

СУЧАСНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

№ 1(4)/2025

НАУКОВИЙ ЖУРНАЛ

ISSN 2788-6603

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА

УДК 004

DOI: <https://doi.org/10.17721/AIT.2025.1>

Подано теоретичні та практичні результати досліджень із програмної інженерії, прикладних інформаційних систем і технологій, мережевих та інтернет-технологій, інженерії знань штучного й обчислювального інтелекту, математичних основ інформаційних технологій, інформаційної аналітики й аналітики даних, машинного навчання та розпізнавання образів, цифрових технологій управління IT-проектами, розумних систем та інтернету речей.

Для науковців, фахівців, студентів.

ГОЛОВНИЙ РЕДАКТОР	Снитюк Віталій, д-р техн. наук, проф. (Україна)
ЗАСТУПНИК ГОЛОВНОГО РЕДАКТОРА	Плющ Олександр, д-р техн. наук, доц. (Україна)
ВІДПОВІДАЛЬНІ СЕКРЕТАРІ	Тменова Наталія, канд. фіз.-мат. наук, доц. (Україна) Федусенко Олена, канд. техн. наук, доц. (Україна)
РЕДАКЦІЙНА КОЛЕГІЯ	Аль-Амморі Алі, д-р техн. наук, проф. (Казахстан); Ахмед-Закі Дархан, д-р техн. наук (Україна); Барабаш Олег, д-р техн. наук, проф. (Україна); Бісікало Олег, д-р техн. наук, проф. (Україна); Бичков Олексій, д-р техн. наук, проф. (Україна); Вовк Володимир, проф. (Велика Британія); Гаврилко Євгеній, д-р техн. наук, проф. (Україна); Гнатієнко Григорій, канд. техн. наук (Україна); Жураковський Богдан, д-р техн. наук, проф. (Україна); Корабльов Микола, д-р техн. наук, проф. (Україна); Корнага Ярослав, д-р техн. наук, доц., проф. (Україна); Коршун Наталія, д-р техн. наук, проф. (Україна); Кравченко Юрій, д-р техн. наук, проф. (Україна); Малахов Євген, д-р техн. наук, проф. (Україна); Марченко Олександр, д-р техн. наук, проф. (Україна); Машков Віктор, канд. техн. наук, доц., проф. (Чеська Республіка); Морозов Віктор, канд. техн. наук, проф. (Україна); Мухін Вадим, д-р техн. наук, проф. (Україна); Плескач Валентина, д-р екон. наук, проф. (Україна); Рот Хуберт, проф. (Німеччина); Самохвалов Юрій, д-р техн. наук, проф. (Україна); Сайко Володимир, д-р техн. наук, проф. (Україна); Семенов Андрій, д-р техн. наук, проф. (Україна); Теленик Сергій, д-р техн. наук, проф. (Україна); Туркін Ігор, д-р техн. наук, проф. (Україна); Циганок Віталій, д-р техн. наук, ст. наук. співроб. (Україна)
Адреса редколегії	факультет інформаційних технологій вул. Богдана Гаврилишина, 24, м. Київ, 04116 ☎ (38044) 481 45 26 e-mail: ait.knu.fit@gmail.com
Затверджено	вченою радою факультету інформаційних технологій 22.09.25 (протокол № 3)
Зареєстровано	Національною радою України з питань телебачення та радіомовлення Рішення № 357 від 15.02.24 Ідентифікатор медіа R30-02758
Засновник і видавець	Київський національний університет імені Тараса Шевченка Видавничо-поліграфічний центр "Київський університет" Свідоцтво про внесення до Державного реєстру ДК № 1103 від 31.10.02
Адреса видавця	ВПЦ "Київський університет" 6-р Тараса Шевченка, 14, м. Київ, 01601 ☎ (38044) 239 32 22, 239 31 58, 239 31 28 e-mail: vrsc@knu.ua

ADVANCED INFORMATION TECHNOLOGY

№ 1(4)/2025

SCIENTIFIC JOURNAL

ISSN 2788-6603

TARAS SHEVCHENKO NATIONAL UNIVERSITY OF KYIV

UDC 004

DOI: <https://doi.org/10.17721/AIT.2025.1>

Theoretical and practical results of research in software engineering, applied information systems and technologies, network and Internet technologies, knowledge engineering, artificial and computational intelligence, mathematical foundations of information technology, information analytics and data analytics, machine learning and pattern recognition, digital management technologies of IT projects, smart systems and the Internet of Things are presented.

For scientists, specialists, students.

EDITOR-IN-CHIEF	Vitaliy Snytyuk, DSc (Engin.), Prof. (Ukraine)
DEPUTY OF EDITOR-IN-CHIEF	Oleksandr Pliushch, DSc (Engin.), Assoc. Prof. (Ukraine)
EXECUTIVE SECRETARIES	Tmienova Nataliia, PhD, Assoc. Prof. (Ukraine) Olena Fedusenko, PhD, Assoc. Prof. (Ukraine)
EDITORIAL BOARD	Ahmed-Zaki Darkhan, DSc (Engin.) (Kazakhstan); Al-Ammouri Ali, DSc (Engin.), Prof. (Ukraine); Barabash Oleg, DSc (Engin.), Prof. (Ukraine); Bisikalo Oleg, DSc (Engin.), Prof. (Ukraine); Bychkov Olexiy, DSc (Engin.), Prof. (Ukraine); Havrylko Yevhen, DSc (Engin.), Prof. (Ukraine); Hnatiienko Hrygorii, PhD (Engin.) (Ukraine); Korablev Mykola, DSc (Engin.), Prof. (Ukraine); Kornaga Yaroslav, DSc (Engin.), Assoc. Prof., Prof. (Ukraine); Korshun Nataliia, DSc (Engin.), Prof. (Ukraine); Kravchenko Yuriy, DSc (Engin.), Prof. (Ukraine); Malakhov Eugene, DSc (Engin.), Prof. (Ukraine); Marchenko Oleksandr, DSc (Engin.), Prof. (Ukraine); Mashkov Viktor, DSc (Engin.), Assoc. Prof., Prof. (Czech Republic); Morozov Victor, PhD (Engin.), Prof. (Ukraine); Mukhin Vadym, DSc (Engin.), Prof. (Ukraine); Pleskach Valentyna, DSc (Econ.), Prof. (Ukraine); Roth Hubert, Prof. (Germany); Saiko Volodymyr, DSc (Engin.), Prof., (Ukraine); Samokhvalov Yuriy, DSc (Engin.), Prof. (Ukraine); Semenov Andriy, DSc (Engin.), Prof. (Ukraine); Telenyk Sergii, DSc (Engin.), Prof. (Ukraine); Tsyganok Vitaliy, DSc (Engin.), Senior Researcher (Ukraine); Turkin Igor, DSc (Engin.), Prof. (Ukraine); Vovk Vladimir, Prof. (Great Britain); Zhurakovskiy Bohdan, DSc (Engin.), Prof. (Ukraine)
Address	Faculty of Information Technology 24, Bohdan Hawrylyshyn str., Kyiv, 04116 ☎ (38044) 481 45 26 e-mail: ait.knu.fit@gmail.com
Approved by	the Academic Council of the Faculty of Information Technology 22.09.25 (protocol № 3)
Registered by	the National Council of Television and Radio Broadcasting of Ukraine Decision № 357 of 15.02.24 Media identifier R30-02758
Founder and publisher	Taras Shevchenko National University of Kyiv Publishing and Polygraphic Center "Kyiv University" Certificate of entry into the State Register ДК № 1103 dated 31.10.02
Address	PPC "Kyiv University" 14, Taras Shevchenka Blvd., Kyiv, 01601 ☎ (38044) 239 32 22, 239 31 58, 239 31 28 e-mail: vpc@knu.ua

Прикладні інформаційні системи та технології

ТКАЧЕНКО Ольга, ЛЕМЕШКО Андрій

Оптимізація обсягу метаданих у сучасних інформаційних системах:
методи, інструменти й алгоритмічні підходи6

ПАВЛЮК Олена

Модель оцінювання робастності та живучості
мобільних роботизованих платформ з урахуванням надійності.....12

ЮРЧУК Ірина, СЕМЕНЧЕНКО Тарас

Метод узгодженого за рухом часового ф'южна для виявлення та відстеження БПЛА22

Мережеві й інтернет-технології

ДУДНІК Андрій, ФЕСЕНКО Андрій, ТОРОШАНКО Олександр, МИКОЛАЙЧУК Віра, БАТРАК Олександр, ВИГОВСЬКИЙ Сергій

Система для пошуку вибухонебезпечних предметів у складі мобільного гусеничного робота
з керуванням із квадрокоптера за допомогою машинного зору.....30

НЕСТЕРЕНКО Микола, МАРІНОВ Антон

Управління SDN-мережею дата-центрів на базі модифікованого підходу DevoFlow38

Огляди й обговорення

ФУРКАЛО Даниїл, ЗУБИК Людмила

Ефективні методи вивчення патернів проектування:
аналіз сучасних тенденцій і освітніх практик.....48

ШЕВЧЕНКО Максим

Аналітичний огляд контентної та колаборативної фільтрації у рекомендаційних системах54

ІНФОРМАЦІЯ ПРО АВТОРІВ67

CONTENTS

Applied information systems and technology

TKACHENKO OIha, LEMESHKO Andriy Optimization of metadata volume in modern information systems: methods, tools and algorithmic approaches.....	6
PAVLIUK Olena Model for assessing the robustness and survival of mobile robotic platforms taking into account reliability	12
YURCHUK Iryna, SEMENCHENKO Taras Motion-consistent temporal fusion for UAV detection and tracking.....	22

Network and internet technologies

DUDNIK Andrii, FESENKO Andrii, TOROSHANKO Olexandr, MYKOLAICHUK Vira, BATRAK Oleksiy, VYHOVSKYY Sergiy A system for detecting explosive objects based on a mobile tracked robot controlled from a quadcopter using mashing vision	30
NESTERENKO Mykola, MARINOV Anton Management of data center SDN based on the modified DevoFlow approach	38

Reviews and discussion

FURKALO Daniil, ZUBYK Liudmyla Effective approached to teaching design patterns: analysis of current trends and educational practices	48
SHEVCHENKO Maksym An analytical review of content-based and collaborative filtering in recommender systems	54
INFORMATION ABOUT AUTHORS	67



ПРИКЛАДНІ ІНФОРМАЦІЙНІ СИСТЕМИ ТА ТЕХНОЛОГІЇ



ОПТИМІЗАЦІЯ ОБСЯГУ МЕТАДАНИХ У СУЧАСНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ: МЕТОДИ, ІНСТРУМЕНТИ Й АЛГОРИТМІЧНІ ПІДХОДИ

Вступ. Дослідження питання оптимізації обсягу метаданих спрямовано на досягнення балансу між достатністю опису й ефективністю системи. Надмірний обсяг метаданих може перевантажувати системи, а недостатній – ускладнювати доступ до даних. Водночас зростаючий обсяг даних ускладнює роботу з метаданими, оскільки їхнє створення, зберігання та оброблення вимагають значних ресурсів. Оптимізація обсягу метаданих стала важливим завданням для організацій, які прагнуть досягти ефективного управління інформацією. Виклики, пов'язані з обсягом метаданих: надмірність, недостатність, дублювання, динамічність даних, невідповідність стандартам.

Методи. Розглянуто теоретичні основи, практичні методи, інструменти (Collibra, Apache Atlas, Talend Metadata Manager, AI-алгоритми) та переваги оптимізації обсягу метаданих (зменшення витрат, покращення продуктивності, підвищення якості даних, гнучкість і масштабованість, покращення аналітики).

Результати. Запропоновано комплексну стратегію оптимізації метаданих, адаптовану для IT-середовища. Показано, що використання систематичного підходу, включаючи аналіз, стандартизацію, автоматизацію та інтеграцію новітніх технологій, дає змогу значно зменшити витрати та покращити керування даними. Представлено алгоритм оптимізації обсягу метаданих, що може бути адаптований для різних сфер застосування, таких як бази даних, системи управління контентом чи великі дані.

Висновки. Запропонований алгоритм враховує: оцінювання корисності метаданих із використанням нормалізації метрик для уніфікації шкал оцінювання та з визначенням корисності кожного елемента метаданих; відбір метаданих (фільтрація, кластеризація); стиснення метаданих; автоматичну оптимізацію з використанням моделей машинного навчання та динамічного налаштування; перевірку й адаптацію. Алгоритм може бути розширений або змінений залежно від специфіки задачі.

Ключові слова: метадані, оптимізація, управління, інформація, динамічність, адаптивність, релевантність, продуктивність, якість, алгоритм, стиснення.

Вступ

Метадані є структурованою інформацією, яка описує, пояснює, визначає або в інший спосіб полегшує пошук, використання і керування даними. У багатьох галузях, таких як інформаційні технології, бібліотечна справа, бізнес-аналітика й управління контентом, обсяг метаданих має вирішальне значення. Надмірний обсяг метаданих може перевантажувати системи, а недостатній – ускладнювати доступ до даних. Оптимізація обсягу метаданих спрямована на досягнення балансу між достатністю опису й ефективністю системи.

Метадані – ключовий компонент управління інформацією в цифрову епоху. Вони забезпечують опис, класифікацію та організацію даних, полегшуючи доступ і використання інформації. Водночас зростаючий обсяг даних ускладнює роботу з метаданими, оскільки їхнє створення, зберігання й оброблення вимагають значних ресурсів.

Оптимізація обсягу метаданих стала важливим завданням для організацій, які прагнуть досягти ефективного управління інформацією. У цій статті ми розглянемо теоретичні основи, практичні методи, інструменти та переваги оптимізації обсягу метаданих.

Аналіз попередніх досліджень показує, що проблема оптимізації метаданих розвивалася поступово та в різних напрямках.

У дослідженнях (Ghemawat et al., 2003), (Dean, & Ghemawat, 2004) увагу направлено на масштабовані системи зберігання й обробки інформації. Google File System та MapReduce створили передумови для управління величезними обсягами метаданих у розподілених обчислювальних середовищах. Сильна сторона цих робіт полягала у практичній придатності для "великих даних", тоді як слабким місцем стала складність інтеграції зі стандартними метамоделями.

Дослідження (Bizer et al., 2009) у напрямі "зв'язаних даних" запропонували нові підходи до інтероперабельності метаданих завдяки семантичним вебтехнологіям. Їхньою перевагою є універсальність і масштабованість, однак неоліком виступає складність упровадження у традиційних системах.

Новий етап розвитку пов'язаний із застосуванням методів машинного навчання та нейронних мереж. Роботи (Mikolov et al., 2013) та (Joulin et al., 2017) показали, що автоматична генерація й відбір релевантних метаданих можуть істотно зменшити їхній обсяг, зберігаючи релевантність для аналітики. Перевагою цього підходу є висока адаптивність і можливість самонавчання, а неоліком – залежність від якості навчальних даних і ресурсомісткість обчислень. Новітні публікації, такі як (Korshun et al., 2023), інтегрують штучний інтелект у процеси керування метаданими в операційних системах, а (Turovsky et al., 2024) пропонують критерії для оцінювання ефективності методів роботи з інформацією, що демонструє міждисциплінарність і прикладну спрямованість сучасних досліджень.

У сучасних інформаційних системах метадані виконують функцію посередника між даними та користувачем, сприяючи інтеграції, пошуку й підвищенню якості інформаційних процесів.

Доречно виокремити такі основні функції метаданих:

- опис даних – інформація про тип, структуру та зміст даних;
- керування даними – полегшення зберігання, пошуку й оброблення великих масивів даних;



- інтеграція даних – підтримка сумісності між різними системами та форматами;
- аналітика – підвищення якості аналізу даних за рахунок доступності релевантної інформації.

З огляду на ці функції, важливо знайти баланс між достатністю метаданих для виконання їхніх завдань і їхнім обсягом, щоб уникнути перевантаження системи.

У процесі забезпечення ефективності використання метаданих виникають виклики, пов'язані з їхнім обсягом:

▪ надмірність – занадто великий обсяг метаданих призводить до перевитрати ресурсів на зберігання, оброблення та передачу інформації;

- недостатність – нестача метаданих ускладнює пошук і управління інформацією, знижуючи ефективність систем;
- дублювання – повторювані метадані створюють плутанину та збільшують вимоги до обчислювальних ресурсів;
- динамічність даних – постійне оновлення інформації вимагає гнучких підходів до управління метаданими;
- невідповідність стандартам – відсутність єдиної структури метаданих ускладнює інтеграцію систем.

Враховуючи основні функції метаданих і виклики щодо їхнього обсягу, сформуємо ключові принципи оптимізації обсягу метаданих:

- релевантність – метадані повинні відповідати конкретним потребам користувачів і завданням систем;
- стислість – скорочення обсягу метаданих до мінімально необхідного рівня для виконання їхньої функції;
- автоматизація – використання технологій для автоматичного створення, оновлення та видалення метаданих;
- стандартизація – впровадження єдиних стандартів метаданих для підвищення сумісності між системами;
- адаптивність – гнучкість у підходах до створення метаданих для різних типів даних і завдань.

Значущість оптимізації метаданих виявляється у її впливі на різні параметри інформаційних систем, основними з яких варто виділити такі:

1) продуктивність системи – занадто великий обсяг метаданих може уповільнювати оброблення запитів, індексацію та пошук; це особливо критично для великих баз даних або хмарних сховищ, де швидкість доступу має вирішальне значення;

2) зручність використання – метадані повинні бути зрозумілими та корисними для кінцевого користувача; надмірно детальний опис може заплутати, тоді як брак інформації може зробити об'єкт важким для ідентифікації;

3) вартість зберігання – хоча зберігання даних стало дешевшим, великий обсяг метаданих у масштабах великих організацій може суттєво збільшити витрати;

4) управління якістю даних – надмірна кількість метаданих ускладнює їхню перевірку й оновлення, що може призводити до помилок і дублювання.

Методи

Досягнення балансу між достатністю опису та продуктивністю системи можливе завдяки застосуванню комплексних методів оптимізації метаданих:

1) аналіз і класифікація даних – перший крок в оптимізації, що передбачає аналіз типів даних і визначення їхніх основних характеристик, для цього необхідно: визначити категорії даних (напр., текстові файли, зображення, відео), встановити ключові атрибути для кожної категорії (назва, дата створення, автор тощо);

2) стандартизація метаданих – використання загальноприйнятих стандартів, таких як Dublin Core, ISO 19115 або METS, що забезпечує єдину структуру даних і зменшує дублювання;

3) автоматизація процесів – сучасні інструменти, такі як AI і машинне навчання, дозволяють автоматично створювати метадані на основі аналізу вмісту файлів, наприклад: розпізнавання об'єктів на зображеннях, аналіз текстів для автоматичного визначення ключових слів;

4) видалення дублікатних даних – інструменти для пошуку та видалення дублікатних метаданих, такі як Talend або Apache Atlas, дозволяють скоротити обсяг інформації, що зберігається;

5) контроль рівня деталізації – диференційований підхід до опису даних, оскільки не всі об'єкти потребують однакового рівня деталізації, наприклад: документи для архівування можуть мати мінімальний набір метаданих, активно використовувані дані потребують детального опису;

6) видалення застарілих метаданих – регулярна перевірка й очищення систем дозволяють видаляти нерелевантну або застарілу інформацію.

Практичну реалізацію оптимізації метаданих забезпечують спеціалізованими інструментами, які автоматизують процеси аналізу, стандартизації та керування даними:

- Collibra – інтегрована платформа для керування метаданими;
- Apache Atlas – система для каталогізації, класифікації та стандартизації метаданих;
- Talend Metadata Manager – інструмент для аналізу й оптимізації метаданих у великих системах;
- AI-алгоритми – інструменти машинного навчання, які автоматизують процес створення й управління метаданими.

Впровадження методів оптимізації метаданих забезпечує значні переваги, пов'язані з продуктивністю, якістю й економічністю управління даними:

- зменшення витрат – скорочення витрат на зберігання й оброблення інформації;
- покращення продуктивності – швидкий доступ до даних і зменшення часу на виконання запитів;
- підвищення якості даних – зменшення кількості помилок і дублювань;
- гнучкість і масштабованість – легкість адаптації систем до нових вимог;
- покращена аналітика – забезпечення доступу до релевантної інформації для аналітичних завдань.

Практичні кейси застосування оптимізації метаданих підтверджують її значущість для корпоративних баз даних, хмарних сервісів та аналітичних платформ:

1) корпоративні бази даних – великі компанії, такі як Google чи Amazon, використовують оптимізацію метаданих для забезпечення швидкого пошуку й ефективного керування інформацією;

2) бібліотеки й архіви – національні бібліотеки впроваджують стандарти Dublin Core для каталогізації книг і цифрових ресурсів;



3) хмарні системи – Dropbox і Microsoft OneDrive автоматично створюють метадані на основі вмісту файлів, зменшуючи обсяг описів без втрати корисності;

4) медіаіндустрія – Netflix оптимізує метадані, щоб покращити рекомендації для користувачів, застосовуючи машинне навчання.

Майбутнє оптимізації метаданих відкриває нові горизонти для підвищення ефективності оброблення, зберігання й аналітики даних, створюючи умови для інтелектуальних систем нового покоління:

- інтеграція штучного інтелекту – застосування AI для автоматичного управління метаданими;
- гіперавтоматизація – поєднання кількох технологій (AI, RPA, NLP) для створення комплексних рішень;
- семантичні метадані – використання онтологій і семантичного вебу для покращення інтерпретації даних;
- енергоефективність – розроблення рішень, які знижують споживання енергії під час оброблення метаданих.

Метадані відіграють важливу роль у функціонуванні інформаційних систем, сприяючи ефективному управлінню даними, покращенню їхньої доступності та підтримці інтеграції між різними платформами. Проте обсяги даних, що постійно зростають, вимагають розроблення стратегій для оптимізації метаданих, які забезпечують баланс між їхньою повнотою, корисністю та продуктивністю систем.

В роботі запропоновано комплексну стратегію оптимізації метаданих, адаптовану для IT-середовища:

1) оцінювання поточного стану метаданих – аналіз існуючих метаданих, мета – зрозуміти поточний стан управління метаданими, ідентифікувати надмірні, дубльовані або нерелевантні метадані, дії – провести аудит метаданих у базах даних, сховищах, системах управління контентом, виявити джерела дублювання та надлишкової деталізації, визначити ключові атрибути метаданих, що найчастіше використовуються, інструменти для аналізу – Apache Atlas, Collibra, Talend Metadata Manager;

2) визначення цілей оптимізації – пріоритети оптимізації (зниження витрат на зберігання й оброблення даних), підвищення продуктивності системи, полегшення пошуку й аналізу даних, підтримка сумісності між різними системами, оцінювання успіху (встановити KPI (ключові показники ефективності), наприклад: час виконання пошукових запитів), скорочення обсягу збережених метаданих, рівень задоволеності користувачів;

3) стандартизація метаданих – вибір стандартів (використання загальноновизнаних стандартів забезпечує узгодженість і спрощує інтеграцію даних), для загальних метаданих – Dublin Core, для просторових даних – ISO 19115, для бібліотек і архівів – MARC (Machine-Readable Cataloging), реалізація – розробити шаблони для метаданих, що відповідають стандартам, переконатися, що всі нові метадані відповідають обраній моделі;

4) автоматизація процесів – збір і генерація метаданих (автоматизація створення й оновлення метаданих, щоб мінімізувати ручну роботу), інструменти – AI для аналізу текстів, зображень і відео, системи ETL (Extract, Transform, Load) для автоматичного імпорту даних, моніторинг та управління (використання інструментів, що постійно перевіряють метадані на дублювання або нерелевантність), приклад – у системах CRM автоматизована генерація метаданих (напр., дата останньої активності клієнта) забезпечує актуальність інформації;

5) контроль рівня деталізації – принципи деталізації (мінімум для архівів – метадані, що забезпечують базову ідентифікацію (назва, дата, автор), розширена деталізація для активних даних – метадані з додатковими атрибутами, зокрема й теги, категорії, технічні характеристики), ієрархія метаданих – створення структури, що дозволяє динамічно додавати або прибирати метадані залежно від контексту використання;

6) видалення дубльованих і застарілих метаданих – очищення системи (регулярно проводьте аналіз метаданих для видалення дублікатної або нерелевантної інформації), автоматизовані інструменти (Talend Data Quality), ручний аудит (залучайте експертів для аналізу важливих наборів даних), приклад – в архівах електронної пошти дублювання тем та авторів листів можна усунути за допомогою інструментів очищення даних;

7) використання семантичних метаданих – це метадані, що не лише описують дані, але й додають контекстуальний сенс за допомогою онтологій або зв'язків між об'єктами, переваги – покращення пошуку за змістом, підвищення точності аналітики, інструменти – RDF (Resource Description Framework), OWL (Web Ontology Language), приклад – в системах електронної комерції семантичні метадані можуть допомогти встановлювати зв'язки між продуктами на основі категорій, рекомендацій або історії покупок;

8) інтеграція із хмарними технологіями – хмарне управління метаданими (хмарні платформи пропонують інструменти для автоматизації та оптимізації метаданих – AWS Glue (Amazon), Google Cloud Data Catalog, Azure Data Factory), переваги – глобальна доступність, масштабованість, мінімізація витрат на локальне зберігання, приклад – організація використовує Google Cloud Data Catalog для автоматизації збору метаданих у розподілених базах даних;

9) забезпечення безпеки метаданих – метадані можуть містити конфіденційну інформацію, наприклад, про користувачів або конфігурації системи, методи захисту – шифрування чутливих метаданих, обмеження доступу до метаданих за допомогою ролей і прав доступу, виявлення аномалій у доступі до метаданих, приклад – в системах управління даними фінансових установ доступ до метаданих обмежений лише для авторизованих співробітників;

10) навчання персоналу – розуміння важливості (персонал має розуміти важливість оптимізації метаданих і володіти базовими навичками управління ними), програми навчання – навчання роботи зі стандартами метаданих, курси з використання інструментів автоматизації, розвиток навичок роботи із семантичними технологіями, приклад – команда IT-відділу проходить тренінг із використання Apache Atlas для управління метаданими.

