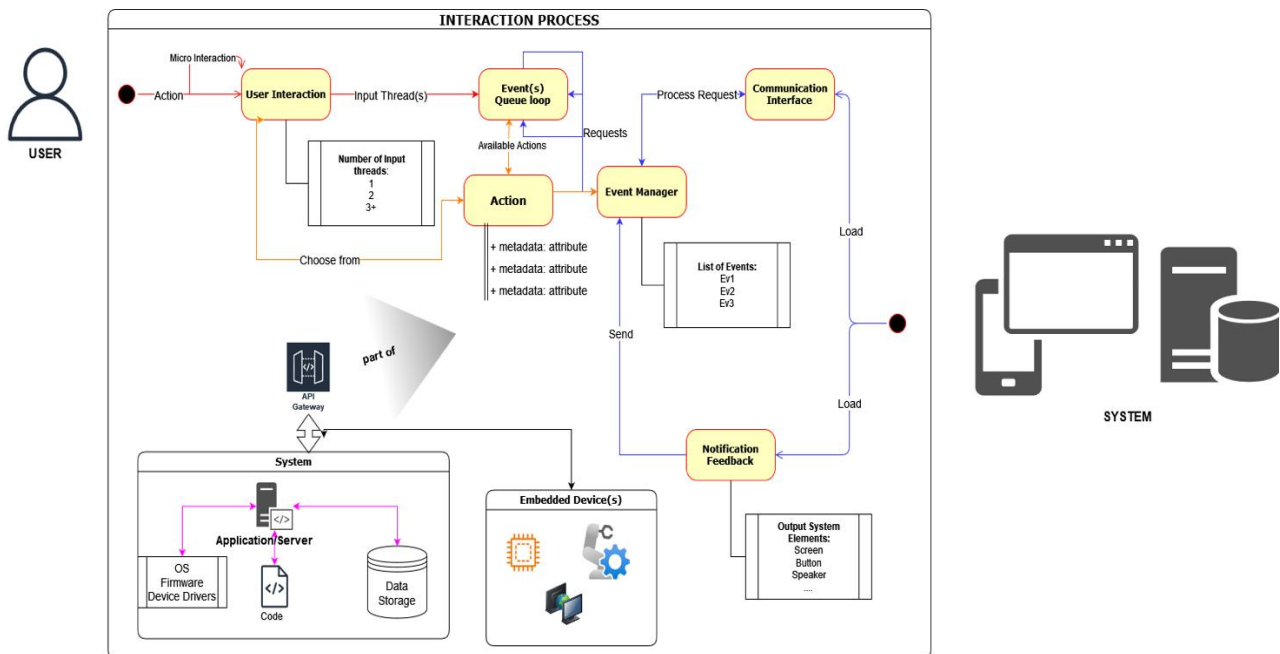


Figure 4 – User Interaction – Feedback Systems and States

Discussion. The modern user interface is a very multilayered and complex system. The user interaction process is no longer limited to simple keyboard types or single-button interactions. There are a great number of widely available interaction devices and output mediums, mobile phone screens, VR glasses, gamepads, and game controllers, but the input-output system is not limited to just the number of devices listed. The user interaction has many steps and stages to it the list of available interaction activities is also great in number. As discussed, the human-computer interaction goes in loops, feedback loops to be precise. The essence of the Human-Computer interaction process revolves around tasks and activities started by the user and processed by the system. The context of use plays an important role in the interaction cycle, as it provides experienced designers with awareness and knowledge of the user, when and how the system will be used. The interaction process itself is made of some steps, components, actors, and systems. Each application has its own interaction and feedback loop. However, implementation can be done based on the system abstraction and certain algorithms, such as feedback loop system and algorithm.

References

1. Fratantonio, Y., Qian, C., Chung, S. P., & Lee, W. (2017, May). Cloak and dagger: from two permissions to complete control of the UI feedback loop. In 2017 IEEE Symposium on Security and Privacy (SP) (pp. 1041-1057). IEEE.
2. Nguyen, J., & Dupuis, M. (2019, September). Closing the feedback loop between UX design, software development, security engineering, and operations. In Proceedings of the 20th Annual SIG Conference on Information Technology Education (pp. 93-98).
3. Novak, D., Nagle, A., & Riener, R. (2014). Linking recognition accuracy and user experience in an effective feedback loop. *IEEE Transactions on Affective Computing*, 5(2), 168-172.

4. Willig, J. H., Krawitz, M., Panjamapirom, A., Ray, M. N., Nevin, C. R., English, T. M., ... & Berner, E. S. (2013). Closing the feedback loop: an interactive voice response system to provide follow-up and feedback in primary care settings. *Journal of medical systems*, 37, 1-9.
5. Ritter, W. (2011). Benefits of subliminal feedback loops in human-computer interaction. *Advances in Human-Computer Interaction*, 2011, 1-11.
6. Chaves, A. P., & Gerosa, M. A. (2021). How should my chatbot interact? A survey on social characteristics in human–chatbot interaction design. *International Journal of Human–Computer Interaction*, 37(8), 729-758.
7. Sridevi, S. (2014). User interface design. *International Journal of Computer Science and Information Technology Research*, 2(2), 415-426.
8. Johnson, J. (2020). *Designing with the mind in mind: simple guide to understanding user interface design guidelines*. Morgan Kaufmann.
9. Persson, H., Åhman, H., Yngling, A. A., & Gulliksen, J. (2015). Universal design, inclusive design, accessible design, design for all: different concepts—one goal? On the concept of accessibility—historical, methodological and philosophical aspects. *Universal access in the information society*, 14, 505-526.
10. Segura, E., Vidal, L., & Rostami, A. (2016). Bodystorming for movement-based interaction design. *Human Technology*, 12(2), 193-251.

Назаренко Володимир Анатолійович

кандидат економічних наук, доцент кафедри комп'ютерних систем, мереж та кібербезпеки,

Національний університет біоресурсів та природокористування України

ORCID: <https://orcid.org/0000-0002-7433-2484>

E-mail: volodnz@nubip.edu.ua

Касаткін Дмитро Юрійович

кандидат педагогічних наук, доцент, завідувач кафедри комп'ютерних систем, мереж та кібербезпеки,

Національний університет біоресурсів та природокористування України,

ORCID: <https://orcid.org/0000-0002-2642-8908>

E-mail: ksim@nubip.edu.ua

СИСТЕМИ КОРИСТУВАЛЬНИЦЬКОЇ ВЗАЄМОДІЇ ТА ЗВОТНОГО ЗВ'ЯЗКУ В ІНТЕРФЕЙСІ КОРИСТУВАЧА ПРИ ПРОЕКТУВАННІ ЦИФРОВИХ СИСТЕМ

Анотація. Область «Взаємодія людини і комп'ютера» відповідає за дослідження нових способів і вдосконалення існуючих механік «користувач-комп'ютер» (користувач-пристрої). Процес взаємодії користувача з комп'ютером відбувається циклічно, і для кожної дії користувача або серії дій системи забезпечують відповідний зворотний зв'язок. Є багато видатних дослідників та дослідницьких робіт у галузі HCI, досліджень UI та інженерії взаємодії. Найбільш поширені роботи в області HCI, інтерфейсу користувача і взаємодії циклів зворотного зв'язку. Незважаючи на те, що область досліджень HCI та взаємодії не є новою, існує багато систем та можливостей, які залишилися невивченими або не отримали достатньої уваги. Серед них сфера цифрового дизайну, яка зосереджена на комплексному проектуванні системи, що включає багаторазове одночасне використання, взаємодію, складні взаємодії та зворотний зв'язок з кількома пристроями. Стаття присвячена розгляду актуальних питань щодо способів взаємодії людини- користувача з комп'ютерними цифровими системами. У центрі взаємодії додатків, програмного забезпечення і систем знаходиться людина – користувач. Причина, по якій ініціюється і інсталується процес взаємодії, полягає в тому, що користувач має одну або кілька цілей, яких він хоче досягти. Така взаємодія відбувається, як правило, через відповідний інтерфейс, який може бути: механічним, цифровим, аналоговим, з сенсорним екраном тощо. Існує кілька типів користувачів, які прямо чи опосередковано впливають на цикл взаємодії, користувач або гравець безпосередньо впливає на системи та взаємодіє з ними, тоді як решта команди зацікавлених сторін відповідає за виробництво та підтримку розгорнутого додатку/системи. Певні завдання та активність взаємодії впливають на те, як користувач буде взаємодіяти з даною системою, можуть додавати додаткові обмеження та обмеження. Диференціація типу користувача та мети використання системи є життєво важливою на ранніх етапах проектування системи, оскільки певні галузі та професійні середовища надають

великого значення меті кінцевого користувача та поетапному потоку взаємодії користувача з системою. Внутрішню комунікацію та логіку обробки систем можна розбити на дві частини – сервіси, орієнтовані на користувача, та модулі внутрішніх служб обробки системи. Для кінцевого користувача відповідного додатку або системи важливим є те, як саме буде реалізований цей процес взаємодії. Тобто, які будуть доступні органи керування, дії для взаємодії та механізми зворотного зв'язку від системи до користувача. Методи дослідження базуються на головних методологіях дизайну та аналізу цифрових продуктів з акцентом на користувальницькому досвіді, дослідженні вимог користувачів, контексту використання застосунку та представлення меню взаємодії з цифровим сервісом. У статті представлено результати детального опису та моделювання головного процесу взаємодії користувача із системою зворотного зв'язку (циклічний процес). Враховано контекст використання застосунку, головні бажання та цілі користувача при взаємодії, а також типи доступних варіантів взаємодії з інтерактивною системою. Представлений у роботі семи-кроковий процес користувальницької взаємодії включає в себе функції, дані, програмні системи та модулі, відповідну архітектуру програмного та апаратного забезпечення та методи внутрішньої комунікації між функціями циклу. Розроблено процес інтерактивної взаємодії, який описує повний цикл від початку до завершення роботи користувача з будь-яким цифровим, аналоговим, обчислювальним пристроєм, додатком, сервісом. Такий процес можна використовувати дизайнерам, розробникам програмного забезпечення, технічних засобів на етапі планування нової інтерактивної системи або удосконалення існуючої.