Представлений алгоритм оптимізації обсягу метаданих може бути адаптований для різних сфер застосування, таких як бази даних, системи управління контентом чи великі дані. На рис. 1 представлено блок-схему алгоритму оптимізації обсягу метаданих.

1. Ініціалізація:

- 1) вхідні дані:
 - колекція метаданих

$$M = \{m_1, m_2, \dots, m_n\},$$

де m_i – окремий елемент метаданих;



- метрики якості метаданих Q, наприклад, релевантність, унікальність, необхідність;
 - обмеження ресурсів – пам'ять, час оброблення, обсяг даних;
- 2) цільова функція:
- зменшення обсягу метаданих |M|, із забезпеченням заданого рівня якості Qmin.
2. Оцінювання корисності метаданих:
- 1) для кожного елемента метаданих mi:
- обчислити метрики Q(mi), включаючи:
 - релевантність – наскільки елемент метаданих відповідає цільовим завданням;
 - частоту використання – як часто елемент запитується чи використовується;
 - час створення / оновлення – чи є дані актуальними;
 - присвоїти ваговий коефіцієнт wi для кожного Q(mi);
- 2) використати нормалізацію метрик для уніфікації шкал оцінювання;
- 3) визначити корисність кожного елемента метаданих:

$$U(m_i) = \sum_{j=1}^k w_j \cdot Q_j(m_i),$$

де w_j – ваговий коефіцієнт, а $Q_j(m_i)$ – оцінка за j -ю метрикою.



Рис. 1. Блок-схема алгоритму оптимізації обсягу метаданих

3. Відбір метаданих:

1) фільтрація:

- видалити дублікати або надлишкові метадані, які мають низькі показники корисності:

$$U(m_i) < \epsilon,$$

де ϵ – порогове значення.



- 2) кластеризація:
 - згрупувати схожі метадані за ключовими ознаками (напр., тематикою чи форматом);
 - залишити репрезентативний елемент для кожного кластера.
4. Стиснення метаданих:
 - 1) алгоритми стиснення:
 - використати методи компресії, такі як:
 - Huffman coding для текстових даних;
 - Run-length encoding (RLE) для повторюваних даних;
 - бінарне кодування для структурованих метаданих;
 - 2) видалення зайвих полів:
 - виключити необов'язкові атрибути, які зрідка використовуються.
5. Автоматична оптимізація:
 - 1) моделі машинного навчання:
 - використати класифікаційні алгоритми для передбачення корисності метаданих на основі їхніх характеристик;
 - застосувати нейронні мережі або дерева рішень для автоматичного відбору релевантних метаданих;
 - 2) динамічне налаштування:
 - оптимізувати обсяг метаданих залежно від запитів користувачів і доступних ресурсів.
6. Перевірка й адаптація:
 - 1) оцінити якість скорочених метаданих за метриками Q;
 - 2) якщо $Q < Q_{\min}$, повернутись до етапу фільтрації або коригування параметрів;
 - 3) забезпечити оновлення метаданих у реальному часі для підтримки актуальності.
7. Вихід – оптимізований набір метаданих M_{opt} , який задовольняє обмеження на обсяг і зберігає необхідний рівень якості. Цей алгоритм може бути розширений або змінений залежно від специфіки задачі (напр., для конкретних доменів чи типів метаданих).

Дискусія і висновки

Оптимізація обсягу метаданих є ключовим фактором успішного управління інформацією в сучасних системах. Баланс між обсягом, якістю та релевантністю метаданих дозволяє досягти ефективності, зменшити витрати та забезпечити зручність для користувачів. У майбутньому розвиток технологій, таких як штучний інтелект і семантичний веб, відкриє нові можливості для вдосконалення процесів оптимізації.

Оптимізація метаданих в ІТ є критичним процесом, який забезпечує ефективність, продуктивність та безпеку інформаційних систем. Використання систематичного підходу, включаючи аналіз, стандартизацію, автоматизацію та інтеграцію новітніх технологій, дозволяє значно зменшити витрати та покращити керування даними.

Впровадження цієї стратегії сприятиме зростанню бізнес-ефективності та конкурентоспроможності організацій у цифрову епоху.

Внесок авторів. Ольга Ткаченко – концептуалізація; Андрій Лемешко – аналіз джерел, підготовка огляду літератури.

Джерела фінансування. Це дослідження не отримало жодного гранта від фінансової установи в державному, комерційному або некомерційному секторах.

Список використаних джерел

- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data – The story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22. <https://doi.org/10.4018/jswis.2009081901>
- Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107–113. <https://doi.org/10.1145/1327452.1327492>
- Ghemawat, S., Gobioff, H., & Leung, S.-T. (2003). The Google file system. In *Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOSP '03)* (pp. 29–43). Association for Computing Machinery. <https://doi.org/10.1145/945445.945450>
- Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2017). Bag of tricks for efficient text classification. In M. Lapata, P. Blunsom, & A. Koller (Eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Vol. 2, Short Papers* (pp. 427–431). Association for Computational Linguistics. <https://doi.org/10.18653/v1/E17-2068>
- Korshun, N., Myshko, I., & Tkachenko, O. (2023). Automation and management in operating systems: The role of artificial intelligence and machine learning. In *Proceedings of the 20th International Scientific Conference Dynamical System Modeling and Stability Investigation (DSMSI 2023: Mathematical Foundations of Information Technologies)* (pp. 59–68). Igor Sikorsky Kyiv Polytechnic Institute; published online in CEUR Workshop Proceedings. <https://ceur-ws.org/Vol-3687/>
- Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. <https://doi.org/10.48550/arXiv.1301.3781>
- Turovsky, O., Tkachenko, O., Ghno, G. S. N., & Abed, A. M. (2024). Selection and substantiation of the system of criteria for evaluating the effectiveness of steganographic methods of hiding information in the image. In *Proceedings of the 35th Conference of Open Innovations Association (FRUCT)* (pp. 755–763). Open Innovations Association (FRUCT). <https://doi.org/10.23919/FRUCT61870.2024.10516392>

References

- Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data – The story so far. *International Journal on Semantic Web and Information Systems*, 5(3), 1–22. <https://doi.org/10.4018/jswis.2009081901>
- Dean, J., & Ghemawat, S. (2004). MapReduce: Simplified data processing on large clusters. *Communications of the ACM*, 51(1), 107–113. <https://doi.org/10.1145/1327452.1327492>
- Ghemawat, S., Gobioff, H., & Leung, S.-T. (2003). The Google file system. In *Proceedings of the 19th ACM Symposium on Operating Systems Principles (SOSP '03)* (pp. 29–43). Association for Computing Machinery. <https://doi.org/10.1145/945445.945450>
- Joulin, A., Grave, E., Bojanowski, P., & Mikolov, T. (2017). Bag of tricks for efficient text classification. In M. Lapata, P. Blunsom, & A. Koller (Eds.), *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Vol. 2, Short Papers* (pp. 427–431). Association for Computational Linguistics. <https://doi.org/10.18653/v1/E17-2068>
- Korshun, N., Myshko, I., & Tkachenko, O. (2023). Automation and management in operating systems: The role of artificial intelligence and machine learning. In *Proceedings of the 20th International Scientific Conference Dynamical System Modeling and Stability Investigation (DSMSI 2023: Mathematical Foundations of Information Technologies)* (pp. 59–68). Igor Sikorsky Kyiv Polytechnic Institute; published online in CEUR Workshop Proceedings. <https://ceur-ws.org/Vol-3687/>



Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). *Efficient estimation of word representations in vector space*. <https://doi.org/10.48550/arXiv.1301.3781>
Turovsky, O., Tkachenko, O., Ghno, G. S. N., & Abed, A. M. (2024). Selection and substantiation of the system of criteria for evaluating the effectiveness of steganographic methods of hiding information in the image. In *Proceedings of the 35th Conference of Open Innovations Association (FRUCT)* (pp. 755–763). Open Innovations Association (FRUCT). <https://doi.org/10.23919/FRUCT61870.2024.10516392>

Отримано редакцією журналу / Received: 25.06.25
Прорецензовано / Revised: 20.09.25
Схвалено до друку / Accepted: 20.09.25

Olga TKACHENKO, DSc (Engin.), Prof.
ORCID ID: 0000-0001-7983-9033
e-mail: olga.tkachenko@knu.ua
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Andriy LEMESHKO, PhD, Assoc. Prof.
ORCID ID: 0000-0001-8003-3168
e-mail: andriy.lemeshko@knu.ua
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

OPTIMIZATION OF METADATA VOLUME IN MODERN INFORMATION SYSTEMS: METHODS, TOOLS AND ALGORITHMIC APPROACHES

Background. *The study of metadata volume optimization aims to achieve a balance between the sufficiency of description and system efficiency. Excessive metadata can overload systems, and insufficient metadata can complicate data access. At the same time, the growing volume of data complicates work with metadata, since its creation, storage and processing require significant resources. Metadata volume optimization has become an important task for organizations seeking to achieve effective information management. Challenges associated with metadata volume: redundancy, insufficiency, duplication, data dynamics, non-compliance with standards.*

Methods. *The paper considers the theoretical foundations, practical methods, tools (Collibra, Apache Atlas, Talend Metadata Manager, AI algorithms) and the benefits of metadata volume optimization (cost reduction, improved productivity, improved data quality, flexibility and scalability, improved analytics).*

Results. *The paper proposes a comprehensive metadata optimization strategy adapted for the IT environment. It is shown that the use of a systematic approach, including analysis, standardization, automation and integration of the latest technologies, allows to significantly reduce costs and improve data management. An algorithm for optimizing the volume of metadata is presented, which can be adapted for various application areas, such as databases, content management systems or big data. The proposed algorithm takes into account the assessment of metadata usefulness using metric normalization to unify evaluation scales and determine the usefulness of each metadata element; metadata selection (filtering, clustering); metadata compression; automatic optimization using machine learning models and dynamic tuning; verification and adaptation. The algorithm can be expanded or modified depending on the specifics of the task.*

Conclusions. *The proposed algorithm takes into account: metadata usefulness assessment using metric normalization to unify evaluation scales and determine the usefulness of each metadata element; metadata selection (filtering, clustering); metadata compression; automatic optimization using machine learning models and dynamic tuning; testing and adaptation. The algorithm can be expanded or modified depending on the specifics of the task.*

Keywords: *metadata, optimization, management, information, dynamism, adaptability, relevance, performance, quality, algorithm, compression*

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.



МОДЕЛЬ ОЦІНЮВАННЯ РОБАСТНОСТІ ТА ЖИВУЧОСТІ МОБІЛЬНИХ РОБОТИЗОВАНИХ ПЛАТФОРМ З УРАХУВАННЯМ НАДІЙНОСТІ

Вступ. Розроблено комплексну модель надійності, робастності та живучості мобільної роботизованої платформи (МРП) із послідовно-паралельною архітектурою. Метою роботи є проведення математичного моделювання й аналізу надійності елементів МРП на основі законів розподілу Вейбулла й експоненціального закону, а також оцінювання впливу резервування та структурної надмірності на загальну стійкість системи до відмов. Встановлено, що критичними елементами МРП є: шасі, двигун, контролер, оскільки вони послідовно з'єднані та мають низький рівень відмовостійкості без резервування, на відміну від паралельно з'єднаних сенсорних і комунікаційних компонентів.

Методи. До методів належать такі: методи розрахунку надійності з використанням послідовно-паралельних схем; оцінювання часу безвідмовної роботи (МТТФ); моделювання з використанням розподілу Вейбулла й експоненціального розподілу; чутливості для аналізу впливу елементів на надійність системи; оцінювання робастності та живучості системи в умовах змін середовища або несправностей.

Результати. На основі побудованих моделей отримано числові значення МТТФ для всіх елементів МРП. Найвищі показники безвідмовної роботи зафіксовано для шасі (135 412 год) та акумуляторної батареї (124 914 год), тоді як найменшу надійність демонструють комунікаційні модулі Wi-Fi (80 000 год) та BLE (70 000 год). Значення МТТФ усієї МРП становить 21 188 год, близько 2,4 року безперервної експлуатації.

Побудовані графіки чутливості показали, що найбільший вплив на загальну надійність мають шасі, контролер і двигун. Вони є пріоритетними для резервування або створення деградованих режимів. Робастність системи за однієї відмови ($\Delta \leq 1$) оцінено через зниження ймовірності безвідмовної роботи. Проте із часом (понад 100 000 год) надійність знижується нижче 5%, що підкреслює важливість обліку зносу компонентів. Для підвищення живучості МРП має працювати за відмови до двох платформ, планувати маршрути за надійністю, підтримувати деградований режим і динамічно перерозподіляти завдання.

Висновки. Запропонована модель дає змогу з високою точністю оцінювати надійність МРП з урахуванням її структури, типу з'єднань і функціональних особливостей елементів. Результати вказують на необхідність резервування критичних компонентів і підвищення робастності системи для забезпечення стійкості до відмов. Представлену методику можна використовувати для планування сервісного обслуговування, прогнозування життєвого циклу й оптимізації роботи МРП у динамічному середовищі.

Ключові слова: мобільна роботизована платформа (МРП), надійність, робастність, послідовно-паралельна структура, чутливість компонентів, оптимізація маршрутів.

Вступ

Все частіше сучасні промислові підприємства використовують мобільні роботизовані платформи (МРП). Такі платформи часто ефективно застосовують у логістиці, автоматизації виробничих процесів для забезпеченні гнучкості виробництва. Проте в таких умовах МРП піддаються підвищеним ризикам щодо відмови компонентів через різноманітні причини, наприклад пил, вібрації, електромагнітні завади тощо. Це негативно впливає на їхню надійність і безперервність роботи всієї системи МРП у внутрішньологістичних системах промислових підприємств.

Традиційні підходи до оцінювання надійності часто не враховують взаємозв'язків між підсистемами, резервування й адаптивні стратегії розгортання роботизованих платформ. У зв'язку із цим постає необхідність у розробленні комплексної моделі оцінювання робастності системи МРП, яка дозволить:

- оцінювати вплив надійності окремих компонентів (датчиків, зв'язку, шасі, обчислювачів) на загальну працездатність;
- забезпечити прийняття інженерних рішень щодо вибору конфігурацій і рівня резервування для досягнення заданої готовності МРП.

Постановка задачі – розробити математичну модель оцінювання робастності та живучості МРП з урахуванням надійності компонентів та умов готовності в промисловому середовищі.

Основні завдання:

1. Проаналізувати архітектуру МРП та її критичні підсистеми (сенсори, виконавчі механізми, комунікації, обчислювальні модулі) з погляду відмовостійкості.
2. Побудувати модель надійності окремої платформи, використовуючи розподіли Вейбулла й експоненціальні моделі для компонентів.
3. Сформулювати математичну модель системи з умовою готовності типу "m з n", що описує робастність і живучість у вигляді ймовірності працездатності за наявності не менше ніж m справних платформ.
4. Розробити програмний засіб для обчислення та візуалізації функцій надійності, інтенсивності відмов, очікуваного часу безвідмовної роботи, а також дослідження впливу параметрів компонентів на загальну робастність і живучість.
5. Провести чисельне моделювання й аналіз сценаріїв роботи МРП у промисловому середовищі з урахуванням характерних значень МТТФ, МТBF, $\lambda(t)$ та інших показників.
6. Сформулювати рекомендації щодо архітектурних рішень, що підвищують робастність і живучість МРП у реальних умовах експлуатації.



Огляд літературних джерел. Публікацію (Luo, 2023) присвячено забезпеченню безпечної та робастної автономії в системах із великою кількістю роботів, що діють у складних і змінних середовищах. Автори пропонують підходи до стійкої взаємодії між гетерогенними роботами та людьми з урахуванням можливих збоїв у зв'язку та відмовами роботів. Наукова новизна полягає у поєднанні безпеки та робастності для досягнення коректної поведінки роботів ще на етапі проектування. В роботі Stetter (2022) розглянуто підхід до підвищення відмовостійкості технічних систем шляхом цілеспрямованого проектування, зокрема й на прикладі автоматизованого транспортного засобу (AGV). Автори описують методичну основу, алгоритми й інструменти, що дозволяють інтегрувати діагностичність, керованість і структурну надмірність уже на етапі розроблення продукту. Наукова новизна полягає у створенні системного підходу до fault-tolerant design, який охоплює як концептуальні, так і геометричні аспекти технічних рішень, сприяючи підвищенню прозорості й ефективності інженерного проектування.

Автори (Fazlollahatabar et al., 2017) моделюють надійність складних систем автономних роботів у виробничих умовах шляхом розроблення методів оцінювання й оптимізації надійності. Їх розробили за допомогою блок-схем надійності, дерев відмов і стохастичних процесів, приділяючи особливу увагу критичним компонентам роботизованих систем. Наукова новизна роботи полягає в інтеграції дерева рішень за критерієм ризику з вартісним аналізом, а також у застосуванні комбінованих підходів для структурно-логічного аналізу складних роботизованих систем. В роботі (Bensaid Amrani et al., 2023) розглянуто прогнозування надійності мехатронних систем на етапі проектування, що є ключовим для зменшення витрат і кількості прототипів. Автори аналізують існуючі кількісні та якісні підходи до моделювання надійності з урахуванням особливостей мехатроніки як динамічної, інтерактивної, а також конфігурувальні системи. Наукова новизна полягає в критичному огляді сучасних методологій прогнозування надійності та виявленні їхніх слабких місць для подальшого розвитку нових підходів і інструментів.

В роботі Betzer et al. (2024) представлено аналіз надійної роботи автономного мобільного робота в умовах невизначеності за допомогою цифрового двійника, який здійснює верифікацію поведінки робота в режимі реального часу. Автори використовують протокол MQTT і систему моніторингу TeSSLa для виявлення відхилень у роботі, пов'язаних із шумами датчиків і змінами навколишнього середовища. Наукова новизна полягає у впровадженні хмарного цифрового двійника як "наглядового" механізму, здатного не лише аналізувати, а і втручатися в управління роботом у разі загрози порушення вимог безпеки чи продуктивності.

У публікації Panchal (2023) запропоновано двофазну методику оцінювання надійності турбінного агрегату цукрового заводу на основі інтуїтивно-нечіткої логіки (Intuitionistic Fuzzy) та підходу Lambda-Tau. Автори застосували вирази серійно-паралельного з'єднання для розрахунку показників надійності з урахуванням функцій належності та неналежності, що дозволяє враховувати розмитість і невизначеність вхідних даних. Наукова новизна полягає у використанні інтуїтивно-нечіткої логіки для точнішого моделювання впливу зміни розкиду значень на доступність системи та планування її технічного обслуговування.

Грунтовний огляд сучасного стану систем керування з відмовостійкістю (FTCS), зокрема і їх типів, методів реалізації та сфер застосування представлено у публікації (Amin, & Hasan, 2019). Автори аналізують активні, пасивні та гібридні підходи до Fault Tolerant Control, а також розглядають типові відмови сенсорів і виконавчих механізмів, питання стабільності та надійності таких систем. Наукова новизна полягає в систематизації новітніх розробок і виявленні актуальних наукових прогалин, що слугує орієнтиром для подальших досліджень у сфері FTCS.

У публікації Maza (2025) розглядається нова проблема керування двосторонніми AGV у разі збоїв сенсорів, зокрема в умовах часткової втрати спостережуваності подій. Автори пропонують подієво-орієнтовану архітектуру керування на основі теорії супервізорного керування (SCT), доповнену діагностичним шаром для підвищення відмовостійкості системи. Наукова новизна полягає у синтезі керуючих автоматів, які враховують інформацію з діагностичних автоматів, що дозволяє уникати тупикових ситуацій навіть за умов сенсорних порушень, а також у використанні стохастичних таймованих автоматів для оцінювання ймовірності безвідмовної роботи. У дослідженні (Purwaningsih et al., 2025) проаналізовано перехід від напівручного транспортування матеріалів до повністю автоматизованої системи на основі AGV у виробничих умовах країн, що розвиваються. За допомогою моделювання дискретних подій і теорії черг оцінено ефективність, витрати та якість продукції для трьох сценаріїв: ручне керування, 2 та 3 AGV. Наукова новизна полягає у кількісному обґрунтуванні переваг впровадження AGV на детермінованих виробничих лініях, що дозволяє підвищити надійність системи й обґрунтовано приймати управлінські рішення. У публікації (Cai et al., 2024) досліджено надійність визначення відносного кута між AGV та ціллю за допомогою методу Angle-of-Arrival із використанням IoT-сенсорів від TI та u-blox відповідно до стандарту Bluetooth 5.1. Проведено порівняльне оцінювання точності сенсорів у симуляції та реальних умовах, після чого кращий варіант інтегровано в систему навігації AGV. Наукова новизна полягає в першому систематичному аналізі застосування AoA-сенсорів для зовнішньої навігації AGV, що відкриває перспективи їхнього використання в складних реальних середовищах. Автори роботи Zheng et al. (2024) розглянули проблему ненадійності Wi-Fi-зв'язку в системах AGV у промисловому середовищі та запропонували альтернативне рішення на основі протоколу WIA-FA. Дослідники розробили нову структуру суперкадрів і стратегію повторної передачі SSER, що забезпечує впорядковану та надійну передачу даних між AGV і мережею. Наукова новизна полягає у впровадженні послідовного розкладу передачі, який динамічно оптимізує фазу підйому та спуску даних, значно покращуючи надійність зв'язку у складних умовах. У статті (Li, & Schulze, 2024) запропоновано підхід до підвищення надійності та доступності AGV-систем у виробничому середовищі за допомогою прогнозного технічного обслуговування на базі нейронної мережі з довготривалою короткочасною пам'яттю (LSTM). Метод дозволяє аналізувати історичні дані, виявляти закономірності та передбачати потенційні відмови або потребу в обслуговуванні компонентів AGV та суміжного обладнання. Наукова новизна полягає у впровадженні AI-методу для проактивного керування технічним станом AGV-систем, що забезпечує мінімізацію простоїв і підвищення ефективності виробництва. У публікації (Yin et al., 2023) розроблено метод планування маршруту для автономних наземних транспортних засобів (AGV), що враховує не лише оптимальність шляху, а і його надійність в умовах складного рельєфу та ймовірних відмов. Автори інтегрували фізично обґрунтоване моделювання динаміки руху з аналізом надійності мобільності AGV, застосували методи сурогатного моделювання й алгоритм



RRT*. Наукова новизна полягає в поєднанні аналізу ймовірнісних відмов мобільності з алгоритмічним плануванням траєкторії, що дозволяє прокладати безпечні маршрути у невизначених позашляхових умовах.

Методи

Надійність МРП описує поведінку окремих її компонентів або всієї системи в нормальних умовах. Робастність враховує структурну надмірність, резервування, адаптацію та стійкість до збоїв. Тому висока робастність системи дає змогу підтримувати задану надійність навіть за виходу з ладу окремих компонентів. Тобто робастність – це архітектурна властивість, яка не гарантує відсутність відмов, але зменшує їхній вплив. Живучість відображає здатність системи зберігати функціональність і ефективність роботи навіть за наявності множинних пошкоджень або тривалого впливу несприятливих умов, забезпечуючи швидке відновлення або адаптацію до нових обставин.

Модель надійності окремої платформи будують на основі основних підсистем МРП. До них належать: шасі, двигун, контролер, які з'єднані послідовно, та камера, лідар, комунікація, які з'єднані паралельно. Якщо в послідовно з'єднаних підсистемах немає резервування, то вони мають низький рівень відмовостійкості. Паралельно з'єднані елементи мають вищу відмовостійкість за рахунок надлишковості. Тому критичними елементами є: двигун, шасі, контролер. Вони потребують структурного резервування або деградованого режиму для підвищення надійності.

За функціональним блоком компоненти МРП можна класифікувати за підсистемами і з'єднанням (табл. 1).

Таблиця 1

Класифікація компонентів МРП за підсистемами і з'єднанням

Компонент	Підсистема	Функціональність	Тип з'єднання	Критичність	Примітка
Акумуляторна батарея	Енергетична	Живлення всіх компонентів МРП	Послідовне	Висока	Без живлення система не функціонує
Шасі	Механічна / мобільність	Основа структури, несна конструкція	Послідовне	Висока	Вихід з ладу = повна зупинка МРП
Двигун	Виконавча	Приводить платформу в рух	Послідовне	Висока	Вихід з ладу двигуна = втрата рухливості
Контролер	Обчислювальна	Керування, прийняття рішень	Послідовне	Висока	Відповідає за процес управління всіма підсистемами
Камера	Сенсорна	Візуальна навігація	Паралельне	Середня	Використовується у парі з лідаром або як резерв
LIDAR	Сенсорна	Обхід перешкод, побудова карти	Паралельне	Середня	Може працювати без камери або навпаки
Wi-Fi / BLE	Комунікаційна	Передача даних, зв'язок із мережею	Паралельне	Середня	Використовуються як дублювальні канали

Розглядається модель надійності МРП середнього класу, тобто це автоматизований транспортний засіб масою близько 600 кг, що складається з елементів із розподілом Вейбулла та експоненціальним розподілом часу до відмови. Елементи, що з'єднані послідовно: акумуляторна батарея, шасі, двигун, контролер. Елементи, які з'єднані паралельно: камера, лідар і комунікація (Wi-Fi, BLE). Архітектура МРП має послідовно-паралельну структуру і зображена на рис. 1.

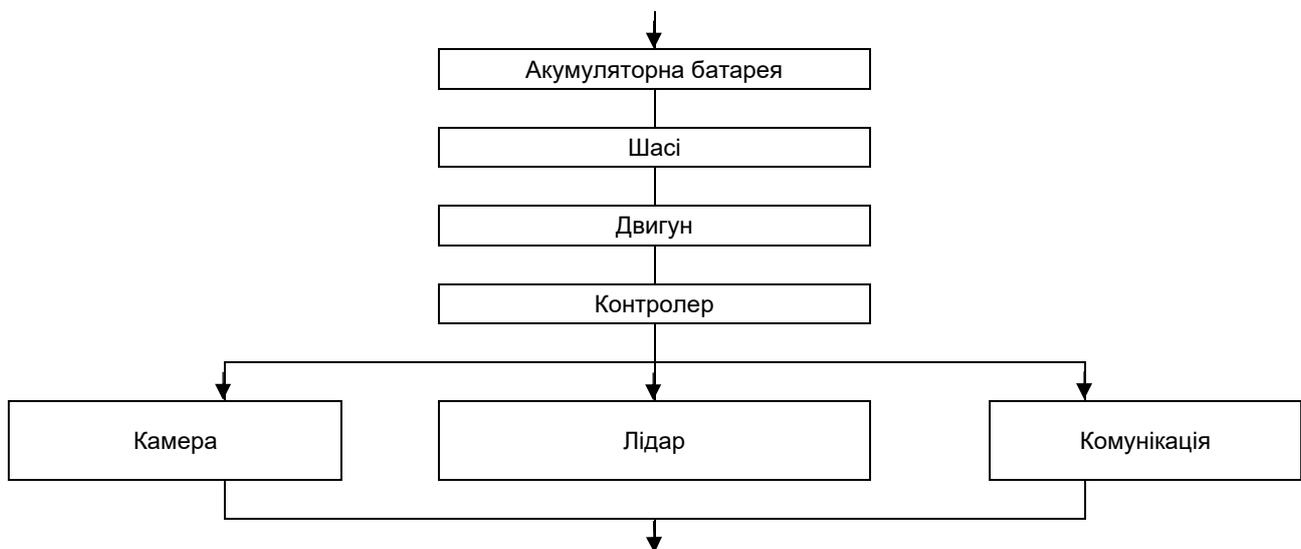


Рис. 1. Архітектура МРП



Закону розподілу Вейбулла підпорядковуються: шасі, контролер та акумуляторна батарея. Зокрема функцію надійності для шасі обчислюють за формулою

$$R_C(t) = \exp[-(\lambda_C t)^{k_C}], \quad (1)$$

де λ_C – параметр масштабу (інтенсивність відмов), обернений до характерного часу безвідмовної роботи; k_C – параметр форми (shape), який визначає характер старіння: $k_C < 1$ – фаза "дитячої смертності" (initial failures); $k_C = 1$ – експоненціальний розподіл (постійна ймовірність відмови); $k_C > 1$ – старіння (failure rate зростає із часом); t – час у годинах (або секундах, залежно від одиниць).

Для контролера функція надійності

$$R_{Ctrl}(t) = \exp[-(\lambda_{Ctrl} t)^{k_{Ctrl}}]. \quad (2)$$

Для акумуляторної батареї функцію надійності обчислюють за формулою

$$R_B(t) = \exp[-(\lambda_B t)^{k_B}]. \quad (3)$$

Експоненціальному закону розподілу підпорядковуються такі елементи МРП: двигун, камера, лідар, Wi-Fi і BLE. Формула для обчислення функції надійності двигуна:

$$R_M(t) = \exp(-\lambda_M t). \quad (4)$$

Функція надійності камери:

$$R_{Cam}(t) = \exp(-\lambda_{Cam} t). \quad (5)$$

Функція надійності лідара:

$$R_{LID}(t) = \exp(-\lambda_{LID} t). \quad (6)$$

Функція надійності Wi-Fi:

$$R_{WiFi}(t) = \exp(-\lambda_{WiFi} t). \quad (7)$$

Функція надійності BLE:

$$R_{BLE}(t) = \exp(-\lambda_{BLE} t). \quad (8)$$

Функцію надійності послідовно з'єднаних елементів обчислюємо, перемножуючи функції надійності кожного з елементів, а саме: акумуляторної батареї, шасі, контролера та двигуна:

$$R_{ser}(t) = R_B(t)R_C(t)R_{Ctrl}(t)R_M(t). \quad (9)$$

Функцію надійності паралельно з'єднаних елементів обчислюємо, перемножуючи функції надійності Wi-Fi і BLE:

$$R_{comm}(t) = 1 - (1 - R_{WiFi}(t))(1 - R_{BLE}(t)). \quad (10)$$

Функцію надійності для паралельно з'єднаних камери, лідара та комунікаційної системи знайдемо за формулою

$$R_{par}(t) = 1 - (1 - R_{Cam}(t))(1 - R_{LID}(t)). \quad (11)$$

Загальна функція надійності МРП матиме вигляд $t \in [0; 2000]$:

$$R_{AGV}(t) = R_{ser}(t)R_{par}(t). \quad (12)$$

Отже загальна функція надійності МРП виглядає так:

$$R_{AGV}(t) = R_B(t)R_C(t)R_M(t)R_{Ctrl}(t)[1 - (1 - R_{Cam}(t))(1 - R_{LID}(t))1 - (1 - R_{WiFi}(t))(1 - R_{BLE}(t))]. \quad (13)$$

Інтенсивність відмов для елементів, що підпорядковуються закону Вейбулла, обчислимо як

$$\lambda_i(t) = k_i \lambda_i^{k_i} t^{k_i-1}. \quad (14)$$

Інтенсивність відмов для елементів, що підпорядковуються експоненціальному закону, розраховуємо за формулою

$$\lambda_i(t) = \lambda_i. \quad (15)$$

Обчислюємо інтенсивність відмов для послідовно з'єднаних елементів: акумуляторної батареї, шасі, контролера та двигуна:

$$\lambda_{ser}(t) = \lambda_B(t) + \lambda_C(t) + \lambda_{Ctrl}(t) + \lambda_M(t). \quad (16)$$

Інтенсивність відмов паралельної частини МРП обчислюємо як

$$\lambda_{par}(t) = \frac{f_{par}(t)}{R_{par}(t)}. \quad (17)$$

Загальну інтенсивність відмов МРП знаходимо за формулою

$$\lambda_{AGV}(t) = \lambda_{ser}(t) + \lambda_{par}(t)\lambda_{par}(t). \quad (18)$$

Частота відмов МРП:

$$a(t) = \int_0^t \lambda_{AGV}(u) du. \quad (19)$$



Імовірність відмови МРП:

$$Q(t) = 1 - R_{AGV}(t). \tag{20}$$

Готовність, враховуючи, що МРП працює без відновлення:

$$A(t) \approx R_{AGV}(t). \tag{21}$$

Очікуваний залишковий час до відмови:

$$T(t) = \frac{\int_t^\infty R_{AGV}(u)du}{R_{AGV}(t)}. \tag{22}$$

Середній час безвідмовної роботи (MTTF) для елементів, що підпорядковуються закону Вейбулла:

$$MTTF_i = \frac{1}{\lambda_i} \Gamma(1 + \frac{1}{k_i}). \tag{23}$$

Середній час безвідмовної роботи (MTTF) для елементів, що підпорядковуються експоненціальному закону:

$$MTTF_i = \frac{1}{\lambda_i}. \tag{24}$$

Середній час безвідмовної роботи (MTTF) для паралельних компонентів МРП:

$$MTTF_{par} = \frac{1}{\frac{1}{MTTF_{Cam}} + \frac{1}{MTTF_{LID}} + \frac{1}{MTTF_{comm}}}. \tag{25}$$

Середній час безвідмовної роботи для комунікаційної системи, яка складається з Wi-Fi і BLE:

$$MTTF_{comm} = (\frac{1}{MTTF_{WiFi}} + \frac{1}{MTTF_{BLE}})^{-1}. \tag{26}$$

Загальний MTTF МРП приблизно такий:

$$MTTF_{AGV} \approx (\frac{1}{MTTF_{ser}} + \frac{1}{MTTF_{par}})^{-1}, \tag{27}$$

де загальний MTTF МРП для паралельно з'єднаних елементів розраховуємо як

$$MTTF_{ser} = MTTF_B + MTTF_C + MTTF_{Ctrl} + MTTF_M. \tag{28}$$

Робастність МРП визначає її здатність зберігати працездатність за наведених нижче умов. Загальна математична модель робастності надійності МРП:

$$R_{AGV}^\Delta(t) = \prod_{i \in V \setminus \Delta} R_i(t) [\prod_{j \in S \setminus \Delta} (1 - R_j(t))], \tag{29}$$

де $V = \{B, C, M, Ctrl\}$ – ключові вузли (батарея, корпус, двигуни, контролер), $S = \{Cam, LID, WiFi, BLE\}$ – сенсорно-комунікаційна система, $\Delta \subseteq V \cup S \cup E$ – порушені компоненти або зв'язки, $R_j(t)$ – функція надійності компонента j .

Умовна робастність враховує, що компоненти з дельта вважаються відмовленими, тобто $R_i(t) = 0, i \in \Delta$. Умову k -робастності описують так: система є k -робастною, якщо за будь-якої множини існує k відмов. Тобто, навіть якщо відмовлять будь-які k компоненти (вузли або канали), загальна надійність системи повинна залишатись вищою за мінімум.

$$R_{AGV}^\Delta(t) \geq R_{min}, \forall \Delta \subseteq V \cup E, |\Delta| \leq k. \tag{30}$$

Чутливість компонентів обчислюємо для оцінювання впливу кожного компонента на загальну надійність:

$$S_i(t) = \frac{\partial R_{AGV}(t)}{\partial R_i(t)} = \frac{R_{AGV}(t)}{R_i(t)}. \tag{31}$$

Компоненти з найбільшими значеннями є найкритичніші, і саме їх потрібно резервувати. Оптимізацію маршрутів МРП здійснюють на основі множини шляхів:

$$P = \{1, 2, \dots, n\}. \tag{32}$$

Результати

На рис. 2 зображено графіки основних характеристик надійності МРП. А саме: інтенсивність відмов, функцію напрацювання на відмову, функцію очікування часу до відмови, імовірність роботи системи, імовірність відмови, готовність МРП.

Згідно з наведеними графіками, розраховано середній час безвідмовної роботи (MTTF) для окремих компонентів автоматизованої мобільної платформи. В табл. 2 наведено середній час безвідмовної роботи кожного з елементів МРП.

Таблиця 2

Середній час безвідмовної роботи кожного елемента МРП

Елемент МРП	Батарея	Шасі	Контролер	Двигун	Камера	Лідар	Wi-Fi	BLE
MTTF (години)	124914	135412	110829	110000	120000	100000	80000	70000

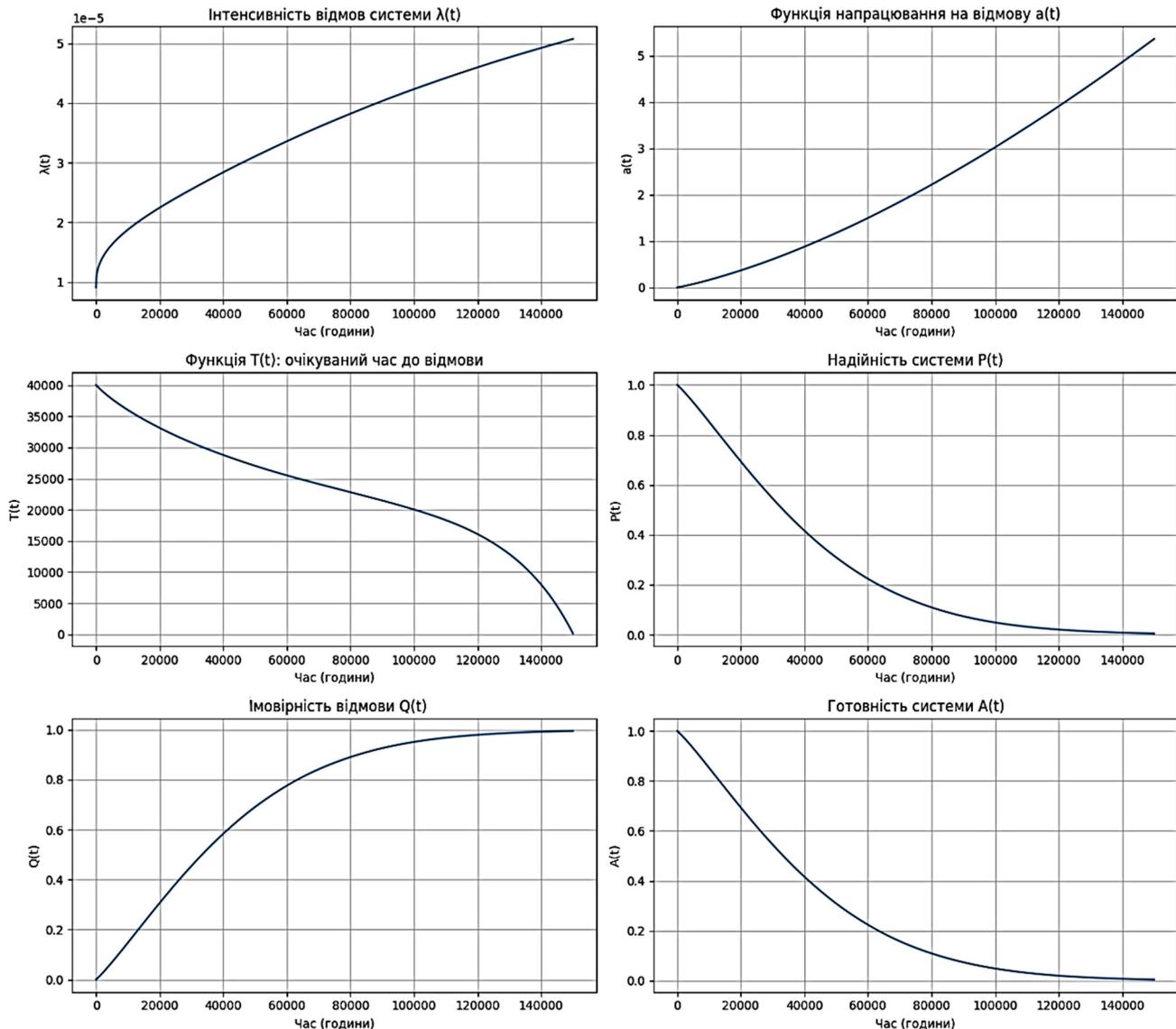


Рис. 2. Графіки основних характеристик надійності МРП

Найдовший МТТФ мають шасі (135 412 год) та акумулятор (124 914 год), що свідчить про їхню високу надійність. Контролер, двигун і камера мають трохи нижчі значення, близько 110 000–120 000 год. Найменшу надійність мають бездротові модулі зв'язку: Wi-Fi – 80 000 год, а BLE – 70 000 год. Через те, що для зв'язку використовують або Wi-Fi, або BLE (умова "або", тобто паралельна робота), МТТФ цієї групи становить лише 37 333 год.

Коли розглядається паралельна група компонентів (камера, LIDAR і зв'язок), загальний МТТФ ще нижчий – лише 22 164 години. Це означає, що в разі відмови будь-якого із цих трьох елементів система втрачає функціональність. У підсумку, орієнтовне МТТФ усієї системи AGV становить близько 21 188 год, що відповідає приблизно 2,4 року безперервної роботи. Це дає змогу оцінити часові межі надійної експлуатації системи та планувати обслуговування чи модернізацію компонентів.

На рис. 3 зображено графіки чутливості елементів МРП та робастності МРП для одночасно однієї відмови її елементів.

Система МРП початково має максимальну надійність 100 %. Через 10 000 год вона знижується до приблизно 85 %, а далі поступово падає, досягаючи близько 0,5 % після 30 000 год і менше 1 % після 140 000 год. Мінімальні значення надійності трохи нижчі за номінальні, що враховує можливі коливання параметрів. Загалом, система надійна на початковому етапі, але із часом її робастність значно знижується.

У табл. 3 представлено розраховані значення часу для номінальних значень без відмов елементів МРП ($R_{nominal}$) та з однією відмовою (R_{min} при $\Delta \leq 1$), а також чутливість компонентів до загальної надійності МРП. При $R_{min} |\Delta| \leq 1$ – це робастність, тобто здатність МРП зберігати працездатність за однієї відмови компонента (напр., втрата зв'язку Wi-Fi або BLE).

На початку роботи системи ($t = 0$ год) її надійність становить 100 %, тобто система повністю працездатна. Із часом надійність поступово знижується. Через 10 000 год вона залишається досить високою – близько 85 % за номінальним значенням і трохи менше з урахуванням варіацій (R_{min}). Далі, зі зростанням часу експлуатації, надійність падає: через 50 000 год вона становить уже близько 30 %, а через 100 000 год – менше 5 %.

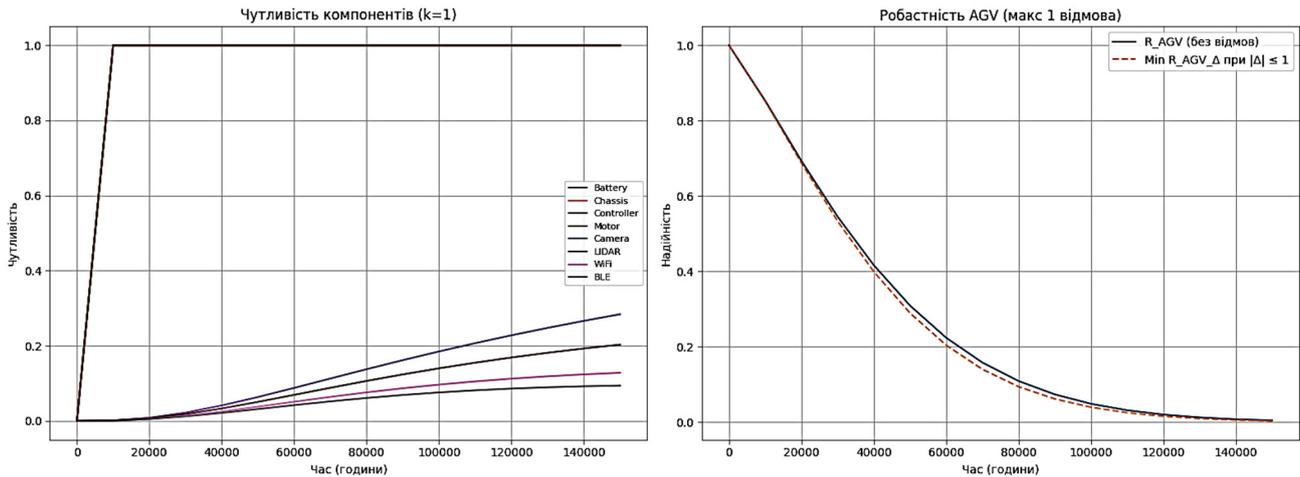


Рис. 3. Графіки чутливості елементів МРП і робастності МРП для одночасно однієї відмови її елементів

Таблиця 3

Значення часу для номінальних значень без відмов елементів МРП, а також чутливість компонентів до загальної надійності МРП за однієї відмови елементів МРП

t (години)	$R_{nominal}$	$R_{min} \mid \Delta \leq 1$	Батарея	Шасі	Контролер	Двигун	Камера	Лідар	Wi-Fi	BLE
0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10000.0	0.853	0.8518	1.0	1.0	1.0	1.0	0.0014	0.0011	0.0009	0.0008
20000.0	0.6935	0.6877	1.0	1.0	1.0	1.0	0.0084	0.0069	0.0054	0.0046
30000.0	0.5454	0.5333	1.0	1.0	1.0	1.0	0.0221	0.018	0.0138	0.0118
40000.0	0.4164	0.3993	1.0	1.0	1.0	1.0	0.0411	0.0331	0.0251	0.0211
50000.0	0.3093	0.2896	1.0	1.0	1.0	1.0	0.0636	0.0506	0.0378	0.0315
60000.0	0.2238	0.2042	1.0	1.0	1.0	1.0	0.0879	0.0693	0.051	0.042
70000.0	0.1581	0.1402	1.0	1.0	1.0	1.0	0.1128	0.0881	0.0639	0.052
80000.0	0.1091	0.0941	1.0	1.0	1.0	1.0	0.1376	0.1064	0.0759	0.0611
90000.0	0.0736	0.0617	1.0	1.0	1.0	1.0	0.1617	0.1238	0.0868	0.069
100000.0	0.0487	0.0397	1.0	1.0	1.0	1.0	0.1849	0.14	0.0966	0.0758
110000.0	0.0316	0.025	1.0	1.0	1.0	1.0	0.2069	0.155	0.1051	0.0815
120000.0	0.0201	0.0155	1.0	1.0	1.0	1.0	0.2279	0.1688	0.1125	0.086
130000.0	0.0126	0.0095	1.0	1.0	1.0	1.0	0.2476	0.1813	0.1187	0.0895
140000.0	0.0078	0.0057	1.0	1.0	1.0	1.0	0.2663	0.1927	0.1239	0.0922
150000.0	0.0047	0.0034	1.0	1.0	1.0	1.0	0.284	0.2031	0.1281	0.094

Мінімальні значення надійності (R_{min}), які враховують можливі коливання параметрів, завжди трохи нижчі за номінальні, але тенденція зниження із часом однакова. До 150 000 год система має дуже низьку ймовірність безвідмовної роботи – менше 0,5 %. Отже, система із часом втрачає свою працездатність, і для забезпечення надійної роботи потрібно планувати профілактичне обслуговування або заміну компонентів до того, як надійність стане критично низькою.

На рис. 4 зображено графіки чутливості елементів МРП та робастності МРП для одночасних двох відмов її елементів.

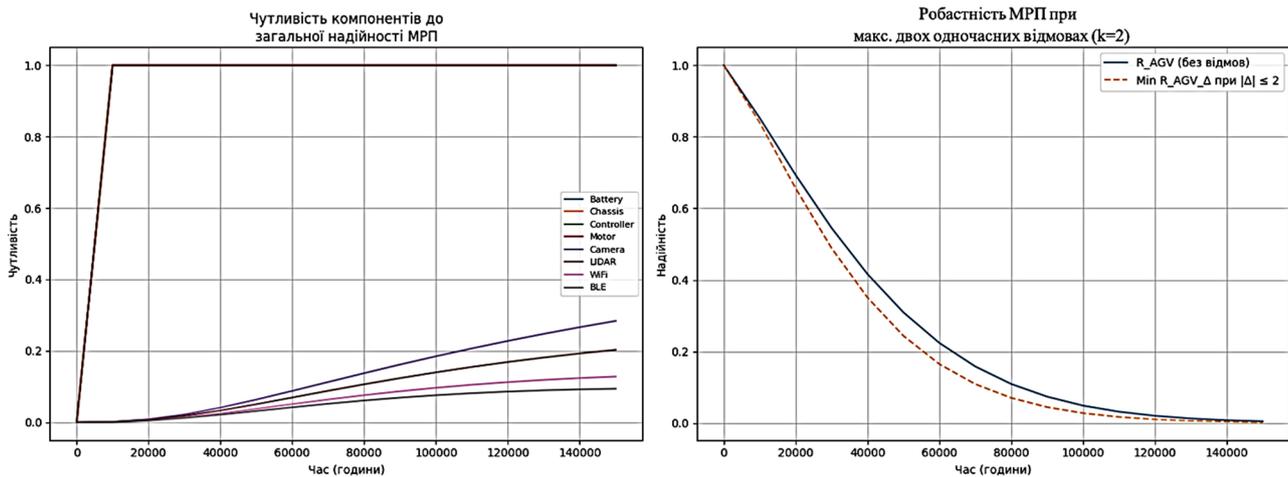


Рис. 4. Графіки чутливості елементів МРП і робастності МРП для одночасних двох відмов її елементів

Система МРП за максимальної кількості відмов (дві одночасні відмови) має високу надійність на початку експлуатації, що становить близько 85 % після 10 000 годин. Із часом надійність поступово знижується і після 150 000 год падає майже до нуля. Мінімальні значення надійності трохи нижчі за номінальні через можливі варіації. Загалом, система стабільна на початковому етапі, але із часом потребує профілактики або заміни для підтримки робастності.

У табл. 4 представлено розраховані значення часу для номінальних значень без відмов елементів МРП ($R_{nominal}$) та з двома одночасними відмовами елементів МРП (R_{min} при $|\Delta| \leq 2$), а також чутливість компонентів до загальної надійності МРП. Якщо $R_{min}|\Delta| \leq 2$, то маємо справу із живучістю, тобто здатністю МРП продовжувати виконання місії навіть за двох відмов.