Ключові слова: HCI, інтерфейс користувача, комп'ютерна інженерія, проектування, дизайн.

УДК 004.056.5:621.3.01:656.1/5

Лакно Валерій Анатолійович*доктор технічних наук, професор, професор кафедри комп'ютерних систем, мереж та кібербезпеки,**Національний університет біоресурсів та природокористування України*ORCID: <https://orcid.org/0000-0001-9695-4543>E-mail: lva964@nubip.edu.ua**Касаткін Дмитро Юрійович***кандидат педагогічних наук, доцент, завідувач кафедри комп'ютерних систем, мереж та кібербезпеки,**Національний університет біоресурсів та природокористування України*ORCID: <https://orcid.org/0000-0002-2642-8908>E-mail: d.kasatkin@nubip.edu.ua**ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ РОЗПІЗНАВАННЯ КІБЕРЗАГРОЗ, ЯК СКЛАДОВА СИСТЕМ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ НА ТРАНСПОРТІ**

***Анотація.** Транспортна галузь, зокрема й залізничний транспорт (ЗТ) все більше використовує комп'ютеризовані системи для управління, що робить його вразливим до кібератак. Для захисту інформаційної та кібернетичної безпеки на транспорті стають важливими інтелектуальні системи розпізнавання загроз. Забезпечення кібербезпеки для інформаційних систем та автоматизованих систем керування на транспорті є важливим завданням для забезпечення національної безпеки України. Це дозволяє запобігти можливим кібератакам, зберегти конфіденційність інформації, забезпечити надійну роботу транспортної системи та зберегти безпеку громадян. Розвиток та впровадження ефективних заходів кібербезпеки є необхідними кроками для забезпечення стійкості і надійності транспортних систем та збереження національної безпеки в Україні. Ця стаття описує метод інтелектуального розпізнавання загроз для інформаційних систем керування на транспорті (на прикладі залізничного транспорту – ЗТ). Метод ґрунтується на дискретних процедурах та використанні логічних функцій та нечітких множин.*

***Ключові слова:** транспорт, автоматизовані системи керування, кібернетична безпека, розпізнавання загроз, інтелектуальні системи.*

Вступ. Забезпечення кібербезпеки для інформаційних систем (ІС) та автоматизованих систем керування (АСК) на транспорті має вирішальне значення в контексті забезпечення національної безпеки України.

По-перше, транспортна система є однією з найважливіших інфраструктурних сфер у країні. Вона забезпечує транспортування людей, вантажів та інших ресурсів, що є необхідними для економічного розвитку та функціонування суспільства. В разі порушення кібербезпеки в транспортних системах можуть виникнути серйозні наслідки, включаючи аварії, перерви у роботі та збитки. Такі події можуть призвести до загрози для громадської безпеки, економічних збитків та навіть загрози для життя і здоров'я громадян. По-друге, ІС та АСК транспортом включають в себе значну кількість критичних інфраструктурних об'єктів, таких як мережі зв'язку, системи диспетчеризації, системи безпеки та контролю, системи керування рухом транспорту тощо. Ці системи містять велику кількість конфіденційної інформації, включаючи дані про рух транспортних засобів, розклади, пасажирську інформацію та інше. Несанкціонований доступ до цих систем може призвести до крадіжки даних, розкриття конфіденційної інформації або навіть зловживання владою. Такі ситуації можуть мати серйозні наслідки для безпеки країни та забезпечення національних інтересів. По-третє, у зв'язку зі швидким розвитком технологій та поширенням Інтернету ризики кібератак на транспортні системи зростають. Зловмисники можуть використовувати різні методи, такі як вторгнення в мережу, фішинг, віруси, шкідливі програми та інші, для отримання несанкціонованого доступу до систем керування та завдання шкоди. Ефективне забезпечення

кібербезпеки вимагає розробки та впровадження сучасних заходів захисту, моніторингу та виявлення вторгнень, контролю доступу та навчання персоналу з питань кібербезпеки.

Залізничний транспорт (ЗТ) є однією з найбільш важливих інфраструктурних галузей економіки України. Він забезпечує безперебійне перевезення вантажів і пасажирів, а також є важливим фактором безпеки держави.

У сучасних умовах транспорт є все більш уразливим до кібератак. Кіберзлочинці можуть використовувати різні методи для порушення роботи транспортної інфраструктури, наприклад, щоб: викликати затримки або скасування рейсів; пошкодити обладнання або інфраструктуру; викрадати або підробляти важливу інформацію; та ін.

Мета статті полягає в аналізі та порівнянні засобів для захисту ІС та АСК транспорту від кібератак, використовуючи комплексні заходи, включаючи: фізичний захист інфраструктури; запровадження стандартів і процедур безпеки; підвищення кваліфікації персоналу.

Результати дослідження та їх обговорення. У статті запропонований метод інтелектуального розпізнавання загроз, який базується на дискретних процедурах та використанні логічних функцій та нечітких множинах, пропонується з метою підвищення ефективності розпізнавання загроз для автоматизованих систем керування на транспорті, наприклад, залізничному, автомобільному та ін.

Одним із ключових елементів захисту ІС та АСК від кібератак є використання інтелектуальних систем кібербезпеки.

Інтелектуальні системи кібербезпеки використовують штучний інтелект (ШІ) для виявлення і реагування на кібератаки. Вони відрізняються від звичайних систем захисту інформації наступними особливостями:

– інтелектуальні системи кібербезпеки можуть виявляти кібератаки набагато швидше, ніж звичайні системи. Це дозволяє швидко реагувати на атаки та запобігти їх поширенню.

– інтелектуальні системи кібербезпеки можуть виявляти кібератаки з більш високою точністю, ніж звичайні системи. Це дозволяє уникати помилкових спрацьовувань і не блокувати законний трафік.

– інтелектуальні системи кібербезпеки можуть адаптуватися до нових загроз і атак. Це дозволяє забезпечувати ефективний захист залізничного транспорту в умовах постійно мінливих кіберзагроз.

Крім зазначених переваг, використання інтелектуальних систем кібербезпеки, наприклад, на ЗТ транспорті може також мати наступні позитивні наслідки:

Зниження витрат на кібербезпеку. Інтелектуальні системи кібербезпеки можуть допомогти оптимізувати витрати на кібербезпеку, оскільки вони дозволяють автоматизувати деякі завдання та підвищити ефективність роботи персоналу.

Покращення іміджу компанії. Впровадження інтелектуальних систем кібербезпеки свідчить про те, що компанія піклується про безпеку своїх систем і даних. Це може позитивно вплинути на імідж компанії і її конкурентоспроможність.

З розвитком інформаційних технологій (ІТ) ЗТ став все більш залежним від них. Це призвело до зростання актуальності проблеми інформаційної безпеки (ІБ) ЗТ, яка є складовою державної безпеки України. ІБ ЗТ полягає в забезпеченні схоронності, конфіденційності та цілісності інформації, яка використовується на залізничному транспорті. Ураховуючи різноманітність потенційних загроз у кіберпросторі, складність їх структури та функцій, а також участь людей в технологічних процесах на ЗТ, забезпечення ІБ ЗТ, зокрема АСК ЗТ, вимагає створення комплексних систем захисту (КСЗІ). КСЗІ повинні містити заходи, спрямовані на захист ІС та АСК ЗТ від усіх видів релевантних кіберзагроз.

Проблеми кібербезпеки в ІС та АСК ЗТ мають важливе значення з кількох причин, зокрема:

1) *Вразливість до кібератак.* ІС та АСК можуть стати об'єктом атак через підключення до мереж, використання застарілих програмних засобів або вразливостей у мережевих протоколах.

2) *Інфраструктурна вразливість*. Збільшення автоматизації на ЗТ призводить до більшої вразливості мереж і систем до атак та збільшує ризик порушення нормальної роботи.

Для України питання захисту інформації та забезпечення інформаційної (кібернетичної) безпеки (далі ІБ) для ІС та АСК ЗТ мають особливе значення. Причина цього полягає в тому, що Україна є великою країною з геополітично важливим розташуванням. Приклади несанкціонованого втручання зловмисників (хакерів) в роботу ІС та АСК ЗТ було наведено у роботах [1, 2, 3]. Разом із тим, завдання визначення ризиків кібератак на ІС та АСК ЗТ, належним чином не розглядається та, у найкращому випадку, підмінюється на етапі проектування СЗІ якісним аналізом надійності системи й можливих наслідків проникнення до неї [3].