Таблиця 4

Значення часу для номінальних значень без відмов елементів МРП, а також чутливість компонентів до загальної надійності МРП за двох відмов елементів МРП

t (години)	$R_{nominal}$	$R_{min} \Delta \leq 2$	Батарея	Шасі	Контролер	Двигун	Камера	Лідар	Wi-Fi	BLE
0.0	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10000.0	0.853	0.8397	1.0	1.0	1.0	1.0	0.0014	0.0011	0.0009	0.0008
20000.0	0.6935	0.6564	1.0	1.0	1.0	1.0	0.0084	0.0069	0.0054	0.0046
30000.0	0.5454	0.489	1.0	1.0	1.0	1.0	0.0221	0.018	0.0138	0.0118
40000.0	0.4164	0.3507	1.0	1.0	1.0	1.0	0.0411	0.0331	0.0251	0.0211
50000.0	0.3093	0.2437	1.0	1.0	1.0	1.0	0.0636	0.0506	0.0378	0.0315
60000.0	0.2238	0.1647	1.0	1.0	1.0	1.0	0.0879	0.0693	0.051	0.042
70000.0	0.1581	0.1087	1.0	1.0	1.0	1.0	0.1128	0.0881	0.0639	0.052
80000.0	0.1091	0.0702	1.0	1.0	1.0	1.0	0.1376	0.1064	0.0759	0.0611
90000.0	0.0736	0.0444	1.0	1.0	1.0	1.0	0.1617	0.1238	0.0868	0.069
100000.0	0.0487	0.0276	1.0	1.0	1.0	1.0	0.1849	0.14	0.0966	0.0758
110000.0	0.0316	0.0169	1.0	1.0	1.0	1.0	0.2069	0.155	0.1051	0.0815
120000.0	0.0201	0.0102	1.0	1.0	1.0	1.0	0.2279	0.1688	0.1125	0.086
130000.0	0.0126	0.006	1.0	1.0	1.0	1.0	0.2476	0.1813	0.1187	0.0895
140000.0	0.0078	0.0035	1.0	1.0	1.0	1.0	0.2663	0.1927	0.1239	0.0922
150000.0	0.0047	0.002	1.0	1.0	1.0	1.0	0.284	0.2031	0.1281	0.094



Робастність ($|\Delta| \leq 1$) МРП визначається як стабільність у разі одиначної відмови. Система лишається повністю працездатною, навіть якщо один компонент виходить із ладу, наприклад, BLE чи камера. Функція надійності основних компонентів (контролер, двигун, шасі, батарея) змінюється від 1.0 до 150 000 год. Це вказує на надлишковість або високу надійність МРП, яка необхідна на промислових підприємствах. Наприклад за 100 000 год $R_{nominal} = 0.0487$ (без відмов), а $R_{min} |\Delta| \leq 1 = 0.0397$. Це означає, що відбулося падіння менш ніж на 20 %, тобто одна відмова практично не впливає на функціонування МРП.

Живучість ($|\Delta| \leq 2$) виглядає, як адаптація до фатальних сценаріїв. Наприклад, при $t = 100\,000$ год $R_{min} |\Delta| \leq 2 = 0.0276$, $R_{min} |\Delta| \leq 2 = 0.0397$, що на 30 % менше. R_{min} падає швидше, ніж за одиначної відмови. Причому, хоча ймовірність виконання завдання суттєво падає, МРП все ще здатна функціонувати, тобто система виживає в режимі деградації. На 140 000 год $R_{nominal} = 0.0078$, $R_{min} |\Delta| \leq 2 = 0.0035$, а це означає, що ймовірність виконання з урахуванням двох відмов удвічі нижча, але все ще ненульова. Тобто МРП може виконати завдання, але для цього повинен від'їхати в безпечну зону, подати сигнал, навіть у випадку відмови, наприклад, BLE + Wi-Fi, або камери + LIDAR. Отже живучість забезпечується за рахунок: деградованих режимів роботи; місцевих автономних рішень і резервного контролю.

Дискусія і висновки

На основі проведених досліджень найменш надійними елементами МРП є: акумуляторна батарея, лідар, Wi-Fi- та BLE-модулі. Акумуляторна батарея має швидке зниження надійності після 50 000 год. Лідар має стрімке зростання ймовірності відмов ($R \approx 0.28$) у разі 150 000 год. Wi-Fi-модуль має схильність до частих збоїв ($R \approx 0.8$ на 100 000 год), а він є критичним для зв'язку. У BLE-модуля функція надійності $R \approx 0.09$ на 150 000 год, тобто є майже гарантована відмова. Інші компоненти (шасі, контролер, двигун, камера) мають дуже високу надійність ($R \approx 1.0$ до 150 000 год) і не вимагають негайного втручання.

Підвищення надійності можна здійснювати на компонентному рівні. А саме, для акумуляторної батареї запровадити моніторинг стану з прогнозом деградації, використовувати LiFePO₄ або твердотільні батареї з вищим MTTF і забезпечити модульну заміну батарей без зупинки платформи. Для лідара: обрати бездзеркальні або твердотільні лідари з меншою кількістю рухомих частин, регулярно калібрувати їх і тримати у вологозахисних корпусках, а також використовувати систему самодіагностики для виявлення часткових відмов. Надійність Wi-Fi- та BLE-модулів можна підвищити так: обрати промислові модулі зв'язку з антивібраційним захистом, запровадити виявлення втрати зв'язку й автоматичний перезапук модулів.

Для підвищення робастності на системному рівні необхідно здійснити резервування критичних компонентів. А саме, встановити подвійний лідар або ф'южн із камерою для компенсації втрати, а також подвійний канал зв'язку (Wi-Fi + BLE або Wi-Fi + LoRa/5G). Механізм перемикання між модулями за втрати Wi-Fi забезпечує автоматичний перехід на BLE або офлайн-режим. Також необхідно прогнозувати збої шляхом побудови функції $\lambda(t)$ та запуску превентивного обслуговування. На рівнях модульної архітектури забезпечити швидку заміну лідару, батареї чи зв'язку без демонтажу всієї платформи.

Щоб підвищити живучість на рівні виконання поставлених перед МРП завдань, необхідно формувати їхню кількість так, щоб за відмови до двох одиниць МРП, система продовжувала працювати, наприклад, із 10 платформ лише 7 мають бути обов'язково справні ($m = 7$, $n = 10$). Планувати маршрути з урахуванням ризиків. Найвищий пріоритет давати платформам із найвищою надійністю $R(t)$ та нижчою $\lambda(t)$. Також МРП повинна мати можливість продовжити роботу у деградованому режимі без зв'язку (локальна навігація, останній маршрут). Центральний диспетчер має перерозподіляти завдання між справними МРП, щойно деякі з них вийдуть із ладу.

Загалом МРП характеризується високою надійністю критичних механічних та обчислювальних модулів, однак сенсори та бездротові комунікації є "вузьким місцем". Для підвищення робастності й живучості слід орієнтуватися на резервування, модульність, адаптивну маршрутизацію та превентивну діагностику. Це дозволить зберігати працездатність навіть у разі часткових або повторних відмов окремих компонентів чи платформ.

Джерела фінансування. Це дослідження не отримало жодного гранта від фінансової установи в державному, комерційному або некомерційному секторах.

Список використаних джерел

- Amin, A. A., & Hasan, K. M. (2019). A review of fault tolerant control systems: Advancements and applications. *Measurement*, 143, 58–68. <https://doi.org/10.1016/j.measurement.2019.04.083>
- Bensaid Amrani, N., Benmoussa, S., Boukhnefer, M., & Habbal, A. (2023). Evaluating the predicted reliability of mechatronic systems: State of the art. *Mechanical Engineering: An International Journal (MEIJ)*, 3(2), 12. <https://doi.org/10.5281/zenodo.7801706>
- Betzer, J. S., Boudjadar, J., Frasheri, M., & Talasila, P. (2024). Digital twin enabled runtime verification for autonomous mobile robots under uncertainty. In *Proceedings of the 28th International Symposium on Distributed Simulation and Real Time Applications (DS-RT 2024)* (pp. 10–17). <https://doi.org/10.1109/DS-RT62209.2024.00012>
- Cai, Z., Zhang, F., Tan, Y., Kessler, S., & Fotner, J. (2024). Integration of an IoT sensor with angle-of-arrival-based angle measurement in AGV navigation: A reliability study. *Journal of Industrial Information Integration*, 42, Article 100707. <https://doi.org/10.1016/j.jii.2024.100707>
- Fazlollahabbar, H., & Akhavan Niaki, S. T. (2017). *Reliability models of complex systems for robots and automation*. CRC Press. <https://doi.org/10.1201/b22491>
- Li, L., & Schulze, L. (2024). Failure prediction of automated guided vehicle systems in production environments through artificial intelligence. *Tehnički glasnik*, 18(2), 268–272. <https://doi.org/10.31803/tg-20240416185206>
- Luo, W. (2023). Towards safe and resilient autonomy in multi-robot systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(13), Article 15449. <https://doi.org/10.1609/aaai.v37i13.26816>
- Maza, S. (2025). Diagnostic-constrained fault-tolerant control of bi-directional AGV transport systems with fault-prone sensors. *ISA Transactions*, 158, 227–241. <https://doi.org/10.1016/j.isatra.2025.01.014>
- Panchal, D. (2023). Reliability analysis of turbine unit using intuitionistic fuzzy Lambda–Tau approach. *Reports in Mechanical Engineering*, 4(1), 47–61. <https://doi.org/10.31181/rme040117042023p>
- Purwaningsih, R., Shintyastuti, A. R., Arvianto, A., & Hapsari, C. A. P. (2025). Evaluating the impact of autonomous material handling on the performance of production system: A simulation approach. *Sinergi*, 29(2), 139–146. <https://doi.org/10.22441/sinergi.2025.2.004>
- Yin, J., Li, L., Mourelatos, Z. P., Liu, Y., Gorsich, D., & Singh, A. (2023). Reliable global path planning of off-road autonomous ground vehicles under uncertain terrain conditions. *IEEE Transactions on Intelligent Vehicles*, 9(1), 1–14. <https://doi.org/10.1109/TIV.2023.3317833>
- Zheng, M., Feng, B., Zhao, Y., & Liang, W. (2024). Ordered and reliable retransmission method for wireless AGV systems based on WIA-FA. *Information and Control*, 53(5), 616–626. <https://doi.org/10.13976/j.cnki.xk.2024.3165>



References

- Amin, A. A., & Hasan, K. M. (2019). A review of fault tolerant control systems: Advancements and applications. *Measurement*, 143, 58–68. <https://doi.org/10.1016/j.measurement.2019.04.083>
- Bensaid Amrani, N., Benmoussa, S., Boukhnifer, M., & Habbal, A. (2023). Evaluating the predicted reliability of mechatronic systems: State of the art. *Mechanical Engineering: An International Journal (MEIJ)*, 3(2), 12. <https://doi.org/10.5281/zenodo.7801706>
- Betzer, J. S., Boudjadar, J., Frasher, M., & Talasila, P. (2024). Digital twin enabled runtime verification for autonomous mobile robots under uncertainty. In *Proceedings of the 28th International Symposium on Distributed Simulation and Real Time Applications (DS-RT 2024)* (pp. 10–17). <https://doi.org/10.1109/DS-RT62209.2024.00012>
- Cai, Z., Zhang, F., Tan, Y., Kessler, S., & Fottner, J. (2024). Integration of an IoT sensor with angle-of-arrival-based angle measurement in AGV navigation: A reliability study. *Journal of Industrial Information Integration*, 42, Article 100707. <https://doi.org/10.1016/j.jii.2024.100707>
- Fazlollahtabar, H., & Akhavan Niaki, S. T. (2017). *Reliability models of complex systems for robots and automation*. CRC Press. <https://doi.org/10.1201/b22491>
- Li, L., & Schulze, L. (2024). Failure prediction of automated guided vehicle systems in production environments through artificial intelligence. *Tehnicki glasnik*, 18(2), 268–272. <https://doi.org/10.31803/tg-20240416185206>
- Luo, W. (2023). Towards safe and resilient autonomy in multi-robot systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 37(13), Article 15449. <https://doi.org/10.1609/aaai.v37i13.26816>
- Maza, S. (2025). Diagnostic-constrained fault-tolerant control of bi-directional AGV transport systems with fault-prone sensors. *ISA Transactions*, 158, 227–241. <https://doi.org/10.1016/j.isatra.2025.01.014>
- Panchal, D. (2023). Reliability analysis of turbine unit using intuitionistic fuzzy Lambda–Tau approach. *Reports in Mechanical Engineering*, 4(1), 47–61. <https://doi.org/10.31181/rme040117042023p>
- Purwaningsih, R., Shintyastuti, A. R., Arvianto, A., & Hapsari, C. A. P. (2025). Evaluating the impact of autonomous material handling on the performance of production system: A simulation approach. *Sinergi*, 29(2), 139–146. <https://doi.org/10.22441/sinergi.2025.2.004>
- Yin, J., Li, L., Mourelatos, Z. P., Liu, Y., Gorsich, D., & Singh, A. (2023). Reliable global path planning of off-road autonomous ground vehicles under uncertain terrain conditions. *IEEE Transactions on Intelligent Vehicles*, 9(1), 1–14. <https://doi.org/10.1109/TIV.2023.3317833>
- Zheng, M., Feng, B., Zhao, Y., & Liang, W. (2024). Ordered and reliable retransmission method for wireless AGV systems based on WIA-FA. *Information and Control*, 53(5), 616–626. <https://doi.org/10.13976/j.cnki.xk.2024.3165>

Отримано редакцією журналу / Received: 06.07.25

Прорецензовано / Revised: 08.07.25

Схвалено до друку / Accepted: 14.09.25

Olena PAVLIUK, DSc (Engin.), Assoc. Prof.
 ORCID ID: 0000-0003-4561-3874
 e-mail: olena.m.pavliuk@plnu.u
 Lviv Polytechnic National University, Lviv, Ukraine

MODEL FOR ASSESSING THE ROBUSTNESS AND SURVIVAL OF MOBILE ROBOTIC PLATFORMS TAKING INTO ACCOUNT RELIABILITY

Background. This publication develops a comprehensive model of reliability, robustness, and survivability of a mobile robotic platform (MRP) with a series-parallel architecture. The aim of the work is to perform mathematical modeling and analysis of the reliability of MRP components based on Weibull and exponential distribution laws, as well as to assess the impact of redundancy and structural redundancy on the overall system resilience to failures. It was established that the critical elements of the MRP are the chassis, motor, and controller, as they are connected in series and have a low level of fault tolerance without redundancy, unlike the sensor and communication components connected in parallel.

Methods. Reliability calculation using series-parallel schemes; evaluation of mean time to failure (MTTF); modeling using Weibull and exponential distributions; sensitivity analysis to evaluate the impact of components on system reliability; assessment of system robustness and survivability under environmental changes or faults.

Results. Based on the constructed models, numerical MTTF values were obtained for all MRP components. The highest fault-free operation times were recorded for the chassis (135,412 hours) and the battery (124,914 hours), while the lowest reliability was demonstrated by the Wi-Fi communication modules (80,000 hours) and BLE modules (70,000 hours). The MTTF for the entire MRP is 21,188 hours, approximately 2.4 years of continuous operation. Sensitivity graphs showed that the chassis, controller, and motor have the greatest impact on overall reliability. These are priorities for redundancy or the creation of degraded modes. The system robustness with one failure ($\Delta \leq 1$) was evaluated through the decrease in probability of fault-free operation. However, over time (beyond 100,000 hours), reliability drops below 5%, highlighting the importance of accounting for component wear. To increase survivability, the MRP should operate with up to two platform failures, plan routes based on reliability, support degraded mode, and dynamically redistribute tasks.

Conclusions. The proposed model enables highly accurate reliability assessment of the MRP considering its structure, connection types, and functional features of components. The results indicate the need for redundancy of critical components and increased system robustness to ensure failure resilience. The presented methodology can be used for maintenance planning, lifecycle forecasting, and optimization of MRP operation in dynamic environments.

Keywords: mobile robotic platform (MRP), reliability, robustness, series-parallel structure, component sensitivity, route optimization.

Автор заявляє про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.



MOTION-CONSISTENT TEMPORAL FUSION FOR UAV DETECTION AND TRACKING

Background. *Detecting and tracking Unmanned Aerial Vehicles (UAVs) in video streams is essential for modern air-space monitoring yet remains challenging because UAVs are small, fast and easily confused with birds or background clutter. Conventional detectors produce noisy, frame-wise boxes, while standard trackers still suffer from false positives and identity switches. The purpose of this study is to stabilize UAV detections by adding a motion-aware temporal fusion method to a mainstream detector-tracker pipeline.*

Methods. *A detection-tracking pipeline was constructed using an RT-DETR (Real-Time Detection Transformer) and ByteTrack baseline, extended with a lightweight, training-free motion-consistent fusion (MCF) method. The method (i) aggregates bounding-box history over five frames, (ii) averages spatial and confidence values, and (iii) penalizes tracks whose short-term velocity or angular change exceeds empirically chosen thresholds. No appearance features or additional learning are required, so the solution runs in real time on a single GPU.*

Results. *Experiments on a labelled UAV-video dataset show that the proposed method increases Multiple Object Tracking Accuracy (MOTA) from 0.533 to 0.591, precision from 73 % to 84 %, and reduces identity switches from 60 to 28 (a 53 % improvement in ID stability). Recall decreases slightly from 90 % to 76 %, reflecting a deliberate trade-off: the system filters unstable or non-UAV motion to improve track consistency and suppress false positives. The evaluation was performed on more than 1,000 video sequences, ensuring robustness across diverse flight environments.*

Conclusions. *The motion-consistent fusion method significantly enhances both accuracy and temporal coherence while adding minor computational cost. It can be added into existing detection-tracking systems and is particularly suited for real-time UAV surveillance applications, though performance may degrade if drones execute extremely abrupt maneuvers outside the predefined motion thresholds.*

Keywords: UAV, Object Tracking, Object Detection, Fusion Algorithm, Motion Filter, RT-DETR, ByteTrack.

Background

Unmanned aerial vehicles (UAVs), commonly named as drones, are increasingly used in both civilian and military applications. As the number of UAV operating in public and private airspace continues to grow, the need for accurate, reliable, and real-time UAV detection systems becomes critical, particularly in video-based monitoring applications.

However, detecting and tracking UAVs in real-world videos is very challenging task. UAVs are typically small, fast-moving, and often visually similar to birds and background objects. Even modern object detectors can produce noisy bounding boxes with volatile confidence values for different frames. Multi-object tracking methods such as ByteTrack can associate detections over time, but they are still vulnerable to false positives, identity switches, and lack semantic understanding of object motion.

This research focuses on the development of a lightweight, real-time fusion method to improve the temporal stability of UAV detections in video. The object of study is a vision-based UAV monitoring system, and the subject is the refinement of frame-level detections through fusion of temporal and motion information.

The goal of this paper is to improve UAV detection accuracy and tracking reliability by introducing a motion-aware temporal fusion module. Special attention is paid to the Multiple Object Tracking Accuracy (MOTA) metric, which combines false positives, missed targets, and identity switches into a single measure of tracking performance.

To achieve this, the following tasks were completed:

- Construction of a detection-tracking pipeline using RT-DETR and ByteTrack;
- Implementation of a temporal fusion module that smooths detection outputs;
- Integration of a motion consistency filter to penalize erratic movement;
- Evaluation of detection and tracking performance.

The scientific novelty of this article lies in the proposed motion-consistent fusion (MCF) mechanism, which extends standard temporal fusion with motion smoothness analysis, which is a method particularly suited for UAVs, that typically follow smooth, linear trajectories. Unlike appearance-based methods, MCF does not require additional training or feature extraction, making it highly practical for real-time UAV tracking.

Experimental results show that the proposed approach improves overall tracking accuracy, indicated by a higher MOTA score, while increasing precision, and significantly lowering identity switches. Recall remains strong, though the method may underperform in cases of highly dynamic or intentionally evasive flight patterns, where the assumption of smooth, UAV-like motion no longer holds.

Related work. The real-time tracking of Unmanned Aerial Vehicles (UAVs) faces bigger challenges than detection alone: extremely small targets, quick moves, frequent disappearances, and camera motion. So, recent research in UAV tracking focuses on two goals: building high-quality benchmarks that reveal these pitfalls and designing trackers that remain robust while still running at video rate on edge hardware.

Jiang et al. introduce Anti-UAV, a large-scale RGB benchmark containing more than 300 video pairs and about 580k manually annotated bounding boxes (Jiang et al., 2021). Beyond its size, the dataset pairs each UAV sequence with an



empty-background counterpart, enabling dual-stream evaluation, and the authors supply a baseline tracker (DFSC) that exploits semantic flow to stabilize appearance drift, showing a significant margin over earlier Siamese architectures.

To support this, Zhao et al. created DUT Anti-UAV, which combines 10 000 labelled still images for detection with 20 RGB videos for tracking (Zhao et al., 2022). The authors demonstrate that fusing a YOLO-style detector with a lightweight correlation-filter tracker markedly boosts precision, underlining the value of joint "detect-then-track" pipelines when the target is tiny or often disappears.

While both datasets rely on visible light, Zhang, P. et al. extend the problem to visible–thermal UAV tracking (VTUAV) by collecting 500 RGB-T sequences totaling 1.7 million 1080 p frame pairs (Zhang, P. et al., 2022). They also created a new tracker called the Hierarchical Multi-modal Fusion Tracker (HMFT), which combines thermal information at different levels. This layered approach reduces tracking failures by about 20% in their tests, showing that heat data can help when lighting is poor.

Transformer backbones increasingly dominate state-of-the-art trackers. Yu et al. present UTTracker, whose unified Transformer feeds four specialized heads: (i) multi-region local search, (ii) global detection for re-entry, (iii) background correction to cope with camera motion, and (iv) a dynamic head tuned for very small targets (Yu et al., 2023). This holistic design secured the 2nd place in the 3rd Anti-UAV Challenge, highlighting that integrating detection, localization, and motion compensation is more effective than alternating them.

Robustness can also be improved through representation learning. Fu et al. propose PRL-Track, which first learns a coarse appearance-aware representation, then progressively refines it via a hierarchical generator (Fu et al., 2024). Despite the two-stage pipeline, PRL-Track maintains 42.6 FPS on an Nvidia Xavier, illustrating that adaptive feature refinement can balance accuracy with real-time constraints on embedded hardware.

Finally, Do et al. tackle deployment at scale with RAMOTS, a real-time multi-object tracker that couples YOLOv8 (Reis et al., 2023) and YOLOv10 (Wang et al., 2024) (versions of the You Only Look Once real-time object detection algorithm) detectors and BYTETrack (Zhang, Y. et al., 2022) / BOTSORT (Aharon, Orfaig, & Bobrovsky, 2022) trackers for fault-tolerant processing (Do et al., 2024). Although aimed at "from-UAV" rather than "against-UAV" scenarios, RAMOTS demonstrates that big-data tooling sustains 28 FPS while managing large video streams—an architectural insight relevant to surveillance networks.

However, existing methods often overlook the specific motion characteristics of UAVs, leading to false positives in cluttered environments. To address these limitations, the method proposed in this paper consists of three main components: a transformer-based detector, an association tracker, and a novel motion-consistent fusion mechanism designed to stabilize trajectories based on flight dynamics.

Methods

The proposed approach comprises three stages: an object detector RT-DETR (Wang et al., 2025), a multi-object tracker ByteTrack (Zhang, Y. et al., 2022), and a custom temporal fusion module that filters and stabilizes detections based on motion consistency.

Detection Stage. The first component of the proposed system is the object detection model, which is responsible for identifying UAV instances in video frames. The RT-DETR (Real-Time Detection Transformer) is used as the backbone detector due to its balance between speed and accuracy, particularly suitable for real-time UAV monitoring applications.

RT-DETR is a transformer-based object detector that directly predicts bounding boxes and class probabilities without the need for anchor generation or post-processing steps such as non-maximum suppression (NMS). It uses a combination of convolutional and attention-based layers to extract rich spatial and contextual information from the input image. The architecture produces a fixed set of object queries, each corresponding to a potential object instance.

At inference time, RT-DETR outputs a set of detections per frame, where each detection includes:

- A bounding box (x, y, w, h) ;
- A class label (in this case, primarily "UAV");
- A confidence score $c \in [0, 1]$.

These frame-wise detections are passed to the tracking and fusion stages for temporal association and refinement.

Tracking stage. To maintain object identity across video frames and enable temporal reasoning, a real-time multi-object tracking method based on ByteTrack is integrated. ByteTrack is a high-performance, association-based tracker that links object detections over time without relying on deep appearance features, making it lightweight and well-suited for UAV tracking in real-time scenarios.

At each frame t , the detector (RT-DETR) produces a set of bounding boxes $\{B_j^t\}$ with associated confidence scores. ByteTrack divides these detections into high-confidence and low-confidence sets using a fixed threshold (e.g., 0.5). It then performs data association between the new detections and previously tracked objects.

Each object track maintains a motion model using a Kalman filter. The position of the object is predicted in the current frame using the following linear state transition model:

$$\dot{x}_t|_{t-1} = A \cdot x_{t-1},$$

where x_{t-1} – is the previous state vector of the object (e.g., center position and velocity), A – is the state transition matrix (typically constant), $\dot{x}_t|_{t-1}$ – is the predicted state before updating with new observations.

To match detections with predicted tracks, ByteTrack computes the Intersection over Union (IoU) between the predicted bounding box B_i and the detected box B_j :

$$IoU(B_i, B_j) = \frac{|B_i \cap B_j|}{|B_i \cup B_j|}.$$

A cost matrix is constructed based on this metric, and the Hungarian algorithm is used to solve the assignment problem, pairing detections with existing tracks by maximizing overall IoU.

Each matched track is updated with the new detection. Unmatched tracks are temporarily retained and predicted forward using the Kalman filter, allowing the system to recover from short-term occlusions. Tracks that remain unmatched for a predefined number of frames are terminated. Importantly, ByteTrack does not modify detection confidence or bounding box geometry, it only propagates and assigns object identities.