Захист від кіберзагроз та кібератак атак звичайних інформаційних систем (ІС) на ЗТ дещо відрізняється від захисту від АСК ЗТ.

Звичайні ІС на ЗТ використовуються для виконання різних завдань, таких як управління фінансами, персоналом, логістикою та маркетингом. Вони, як правило, мають меншу критичність для безпеки руху поїздів, ніж АСК ЗТ.

Автоматизовані системи керування (АСК) на ЗТ використовуються для управління рухом поїздів, залізничною інфраструктурою та іншими критичними системами. Вони мають високий рівень критичності для безпеки руху поїздів, тому до їхнього захисту пред'являються особливі вимоги.

Основні відмінності між захистом від кіберзагроз та кібератак ІС на ЗТ та захистом АСК ЗТ є:

Критичність. АСК ЗТ мають вищий рівень критичності для безпеки руху поїздів, ніж звичайні ІС. Тому до їхнього захисту пред'являються особливі вимоги.

Вплив на безпеку руху поїздів. Кіберзагрози та кібератаки, спрямовані на АСК ЗТ, можуть призвести що найменш до серйозних порушень безпеки руху поїздів. Тому захист АСК ЗТ від кіберзагроз та кібератак є життєво важливим для забезпечення безпеки ЗТ.

Складність. АСК ЗТ, як правило, мають більш складну архітектуру, ніж звичайні ІС. Тому їхній захист вимагає більших ресурсів.

Для ІС та АСК ЗТ характерними є наступні види елементів: бортові засоби (наприклад: електронні датчики, які вимірюють різні фізичні величини; камери, які забезпечують відеоспостереження за навколишнім середовищем; засоби обробки інформації; системи комп'ютерного зору, які використовують для обробки даних, отриманих відеокамер; системи розпізнавання образів, які використовують для розпізнавання об'єктів на залізничному шляху; системи радіозв'язку, які використовуються для обміну інформацією між рухомим складом та іншими об'єктами на залізничному шляху; та ін.); засоби, що встановлюються на стаціонарні об'єкти інфраструктури ЗТ (наприклад, системи: автоблокування, сигналізації та ін.); дистанційно керовані виконавчі та індикаційні пристрої (наприклад, системи: управління гальмуванням; управління дверима; управління освітленням та ін.); сервери для опрацювання та зберігання інформаційних масивів (ІМ); ситуаційні центри, диспетчерські центри керування ЗТ; засоби забезпечення різноманітного зв'язку; інформаційно-телекомунікаційні системи та ін.

Неповнота інформації про загрози ІБ для АСК ЗТ може проявлятися двома способами. По-перше, ми можемо не мати повної інформації про структуру об'єкта кібератаки, навіть на базовому рівні. Особливо якщо цей об'єкт має складну архітектуру. По-друге, ми можемо не бути в змозі повністю спостерігати за об'єктом кібератаки й розпізнавати всі кіберзагрози та атаки, які він може зазнати.

Для побудови ефективної СЗІ АСК ЗТ, вибору та впровадженню адекватних технічних засобів кібербезпеки повинен передувати опис, аналіз і моделювання загроз й уразливостей АСК ЗТ. Також потрібен ретельний розрахунок й аналіз ризиків ІБ для АСК ЗТ. Отже, очевидним є те, що спочатку кожна загроза повинна бути впізнана й ідентифікована.

Сучасні системи виявлення та протидії кібератакам (СВАП) на залізничному транспорті у країнах ЄС, США, Японії, Південної Кореї використовують методи, які є ефективними лише

в тому випадку, якщо відомо, які кібератаки або загрози ІБ можуть бути реалізовані зловмисниками.

Саме тому, нові дослідження, пов'язані з методами інтелектуального розпізнавання загроз, мають вирішальне значення для забезпечення ІБ в ІС або АСК ЗТ. Це, зокрема обумовлено можливостями забезпечити:

Реактивний та прогностичний підхід. Методи інтелектуального розпізнавання дозволяють не лише реагувати на вже відомі загрози, але й прогнозувати нові атаки на основі аналізу попередніх моделей або аномальних відхилень, що підвищує рівень захисту ІС або АСК ЗТ;

Виявлення нових загроз. Застосування інтелектуального розпізнавання дозволяє виявляти навіть раніше невідомі атаки, загрози або аномалії в поведінці ІС або АСК ЗТ, що важливо для запобігання їхнього розвитку та поширення;

Автоматизація аналізу великих обсягів даних. Методи інтелектуального розпізнавання дозволяють ефективно обробляти та аналізувати великі обсяги інформації, що важливо для вчасного виявлення аномалій та загроз для ІС або АСК ЗТ;

Підвищення швидкості реагування. Завдяки автоматизації процесу аналізу та розпізнавання, можливо швидше реагувати на потенційні загрози та вчасно приймати заходи щодо їх запобігання чи мінімізації впливу;

Адаптація до змін у загрозах ІБ. Методи інтелектуального розпізнавання дозволяють ІС або АСК ЗТ адаптуватися до нових видів атак та загроз, та змінювати стратегії захисту в реальному часі.

Оцінка можливості реалізації конкретної загрози для ІБ ІС або АСК ЗТ залежить від багатьох факторів. Загроза, яка може завдати шкоди, є потенційно небезпечною і тому повинна бути оцінена. Оцінка загрози для ІС та/або АСК ЗТ передбачає формування моделі загроз, яка описує потенційні способи реалізації загрози та її наслідки.

Методи та моделі. Нижче зупинимося на запропонованому у [1, 2, 3] методі і моделі інтелектуального розпізнавання загроз для безпеки ІС та АСК ЗТ, заснованих на побудові покриттів класів загроз ІБ.

Нехай існує ряд загроз об'єкту інформаційної безпеки (ОІБ у нашому випадку АСК ЗТ) (загальна класифікація загроз наведена у [1]). Рівень загрози для АСК ЗТ залежить від того, наскільки система захищена. Фактори, які можуть призвести до порушення безпеки АСК ЗТ – це фактори ризику. Фактори, які сприяють підвищенню безпеки АСК ЗТ, будемо називати факторами захищеності.

Як було показано в роботах [2, 3] апарат нечіткої логіки виявляється перспективним напрямом для створення систем інтелектуального розпізнавання загроз ІБ ІС або АСК ЗТ. Цьому сприяють наступні фактори [2, 3]:

Робастність до невизначеності. Нечітка логіка дозволяє враховувати та обробляти нечіткі або невизначені дані, що дозволяє аналізувати та розпізнавати аномальні патерни, коли точні значення невідомі або неоднозначні;

Адаптивність до змінних умов. Системи розпізнавання загроз ІБ, що базуються на нечіткій логіці, можуть легше адаптуватися до змінних умов роботи та нових видів загроз, що важливо у контексті кібербезпеки, де атаки постійно еволюціонують.

Менша чутливість до шуму в даних. Нечітка логіка може працювати ефективно при наявності шуму або неточностей у вхідних даних, що дозволяє покращити точність розпізнавання загроз при обробці реальних даних з ІС або АСК ЗТ.

Використання апарату нечіткої логіки вимагає математичного моделювання нечітких множин, розробки нечітких правил та алгоритмів, які можуть бути використані для аналізу даних та прийняття рішень у системах інтелектуального розпізнавання загроз. Для кожного із співвідношень дерева висновку в [1] побудовані нечіткі бази знань (див. табл. 1, показано частину бази знань), які представляють сукупність нечітких правил «ЯКЩО-ТОДІ», що визначають взаємозв'язок між вхідними та вихідною змінними, див. табл. 1. За нечіткими базами знань складені логічні рівняння, див. табл. 2.

Таблиця 1 – Фрагмент бази знань для розпізнавання загроз ІБ для ІС та АСК ЗТ (об'єкту інформаційної безпеки (ОІБ), аналогічно складається для інших ІС на транспорті)