While ByteTrack is highly effective for frame-to-frame association, it is sensitive to detection noise and does not enforce global consistency across time. To deal with these challenges, a temporal fusion method is introduced in the next section that smooths detection results and penalizes unusual motion patterns, which are rare in typical UAV trajectories.

Temporal Fusion Method. Although the integration of ByteTrack provides temporal identity assignment and object association across frames, it does not directly improve the quality or stability of detection results. In real-world UAV tracking scenarios, especially under challenging conditions (e.g., occlusions, fast motion, or cluttered backgrounds), frame-by-frame detections can have high variance in both confidence scores and bounding box locations. To address these limitations, in this paper a lightweight temporal fusion method is introduced that aggregates detection information over time for each tracked object.

The core idea of the proposed fusion method is to keep a fixed-length history buffer for each object track, indexed by its unique track ID. This buffer stores the most recent N detections associated with the object, including:

- The bounding box coordinates $B_i^t = (x, y, w, h)$,
- The detection confidence c_i^t .

At each frame, the fusion method performs the following steps for every active track:

Step 1. Update the buffer with the current detection (bounding box and confidence).

Step 2. Remove the oldest entry if the buffer exceeds the maximum length N .

Step 3. Compute a fused bounding box by averaging the coordinates across all buffer entries:

$$\bar{B}_i = \frac{1}{N} \sum_{t=t-N+1}^t B_i^t.$$

Step 4. Compute a fused confidence score as the mean of the stored scores:

$$\bar{c}_i = \frac{1}{N} \sum_{t=t-N+1}^t c_i^t.$$

Step 5. Estimate the detection stability by calculating the standard deviation of the bounding box center positions across the buffer. If the standard deviation exceeds a defined threshold, the detection is flagged as unstable and may be suppressed.

This temporal fusion approach reduces jitter in bounding box placement and filters out short-lived, noisy detections. The averaging strategy also increases the reliability of confidence scores, allowing the system to more robustly handle volatile detections that may otherwise be discarded or mismatched.

The fusion process is non-parametric and computationally cheap, requiring only arithmetic operations over a small buffer. It can be applied in real time and is fully compatible with online tracking systems.

In the next section, this fusion mechanism is extended by incorporating motion consistency analysis to further penalize erratic movement patterns that are unlikely to originate from UAVs.

Motion Consistency Filtering. To further improve the robustness of UAV tracking, a motion consistency filtering mechanism is proposed to enhance the temporal fusion method by incorporating dynamic constraints based on typical UAV movement patterns. While the fusion module stabilizes detections over time, it does not consider the physical plausibility of an object's trajectory. In contrast, motion consistency filtering evaluates whether an object is moving in a smooth, UAV-like manner and penalizes tracks that have unnatural behavior.

UAVs typically have controlled, smooth motion, often with steady direction or hovering behavior. In contrast, false positives such as birds, background clutter, or detection noise tend to result in abrupt position changes or inconsistent movement between frames. The proposed method exploits this behavioral distinction to assess and filter object tracks.

For each object track, the short-term trajectory is analyzed over the last N frames using its center point positions $p_i^t = (x_t, y_t)$.

The following parameters are computed:

- Velocity change (Δv):

$$\Delta v = \|p_i^t - 2p_i^{t-1} + p_i^{t-2}\|;$$

- Angular deviation ($\Delta\theta$):

$$\Delta\theta = \arccos\left(\frac{(p_i^{t-1} - p_i^{t-2}) \cdot (p_i^t - p_i^{t-1})}{\|p_i^{t-1} - p_i^{t-2}\| \cdot \|p_i^t - p_i^{t-1}\|}\right);$$

- Motion consistency score (MCS):

$$MCS = \alpha \cdot \Delta v + \beta \cdot \Delta\theta.$$

If the motion consistency score exceeds a predefined threshold τ , the corresponding detection is flagged as inconsistent and either penalized (by reducing its confidence) or temporarily ignored by the tracking system.

This filtering mechanism helps suppress false positives caused by fast-moving distractors, reduce identity switches in noisy scenes, and prioritize detections that exhibit UAV-like motion patterns. Importantly, this enhancement is model-free and fully compatible with real-time deployment.

Fusion Algorithm. The final stage of the system combines the results of detection, tracking, temporal fusion, and motion consistency filtering to produce reliable, temporally stable UAV detections with unique identifiers. This stage determines which object tracks are retained and presented as output at each frame.

For each frame, the system performs the following steps:

1. Detection and tracking: The RT-DETR model generates bounding boxes and confidence scores, which are associated across frames using ByteTrack. Each detection is linked to a persistent object track with a unique ID.

2. Temporal fusion: For each active track, a fused bounding box and smoothed confidence score are computed by aggregating the track's detection history over the last N frames. This reduces jitter and stabilizes detection output.



3. Motion consistency filtering: The short-term motion pattern of each track is analyzed to compute a motion consistency score. Tracks with erratic or physically implausible motion are flagged and penalized by lowering their confidence or suppressing their output altogether.

4. Decision thresholding: Tracks that meet all criteria – including minimum confidence, temporal stability, and motion consistency – are passed to the final output stage. Each retained track is represented as:

$$Output = (B_i, \bar{c}_i, track_id),$$

where B_i is the fused bounding box, \bar{c}_i is the smoothed confidence score, and $track_id$ is the assigned identity.

To formalize the decision-making process, we define a valid output set O_{final} at frame t . A track T_i is included in the final output only if it satisfies both spatial stability and motion consistency constraints. This can be expressed using an indicator function:

$$O_{final} = \{ (\hat{B}_i, \hat{C}_i, id_i) \mid i \in \mathcal{T}_t, \sigma(H_i^{bbox}) \leq \sigma_{th} \wedge S_{motion}(i) \leq \tau \}.$$

where O_{final} – is the set of validated output tracks; \mathcal{T}_t – represents the set of all candidate tracks at frame; \hat{B}_i and \hat{C}_i – denote the fused bounding box and confidence score; $\sigma(H_i^{bbox})$ – is the spatial stability metric (standard deviation of the history buffer); $S_{motion}(i)$ – is the motion consistency score; σ_{th} and τ are the stability and motion thresholds, respectively.

The full logic of the fusion and filtering process is showed in Algorithm 1 below.

Algorithm 1. Fusion strategy of tracker

Input:

- T_k – set of tracks at frame k
- H – history buffer per track ID
- N – history length (e.g., 5 frames)
- σ_{thresh} – stability threshold
- θ_{max} – max angle deviation
- v_{max} – max velocity change

Output:

T_k^{fused} – filtered, fused tracks for frame k

```

1:   for each track t in  $T_k$  do
2:       id ← t.id
3:       bbox ← t.bbox
4:       conf ← t.confidence
5:       if H[id] does not exist then
6:           initialize H[id] ← empty list
7:       end if
8:       append (bbox, conf) to H[id]
9:       if length(H[id]) > N then
10:          remove the oldest entry from H[id]
11:       end if
12:       bboxes ← list of all b from (b, ) in H[id]
13:       confs ← list of all c from (, c) in H[id]
14:       fused_bbox ← mean(bboxes)
15:       fused_conf ← mean(confs)
16:       bbox_std ← standard_deviation(bboxes)
17:       compute motion_score from recent bboxes
18:         – velocity ← center_distance( $b_n, b_{n-1}$ )
19:         – angle_change ← angle( $center_{n-2} \rightarrow center_{n-1} \rightarrow center_n$ )
20:         – velocity ← center_distance( $b_n, b_{n-1}$ )
21:       if bbox_std <  $\sigma_{thresh}$  and motion_score < threshold then
22:           mark track t as "stable"
23:           output (fused_bbox, fused_conf, id)
24:       else
25:           penalize fused_conf
26:       end if
27:   end for

```

The proposed fusion algorithm introduces a novel enhancement over classical temporal fusion techniques by incorporating a lightweight motion consistency filter. Traditional fusion methods typically rely on statistical averaging (e.g., exponential moving averages or buffer smoothing) across historical detections to reduce jitter and noise. While effective in improving spatial stability, these approaches are blind to the dynamics of object motion and may average out valid but fast movements, or fail to suppress outliers caused by false positives.

In contrast, the proposed method evaluates each fused detection not only based on temporal coherence but also on motion plausibility, derived from short-term velocity and angular trajectory consistency. By explicitly penalizing tracks with erratic movement patterns, such as sudden jumps, sharp turns, or non-UAV-like acceleration, the proposed approach suppresses false positives (e.g., birds, clutter, background detections) without compromising on real-time performance.

This is considered particularly critical in UAV surveillance applications, where:

- Most UAV have smooth, predictable motion patterns.
- False positives often have inconsistent or unnatural motion.
- The system must be efficient and non-reliant on deep appearance models (e.g., no Re-ID or embeddings).

This fusion strategy improves detection precision while maintaining tracking integrity, leading to fewer identity switches, lower false positives, and more consistent trajectories.

Results

The system was evaluated on the (Jiang et al., 2021; Zhang, P. et al., 2022; Zhao et al., 2022) UAV video dataset, which contains more than 1000 video sequences annotated with bounding boxes and object identities. The dataset includes a diverse range of scenes such as urban areas, vegetation, and open skies, with variable lighting and UAV altitudes.

The detection method uses a pretrained RT-DETR model fine-tuned on UAV-specific data. Tracking is performed with ByteTrack, configured with a high-confidence threshold of 0.5 and a Kalman filter for state propagation.

Our temporal fusion method maintains a fixed-length history buffer $N = 5$ and applies motion filtering using two features: center velocity change and angle deviation. Tracks are penalized if their standard deviation exceeds $\sigma_{thresh} = 20$ pixels or if their motion score exceeds a dynamic threshold. All experiments were run on an NVIDIA RTX 3070 ti.

We evaluate performance using:

- MOTA as an overall indicator of tracking accuracy,
- Precision and Recall for detection quality,
- Identity Switches (ID Switches) for measure identity consistency during tracking.

Table 1

Result comparison of fusion configurations

Approach	MOTA	Precision	Recall	ID Switches
RT-DETR only	0.431	0.682	0.824	-
+ ByteTrack	0.480	0.736	0.806	57
+ Fusion (baseline)	0.533	0.733	0.900	60
+ MCF (our approach)	0.591	0.838	0.755	28

Table 1 presents a comparison of detection and tracking performance across different system variants with MOTA used as the primary measure of overall tracking accuracy. The baseline RT-DETR detector achieves a MOTA of 0.431, along with precision of 68% and recall of 82%, but lacks any temporal consistency or object ID tracking.

Adding ByteTrack improves tracking performance to a MOTA of 0.480, driven by gains in precision (up to 74%) and tracking consistency, though 57 identity switches still occur.

The baseline fusion approach further raises MOTA to 0.533, with recall peaking at 90%, confirming that temporal smoothing helps recover missed detections, but identity stability remains limited (60 ID switches).

The proposed method, Motion-Consistent Fusion (MCF), achieves the best overall results:

- MOTA increases to 0.591, the highest score among all tested variants,
- Precision improves to 84%,
- Recall decreases slightly to 76%,
- ID switches are reduced to just 28.

This trade-off, higher MOTA and precision at the cost of a small recall drop-reflects the effect of motion consistency filtering, which suppresses detections with erratic or non-UAV motion. While a few true but noisy detections may be excluded, the result is a more stable and reliable tracking system, with improved resistance to clutter and false positives.

Figure 1 illustrates typical frames where the proposed method smooths detections and eliminates false tracks.

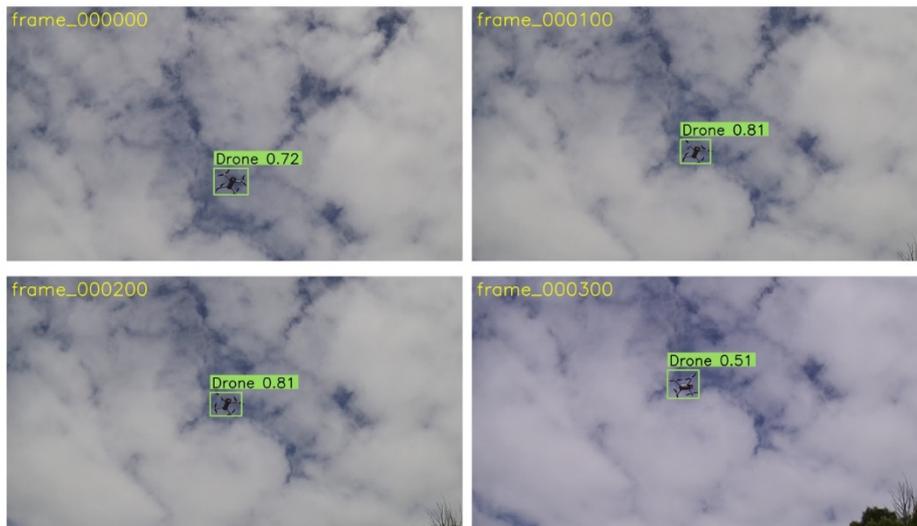


Fig. 1. Examples of the motion-consistent fusion (MCF) results



Discussion and conclusions

In this paper, a lightweight and effective temporal fusion framework for UAV detection and tracking in video streams is presented. Based on a standard detection-tracking pipeline with RT-DETR and ByteTrack, a novel motion-consistent fusion (MCF) method is proposed, which improves detection stability by combining temporal smoothing with motion plausibility analysis.

The proposed method introduces two key ideas:

1. **Temporal aggregation** of bounding boxes and confidence scores over a fixed-length buffer, reducing frame-to-frame jitter.
2. **Motion consistency filtering**, which penalizes tracks exhibiting erratic movement patterns unlikely to represent UAV behavior.

Experimental results on a UAV video dataset demonstrate that the proposed approach improves overall tracking performance, achieving a higher MOTA score, increased precision, and significantly fewer identity switches. While recall decreases slightly, this trade-off helps suppress false positives and ensures more stable, UAV-like tracking behavior. MCF increases MOTA from 0.533 to 0.591 – the highest among all tested variants, while maintaining real-time operation, requiring no additional training, and integrating seamlessly with existing detection and tracking systems.

Overall, the proposed method offers a practical and generalizable improvement for UAV detection systems, particularly in real-world conditions where false positives and unstable detections are common.

Authors' contribution: Iryna Yurchuk – general supervision, methodology verification, review and editing of the manuscript, and final approval of the manuscript; Taras Semenchenko – conceptualization of the study, literature review, design and implementation of the motion-consistent fusion algorithm, experimental set-up and evaluation, analysis of results, drafting and final approval of the manuscript.

Sources of funding. This study did not receive any grant from a funding institution in the public, commercial, or non-commercial sectors.

References

- Aharon, N., Orfaig, R., & Bobrovsky, B.-Z. (2022). *BoT-SORT: Robust associations multi-pedestrian tracking*. arXiv. <https://doi.org/10.48550/arXiv.2206.14651>
- Do, N.-T., Nguyen, N. N.-Y., Nguyen, D.-P., & Do, T.-H. (2024). Ramots: A real-time system for aerial multi-object tracking based on deep learning and big data technology. In *2024 16th International Conference on Knowledge and System Engineering (KSE)* (pp. 1–6). VNU University of Engineering and Technology. <https://doi.org/10.1109/KSE63888.2024.11063545>
- Fu, C., Lei, X., Zuo, H., Yao, L., Zheng, G., & Pan, J. (2024). Progressive representation learning for real-time UAV tracking. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 5072–5079). School of Electrical and Electronic Engineering, Nanyang Technological University. <https://doi.org/10.1109/IROS58592.2024.10803050>
- Jiang, N., Wang, K., Peng, X., Yu, X., Wang, Q., Xing, J., Li, G., Zhao, J., Guo, G., & Han, Z. (2021). *Anti-UAV: A large multi-modal benchmark for UAV tracking*. arXiv. <https://doi.org/10.48550/arXiv.2101.08466>
- Reis, D., Kupec, J., Hong, J., & Daoudi, A. (2023). *Real-time flying object detection with yolov8*. arXiv. <https://doi.org/10.48550/arXiv.2305.09972>
- Wang, A., Chen, H., Liu, L., Chen, K., Lin, Z., Han, J., & Ding, G. (2024). Yolov10: Real-time end-to-end object detection. *Advances in Neural Information Processing Systems*, 37, 107984–108011. https://proceedings.neurips.cc/paper_files/paper/2024/hash/c34ddd05eb089991f06f3c5dc36836e0-Abstract-Conference.html
- Wang, S., Xia, C., Lv, F., & Shi, Y. (2025). Rt-detr3: Real-time end-to-end object detection with hierarchical dense positive supervision. In *2025 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)* (pp. 1628–1636). Johns Hopkins University. <https://doi.org/10.1109/WACV61041.2025.00166>
- Yu, Q., Ma, Y., He, J., Yang, D., & Zhang, T. (2023). A unified transformer based tracker for anti-UAV tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops* (pp. 3036–3046). IEEE Computer Society; Computer Vision Foundation. https://openaccess.thecvf.com/content/CVPR2023W/Anti-UAV/html/Yu_A_Unified_Transformer_Based_Tracker_for_Anti-UAV_Tracking_CVPRW_2023_paper.html
- Zhang, P., Zhao, J., Wang, D., Lu, H., & Ruan, X. (2022). Visible-thermal UAV tracking: A large-scale benchmark and new baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 8886–8895). IEEE Computer Society; Computer Vision Foundation. https://openaccess.thecvf.com/content/CVPR2022/html/Zhang_Visible-Thermal_UAV_Tracking_A_Large-Scale_Benchmark_and_New_Baseline_CVPR_2022_paper.html
- Zhang, Y., Sun, P., Jiang, Y., Yu, D., Weng, F., Yuan, Z., Luo, P., Liu, W., & Wang, X. (2022). ByteTrack: Multi-object tracking by associating every detection box. In S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, & T. Hassner (Eds.), *Computer Vision – ECCV 2022* (pp. 1–21). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-20047-2_1
- Zhao, J., Zhang, J., Li, D., & Wang, D. (2022). Vision-based anti-uav detection and tracking. *IEEE Transactions on Intelligent Transportation Systems*, 23(12), 25323–25334. <https://doi.org/10.1109/TITS.2022.3177627>

Отримано редакцією журналу / Received: 20.07.25
 Прорецензовано / Revised: 20.09.25
 Схвалено до друку / Accepted: 20.09.25



Ірина ЮРЧУК, канд. фіз.-мат. наук, доц.
ORCID ID: 0000-0001-8206-3395
e-mail: i.a.yurchuk@gmail.com
Київський національний університет імені Тараса Шевченка, Київ, Україна

Тарас СЕМЕНЧЕНКО, асп.
ORCID ID: 0009-0007-3259-7007
e-mail: taras.semenchenko@knu.ua
Київський національний університет імені Тараса Шевченка, Київ, Україна

МЕТОД УЗГОДЖЕНОГО ЗА РУХОМ ЧАСОВОГО Ф'ЮЖНА ДЛЯ ВИЯВЛЕННЯ ТА ВІДСТЕЖЕННЯ БПЛА

Вступ. *Виявлення та відстеження безпілотних літальних апаратів (БПЛА) у відеопотоках є критично важливим завданням сучасного моніторингу повітряного простору. Водночас воно залишається складним через малі розміри, швидкий рух БПЛА та їхню схожість із птахами чи фоновими об'єктами. Типові детектори формують нестабільні результати, які залежать від кадрів, а класичні треки часто дають хибні спрацьовування та помилки детекції. Метою цієї роботи є стабілізація детекцій БПЛА додаванням до стандартного детекції-трекінг пайплайну спеціального методу часово-просторового ф'южна, чутливого до характеру руху об'єкта.*

Методи. *Базову систему RT-DETR + ByteTrack розширено за допомогою легкого методу узгодженого з рухом згладжування (MCF), який не залежить від навчання. Цей метод: (i) агрегує історію обмежувальних рамок за останні п'ять кадрів, (ii) усереднює просторові координати та рівні довіри, (iii) штрафус знайдені об'єкти, у яких короточасні зміни швидкості або кута перевищують емпірично обрані порогові. Жодні ознаки зовнішності чи додаткове навчання не потрібні, тож рішення працює в реальному часі на одному GPU.*

Результати. *Експерименти на розміченому наборі відео з БПЛА показують, що запропонований метод підвищує MOTA з 0.533 до 0.591, Precision – із 73 % до 84 %, а кількість помилок ідентифікації зменшується із 60 до 28 (покращення на 53 % у стабільності ідентифікації). Recall трохи знижується з 90 % до 76 %, що відображає свідомий компроміс: система відфільтровує нестабільні або нехарактерні для БПЛА траєкторії, щоб покращити точність і зменшити кількість хибних спрацьовувань. Оцінювання проведено на понад 1 000 відеозаписах, що забезпечує надійність результатів у різноманітних умовах польоту.*

Висновки. *Запропонований метод ф'южна суттєво покращує як точність, так і стабільність результатів у послідовності кадрів відеовідстеження, практично не збільшуючи обчислювальні витрати. Його можна легко інтегрувати у вже наявні системи детекції та трекінгу. Метод особливо ефективний для застосування в реальному часі, хоча його продуктивність може знижуватися у випадках різких або непередбачуваних маневрів БПЛА поза межами заздалегідь визначених параметрів руху.*

Ключові слова: *БПЛА, трекінг об'єктів, детекція об'єктів, алгоритм ф'южна, фільтр руху, RT-DETR, ByteTrack.*

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.



МЕРЕЖЕВІ Й ІНТЕРНЕТ-ТЕХНОЛОГІЇ



^{1,2} Андрій ДУДНИК, д-р техн. наук, доц.
ORCID ID: 0000-0003-1339-7820
e-mail: andrii.dudnik@knu.ua

³ Андрій ФЕСЕНКО, канд. техн. наук, доц.
ORCID ID: 0000-0001-5154-5324
e-mail: andrii.fesenko@npp.kai.edu.ua

¹ Олександр ТОРОШАНКО, канд. техн. наук
ORCID ID: 0000-0002-2354-0187
e-mail: oleksandr.toroshanko@knu.ua

¹ Віра МИКОЛАЙЧУК, д-р філософії
ORCID ID: 0000-0002-2532-5771
e-mail: viramykolaichuk@knu.ua

² Олексій Батрак, асп.
ORCID ID: 0009-0007-8248-5935
e-mail: batrachelli@gmail.com

⁴ Сергій Виговський, асп.
ORCID ID: 0009-0007-7868-4923
e-mail: vigovsky.sa@gmail.com

¹ Київський національний університет імені Тараса Шевченка, Київ, Україна
² Відкритий міжнародний університет розвитку людини "Україна", Київ, Україна
³ Державне некомерційне підприємство Державний університет "Київський авіаційний інститут", Київ, Україна
⁴ Приватне акціонерне товариство "Вищий навчальний заклад «Міжрегіональна академія управління персоналом», Київ, Україна

СИСТЕМА ДЛЯ ПОШУКУ ВИБУХОНЕБЕЗПЕЧНИХ ПРЕДМЕТІВ У СКЛАДІ МОБІЛЬНОГО ГУСЕНИЧНОГО РОБОТА З КЕРУВАННЯМ ІЗ КВАДРОКОПТЕРА ЗА ДОПОМОГОЮ МАШИННОГО ЗОРУ

Вступ. Результати, описані в цій статті, є початком досліджень команди проєкту у сфері розроблення роботизованих систем виявлення вибухонебезпечних предметів. Дослідження проводились у лабораторних умовах і в них розглянуто лише загальну концепцію та визначення напрямів подальших досліджень. Подальші дослідження заплановано проводити в реальних умовах або наближених до них. Запропоновано систему виявлення вибухонебезпечних предметів на базі імпульсного металошукача K158, мікроконтролера та мобільної гусеничної платформи Keyestudio Mini Tank V3.0. Запропоновано інтеграцію з квадрокоптером, оснащеним комп'ютерним зором, що дає змогу здійснювати точне наведення та дистанційне керування. Описано структуру системи, принцип взаємодії компонентів, логіку прийняття рішень і математичну модель функціонування.

Методи. Використано методи моделювання функціональних підсистем, комп'ютерного осцилографічного аналізу електричних сигналів, математичного моделювання кінематики гусеничного робота в координатах XY, а також алгоритмічного проєктування автомата керування. Для перевірки взаємодії з квадрокоптером застосовано метод віртуального тестування у симуляційному середовищі з використанням машинного зору. Реалізацію оброблення сигналів здійснено засобами програмування мікроконтролера на основі порогової фільтрації.

Результати. Результати демонструють чітке розділення сигналів за наявності та відсутності металу. У фоновому режимі фіксується стабільна нульова лінія, а при виявленні цілі — серія імпульсів із частотою 125–150 Гц і струмом до 625 мА. Це забезпечує достовірну ідентифікацію об'єктів у реальному часі. Усі вузли, включно з модулем живлення та бездротового зв'язку, показали стабільну роботу. Наведення за допомогою машинного зору досягло точності ± 20 см. Загалом система продемонструвала надійність і ефективність у лабораторних умовах.

Висновки. Запропонована система виявлення вибухонебезпечних об'єктів об'єднує переваги аналогового сенсора, цифрової обробки сигналу, мобільної платформи та повітряної координації. Простота, автономність і низька вартість роблять її придатною для застосування у зонах ризику.

Ключові слова: вибухонебезпечні предмети, Arduino, імпульсний металошукач, гусеничний мобільний робот, квадрокоптер, машинний зір, автономне керування, виявлення металів.