Класи загроз ІБ	Атрибути		Ознаки $\{P_{ax1}, \dots, P_{axn}\}$	Універсум	Терми для лінгвістичної оцінки	
Можливі загрози ІС та АСК ЗТ	Відомі загрози	KL_1	Відмова в обслуговуванні	1 – Стандартні компоненти АСК ЗТ не функціонують; 2 – ін.	[0,1], у. о.	<ul style="list-style-type: none"> не представляє значної загрози для ІБ АСК ЗТ (не критичний – нкр), представляє значну загрозу (критичний – кр)
		KL_2	Викрадення компонентів	1 – фактичні ознаки (наприклад, поява конфіденційної інформації у ЗМІ); 2 – необ'єктивні ознаки; 3 – ін.	[0,1], у. о.	<ul style="list-style-type: none"> знайдені в рамках моніторингу ІБ (виявлені – в), не повністю вивчена (частково не зафіксовані / невиявлені – чв), не знайдені (нв)
		KL_3	Привласнення особистості	1 – фактичні ознаки (наприклад, було виявлено спроби несанкціонованого доступу до АСК ЗТ під чужим обліковим записом.); 2 – необ'єктивні ознаки; 3 – ін.	[0,1], у. о.	<ul style="list-style-type: none"> знайдені в рамках моніторингу ІБ (в), не повністю вивчені (чв), не знайдені (нв)
Інше						
Стани ІС або АСК в контексті забезпечення ІБ						
<p>S_1 – встановлене ПЗ та оновлення до нього; S_2 – в АСК ЗТ присутні мережеві сервіси; S_3 – АСК ЗТ підтримує багатозадачність; S_4 – підтримка багатокористувацького режиму; S_5 – встановлені пристрої введення / виводу; S_6 – наявність пристроїв «гарячої заміни»; S_7 – наявність зовнішніх каналів зв'язку; S_8 – наявність системи відеоспостереження, яка з'єднана з АСК ЗТ; S_9 – наявність системи супутникової навігації, яка з'єднана з АСК ЗТ; S_{10} – наявність ЗЗІ.</p>						
Методи протидії загрозам та атакам						
<p>D_1 – ідентифікація і аутентифікація; D_2 – блокування безконтрольного доступу; D_3 – захист від шкідливого програмного забезпечення; D_4 – контроль цілісності даних; D_5 – знищення залишкових даних; D_6 – захист ПЗ та інформаційних масивів ІС та АСК ЗТ від дослідження; D_7 – резервування інформації ІС та АСК ЗТ; D_8 – відновлення і самовідновлення компонентів ІС та АСК ЗТ; D_9 – перевірка сертифіката безпеки; D_{10} – блокування запуску ПЗ; D_{11} – криптографічний захист; D_{12} – ін.</p>						

Таблиця 2 – Фрагмент бази знань у вигляді правил

Правила			
<i>PR1</i>	$IF (KL_1 \vee S_1 \vee S_7) THEN D_2$	<i>PR10</i>	$IF (KL_9 \vee S_1) THEN D_8$
<i>PR2</i>	$IF (KL_1 \vee S_2 \vee S_7) THEN D_2$	<i>PR11</i>	$IF (KL_{10} \vee S_7) THEN D_7$
<i>PR3</i>	$IF (KL_2 \vee S_4 \vee S_7) THEN D_2$	<i>PR12</i>	$IF (KL_{11} \vee S_4 \vee S_7) THEN D_2$
<i>PR4</i>	$IF (KL_3 \vee S_4) THEN D_8$	<i>PR13</i>	$IF (KL_{11} \vee S_5) THEN D_7$
<i>PR5</i>	$IF (KL_4 \vee S_1) THEN D_4$	<i>PR14</i>	$IF (KL_{11} \vee S_4 \vee S_7) THEN D_2$
<i>PR6</i>	$IF (KL_5 \vee S_3) THEN D_6$	<i>PR15</i>	$IF (KL_5 \vee S_7) THEN D_2$
<i>PR7</i>	$IF (KL_6 \vee S_1) THEN D_3$	<i>PR16</i>	$IF (KL_5 \vee KL_8 \vee KL_{11} \vee S_5 \vee S_7) THEN D_{10}$
<i>PR8</i>	$IF (KL_7 \vee S_4 \vee S_5 \vee S_6 \vee S_7) THEN D_5$	<i>PR17</i>	$IF (KL_3 \vee KL_9 \vee KL_4) THEN D_9$
<i>PR9</i>	$IF (KL_8 \vee S_4) THEN D_1$	<i>PR18...PR_{nm}</i>	<i>Інші</i>

Якщо істинність умови логічного рівняння більша за нуль, тоді правило активується. В базах знань для об'єднання умов в правилах використовуються нечіткі логічні операції, такі як нечітка кон'юнкція, нечітка диз'юнкція, нечітка відмова та інші.

Запропоновано метод, дозволяє формувати вирішальні правила для дискретних процедур розпізнавання загроз, наприкладі, АСК ЗТ. Метод базується на аналізі критичності окремих елементів АСК ЗТ.

Суть методу полягає у визначенні кон'юнкцій за покриттям класів загроз ІБ, та який відрізняється від наявних методів застосуванням дискретних процедур із використанням апарату логічних функцій та нечітких множин ознак атак на АСК ЗТ, що дозволяє створювати ефективні програмні, аналітичні та схемотехнічні рішення для систем кібернетичної безпеки АСК ЗТ.

Також запропоновано метод формування вирішального правила для дискретних процедур розпізнавання загроз ІБ АСК ЗТ, який базується на процедурі аналізу критичності окремих елементів АСК ЗТ, та на відміну від наявних, забезпечує можливість виконувати інтелектуальне розпізнавання загрози з мінімальною кількістю помилок та урахувати нетипові ознаки кібератак.

Висновки і перспективи Будь який транспорт, використовує значну кількість комп'ютеризованих систем для управління, що робить його уразливим перед кібератаками. Інтелектуальні системи можуть моніторити та виявляти потенційні загрози безпеці автоматизованих систем керування на транспорті. У сфері кібернетичної безпеки (КБ), інтелектуальні системи розпізнавання загроз стають ключовим елементом захисту. Ці системи дозволяють автоматично виявляти аномальну поведінку та потенційні загрози в мережах та автоматизованих системах керування на транспорті, сприяючи швидкому реагуванню та запобіганню можливим атакам. Запропонований у статті метод інтелектуального розпізнавання загроз, який базується на дискретних процедурах та використанні логічних функцій та нечітких множинах, пропонується з метою підвищення ефективності розпізнавання загроз для автоматизованих систем керування на транспорті, наприклад, залізничному, автомобільному та ін.

Список використаних джерел

1. Lakhno V.A., Petrov O.S., Hrabariev A.V., Ivanchenko Y.V., Beketova G.S. Improving of information transport security under the conditions of destructive influence on the information-communication system (2016) *Journal of Theoretical and Applied Information Technology*, 89 (2), pp. 352 – 361.
2. Lakhno V., Mohylnyi H., Donchenko V., Smahina O., Pyroh M. A model developed for teaching an adaptive system of recognising cyberattacks among nonuniform queries in information systems (2016) *Eastern-European Journal of Enterprise Technologies*, 4 (9), pp. 27 – 36. DOI: 10.15587/1729-4061.2016.73315
3. Lakhno V. Creation of the adaptive cyber threat detection system on the basis of fuzzy feature clustering (2016) *Eastern-European Journal of Enterprise Technologies*, 2 (9), pp. 18 – 25. DOI: 10.15587/1729-4061.2016.66015

Lakhno Valeriy

Doctor of Technical Sciences, Professor of the Department of Computer systems, networks and cybersecurity,

National University of Life and Environmental Sciences of Ukraine,

ORCID: <https://orcid.org/0000-0001-9695-4543>

E-mail: lva964@nubip.edu.ua

Kasatkin Dmytro

PhD, Associate Professor, Head of the Department of Computer systems, networks and cybersecurity,
National University of Life and Environmental Sciences of Ukraine

ORCID: <https://orcid.org/0000-0002-2642-8908>

E-mail: d.kasatkin@nubip.edu.ua

INTELLIGENT CYBER THREATS RECOGNITION SYSTEMS AS A COMPONENT OF THE INFORMATION SECURITY SYSTEM IN TRANSPORT

Abstract. The transport industry, in particular the railway transport (RT), is increasingly using computerized systems for management, which makes it vulnerable to cyber attacks. Intelligent threat recognition systems are becoming important for the protection of information and cyber security in transport. Ensuring cyber security for information systems and automated control systems in transport is an important task for ensuring the national security of Ukraine. This makes it possible to prevent possible cyber attacks, preserve the confidentiality of information, ensure the reliable operation of the transport system and preserve the safety of citizens. The development and implementation of effective cyber security measures are necessary steps to ensure the stability and reliability of transport systems and the preservation of national security in Ukraine. This article describes the method of intelligent recognition of threats for information systems of transport management (on the example of railway transport – TZ). The method is based on discrete procedures and the use of logical functions and fuzzy sets.

Keywords: transport, automated control systems, cyber security, threat recognition, intelligent systems.

УДК 004.4

Касаткін Дмитро Юрійович

кандидат педагогічних наук, доцент, завідувач кафедри комп'ютерних систем, мереж та кібербезпеки,

Національний університет біоресурсів та природокористування України

ORCID: <https://orcid.org/0000-0002-2642-8908>

E-mail: d.kasatkin@nubip.edu.ua

Волошин Семен Михайлович

кандидат технічних наук, доцент, доцент кафедри комп'ютерних систем, мереж та кібербезпеки,

Національний університет біоресурсів та природокористування України

ORCID: <https://orcid.org/0000-0002-4913-7003>

E-mail: voloshyn@nubip.edu.ua

Гусєв Борис Семенович

кандидат технічних наук, доцент, доцент кафедри комп'ютерних систем, мереж та кібербезпеки,

Національний університет біоресурсів та природокористування України

ORCID: <https://orcid.org/0000-0003-1658-7822>

E-mail: gusevbs@gmail.com

Матієвський Володимир Валерійович

старший виклад кафедри комп'ютерних систем, мереж та кібербезпеки,

Національний університет біоресурсів та природокористування України

ORCID: <https://orcid.org/0000-0002-1954-8493>

E-mail: m_vv@outlook.com

АЛГОРИТМИ РОЗРОБКИ БАЗИ ЗНАНЬ ДЛЯ ВДОСКОНАЛЕННЯ СИСТЕМ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ ПРИ ВИРІШЕННІ ПРОБЛЕМ КІБЕРБЕЗПЕКИ

Анотація. У статті представлено розробку модульної системи підтримки прийняття рішень (СППР) для кібербезпеки, спрямовану на підвищення захищеності критично важливих комп'ютерних систем (КВКС). В основі системи лежить модель підсистеми нечіткого виводу (НечВ), яка, використовуючи дані з датчиків та систем SIEM, здатна виявляти ознаки загроз, аномалій та атак шляхом фазифікації вхідних значень. Розроблений алгоритм формування бази знань про типові та аварійні ситуації дозволяє системі не тільки ефективно реагувати на відомі загрози, але й аналізувати непередбачувані ситуації. Застосування модуля НечВ дає змогу створити багатопараметричний образ уразливості КВКС, що забезпечує більш комплексну та точну оцінку їх захищеності.