Вступ

Сучасні воєнні дії, гуманітарні місії та пошуково-рятувальні операції у зонах підвищеної небезпеки супроводжуються високими ризиками для життя персоналу, особливо у виявленні вибухонебезпечних об'єктів (ВНО), таких як протипіхотні або протитанкові міни, нерозірвані снаряди тощо (Невлюдов та ін., 2023). У зв'язку із цим постає нагальна потреба у впровадженні дистанційних, автономних систем, які здатні забезпечити безпечне обстеження територій без безпосередньої участі людини.

Інтеграція доступних сенсорів, зокрема й імпульсних металошукачів, на базі мікроконтролерів Arduino у мобільні роботизовані платформи, створює передумови для побудови компактних, бюджетних і адаптивних пристроїв для виявлення ВНО. Arduino забезпечує гнучкість у налаштуванні, можливість оброблення аналогових сигналів і керування периферійними модулями у реальному часі.

Використання квадрокоптера як повітряного наглядного модуля з функцією машинного зору суттєво розширює можливості такої системи. Зокрема, дрон дозволяє здійснювати повітряну розвідку, формувати маршрут для наземного



робота, точно координувати його рух, а також оперативно передавати відео- та координатну інформацію. У сукупності це забезпечує комплексний підхід до дистанційного моніторингу та виявлення загроз без ризику для оператора.

Зростаюча зацікавленість у створенні роботизованих систем для виявлення вибухонебезпечних предметів із використанням платформ Arduino, сенсорів і засобів автономного керування, набула великої популярності, у зв'язку з російським вторгненням (Олексенко та ін., 2024). Наприклад, у працях (Mujiaro et al., 2021) та (Jeyagopi et al., 2022) розглянуто мобільні роботи-металощукачі на базі Arduino, однак відсутня підтримка віддаленого управління або візуальної навігації. У роботі (Abdulmajeed, & Hussein, 2015) досліджуються параметри впливу на якість виявлення металів, але немає механізмів адаптації до складних польових умов. У статті (Ravi Kiran et al., 2024) вперше запропоновано поєднання роботизованого шасі з камерою спостереження, проте оброблення відео не інтегрується у логіку автономного керування. Дослідження українських авторів, таких як (Солодчук, & Войтенко, 2022; Кузавков, 2022; Міщук, & Фесенко, 2024; Решетняк, 2024; Савін, 2024; Шовкоштитий, & Василенко, 2024), детально розглядають окремі аспекти: створення віртуального середовища, вибір сенсорів, використання БПЛА, реалізацію машинного зору, однак вони не пропонують повноцінної платформи з поєднанням усіх компонентів в єдину систему виявлення ВНО. Відмінністю запропонованого дослідження є повна інтеграція сенсорної частини (K158), цифрової обробки сигналу (Arduino), автономного управління та візуальної координації (квадрокоптер). Також проведено моделювання характеристик сигналу, побудовано автомати логіки та реалізовано адаптивну реакцію на виявлення цілі. Отже, ця робота не лише синтезує ідеї попередніх досліджень, а і пропонує цілісну систему для практичного застосування у зонах ризику.

Метою дослідження є розроблення та практична реалізація автономної мобільної системи для виявлення вибухонебезпечних предметів на основі імпульсного металощукача K158, інтегрованого з мікроконтролером Arduino, системами цифрової обробки сигналів і дистанційного візуального керування за допомогою квадрокоптера з комп'ютерним зором. Запропоноване рішення має забезпечити точне, безпечне й ефективне виявлення металевих об'єктів у складних польових умовах із можливістю автоматизованої фіксації та передачі координат виявлених цілей.

Постановка задачі. Для досягнення мети дослідження необхідно розв'язати такі завдання.

Провести аналіз сучасних підходів до побудови систем виявлення вибухонебезпечних об'єктів, зокрема з використанням Arduino, роботизованих платформ і безпілотних літальних апаратів.

Розробити електричну схему підключення імпульсного металощукача K158 до мікроконтролера Arduino та забезпечити коректне зчитування вихідного сигналу.

Побудувати математичні моделі функціонування сенсорної частини, логіки обробки сигналів і кінематики гусеничного робота.

Інтегрувати систему керування рухом гусеничної платформи з візуальним наведенням через квадрокоптер із машинним зором.

Методи

У дослідженні застосовано комплексний методологічний підхід, що включає етапи моделювання, аналітичної верифікації, прототипування й експериментального тестування в умовах, наближених до реального середовища. Початково побудовано електричну принципову схему металощукача з урахуванням параметрів імпульсного генератора NE555, транзисторних ключів, компаратора LM393 та резонансних характеристик котушки. Виконано цифрову симуляцію форми сигналу з використанням програмного осцилографа на основі даних, отриманих у середовищі Arduino IDE, зокрема за допомогою Serial Plotter і зовнішніх інструментів аналізу.

Затухаючий коливальний сигнал сенсора:

$$v(t) = Ae^{-at} \cos(\omega_1 t + \varphi),$$

де A – амплітуда, a – коефіцієнт затухання, ω_1 – кругова частота сигналу, φ – початкова фаза.

Програмне забезпечення мікроконтролера реалізовано мовою C/C++ із використанням бібліотек аналого-цифрового перетворення. Пороговий аналіз вхідного сигналу на аналоговому піні A0 дозволив ідентифікувати імпульси, що генеруються в разі виявлення металу. Для уточнення логіки сформовано математичну модель фільтрації та підрахунку імпульсів на часовому інтервалі.

Порогова фільтрація сигналу:

$$S(t) = \begin{cases} 1, & \text{if } v(t) > T_{thr}, \\ 0, & \text{otherwise} \end{cases}$$

де $v(t)$ – вхідний сигнал (напруга) на аналоговому вході мікроконтролера у момент часу t . Це сигнал, сформований внаслідок індукції в котушці після імпульсу; T_{thr} – порогове значення напруги, яке встановлюється програмно (напр., 1.5–2.0 В) і визначає межу, вище якої сигнал вважається значущим; $S(t)$ – вихід логічного сигналу після порівняння. Може набувати двох значень:

111 – сигнал перевищив поріг → потенційна наявність металу,

000 – сигнал не перевищив поріг → метал не виявлено.

Умова виявлення металу:

$$S_f(t) > T_{metal}$$

де $S_f(t)$ – інтегрований пороговий вихід за певний інтервал часу.

Модель руху мобільної платформи побудовано на основі диференціальних кінематичних рівнянь для гусеничних роботів у координатах X-Y. Проведено оцінювання точності позиціонування на основі даних із комп'ютерного зору, отриманих із відео з квадрокоптера, та синхронізовано траєкторію робота із цілями, видимими в кадрі.

Кінематика руху в координатах X-Y:

$$\begin{cases} \dot{x}(t) = v(t) \cos(\theta(t)), \\ \dot{y}(t) = v(t) \sin(\theta(t)), \\ \dot{\theta}(t) = \omega(t), \end{cases}$$

де $v(t)$ – лінійна швидкість, $\omega(t)$ – кутова швидкість, $\theta(t)$ – орієнтація платформи.



Розроблено моделі для всіх основних підсистем: сенсорна частина (затухаючі коливання у котушці), цифрова обробка сигналу Arduino (порогова фільтрація), динаміка руху робота (в системі координат X-Y), передача координат з дрона (машинний зір), та автомат переходів станів (рух → сканування → виявлення).

Умова досягнення цілі:

$$d_t = \sqrt{(x - x_t)^2 + (y - y_t)^2} < \varepsilon,$$

де x, y – поточні координати мобільної платформи (гусеничного робота); x_t, y_t – координати цільової точки (цілі або зони сканування, напр., вказаної з квадрокоптера); d_t – фактична відстань між поточним положенням робота та заданою точкою; ε – допустима похибка (порогове значення), в межах якої вважається, що робот "досяг цілі".

Автомат переходів станів (логіка керування):

$$\text{State}(t) = \begin{cases} \text{Move to target,} & \text{if } o(t) = 1, \\ \text{Scan,} & \text{if } S_f(t) > T_{metal} \\ \text{Dodge obstacles,} & \text{if obstacle detected,} \\ \text{Wait for coordinates,} & \text{otherwise.} \end{cases}$$

Це поточний функціональний стан системи у момент часу t , який визначає, що саме виконує платформа:

- Move to target – рухається до вказаних координат;
- Scan – зупиняється й аналізує сигнал металошукача;
- Dodge obstacles – ухиляється від перешкоди;
- Wait for coordinates – очікує на оновлення даних.

Умови переходу між станами:

1. $o(t) = 1$: платформа отримала координати цілі (напр., із квадрокоптера) – переходить у стан "рух до цілі".
2. $S_f(t) > T_{metal}$: зафіксовано перевищення сигналу над порогом – робот зупиняється для сканування.
3. Obstacle detected: виявлено перешкоду – переходить до режиму обходу.
4. Інакше: якщо немає координат або подій – залишається в режимі очікування.

Для кожної моделі побудовано графіки, що демонструють поведінку сигналів і зміни станів у часі (рис. 1).

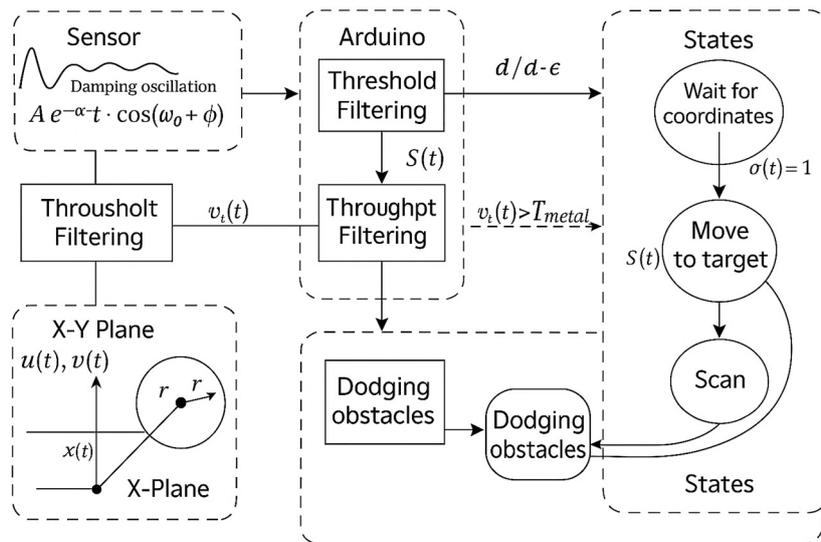


Рис. 1. Математична модель системи

Інтеграцію апаратного забезпечення з квадрокоптером виконано через віртуальне середовище емуляції (Processing, Gazebo), що дозволило протестувати обмін координатами та взаємодію між вузлами системи. Для тестування поведінки системи у разі виявлення ВНО реалізовано автомат переходів станів із програмним керуванням реакціями.

Побудовані графіки напруги та струму відображають характерні відмінності між станами "металу немає" та "метал виявлено", підтверджуючи коректність електричної частини. Графіки верифіковано шляхом порівняння теоретичних імпульсних характеристик з емпірично зафіксованими значеннями в лабораторному середовищі.

Реалізовано програмну логіку прийняття рішень у вигляді автомата керування, здатного виконувати дії в реальному часі у відповідь на виявлення металевих об'єктів.

Проведено експериментальну перевірку працездатності пристрою та оцінено точність, стабільність і адаптивність його роботи в умовах імітації польових ситуацій.

Будова та принцип роботи системи. Система побудована на основі імпульсного металошукача K158, що складається з генератора імпульсів (NE555), транзисторних ключів VT1/VT3, котушки L1 та приймального каскаду на LM393. Виявлення металевих об'єктів відбувається шляхом аналізу зворотних імпульсів у присутності металу. Без металу: напруга на SP1 близька до 0 В, струм відсутній. З металом: імпульси 3–5 В, частота близько 140 Гц, піковий струм до 625 мА. Ця система вмонтована в корпус гусеничного робота Keyestudio Tank, що забезпечує автономне переміщення місцевості (рис. 2).

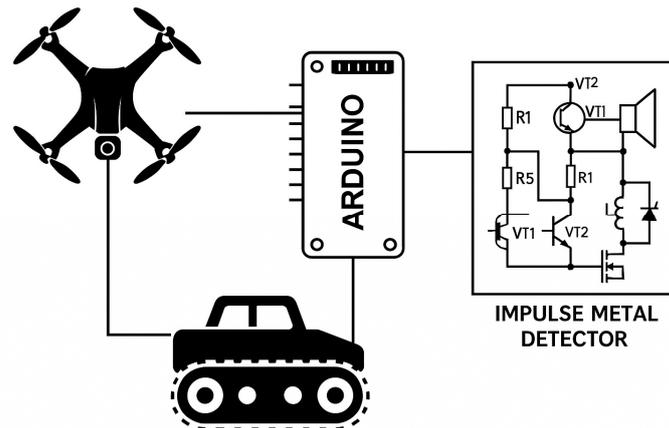


Рис. 2. Загальна структура досліджуваної системи

Металошукач підключено до мікроконтролера Arduino, встановленого на мобільному шасі. Сигнал із точки SP1 (вихід компаратора) зчитується на аналоговий вхід Arduino, який визначає наявність об'єкта за параметрами сигналу. Пристрій виконує роль виконавчого елемента: у разі виявлення ВНО передає координати та команду зупинки або об'їзду через бездротовий модуль. Платформа керується квадрокоптером, оснащеним камерою з комп'ютерним зором, який визначає маршрут і контролює зону огляду.

Блок-схема керуючого автомата. Сформовано блок-схему, яка ілюструє логіку роботи всієї системи: отримання координат із дрона, навігація робота, активація металошукача, оброблення сигналу, прийняття рішення. Також додано гілки для реакції на перешкоди (зміна напрямку руху) та логіку повернення до траєкторії після обходу. Це забезпечує адаптивність системи у реальних умовах місцевості (рис. 3).

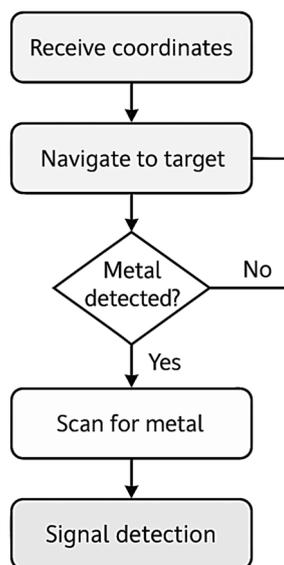


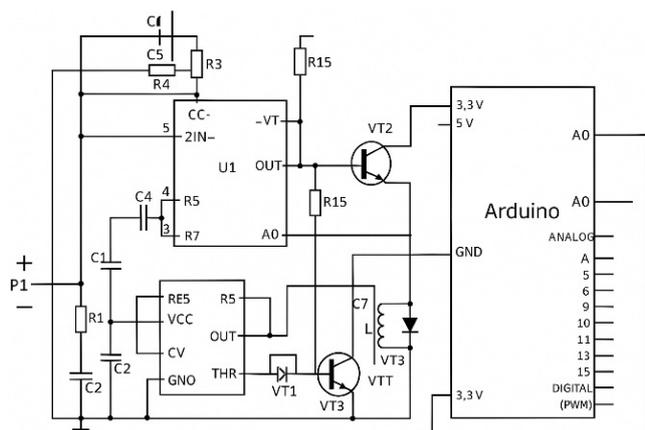
Рис. 3. Блок-схема керуючого автомата

Постановка експерименту. Досліджуваний прилад являє собою функціональне з'єднання імпульсного металошукача на базі модуля K158 із мікроконтролером Arduino Nano. Система реалізована на макетній платі та живиться від трьох акумуляторів типу 18650, з'єднаних послідовно, що забезпечує достатню напругу для стабільної роботи всіх компонентів.

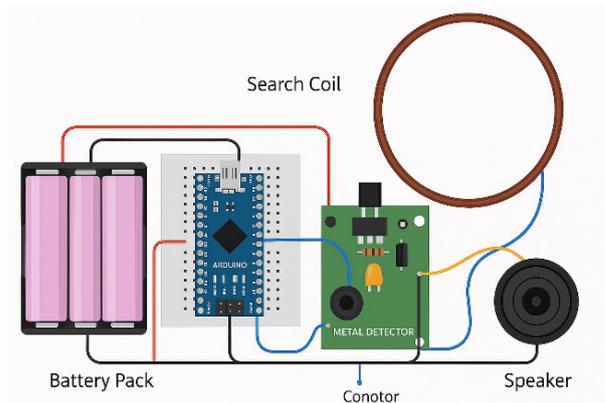
Модуль Arduino Nano встановлено на макетній платі. Його живлення подається безпосередньо з акумуляторного блока, через лінії VCC (червоний провід) та GND (чорний провід). З модуля металошукача до аналогового входу Arduino (A0) підведено сигнальний провід (синій), який передає оброблений імпульсний сигнал після компаратора LM393. Це дозволяє мікроконтролеру аналізувати виявлення металу програмно.

Пошукова котушка з мідного проводу під'єднана до плати металошукача. Її завдання – формувати електромагнітне поле й уловлювати його зміни у присутності металевих об'єктів. Вихідний каскад плати з'єднаний із динаміком (чорний і жовтий дроти), який генерує звуковий сигнал у разі виявлення металу (рис. 4).

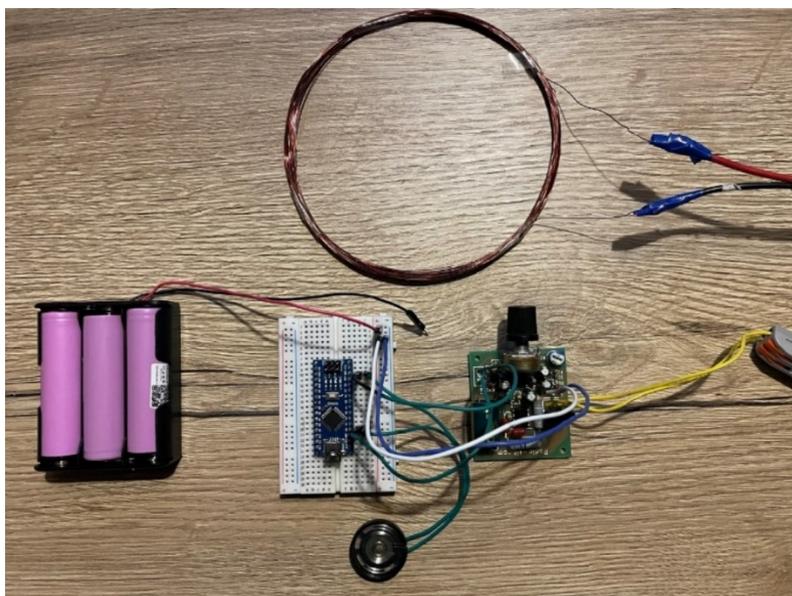
Кожен компонент схеми чітко виконує свою функцію: блок живлення – енергопостачання; металошукач – генерування і фіксація сигналів; Arduino – цифровий аналіз і можливе прийняття рішень. Така структура дозволяє легко інтегрувати схему в мобільну платформу для практичного застосування в умовах пошуку вибухонебезпечних предметів.



а



б



в

Рис. 4. Пристрій пошуку вибухонебезпечних предметів:
а – електрична схема, б – функціональна схема, в – зовнішній вигляд розробленого пристрою

Результати

В ході експериментального дослідження імпульсного металошукача, інтегрованого з мікроконтролером Arduino Nano, було отримано характеристики, зображені на рис. 5.

На першому графіку (рис. 5а) видно чіткі імпульсні сплески амплітудою до ~500 відліків аналого-цифрового перетворювача (АЦП), що різко виділяються на тлі фонових коливань (~300–350). Імпульси виникають коротко та різко, без тривалого затухання. Це типовий сигнал для дрібного металевих об'єкта з високою провідністю, розташованого дуже близько до центра котушки. Такий сигнал легко піддається пороговій фільтрації, і його можна використовувати як "еталонний" для класифікації дрібних цілей.

Другий графік (рис. 5б) демонструє високочастотну пілоподібну форму сигналу, із чіткими чергуваннями мінімумів і максимумів у діапазоні близько 310–390 відліків АЦП. Така поведінка характерна для великого об'єкта з високою провідністю, який щільно перекриває магнітне поле. Незважаючи на те, що алюміній має низьку магнітну проникність, його площа дозволяє генерувати стабільний індукційний зворотний сигнал. Зубчаста форма сигналу дозволяє ідентифікувати цей тип об'єкта за частотною характеристикою.

На третьому графіку (рис. 5в) спостерігається менша амплітуда (~300–400), але форма сигналу стає ширшою та менш стабільною. Імпульси не мають чіткої форми – деякі піки затуплені або розтягнуті. Це свідчить про слабший індукційний вплив об'єкта на магнітне поле, оскільки предмет міститься за межами оптимальної зони чутливості котушки. У такому випадку для впевненого виявлення бажано накопичувати статистику кількох імпульсів або аналізувати середнє відхилення.

На графіку (рис. 5г) видно серію імпульсів середньої амплітуди (~600–800), які розташовані нерівномірно, з окремими високими сплесками. Після кожного піка сигнал повертається до рівня близько 300–350. Це типовий вигляд сигналу для великого металевих об'єкта, який перебуває на межі чутливості пристрою. Важливо зазначити, що навіть



на такій значній відстані алюмінієвий предмет усе ще дає стабільні викиди, що свідчить про добру чутливість схеми до великих цілей. Однак для оброблення таких сигналів Arduino потребує або інтеграції значень у часі, або класифікації за фрагментами.



Рис. 5. Сигнатури досліджень металевих предметів:
а – дрібний сталевий предмет впритул, **б** – алюмінієвий предмет впритул,
в – дрібний сталевий предмет на відстані 4 см, **г** – великий металевий об'єкт на межі чутливості пристрою

На основі експериментальних даних, отриманих у процесі випробувань металошукача з імпульсною схемою на основі NE555, LM393 і Arduino Nano, можна зробити висновок, що система здатна виявляти дрібні металеві об'єкти на відстані до 2–3 см з амплітудою аналогового сигналу 450–520 відліків АЦП. У цьому режимі фіксується чіткий короточасний імпульс тривалістю 5–10 мс, що дає змогу точно застосувати просту порогову обробку без додаткової фільтрації. У разі збільшення відстані до 3–4 см, амплітуда сигналу зменшується до 300–400 відліків АЦП і спостерігається зміщення форми імпульсу, що потребує цифрового згладжування або кумулятивного обліку кількох сплесків для забезпечення достовірного виявлення.

При виявленні великих об'єктів система демонструє відчутну індукційну чутливість на відстанях до 20 см. Амплітуда сигналу при цьому досягає 750–800 відліків АЦП, форма сигналу має регулярний зубчастий характер із частотою коливань близько 12–15 Гц. Це дає змогу не лише виявити наявність об'єкта, а і здійснити типову класифікацію об'єкта за його розмірами та провідністю. Така форма сигналу є характерною для великих цілей із великою площею відбиття, виготовлених із провідних металів (алюміній, сплави на його основі).

Дискусія і висновки

Отримані сигнатури сигналів свідчать про чітку залежність форми й амплітуди відгуку імпульсного металошукача від геометричних розмірів об'єкта, його електропровідності та відстані до котушки.

Для підвищення надійності виявлення об'єктів різних класів рекомендовано впровадити двоступеневу систему обробки сигналу:

- пороговий аналіз з порогом $T_{thr} \approx 400T$ – для фіксації дрібних металевих об'єктів,
- інтегральний аналіз або фільтрація за кількістю імпульсів у часовому вікні (напр., ≥ 5 сплесків протягом 500 мс) – для великих або віддалених цілей,
- також рекомендовано зберігати еталонні форми сигналів у EEPROM або передавати їх по UART/LoRa/Wi-Fi для віддаленої класифікації. Варто зазначити, що пристрій стабільно працює при живленні від 3×18650 (12.6 В), а середній робочий струм у режимі сканування не перевищує 150–180 мА, що забезпечує автономність до 8 год.

З урахуванням зазначених вище результатів, система підтверджує свою практичну ефективність у задачах пошуку вибухонебезпечних об'єктів, дозволяючи виявляти як дрібні об'єкти з відстані до 3 см, так і великі металеві корпуси з дистанції до 20 см з імовірністю виявлення, що наближається до 95 %, для фонових рівнів помилкових спрацьовувань, менше 5 % за умови фіксованого порога.

У ході подальших досліджень будуть розроблені та застосовані методики пошуку металевих предметів на більшій глибині, а також практично досліджена концепція використання роботизованого наземного сегменту, що керується з квадрокоптера.



Внесок авторів. Андрій Дуднік – концептуалізація, постановка задачі дослідження; Андрій Фесенко – методологія, побудова математичної моделі; Олександр Торошанко – формальний аналіз, підбір технічних компонентів для експерименту; Віра Миколайчук – методологія, розроблення методу інтеграції імпульсного металошукача K158 з платформою Arduino; Олексій Батрак – інженерія, проектування та виготовлення імпульсного металошукача K158; Сергій Виговський – програмне забезпечення, валідація даних, проведення експерименту.

Джерела фінансування. Це дослідження не отримало жодного гранта від фінансової установи в державному, комерційному або некомерційному секторах.