Ключові слова: критично важливі комп'ютерні системи, кібербезпека, система підтримки прийняття рішень, багатопараметричний "образ", оцінка захищеності.

Вступ. У процесі експлуатації критично важливих комп'ютерних систем (КВКС) одним із пріоритетних завдань опрацювання даних, що надходять від різних пристроїв, які входять до структури комплексних систем захисту інформації (далі – КСЗІ, для яких мається на увазі, перш за все, апаратно-програмні складові), є отримання відомостей про стан компонентів захисту. Ефективність і безпомилковість під час оперативного оцінювання ступеня захищеності КСЗІ може ускладнитися через вплив наведених нижче факторів. Перший – дані, що надходять (від датчиків SIEM, мультиагентних систем, сенсорів, що визначають наявність загроз, кібератак, аномалій, далі запроваджено аббревіатуру датчиків підсистеми – ДатП), можуть бути різними за своїми параметрами. Другий – у процесі отримання даних можливий вплив зовнішніх впливів, що впливають на автентичність характеристик, що відстежуються.

Третє – реакція на деструктивні втручання обмежені часовими рамками, при цьому залишається мінімальний час для результатів аналізу. Четверте – можливі ситуації, коли комбінація оцінюваних параметрів у КВКС призводить до "нечіткості" під час ухвалення рішень з оцінки поточних станів захищеності КВКС (на відміну від типових). З огляду на перераховані причини, оперативні та ефективні рішення в процесі аналізу складних таргетованих кібератак на КВКС і відповідних процедур ухвалення рішень, вимагають застосування спеціальних аналітичних систем. Цілком очевидно, що подібні системи мають бути засновані на сучасних методах автентифікації станів КВКС і КСЗІ. Також доцільно залучити потенціал інтелектуалізованих адаптивних систем підтримки ухвалення рішень (СППР) у задачах кібербезпеки (КрБ), а також і розпізнавання загроз, аномалій та кібератак [1].

У період лавиноподібного зростання складності та кількості кібератак на КВКС [2], і зростання чисельності параметрів, що надходять від сенсорних ДатП КСЗІ, виникла необхідність впровадження в структури КСЗІ адаптивних експертних (АЕС) і СППР. Це необхідно для комплексного багатокритеріального аналізу даних від ДатП КСЗІ, які формують дані для оцінювання захищеності КВКС.

Зауважимо, що одним із найрезультативніших методів у розв'язанні цього класу завдань є метод, що передбачає побудову АЕС і СППР на базі теорії нечітких множин (ТНМ). Можливе також використання апарату нечіткої логіки (НЛ) [1, 3, 4].

Використання АЕС і СППР здатне мінімізувати вплив "людського фактора" на якість прийнятих рішень. Крім того, зростає швидкість ухвалення рішення. Скорочуються можливі ситуації, пов'язані з відволіканням персоналу служб інформаційної безпеки на рутинні роботи. Також зрештою, скорочується собівартість володіння подібними комплексами.

Огляд літератури та постановка проблеми останніми роками ускладнилися сценарії реалізації кібератак [2, 5]. Фіксується зростання кількості фіксованих у КВКС аномалій та інших спроб несанкціонованого втручання в роботу складних цифрових систем [2, 6]. У цих умовах виник напрямок досліджень з інтелектуалізації процедур підтримки прийняття рішень у процесі розпізнавання загроз, кібератак та аномалій. Аналіз наявного світового досвіду [2, 4-8] підтверджує, що екстенсивний підхід до розв'язання завдань кібербезпеки КВКС коштом нарощування засобів і заходів із захисту інформації (ЗІ), часто не приводять до очікуваного результату. Перспективним напрямком досліджень стали роботи, присвячені створенню інтелектуалізованих систем підтримки ухвалення рішень (СППР) [1, 4] та експертних систем (ЕС) [3, 8] у задачах оцінювання захищеності об'єктів інформатизації. Ці дослідження ще не завершені.

У [1, 3, 4, 8] проаналізовано досвід упровадження комерційних СППР та ЕС для задач аналізу загроз, атак і аномалій. Зазначено, що комерційні системи мають закритий характер, і їх придбання окремими компаніями або організаціями пов'язане зі значними фінансовими витратами.

Таким чином, з огляду на полеміку в роботах [1, 2, 8], видається релевантним завдання з розроблення нових і вдосконалення наявних моделей та алгоритмів для адаптивних СППР, що задіяні в процесах опрацювання даних від різноманітних ДатП підсистем кібербезпеки і (КрБ) захисту інформації (ЗІ) у КВКС.

Мета дослідження – розробити нові або вдосконалити наявні моделі та алгоритми для адаптивних СППР, що задіяні в процесах аналізу даних від ДатП підсистем кібербезпеки та захисту інформації в КВКС.

У рамках статті розглянуто завдання з розроблення:

модуля "нечіткого логічного висновку" для експертного вивчення даних від ДатП КВКС; алгоритму, що формує базу знань (БЗ) про типові та непередбачувані ситуації (НепС) у КВКС (що забезпечує експертний аналіз ступеня захищеності КВКС).

Моделі та методи загальна структура розроблюваної модульної системи підтримки ухвалення рішень у задачах кібербезпеки наведена на рис. 1.

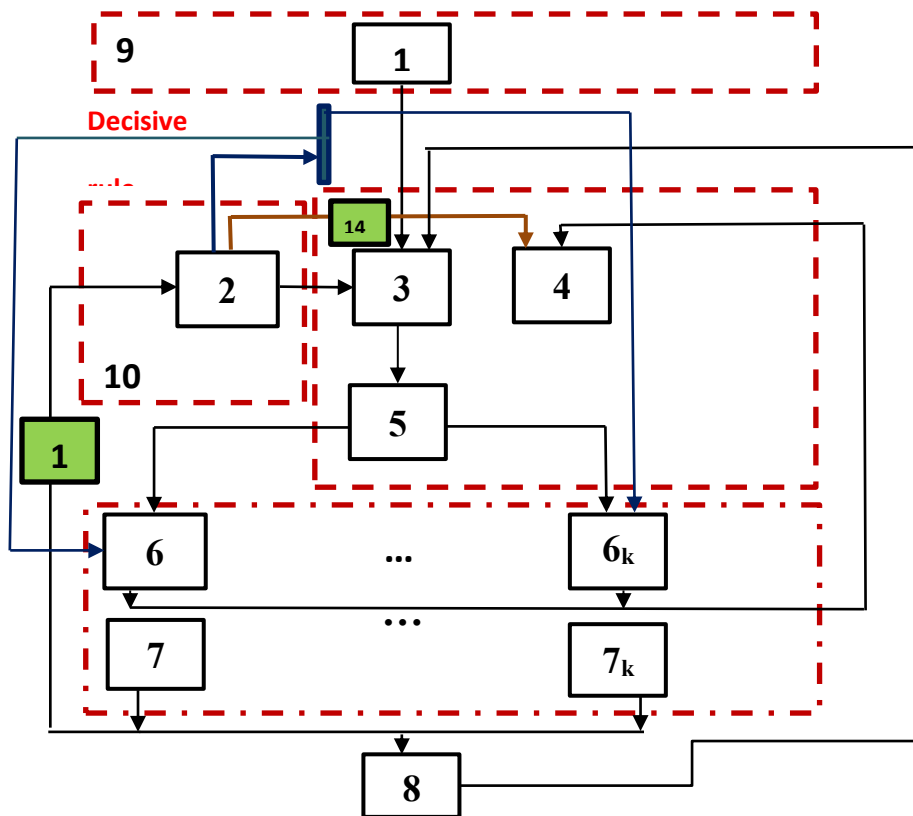


Рисунок 1 – Загальна структура модульної системи підтримки прийняття рішень у задачах кібербезпеки

Позначення, прийняті на схемі: 1 – пристрій введення; 2 – сервер; 3 – модуль візуалізації; 4 – модуль дефазифікації; 5 – модуль фазифікації; 6 (6_k) – модулі підсистеми нечіткого виведення; 7 (7_k) – пристрої виведення; 8 – модуль виведення результатів аналізу та рекомендацій щодо виходу з непередбачуваних ситуацій; 9 – модуль первинного опрацювання інформації, що надходить від датчиків, мультиагентних систем, сенсорів, які визначають наявність загроз, кібератак, аномалій; 10 – модуль сервера з базами знань (БЗ); 11 – модуль аналізу основних параметрів функціонування КВКС з інтегрованою оцінкою захищеності; 12 – модулі співробітників служб захисту інформації та кібербезпеки КВКС (за кількістю підсистем КВКС); 13 – нові правила (рекомендації), які додають до БЗ; 14 – рекомендації з виходу з непередбачуваних ситуацій, пов'язаних із КрБ КВКС.

Модуль нечіткого логічного виводу (НЛВ) та алгоритми, що формують БЗ типових (еталонні) та непередбачених ситуацій для СППР з аналізу захищеності КВКС описані нижче.

Модулі НЛВ призначені для реалізації системи нечіткого виводу (НечВ) на рис. 1 позначені як 6–6_k. Ґрунтуючись на правилах нечіткого виводу (НечВ), за вхідними значеннями ДатП $\{Rd(R_{i,n_i}(t)), m(R_{i,n_i}(t))\} (i=\overline{1,k}; n_i=\overline{1,N_i})$, визначаються вихідні значення

$$\{(Rdc(R_{i,n_i}^j(t)), m(Rdc_{i,n_i}^j(t))\} (i=\overline{1,k}; n_i=\overline{1,N_i}; j = \overline{1,J}).$$

При цьому вважаємо, що вхідні значення були отримані як результат процедури фаифікації у відповідному модулі (модуль 5, рис. 1). Кожний елемент вихідних значень, своєю чергою, характеризує наявність (відсутність) ознаки НепС (далі введено позначення – EmS). Тоді конкретну ознаку непередбаченої ситуації (j), наприклад, що виникла як наслідок кібератаки на КВКС, можна описати за допомогою змінних, що описані нижче.

Вводяться такі змінні:

$Md_{i,n_i}^j(t)$ – характеристика дискретного стану n_i -го *Param*, наприклад, кількість байтів від джерела до адресату, кількість байтів відповіді клієнту, ознаки з'єднання, кількість "root" доступів, кількість операцій створення файлів, кількість запитів на надання оболонки,

кількість операцій на доступ до контролю файлів та ін. (приймаємо за KDD 99) [11, 12]. Ця змінна може набувати одного з наступних значень, див. рис. 2:

- «1» – відхилення від норми зверху;
- «0» – типова ситуація (або норма);
- «-1» – відхилення від норми знизу;
- «2» – відхилення від норми знизу або зверху;

$m(Md_{i,n_i}^j(t))$ – експериментальна оцінка ступеня експлікованості впливу n_i -ї величини (параметр – *Param*) (i) підсистеми КВКС на виникнення j -ї *EmS* (наприклад, аномалії в мережі КВКС), у кожний момент часу (t).

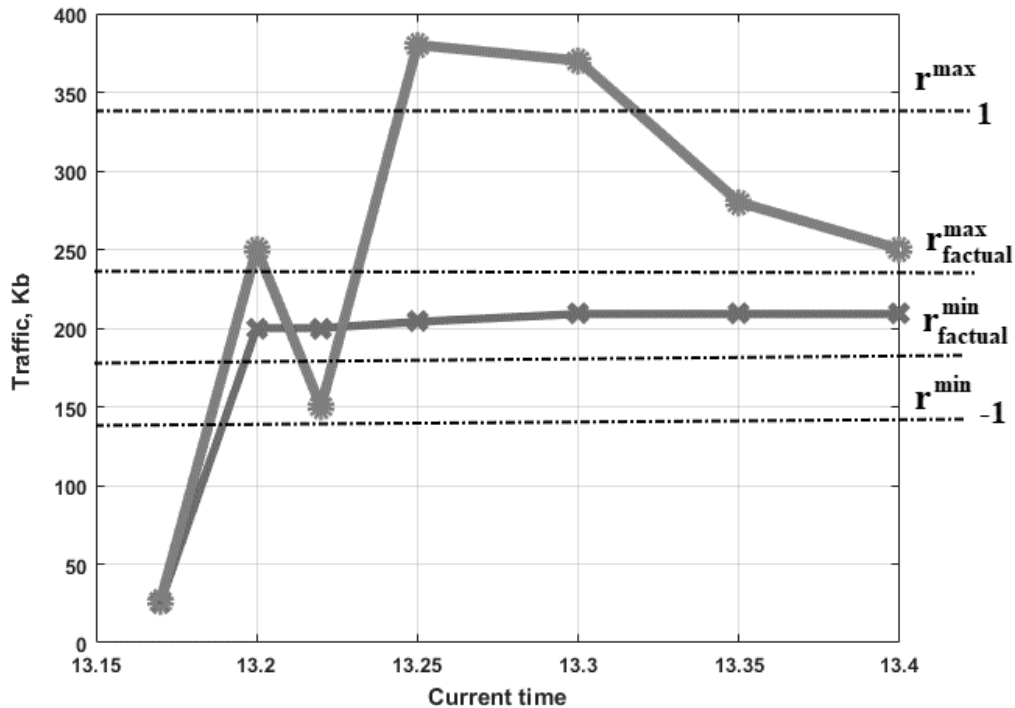


Рисунок 2 – Формування змінної для модуля підсистеми нечіткого виведення

Позначення: "0" – типова ситуація (або норма); "1" – відхилення від норми зверху;
"-1" – відхилення від норми знизу; "2" – відхилення від норми знизу або вгору

Правила виявлення j -ї *EmS* можна описати таким чином:

$$\text{if } RD(R_{i,n_i}(t)) \vee Md_{i,n_i}^j(t) = 1 \text{ that is } (j) \\ \text{зі ступенем експлікованості } m(Md_{i,n_i}^j(t)); \quad (1)$$

$$\text{if } RD(R_{i,n_i}(t)) \vee Md_{i,n_i}^j(t) = -1 \text{ that is } (j) \\ \text{зі ступенем експлікованості } m(Md_{i,n_i}^j(t)); \quad (2)$$

$$\text{if } (RD(R_{i,n_i}(t)) = 1 \vee Md_{i,n_i}^j(t) = -1) \vee (Md_{i,n_i}^j(t) = 2) \text{ that is } (j) \\ \text{зі ступенем експлікованості } m(Md_{i,n_i}^j(t)); \quad (3)$$

На рис. 2 показано приклад формування змінних для модуля підсистеми нечіткого виведення під час наповнення БЗ СППР даними щодо трафіку протоколу TCP (для конкретної

КВКС на прикладі атаки TCP-flood). На графіку, показаному лінією із двома штрихами, задано фактичне середнє порогове значення трафіку. Якщо протягом від 30 секунд буде зареєстровано відхилення від норми (червона лінія), до БЗ СППР заноситься відповідна інформація, яка описує аномалію або загрозу. Якщо протягом заданого проміжку середні значення метрик перевищать порогові значення, то аналітику буде виведено повідомлення про відхилення від норми зверху або знизу ("1" або "-1"). Інформація про відхилення також доступна для перегляду адміністратору мережі КВКС.

У кожен момент (t) для виявлення EmS за ознакою (j) може змінюватися тільки одне з перерахованих трьох правил (1) – (3).

Вважаємо, що кожне правило, описане в БЗ СППР з розпізнавання загроз, аномалій і кібератак в КВКС, тобто j -й EmS , може зіставити значення відхилень $R_{i,n_i}(t)$ в момент (t) для n_i -го $Param$ i -ї підсистеми КВКС зі ступенем експлікованості впливу n_i -го параметра на EmS , яка дорівнює $m(Md_{i,n_i}^j(t))$;

Як наслідок використання правил (1) – (3) для n_i -го $Param$ та i -ї підсистеми КВКС можна сформулювати (J) вихідних значень тобто .

$$\{(Rdc(R_{i,n_i}^j(t)), m(Rdc_{i,n_i}^j(t)))\} (i=\overline{1,k}; n_i=\overline{1,N_i}; j = \overline{1,J}).$$

Зауважимо, що для фіксованих (наприклад, критичних номерів j -х EmS) вважаємо, що:

- якщо є активне правило з (1) – (3) то можлива j -а EmS , тобто

$$Rdc_{i,n_i}^j(t) = m(R_{i,n_i}^j(t)) \vee m(Rdc_{i,n_i}^j(t)) = m(Md_{i,n_i}^j(t));$$

- якщо немає активних правил з (1)-(3) то не можлива j -та EmS , тобто

$$Rdc_{i,n_i}^j(t) = 0 \vee m(Rdc_{i,n_i}^j(t)) = m(Md_{i,n_i}^j(t)).$$

У модулі дефазифікації (J) вихідним значенням

$$\{(Rdc(R_{i,n_i}^j(t)), m(Rdc_{i,n_i}^j(t)))\} (i=\overline{1,k}; n_i=\overline{1,N_i}; j = \overline{1,J})$$

ставиться числове значення (нечітке). За даним значенням у подальшому за допомогою СППР визначаємо набір рекомендацій щодо виходу з непередбачених ситуацій (EMS).