Список використаних джерел

- Кузавков, В. (2022). Застосування методів технічного діагностування при розв'язанні завдань кібернетичного захисту. *Ukrainian Information Security Research Journal*, 24(1), 29–36. <https://doi.org/10.18372/2410-7840.24.16861>
- Мишук, В. В., & Фесенко, Г. В. (2024). Аналіз методів та засобів комп'ютерного зору для мобільних систем пошуку вибухонебезпечних предметів. *Elektronnoe Modelirovanie*, 46(1). <https://doi.org/10.15407/emodel.46.01.090>
- Невлюдов, І. Ш., Янушкевич, Д. В., Толкунов, І. О., Попов, І. І., & Іванець, Г. В. (2023). Обґрунтування необхідності створення робототехнічних комплексів для гуманітарного розминування. *Проблеми надзвичайних ситуацій*, 2(38), 17–38. <https://doi.org/10.52363/2524-0226-2023-38-2>
- Олексенко, О. О., Місюк, Г. В., Ікаєв, Д. Р., Коршок, В. М., & Палка, В. М. (2024). Основні тенденції застосування безпілотних літальних апаратів у російсько-українській війні. *Збірник наук. пр. Державного науково-дослідного інституту випробувань і сертифікації озброєння та військової техніки*, 3(21), 99–105. <https://doi.org/10.37701/dndivsovt.21.2024.12>
- Решетняк, М. С. (2024). *Розробка системи автоматизації керування металошукачем з використанням Arduino : пояснювальна записка до атестаційної роботи здобувача вищої освіти на першому (бакалаврському) рівні, спеціальність 151 – Автоматизація та комп'ютерно-інтегровані технології*. М-во освіти і науки України; Харків. нац. ун-т радіоелектроніки. <https://openarchive.nure.ua/handle/document/27821>
- Савін, В. (2024). *Удосконалення методу виявлення вибухонебезпечних об'єктів: кваліфікаційна робота*. Харківський національний університет радіоелектроніки. <https://openarchive.nure.ua/handle/document/25976>
- Солодчук, М., & Войтенко, В. (2022). Вибір бортового детектора об'єктів для БПЛА. У *Створення та модернізація ОБТ для потреб ЗС України* (с. 280–282). Лундський університет.
- Шовкошитний, І., & Василенко, О. (2024). Вибір показників оцінювання ефективності застосування роїв ударних безпілотних літальних апаратів для ураження нестационарних групових цілей. *Повітряна міць України*, 2(7), 61–72. <https://doi.org/10.33099/2786-7714-2024-2-7-61-72>
- Abdulmajeed, W. R., & Hussein, M. A. (2015). Factors effect on metal detecting system using mobile robot. *International Journal of Computer Applications*, 126(9), 43–46. <https://doi.org/10.5120/ijca2015906194>
- Jeyagopi, R., Chan, C., & Ma'arof, M. (2022). Design Metal Detecting Arduino Remote Control Robot Vehicle Controlled via Bluetooth. *Journal of Innovation and Technology*, 2022. <https://iuojs.intimal.edu.my/index.php/joit/article/view/287>
- Mujiarto, B. P., Sambas, A., & Haerudin, I. (2021). Design of Arduino-based metal detector robot. ResearchGate. https://www.researchgate.net/publication/355840656_Design_of_Arduino-Based_Metal_Detector_Robot
- Ravi Kiran, B., Padmini Rani, S., Venkatesh, S., Chandra Kanth, Y., Sasank Sai Krishna, M., & Siva Prasad, B. (2024). Metal detector robot using surveillance camera. *International Journal of Research Publication and Reviews*, 5(5). <https://ijrpr.com/uploads/V5ISSUE5/IJRPR27963.pdf>

References

- Abdulmajeed, W. R., & Hussein, M. A. (2015). Factors effect on metal detecting system using mobile robot. *International Journal of Computer Applications*, 126(9), 43–46. <https://doi.org/10.5120/ijca2015906194>
- Jeyagopi, R., Chan, C., & Ma'arof, M. (2022). Design metal detecting Arduino remote control robot vehicle controlled via Bluetooth. *Journal of Innovation and Technology*, 2022. <https://iuojs.intimal.edu.my/index.php/joit/article/view/287>
- Kuzavkov, V. (2022). Application of methods of technical diagnosis in solving problems of cybernetic protection. *Ukrainian Information Security Research Journal*, 24(1), 29–36 [in Ukrainian]. <https://doi.org/10.18372/2410-7840.24.16861>
- Mishchuk, V. V., & Fesenko, H. V. (2024). Analysis of Computer Vision Methods and Means for Explosive Ordnance Detection Mobile Systems. *Elektronnoe modelirovanie*, 46(1), 90–111 [in Ukrainian]. <https://doi.org/10.15407/emodel.46.01.090>
- Mujiarto, B. P., Sambas, A., & Haerudin, I. (2021). Design of Arduino-based metal detector robot. ResearchGate. https://www.researchgate.net/publication/355840656_Design_of_Arduino-Based_Metal_Detector_Robot
- Nevlyudov, I., Yanushkevich, D., Tolkunov, I., Popov, I., & Ivanets, H. (2023). Justification of the need to create modern robotic and technical complexes for humanitarian demining. *Problems of Emergency Situations*, 2(38), 17–38 [in Ukrainian]. <https://doi.org/10.52363/2524-0226-2023-38-2>
- Oleksenko, O., Misiuk, H., Ikaiev, D., Korshok, V., & Palka, V. (2024). MAIN TRENDS IN THE USE OF UNMANNED AIRCRAFT IN THE RUSSIAN-UKRAINIAN WAR. Scientific works of State Scientific Research Institute of Armament and Military Equipment Testing and Certification, 3(21), 99–105 [in Ukrainian]. <https://doi.org/10.37701/dndivsovt.21.2024.12>
- Ravi Kiran, B., Padmini Rani, S., Venkatesh, S., Chandra Kanth, Y., Sasank Sai Krishna, M., & Siva Prasad, B. (2024). Metal detector robot using surveillance camera. *International Journal of Research Publication and Reviews*, 5(5). <https://ijrpr.com/uploads/V5ISSUE5/IJRPR27963.pdf>
- Reshetnyak, M. (2024). *Development of a metal detector automation system using Arduino (Explanatory note to the certification work of a higher education applicant at the first (bachelor's) level)*. Ministry of Education and Science of Ukraine; Kharkiv. National University of Radio Electronics [in Ukrainian]. <https://openarchive.nure.ua/handle/document/27821>
- Savin, V. (2024). *Improvement of the method for detecting explosive objects: Qualification work*. OpenArchive NURE. Kharkiv National University of Radio Electronics [in Ukrainian]. <https://openarchive.nure.ua/handle/document/25976>
- Shovkoshytnyi, I., & Vasylenko, O. (2024). Selection of indicators for assessing the effectiveness of the use of swarms of striking unmanned aerial vehicles to defeat non-stationary group targets. *Air Power of Ukraine*, 2(7), 61–72 [in Ukrainian]. <https://doi.org/10.33099/2786-7714-2024-2-7-61-72>
- Solodchuk, M., & Voitenko, V. (2022). Selection of an onboard object detector for UAVs. In *Development and Modernization of Military Equipment for the Needs of the Armed Forces of Ukraine* (pp. 280–282). Lund University [in Ukrainian].

Отримано редакцією журналу / Received: 18.04.25
Прорецензовано / Revised: 10.05.25
Схвалено до друку / Accepted: 03.06.25



^{1,2}Andrii DUDNIK, DSc (Engin.), Assoc. Prof.
ORCID ID: 0000-0003-1339-7820
e-mail: andrii.dudnik@knu.ua

³Andrii FESENKO, PhD (Engin.), Assoc. Prof.
ORCID ID: 0000-0001-5154-5324
e-mail: andrii.fesenko@npp.kai.edu.ua

¹Olexandr TOROSHANKO, PhD (Engin.)
ORCID ID: 0000-0002-2354-0187
e-mail: oleksandr.toroshanko@knu.ua

¹Vira MYKOLAYCHUK, PhD
ORCID ID: 0000-0002-2532-5771
e-mail: viramykolaichuk@knu.ua

²Oleksiy BATRAK, PhD Student
ORCID ID: 0009-0007-8248-5935
e-mail: batrachelli@gmail.com

⁴Sergiy VYHOVSKYY, PhD Student
ORCID ID: 0009-0007-7868-4923
e-mail: vigovsky.sa@gmail.com

¹Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

²Higher Education Institution "Open International UNIVERSITY of Human Development «Ukraine»", Kyiv, Ukraine

³State Non-Commercial Company "State University «Kyiv Aviation Institute»", Kyiv, Ukraine

⁴Private Joint-Stock Company "Higher education institution «Interregional Academy of Personnel Management»", Kyiv, Ukraine

A SYSTEM FOR DETECTING EXPLOSIVE OBJECTS BASED ON A MOBILE TRACKED ROBOT CONTROLLED FROM A QUADCOPTER USING MASHING VISION

Background. *The results presented in this article mark the beginning of the project team's research in this direction. The experiments were conducted under laboratory conditions and focus on the general concept and identification of directions for further development. Future research is planned to be conducted in real or near-real conditions. The work proposes a system for detecting explosive objects based on the K158 pulse metal detector, a microcontroller, and the Keyestudio Mini Tank V3.0 mobile tracked platform. An integration with a quadcopter equipped with computer vision is proposed, enabling precise targeting and remote control. The system architecture, component interaction principles, decision-making logic, and a mathematical model of operation are described.*

Methods. *The research employed methods of functional subsystem modeling, computer-based oscillographic analysis of electrical signals, mathematical modeling of the tracked robot's kinematics in XY coordinates, and algorithmic design of a control automaton. To test the interaction with the quadcopter, virtual testing was conducted in a simulation environment using machine vision. Signal processing was implemented through microcontroller programming based on threshold filtering.*

Results. *The results demonstrate a clear differentiation between signals in the presence and absence of metal. In the background mode, a stable zero line is observed, while detection of a target produces a series of pulses at a frequency of 125–150 Hz and a current of up to 625 mA. This ensures reliable real-time object identification. All components, including the power and wireless communication modules, showed stable operation. Machine vision-based targeting achieved an accuracy of ± 20 cm. Overall, the system demonstrated reliability and efficiency in laboratory conditions.*

Conclusions. *The proposed system for detecting explosive objects combines the advantages of analog sensing, digital signal processing, a mobile platform, and aerial coordination. Its simplicity, autonomy, and low cost make it suitable for use in high-risk areas.*

Keywords: *explosive objects, Arduino, pulse metal detector, tracked mobile robot, quadcopter, computer vision, autonomous control, metal detection.*

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.



УПРАВЛІННЯ SDN-МЕРЕЖЕЮ ДАТА-ЦЕНТРІВ НА БАЗІ МОДИФІКОВАНОГО ПІДХОДУ DEVOFLOW

Вступ. Досліджено особливості реалізації рівня управління SDN-мереж під час обслуговування різних за обсягом потоків дата-центрів на основі технології Hedera, протоколу OpenFlow та підходу DevoFlow. Розподіл функцій оброблення, збору статистичних даних та періодичності оновлення статистичних даних впливають як на обсяг службової інформації, так і на ефективність та оперативність управління щодо розподілу вхідного трафіка для балансування навантаження. Метою роботи є модифікація підходу Hedera за рахунок перерозподілу функцій класифікації потоків та адаптивного управління залежно від обсягу потоку на рівні управління, а також розроблення алгоритму визначення допустимого часу надсилання статистичних даних від комутаторів на SDN-контролер.

Методи. Використано метод системного аналізу та декомпозиції для дослідження складних систем, методи збору даних про стан мережі, а також евристичні правила для визначення обсягу службового трафіка.

Результати. У результаті проведеного аналізу особливостей побудови і роботи таких технологій, як Hedera та DevoFlow, запропоновано модифіковану архітектуру побудови рівня управління та визначено основні функції програмних модулів, які дають змогу підвищити ефективність роботи SDN-контролерів. А саме, завдяки зміні порядку збору й оброблення службової інформації на обладнанні SDN-мережі дата-центрів і поверненню до централізованого типу управління.

Визначено підходи щодо класифікації вхідних потоків та їх розбиття на малі "mice-flows", середні "medium-flows" і великі потоки "elephant-flows", що дозволяє в подальшому використовувати багатошляхову маршрутизацію для ефективнішого використання ресурсів мережі. Також у цій статті запропоновано алгоритм визначення допустимого часу надсилання статистики від комутаторів (метод push-based), який враховує динаміку зміни навантаження в лініях у мережі, кількість активних сесій і базується на евристичному правилі обмеження обсягу службового трафіка.

Висновки. Розроблено модифіковану архітектуру рівня управління на базі підходу DevoFlow та визначено основні функції програмних модулів SDN-контролера мережі дата-центрів. Розроблено аналітичні залежності для визначення допустимого часу надсилання статистичних даних від комутаторів на SDN-контролер, які враховують поточний стан завантаженості мережі. А також проведено ряд експериментів для обґрунтування обраного методу отримання статистичних даних і підтвердження достовірності запропонованого алгоритму допустимого часу надсилання статистичних даних від комутаторів на SDN-контролер.

Ключові слова: Software-defined Networking, дата-центр, OpenFlow, Hedera, DevoFlow, elephant-flows, mice-flows, SDN-контролер, OpenFlow Switch, таблиця потоків.

Вступ

Сучасний розвиток мереж SDN вказує на широке застосування цієї технології для розгортання мережевої інфраструктури розподілених центрів обробки даних. Причому динамічна природа додатків і сервісів, а також різноманітність технологій транспортних мереж вимагають упровадження додаткових механізмів, що дозволяють ефективно використовувати наявні мережеві ресурси та забезпечувати необхідні показники якості обслуговування QoS.

Зазвичай основною перевагою SDN-мережі є централізований підхід щодо управління мережевими ресурсами у процесі розподілу різноманітних потоків трафіка за рахунок агрегації даних про поточний стан обладнання з урахуванням усієї топології мережі. Тобто, при розрахунку шляхів передачі (маршрутів) SDN-контролер має повнішу статистику про стан усіх елементів мережі (Globa et al., 2019, с. 76–100; Romanov, & Mankivskyi, 2019, с. 683–688).

Однак під час обслуговування потоків різних обсягів і часу тривалості сесії, виникають складності щодо їхнього розподілу між доступними шляхами передачі (оптимальності формування таблиць потоків), оскільки це напряму впливає на ефективність використання мережевого ресурсу та завантаження ліній передачі.

В роботах (Costa et al, 2021, vol. 147; Yu et al., 2018, с. 251–258), розглянуто такі рішення, як Hedera та Devolved OpenFlow (DevoFlow), які націлені на збільшення пропускної спроможності мережі дата-центру за рахунок класифікації (розбиття) вхідного трафіка на великі "elephant-flows" та малі "mice-flows" потоки з метою використання протоколів багатошляхової маршрутизації для балансування навантаження великих потоків. Але Hedera обмежується лише внутрішньою мережею дата-центру й аналізує обсяг потоків лише на граничних комутаторах (edge-switches). Особливість підходу DevoFlow (Costa et al., 2021, vol. 147) полягає також у розділенні потоків на "elephant-flows" і "mice-flows" та обробці великих потоків із залученням контролера, й обробці малих потоків на комутаторах без додаткового звернення до SDN-контролера.

Однак застосування моделі DevoFlow хоч і дає змогу зменшити час для визначення маршруту передачі та зменшує навантаження на сам контролер, але цей фактор може призвести до максимального завантаження окремих маршрутів (ліній передачі) і як наслідок – до перевантаження окремих сегментів мережі. Це пояснюється тим, що часткове перенесення рівня управління на комутатори (OpenFlow Switch) приводить до втрати повної картини щодо топології та стану функціонування мережі. Тому в цій статті пропонується модифікація підходу DevoFlow з перерозподілом функцій класифікації потоків та адаптивного управління залежності від обсягу потоку саме на SDN-контролері. Розглянемо детальніше наведені вище технології та особливості їхньої роботи.



Постановка задачі. Почнемо з технології Hedera, яку створено для ефективного використання пропускну здатності в центрі обробки даних. Реалізація Hedera полягає в періодичному збиранні статистики (кожні 5 секунд) із граничних комутаторів для виявлення "elephant-flows". Спочатку комутатори відправляють новий потік, використовуючи свої стандартні правила зіставлення потоків, по одному зі шляхів з однаковою вартістю. Якщо обсяг потоку досягає порогового значення, тоді він маркується як elephant-flow. Поріг, за замовчуванням, становить 10 % потужності мережевого інтерфейсу (NIC). У цей момент центральний планувальник Hedera використовує глобальну інформацію про стан мережі і розраховує найкращий шлях для потоку. Причому, для обчислення шляху передачі визначається його доступна пропускну здатність, щоб він міг вмістити потік.

Зазначимо, що DevoFlow являє собою модифікацію моделі OpenFlow, в якій відбувається деякий розрив між централізованим керуванням і централізованою видимістю стану всієї мережі. А саме, DevoFlow використовує агресивне використання існуючих OpenFlow-правил на комутаторах для "mice-flows", щоб зменшити обмін службовими повідомленнями між контролером і комутаторами. Отже, за відповідних умов, комутатори можуть приймати рішення про маршрутизацію малих за обсягом потоків локально, в той час як контролер здійснює загальне управління мережею і маршрутизує великі потоки.

Для розуміння запропонованих удосконалень та логіки роботи SDN-мережі зазвичай потрібно приділити увагу протоколу OpenFlow. Відповідно, OpenFlow є одним із ключових протоколів у концепції програмно-конфігурованих мереж (SDN), який дозволяє розділити площину керування та площину пересилання трафіка (Wang et al., 2023, с. 4377–4393). Основна ідея OpenFlow полягає в тому, що комутатори отримують інструкції від централізованого контролера, який визначає маршрутизацію пакетів на основі глобальної інформації про мережу (рис. 1). Це забезпечує гнучке управління потоками, спрощує налаштування політик і дає змогу централізовано контролювати трафік, що є критично важливим для сучасних дата-центрів і мереж із високими вимогами до якості обслуговування (QoS).

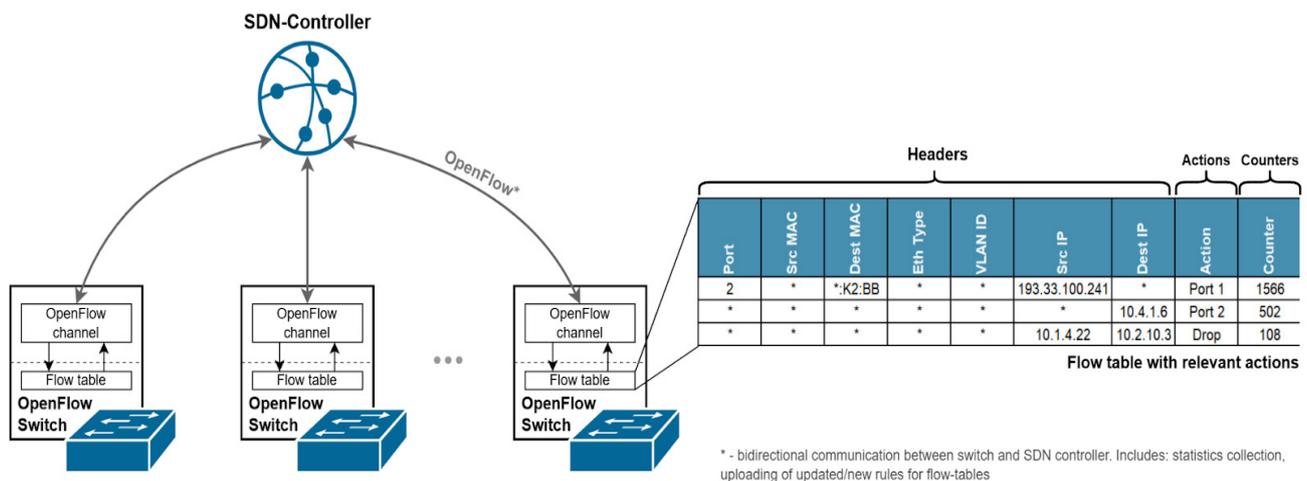


Рис. 1. Модель роботи протоколу OpenFlow і структура таблиці потоків

Перевагами централізованого контролю, базою для якого став підхід OpenFlow, є:

- глобальне управління політиками без складної конфігурації на рівні комутаторів, а точніше, навіть відсутність необхідності налаштовувати комутатор;
- можливість здійснювати оптимальну маршрутизацію потоків, особливо в мережах із гарантіями якості обслуговування (QoS) (Romanov et al., 2021, с. 33–40; Shirmarz, & Ghaffari, 2020, с. 7545–7593);
- гнучке управління потоками та безпека;
- можливість глобального бачення мережі, де контролер має "перед очима" весь процес передачі в мережі, що дозволяє завчасно реагувати на перевантаження, змінювати маршрути в реальному часі та, завдяки цьому, ефективніше використовувати мережеві ресурси. (насправді майже кожен SDN-контролер може задіяти процес динамічного переадресування потоків, на базі використання завжди оновлених даних про навантаження на мережевих з'єднаннях).

За використання традиційного OpenFlow, потоки можуть бути розділені на:

- безпекові (security-sensitive flows), які обробляються централізовано для підтримки безпеки;
- значущі потоки (significant flows), що керуються централізовано для покращення QoS та уникнення перевантажень у мережі;
- звичайні потоки (normal flows), які можуть бути оброблені локально комутаторами на базі заздалегідь прописаних правил, чи історично збереженими політиками передачі (ці дані очищуються досить швидко).

Однак, маючи досить суттєві переваги, що роблять мережу не дуже великих масштабів досить робочою і ефективною у використанні, централізована модель OpenFlow має суттєві недоліки, що серйозно обмежують її масштабованість (Tennakoon et al., 2018, с. 1043–1050).

По-перше, задіяння контролера для кожного нового сеансу передачі серйозно навантажує сам контролер. Наприклад, один NOX-контролер може обробляти до 30 000 нових потоків за секунду, але цього може бути недостатньо для великих за обсягом трафіка дата-центрів, із сотнями, а то й тисячами користувачів (Rout, Patra, & Sahoo, 2017, с. 543–551; Romanov, Nesterenko, & Mankivskiy, 2021, с. 159–182). Якщо взяти кластер, який має в підпорядкуванні 1500 серверів, то в такій мережі може надходити навіть до 100 000 нових потоків щосекунди, що вимагає, як мінімум, використання кількох контролерів для ефективної роботи мережі в цілому.



По-друге, маємо високі витрати на мережеву взаємодію. Використання контролера для налаштування потоків збільшує мережеві затримки в мережі. Наприклад, для N комутаторів потрібно створити та передати вдвічі більше ($2N$) записів у таблиці потоків, а також передати додаткові контрольні пакети, що, знову ж таки, створює додаткове службове навантаження.

По-третє, виникають обмеження продуктивності комутаторів, оскільки фізичні комутатори мають суттєві обмеження по ресурсах, які вводять в дію для оброблення потоків. Наприклад, традиційні комутатори можуть обробляти до 146 нових потоків за секунду, що може стати тим самим проблемним місцем (вузьким місцем / bottleneck) у великих за масштабом мережах.

Також варто зазначити, що OpenFlow-правила зберігаються в TCAM-пам'яті (Ternary Content Addressable Memory), яка є хоч дуже швидкою, але дорогою апаратною пам'яттю та використовує багато місця на ASIC (застосування цього підходу дозволяє збільшити продуктивність оброблення на апаратному рівні). Тобто OpenFlow вимагає більше ресурсів на кожен запис на відміну від наприклад, звичайного Ethernet трафіка, де використовується лише хеш-пам'яті і працює значно економніше. До порівняння, Ethernet forwarding потребує лише 60 біт на один запис (48-бітова MAC-адреса + 12 біт VLAN ID), дозволяючи зберігати до 64 000 записів (Rout, Patra, & Sahoo, 2017, с. 543–551). В той час як OpenFlow потребує 288 біт на правило, що суттєво зменшує можливу кількість таких записів до ~1500.

Відповідно правила OpenFlow генеруються та зберігаються на рівні кожного потоку, а саме для кожного нового клієнта, який ініціює з'єднання, необхідно створити окреме правило. Наприклад, середній ToR-комутатор (Top-of-Rack switch) може обслуговувати близько 10 000 потоків одночасно, що просто фізично перевищує кількість правил, які можна було б зберігати в TCAM-пам'яті. Тому жоден із наявних і запропонованих підходів OpenFlow не є ідеальним, оскільки у великих мережах дата-центрів цей процес може спричинити появу bottleneck на деяких ділянках мережі, саме тому сучасні рішення все частіше використовують або гібридні моделі, або розподілену аналітику, щоб мати баланс між продуктивністю й ефективністю (Gilliard et al., 2024, vol. 35).

Розроблення підходу DevoFlow націлено на часткове повернення рівня управління над потоками назад на комутатори, однак централізований контроль і видимість важливих потоків лишається на контролерах, що дає змогу зменшити навантаження на SDN-контролер. Зазначена технологія модифікує модель відкритого потоку у такий спосіб, аби перерозподілити якомога більше рішень на комутатори так, щоб їх можна було просто, економічно й ефективно реалізувати на апаратному рівні. Також це може дозволити зменшити кількість записів TCAM-пам'яті та надати нові механізми для ефективного виявлення QoS-значущих потоків.

Згідно з моделлю DevoFlow є кілька методів ідентифікації великих потоків:

- по таймерах на комутаторах (Timeouts). Якщо потік існує довше певного часу (напр., більше ніж 1 секунда), то він починає вважатися великим (elephant flow);
- по перевищенню заданого порогу обсягу переданих даних (Byte Threshold). Якщо потік передав більше зазначеного обсягу трафіка (напр., 10 МБайт), тоді такий потік маркується як elephant flow. Відповідно це й дозволяє фільтрувати маленькі запити HTTP, DNS і сфокусуватись на FTP, стрімінгу;
- вибірка або семплінг (Packet Sampling), де комутатор періодично аналізує пакети (напр., кожен 100-й пакет). Якщо для певного потоку фіксується велика кількість пакетів за короткий час, він маркується як elephant flow;
- тригери на основі затримки або перевантаження каналів зв'язку (Load-Based Triggers). Якщо канал мережі використовується більш ніж на 70 %, комутатор шукає потоки, які безпосередньо спричиняють таке навантаження, та класифікується як elephant flow.