Вагові коефіцієнти, наприклад, для j -їй EMS , визначені таким чином:

$$\zeta_i^j = \frac{\sum_{n_i=1}^{N_i} m(Rdc_{i,n_i}^j(t)) \cdot Rdc_{i,n_i}^j(t)}{\sum_{n_i=1}^{N_i} m(Rdc_{i,n_i}^j(t))}. \tag{4}$$

Вважаємо, що $EMS(j^*)$ відбулася в i -їй підсистемі КВКС, якщо:

$$\zeta_i^{j^*}(t) = \max_{j=\overline{1,J}}(\zeta_i^j(t)) \vee \zeta_i^{j^*}(t) \geq thv_i, \tag{5}$$

де thv_i – порогове значення ступеня виявлення (наприклад, одноразово виявлено, частково виявлено, не виявлено) EMS . Вважаємо, що $thv_i \in [0,1]$.

Якщо $\xi_i^{j*}(t) < thv_i$ то EmS в i -й підсистемі на даний момент (t) не ідентифіковано. У такій ситуації слід розпочати інтерактивний аналіз експерта з КрБ і СППР. При цьому можна задіяти алгоритм формування БЗ для системи НечВ, див. рис. 1.

Алгоритм, що формує БЗ для типових (еталонних) і непередбачуваних ситуацій для СППР описано далі.

Початкові дані для підсистеми нечіткого виводу включають:

- допустимі межі $[r_{i,n_i}^{\min}, r_{i,n_i}^{\max}]$, що визначають типові режими функціонування КВКС;
- функцію приналежності до непередбачених режимів у КВКС для параметрів, що відстежуються, для ситуацій відхилення їх від допустимих меж.

Персональне формування БЗ для типових режимів роботи КВКС наповнюється з урахуванням інструкцій для штатного поведінки КВКС.

За підсумками роботи програмної та апаратної складових КСЗІ відбувається заповнення БЗ СППР фактичними даними про штатні режими роботи системи.

Вважаємо: функція $m(R_{i,n_i}(t))$ для фактичних допусків – це кусочно-лінійна функція, за допомогою якої встановлюють ступінь експлікованості (приналежності):

Якщо «0», то $Param$ перебуває у фактичній зоні межі, що допускається;

Якщо «1», то $Param$ перебуває поза розрахункової області межі, що допускається.

Порогові значення $tv_{q_{n_i}}(t), q_{n_i} = \overline{1, Q_{n_i}}$ розраховувалися при запиті до БЗ з використанням наступних формул:

$$tv_{q_{n_i}}(t) = tv_{Q_{n_i}} \cdot \frac{(q_{n_i} - 1)^{st_{n_i}}}{(Q_{n_i} - 1)^{st_{n_i}}}, \quad (6)$$

$$st_{n_i} = u \cdot \frac{1}{su}$$

де st – ступінь заповнення БЗ для конкретного параметра; u – кількість тестів, які враховані в статистиці; su – кількість тестів СППР, у яких БЗ вважається заповненою (фрагмент БЗ прототипу СППР «Аналізатор загроз» показано на рис.3 [11]).

Процедуру формування БЗ, що включає правила для модуля нечіткого виводу в процесі автоматизованої обробки показань від датчиків SIEM, мультиагентних систем, сенсорів, що визначають наявність загроз, кібератак, аномалій описано нижче.

Нове або існуюче правило виявлення непередбаченої ситуації в КВКС за значенням n_i -го параметра $n_i = \overline{1, N_i}$ для підсистеми ($i = \overline{1, k}$) КВКС формуємо на основі

$$Rd(R_{i,n_i}^j(t), Md_{i,n_i}^j(t)),$$

Далі слід визначити ступені експлікованості впливу $m(Md_{i,n_i}^j(t))$ n_i -параметра на виникнення $j = \overline{1, J} EMS$.

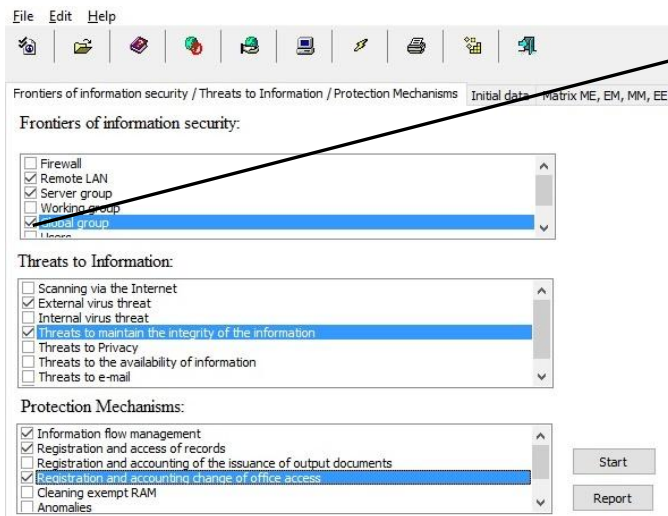
Використовуються правила, описані виразами (1)-(3).

Початкові величини для виявлення $j = \overline{1, J} EMS$:

$Md_{i,n_i}^j(t)$ – параметри, що характеризують дискретні стани n_i – го параметра ($n_i = \overline{1, N_i}$); для підсистеми ($i = \overline{1, k}$) КВКС, які впливають на виникнення ($j = \overline{1, J}$) EmS ;

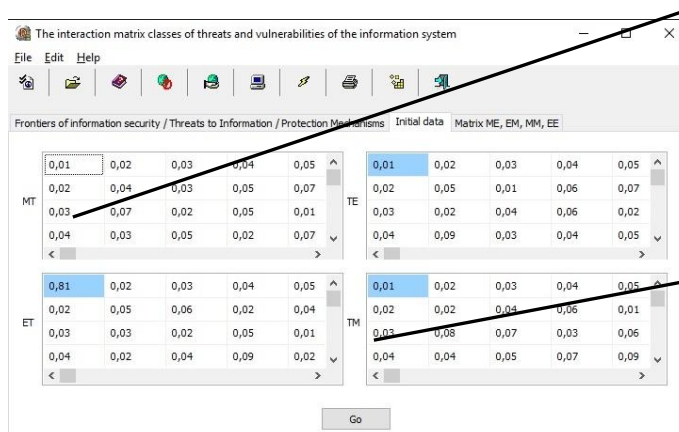
$m(Md_{i,n_i}^j(t))$ – параметри, які характеризують експертне оцінювання ступеня експлікованості впливу n_i -го параметра на появу ($j = \overline{1, J}$) EmS .

Отримані значення заносилися в БЗ СППР "Аналізатор загроз" [12] – рис. 3 – при первинному описі експертами. Також можливе занесення даних у БЗ під час автоматичного/автоматизованого збирання показань від датчиків SIEM, мультиагентних систем, сенсорів, що визначають наявність загроз, кібератак, аномалій.



Список контрольованих параметрів з інформацією від ДатП комплексної системи захисту інформації КВКС

a)



Матриці дефазифікації та фазифікації контрольованих параметрів захищеності КВКС

Матриці, що описують контрольовані ознаки з урахуванням їх інформативності кожної ознаки

б)

Рисунок 3 – Інтерфейс СППР "Аналізатор загроз"

Якщо експерт виявив нову непередбачувану (позаштатну) ситуацію із захищеністю КВКС, він, має можливість відреагувати на повідомлення віконного інтерфейсу СППР. Після чого експерт робить запис у БЗ. Новий запис характеризує поточний стан підсистем, які ідентифікували нештатні ситуації.

При цьому оцінювання фахівцями виконано на підставі самостійного візуального відстеження показань від датчиків SIEM, мультиагентних систем, сенсорів, що визначають наявність загроз, кібератак, аномалій (червоний колір) і за допомогою СППР (синій колір), як показано на рис. 4. Еталонне значення оцінюваного параметра захищеності прийнято рівним 1 [1, 11, 13]. Якщо оцінка параметра дорівнює 0 – захист відсутній. Аналогічне оцінювання виконано для показника захищеності серверів КВКС, що проілюстровано на рис. 5.

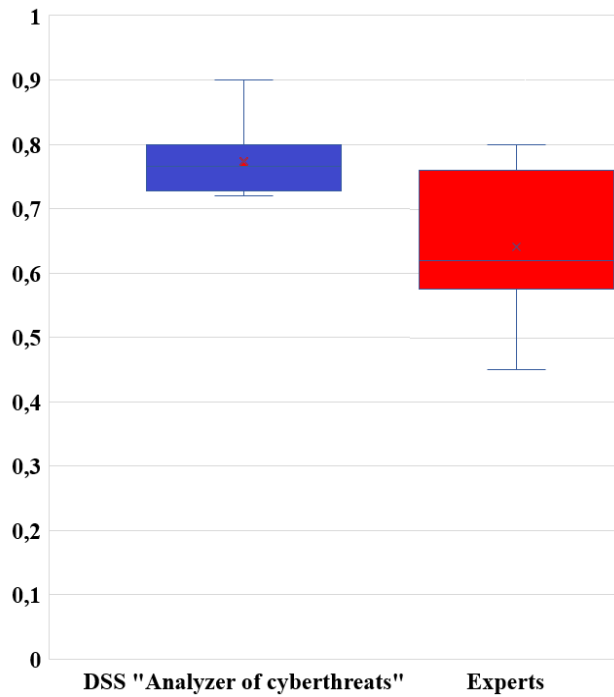


Рисунок 4 – Результати оцінювання експертами самостійно та за допомогою СППР "Аналізатор загроз" ступеня захищеності КВКС

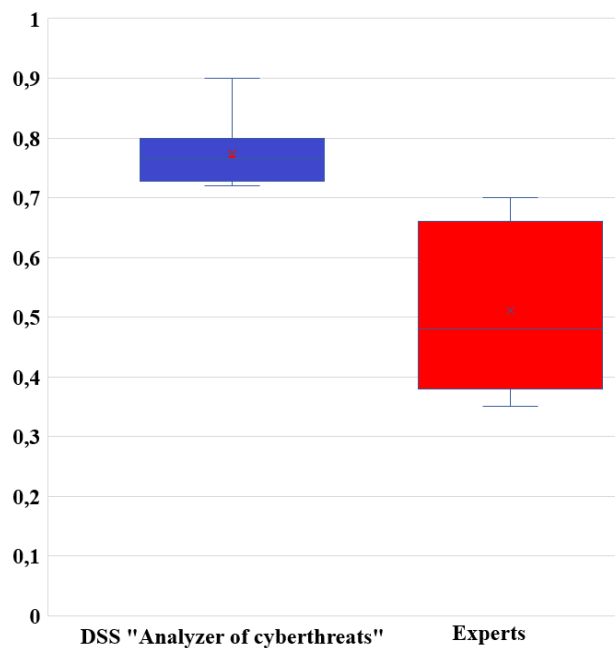


Рисунок 5 – Результати оцінювання експертами самостійно та за допомогою СППР "Аналізатор загроз" ступеня захищеності серверів КВКС

На рис. 4 і 5 видно, що розбіжність у думці експертів, які використовували СППР "Аналізатор загроз", приблизно на 14-18 % менша, ніж для варіанта оцінювання без використання СППР.

Під час тестування прототипу СППР також апробовано механізми взаємодії експертів і СППР "Аналізатор загроз" у процесі синтезу керівних правил у задачах оцінювання ступеня захищеності КВКС.

Висновки. У роботі вперше запропоновано: модель для модуля «нечіткого логічного виводу», який призначений для реалізації підсистеми нечіткого виводу (НечВ). На основі правил нечіткого виводу (НечВ) за вхідними значеннями ДатП визначаються вихідні значення для оцінювання за допомогою СППР ступеня захищеності КВКС. Модель заснована на припущенні, що вхідні значення для підсистеми НечВ були отримані як результат процедури фазифікації у відповідному модулі. Кожен елемент вихідних значень характеризує наявність (відсутність) ознаки непередбаченої ситуації, пов'язаної з аномаліями, атаками чи іншими спробами несанкціонованого втручання у роботу КВКС; алгоритм формування бази знань непередбачених та типових ситуацій у КВКС. Алгоритм відрізняється від відомих тим, що дозволив сформувати сукупність випадків типових варіантів реагування на загрози, аномалії та атаки у КВКС, а також правила виводу для автентифікації непередбачених ситуацій, які пов'язані з цілеспрямованим деструктивним впливом на КВКС.

Показано, що запропоновані моделі та алгоритми експертного вивчення даних від ДатП, відрізняються від відомих тим, що у них застосовуються елементи теорій експертних систем, нечітких множин, а також апарат нечіткої логіки. Використання для модуля «нечіткого логічного виводу» дозволяє забезпечити відображення стану найбільш уразливих компонентів КВКС як багатопараметричний образ (БПВ). Отриманий БПО використовується в СППР для якісної оцінки процесів, що протікають у КВКС. При цьому бралася до уваги можлива «нечіткість» в оцінюванні ситуації із захищеністю КВКС.

Список використаних джерел

1. Lakhno, V., Boiko, Y., Mishchenko, A., Kozlovskii, V., & Pupchenko, O. (2017). Development of the intelligent decision-making support system to manage cyber protection at the object of informatization. *Eastern-European Journal of Enterprise Technologies*, (2(9)), pp. 53–61.
2. Iasiello, E. (2013, June). Cyber attack: A dull tool to shape foreign policy. In *Cyber Conflict (CyCon)*, 2013 5th International Conference on (pp. 1–18). IEEE.
3. Goztepe, K. (2012). Designing Fuzzy Rule Based Expert System for Cyber Security, *International Journal of Information Security Science*, 1(1), pp. 13–19.
4. Akhmetov, B., Lakhno, V., Boiko, Y., Mishchenko, A. (2017). Designing a decision support system for the weakly formalized problems in the provision of cybersecurity, *Eastern-European Journal of Enterprise Technologies*, 1(2 (85)), pp. 4–15.
5. Hu, X., Xu, M., Xu, S., & Zhao, P. (2017). Multiple cyber attacks against a target with observation errors and dependent outcomes: Characterization and optimization. *Reliability Engineering & System Safety*, 159, pp. 119–133.
6. Yang, Y., Xu, H. Q., Gao, L., Yuan, Y. B., McLaughlin, K., & Sezer, S. (2017). Multidimensional intrusion detection system for IEC 61850-based SCADA networks. *IEEE Transactions on Power Delivery*, 32(2), pp. 1068–1078.
7. Wong, K., Dillabaugh, C., Seddigh, N., & Nandy, B. (2017). Enhancing Suricata intrusion detection system for cyber security in SCADA networks. In *Electrical and Computer Engineering (CCECE)*, 2017 IEEE 30th Canadian Conference on (pp. 1–5). IEEE.
8. Elhag, S., Fernández, A., Bawakid, A., Alshomrani, S., & Herrera, F. (2015). On the combination of genetic fuzzy systems and pairwise learning for improving detection rates on intrusion detection systems. *Expert Systems with Applications*, 42(1), pp. 193–202.
9. Moustafa, N., & Slay, J. (2016). The evaluation of Network Anomaly Detection Systems: Statistical analysis of the UNSW-NB15 data set and the comparison with the KDD99 data set. *Information Security Journal: A Global Perspective*, 25(1-3), pp. 18–31.
10. Villaluna, J. A., & Cruz, F. R. G. (2017). Information security technology for computer networks through classification of cyber-attacks using soft computing algorithms. In *Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment and Management (HNICEM)*, 2017 IEEE 9th International Conference on (pp. 1–6). IEEE.
11. Lakhno, V., Kazmirchuk, S., Kovalenko, Y., Myrutenko, L., & Zhmurko, T. (2016). Design of adaptive system of detection of cyber-attacks, based on the model of logical procedures and the

- coverage matrices of features. *Eastern-European Journal of Enterprise Technologies*, (3 (9)), pp. 30–38. DOI: 10.15587/1729-4061.2016.71769.
12. Lakhno, V., Tkach, Y., Petrenko, T., Zaitsev, S., & Bazylevych, V. (2016). Development of adaptive expert system of information security using a procedure of clustering the attributes of anomalies and cyber attacks. *Eastern-European Journal of Enterprise Technologies*, (6 (9)), pp. 32–44. DOI: 10.15587/1729-4061.2016.85600
13. Zhang, Y., Wang, L., Xiang, Y., & Ten, C. W. (2016). Inclusion of SCADA cyber vulnerability in power system reliability assessment considering optimal resources allocation. *IEEE Transactions on Power Systems*, 31(6), pp. 4379–4394.

Kasatkin Dmytro,

Candidate of Pedagogical Sciences, Head of the Department of Computer Systems, Networks and Cybersecurity,

National University of Life and Environmental Sciences of Ukraine

ORCID <https://orcid.org/0000-0002-2642-8908>

E-mail: d.kasatkin@nubip.edu.ua

Voloshyn Semen,

Candidate of Engineering Sciences, Associate Professor of the Department of Computer Systems, Networks and Cybersecurity,

National University of Life and Environmental Sciences of Ukraine

ORCID <https://orcid.org/0000-0002-4913-7003>

E-mail: voloshyn@nubip.edu.ua

Gusev Borys,

Candidate of Engineering Sciences, Associate Professor of the Department of Computer Systems, Networks and Cybersecurity,

National University of Life and Environmental Sciences of Ukraine

ORCID <https://orcid.org/0000-0003-1658-7822>

E-mail: gusevbs@gmail.com

Matiievskiy Volodymyr,

Senior Lecturer of the Department of Computer Systems, Networks and Cybersecurity,

National University of Life and Environmental Sciences of Ukraine

ORCID <https://orcid.org/0000-0002-1954-8493>

E-mail: m_vv@outlook.com

ALGORITHMS FOR THE DEVELOPMENT OF A KNOWLEDGE BASE TO ENHANCE DECISION SUPPORT SYSTEMS IN ADDRESSING CYBERSECURITY CHALLENGES

Annotation. This article presents the development of a modular decision support system (DSS) for cybersecurity, aimed at enhancing the protection of critical computer systems (CCS). The system is based on a fuzzy logic inference subsystem (FIS) model that utilizes data from sensors and SIEM systems to detect signs of threats, anomalies, and attacks through fuzzification of input values. A developed algorithm for forming a knowledge base of typical and emergency situations allows the system not only to effectively respond to known threats but also to analyze unforeseen situations. The application of the FIS module enables the creation of a multi-parameter image of CCS vulnerability, which ensures a more comprehensive and accurate assessment of their security.

Keywords: critical computer systems, cybersecurity, decision support system, security assessment.