Коли виявляється великий потік (elephant flow) вже контролер обчислює доступні шляхи і перенаправляє трафік по найменш завантаженому альтернативному маршруту. Механізми виявлення потоків, що використовуються в Hedera та DevoFlow, мають досить високі накладні витрати як апаратних так і мережевих ресурсів, що особливо критично для менш продуктивного обладнання рівня інфраструктури SDN-мережі.

Методи

З огляду на зазначене вище в цій роботі, на основі методів аналізу складних систем, тобто аналізу порядку взаємодії базових елементів SDN-мережі за різними технологіями та протоколами, була запропонована архітектура модифікованого підходу DevoFlow. Також за рахунок декомпозиції визначено функціональні можливості програмних модулів, які дозволяють обробляти дані про поточний стан мережі та здійснювати класифікацію вхідних потоків за обсягом із метою балансування навантаження й уникнення аварійних ситуацій.

Зауважимо, що існують проблеми із збиранням статистики потоків, оскільки модель OpenFlow підтримує два методи отримання статистики:

- перший із них push-based, де контролер отримує повідомлення про всі зміни (тобто комутатори автоматично надсилають статистику на контролер, щойно в потоці відбувається певна подія), що створює суттєве навантаження в мережі.
- другий метод – pull-based, в якому контролер періодично запитує дані, що знижує навантаження мережі, але разом із цим падає ефективність збору статистики через свою неоперативність і затримки (оскільки дані оновлюються не миттєво).

Реалізація функції збору статистики напряму пов'язана з кількома ключовими обмеженнями – як з боку мережевих пристроїв (комутаторів), так і з боку самого контролера. Якщо розглядати метод push-based, то кожен новий потік, що надходить, або зміна стану потоку (таймери, лічильники) активує надсилання службових повідомлень до контролера. Для додаткового оброблення такого обсягу службового трафіка контролер потребує високої продуктивності CPU та достатньої кількості пам'яті.

Якщо говорити про pull-based метод, то запити контролером охоплюють усі активні потоки (навіть ті, в яких нічого не змінилося). Відповідно, комутатори повинні формувати відповіді по кожному потоку, обробляючи ці додаткові операції на своїх і так досить обмежених обчислювальних ресурсах (ASIC або CPU). А такі відповіді займають значний обсяг пам'яті та пропускної здатності каналів мережі, особливо це критично в мережах дата-центрів із тисячами потоків.

Як результат, збір статистики займає або занадто багато ресурсів, що може знижувати продуктивність мережі, або збільшення часу періоду опитування зменшує ефективність роботи контролера у реальному часі у випадку



перенавантажень. Тому жоден із наявних і запропонованих підходів OpenFlow не є ідеальним, оскільки у великих мережах дата-центрів цей процес може спричинити появу bottlenecks на деяких ділянках мережі, саме тому сучасні рішення все частіше використовують або гібридні моделі, або розподілену аналітику, щоб мати баланс між продуктивністю й ефективністю.

Відповідно, у розробленні алгоритму визначення допустимого часу надсилання статистики від комутаторів до SDN-контролера, на основі методу push-based, були використані евристичні правила обслуговування службового трафіка з урахуванням поточних показників мережі.

Результати

На відміну від запропонованих рішень в роботах (Costa et. al., 2021, vol. 147; Yu et. al., 2018, с. 251–258), щодо порядку роботи мережевого обладнання на основі DevoFlow, в цій статті пропонується змінити порядок збору й оброблення службової інформації про характеристики потоків. Тобто комутатори (OpenFlow Switch) в загальному працюють так:

- комутатори не приймають рішень, щодо розрахунку шляхів для передачі даних, а здійснюють комутацію пакетів лише на основі таблиць потоків;

- комутатори збирають статистичні дані (напр., кількість переданих пакетів або байтів) для кожного потоку, використовуючи відповідні поля в таблицях потоків (counters);

- комутатори самостійно відправляють дані на контролер (push-based). Поки сесія передачі потоку активна (на період передачі), статистичні дані на лічильниках (counters) постійно оновлюються, а їхні значення не обмежуються.

Також на SDN-контролер покладають такі завдання:

- збір, узагальнення та підтримка актуальних статистичних даних за кожен потік;

- збір, узагальнення та підтримка актуальних даних про поточну топологію мережі (лінії та активне мережеве обладнання);

- збір, узагальнення та підтримка актуальних даних про поточний стан функціонування всієї мережі (напр., розрахунок завантаження ліній зв'язку);

- класифікація потоків. Тобто, обрахунок обсягів потоків і присвоєння відповідним потокам міток: малий "micro-flows", середній "medium-flows" і великий потік "elephant-flows". У свою чергу, потік – це сесія передачі даних між відправником та отримувачем у мережі.

- оперативний перерозподіл навантаження для значущих потоків (напр., після ідентифікації elephant flow перерахунок і визначення альтернативних шляхів передачі);

- передача отриманих у процесі розрахунків маршрутної інформації OpenFlow-правил на комутатори та їх подальше корегування залежно від одержаних поточних статистичних даних.

На рис. 2 представлено архітектуру й основні компоненти SDN-мережі згідно з удосконаленим підходом DevoFlow.

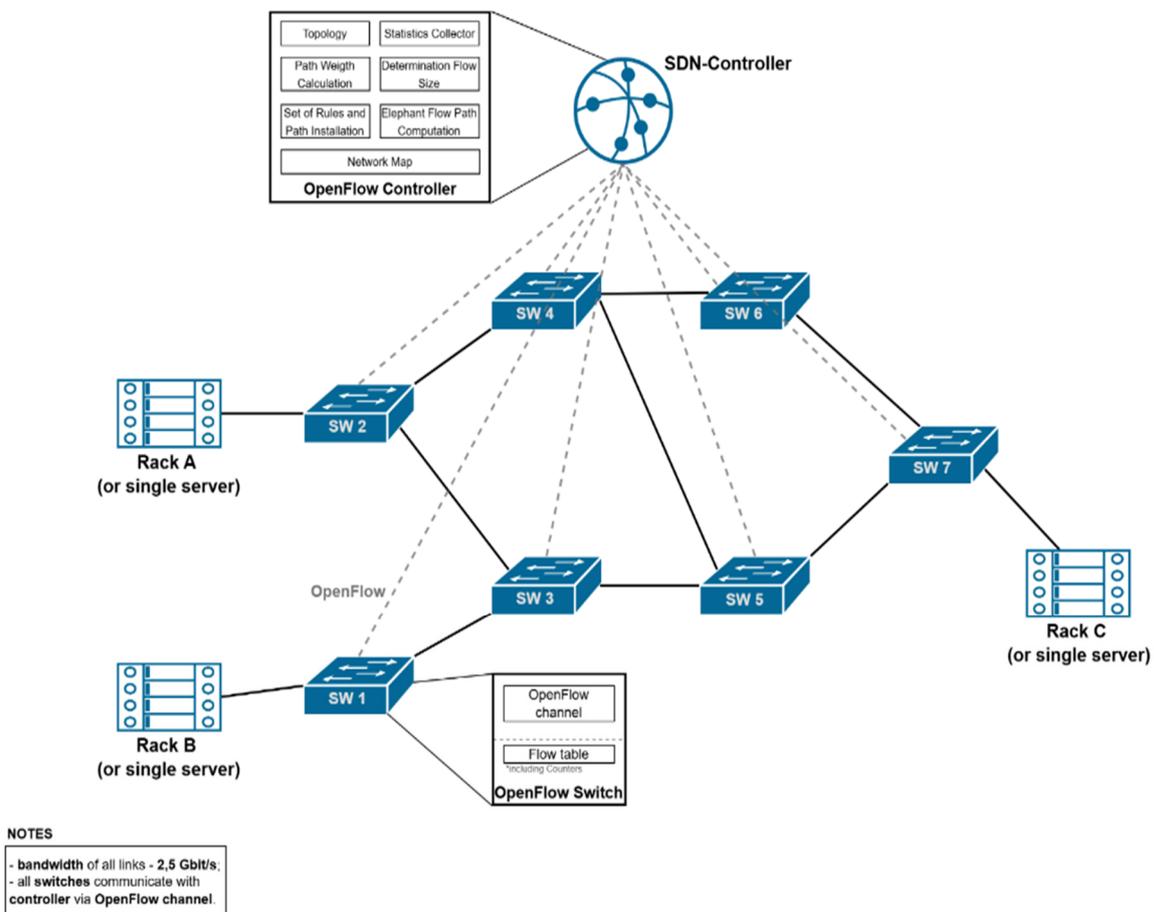


Рис. 2. Варіант архітектури SDN-мережі дата-центру на основі вдосконаленого підходу DevoFlow

Центральним компонентом зазвичай є SDN-контролер, що включає:

- програмно-функціональний блок топології (Topology), який зберігає інформацію про всі наявні в мережі вузли (комутатори) та канали;
- блок збору статистики (Statistics collector), який відстежує завантаження каналів, періодично збираючи (агрегуючи) інформацію про завантаженість портів комутатора;
- блок обчислення ваги шляху (Path weight calculation), який розраховує найкоротші шляхи передачі для вхідних потоків між будь-якою парою джерело-одержувач;
- блок визначення розміру потоку (Determination flow size), який розраховує (класифікує) розмір потоку залежно від порогових значень та обраних параметрів ідентифікації;
- блок обчислення шляхів для великих "elephant flow" потоків (Elephant flow path computation), який розраховує шляхи для потоків "elephant flow" у процесі використання алгоритмів багатошляхової маршрутизації наприклад ECMP;
- блок пересилання та контролю доставки таблиць потоків (Set of rules and path installation), який відправляє правила передачі трафіка (flow table) на комутатор;
- блок контролю функціонування мережі (Network map) – узагальнена інформація про поточний стан усієї мережі (напр., поточна матриця циркулюючого трафіка, прокладені (актуальні) маршрути тощо).

Тепер розглянемо в динаміці застосування багатошляхового протоколу ECMP для балансування навантаження у мережі у випадку ідентифікації "elephant flow". На рис. 3 представлено приклад розподілу навантаження з урахуванням запропонованих модифікацій.

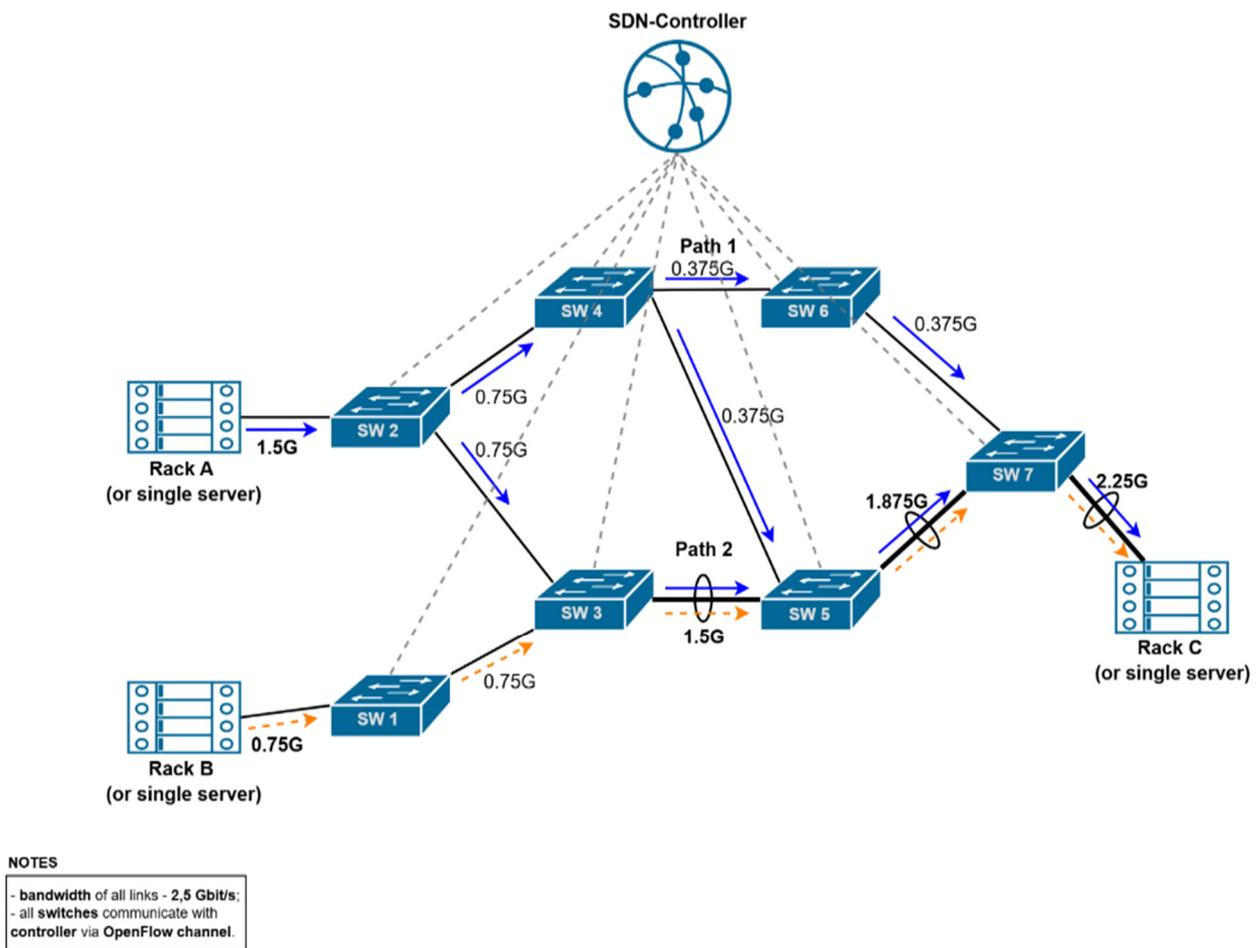


Рис. 3. Розподіл навантаження для модифікованого DevOfFlow для великих потоків (elephant flows) на основі протоколу ECMP

Згідно з прикладом і наведеною топологією, відбувається передача трафіка із двох серверів (із сервера А зі швидкістю 1,5 Гбіт/с та із сервера В зі швидкістю 750 Мбіт/с) на сервер С, бітова швидкість усіх каналів зв'язку в мережі – 2,5 Гбіт/с. Особливість ECMP полягає в тому, що на кожному комутаторі відбувається рівноцінний (навантаження ділиться навпіл) розподіл вхідного трафіка між шляхами передачі з однаковою ціною (приймаємо однаково кількість транзитів), якщо ціна одного шляху менша за ціну інших, то розподіл навантаження не відбувається.

Під час передачі трафіка із сервера В на С, його маршрут буде без будь-яких розподілів, оскільки найвигіднішим (мінімальною кількістю транзитів) є маршрут через:

SW1 – SW3 – SW5 – SW7.



За передачі трафіка із сервера А до сервера С, існує три рівноцінні маршрути, а саме:

SW2 – SW4 – SW6 – SW7 (Path 1);

SW2 – SW3 – SW5 – SW7 (Path 2);

SW2 – SW4 – SW5 – SW7.

При надходженні потоку на комутатор SW2, контролер має відомості, що є два рівноцінні шляхи передачі (через SW3 та SW4) і тому рівномірно розподіляє трафік між лініями зв'язку цих комутаторів із метою балансування навантаження. Далі розглянемо окремо випадки передачі шляхами 1 та 2:

Path 1. Потік надходить на SW4, контролер має відомості, що має 2 рівноцінні шляхи передачі (через SW5 та SW6) і тому рівномірно розподіляє вхідний трафік між інцидентними лініями цих комутаторів. Частина трафіка після попереднього поділу через SW6 та SW7 доставляється до сервера С.

Path 2. Потік, який проходить по даному шляху об'єднується з потоком від сервера В (в лініях зв'язку між SW3 – SW5, а також SW5 – SW7). Зазвичай потоки об'єднуються на лінії між SW7 і сервером С.

Подальше вдосконалення спрямовано на розроблення аналітичної залежності (алгоритму) для визначення мінімально допустимого часу надсилання статистичних даних (поточне значення лічильників) від комутаторів на контролер. Загальний алгоритм розрахунку допустимого періоду надсилання такий:

- аналізуються поточні значення швидкості на всіх лініях зв'язку $\|V_{linkij}\|$ мережі й обирається найменша швидкість передачі $\min(V_{link})$;
- від визначеної мінімальної швидкості розраховується значення, що становить не більше 10 %, на передачу службової інформації, причому отримане значення ділиться на кількість активних сесій (NAS): $V_{threshold} \leq \left[\frac{(V_{link}) \times 0,1}{NAS} \right]$;
- розраховується допустимий час ($T_{allowable}$) надсилання статистики від комутаторів до SDN-контролера згідно з аналітичним виразом:

$$T_{allowable} = \frac{1}{\frac{v_{threshold}}{L_{frame}}}, \quad (1)$$

де L_{frame} – обсяг (розмір) кадру, вміст якого включає службову інформацію, тобто необхідні значення лічильників.

Для проведення експериментів і верифікації запропонованих рішень використано програмний продукт MiniEdit для створення імітаційної моделі сегменту SDN-мережі. По суті, MiniEdit являє собою графічний редактор, в якому користувачу надається можливість власноруч побудувати свою мережу, без прямої роботи з кодом, оскільки більшість параметрів можна налаштувати із самого редактора. MiniEdit є частиною програмного продукту Mininet, що представляє собою віртуальну платформу для емуляції сегментів SDN-мережі, де можна використовувати віртуальні хости, OpenFlow комутатори та контролери (як вбудовані, що завантажуються за замовчуванням із пакетами Mininet, так і з можливістю підключення зовнішніх різних розробників).

Вихідними даними щодо створення імітаційної моделі стала топологія з мережевими параметрами, яка представлена на рис. 3.

Для початку виконуємо побудову нашого сегменту SDN-мережі шляхом наповнення мережевого "полотна" в MiniEdit елементами мережі (контролер, віртуальні OpenFlow комутатори, віртуальні хости). В нашому випадку для емуляції роботи кластерів серверів (Rack А, В та С) використано віртуальні хости (h1, h2, h3). Як базовий, був використаний зовнішній контролер від Hewlett Packard (HPE VAN SDN Controller). Скомпонована схема SDN-мережі для проведення експериментів виглядатиме таким чином (рис. 4).

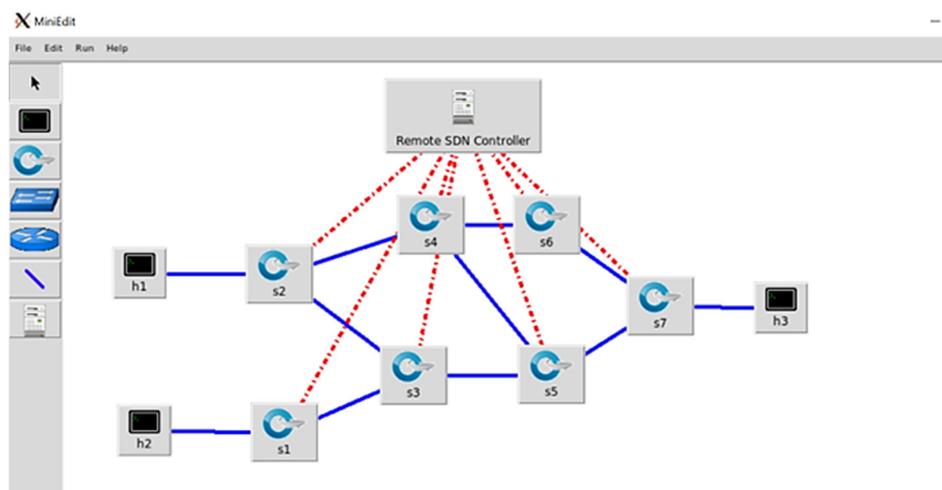


Рис. 4. Візуалізація топології сегменту SDN-мережі дата-центру в MiniEdit

Наступним етапом налаштування є конфігурація контролера для підключення до Mininet (де розгорнута топологія мережі: OpenFlow комутатори, хости, лінії зв'язку), а також налаштування необхідних мережевих параметрів у лініях зв'язку (delay, bandwidth та ін.), що показано на рис. 5.

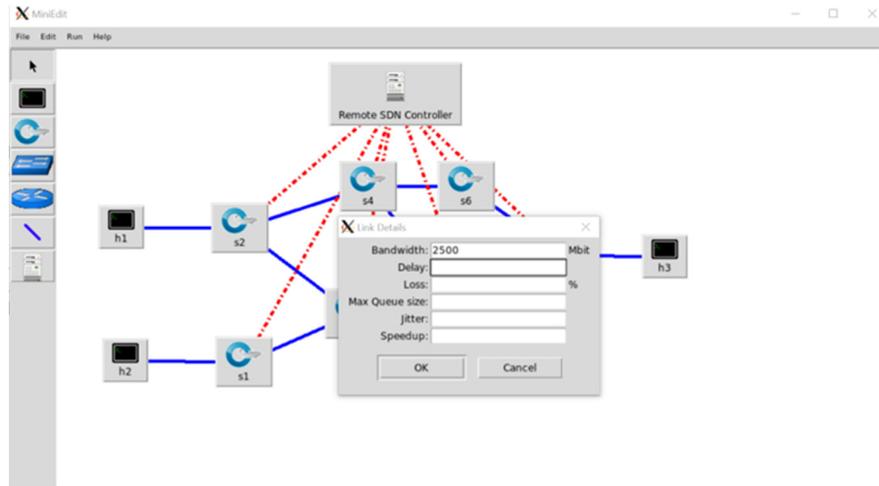


Рис. 5. Порядок налаштування параметрів ліній зв'язку в MiniEdit

В результаті проведених експериментів у процесі зміни величини навантаження в мережі та використанні різних методів збору статистичних даних про характеристики потоків, була зібрана статистика, що представлена на рис. 6. А саме, порівнювали такі методи: метод pull-based (контролер відправляє запит на комутатор із метою отримати статистичні дані) та метод push-based (комутатор самостійно з деякою періодичністю надсилає статистичні дані на контролер).

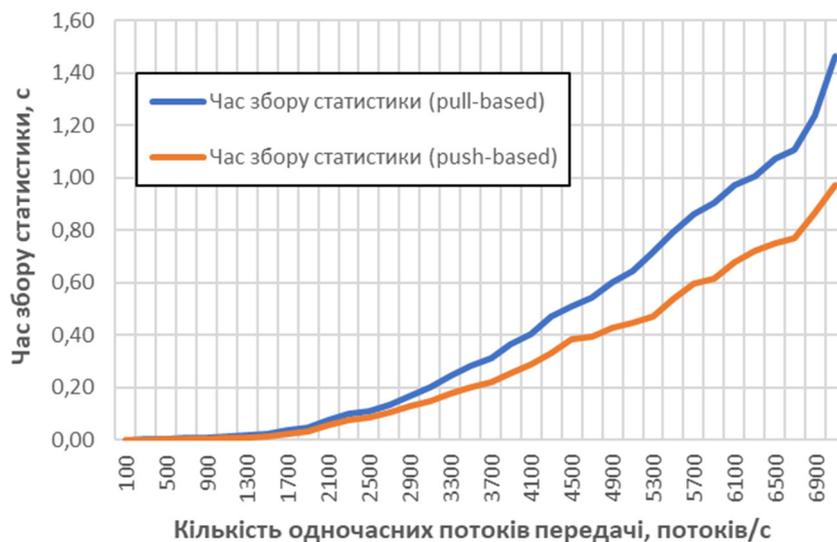


Рис. 6. Залежність часу отримання статистики від розмірності таблиць потоків для методів pull- та push-based

Аналіз отриманих залежностей свідчить про те, що зі збільшенням навантаження в мережі метод push-based демонструє кращі результати, щодо часу отримання статистики, ніж метод pull-based, та відповідно дозволяє отримувати частіше актуальну інформацію про стан мережі. Відповідно push-based метод пропонується використовувати на рівні управління в модифікованій архітектурі на базі підходу DevonFlow.

Подальші експерименти на імітаційній моделі сегменту мережі SDN (див. рис. 4) були направлені на перевірку запропонованої аналітичної моделі допустимого часу надсилання статистики на контролер (аналітичний вираз 1).

На рис. 7 графічно узагальнено отримані експериментальні дані порівняно з аналітичною моделлю для однакових вхідних даних.

Згідно з аналізом результатів проведених експериментів (рис. 6, 7) можемо відмітити таке:

- метод отримання статистичних даних push-based є в рази швидшим за метод pull-based, що підтверджується експериментом на імітаційній моделі SDN-мережі при емуляції різних рівнів навантаження. Посилаючись на залежності на рис. 6, можна побачити, що при передачі 4500 потоків за секунду в мережі, час збору статистики методом pull-based становить 0,51 секунди, в той час як час збору методом push-based – 0,38 секунди, що приблизно в 1,3 раза швидше;

- результати експерименту, зображені на рис. 7, підтверджують достовірність запропонованої аналітичної моделі допустимого часу надсилання статистики на контролер (аналітичний вираз 1), а сама модель враховує залежність зміни інтенсивності надсилання статистичних даних від завантаженості елементів мережі.

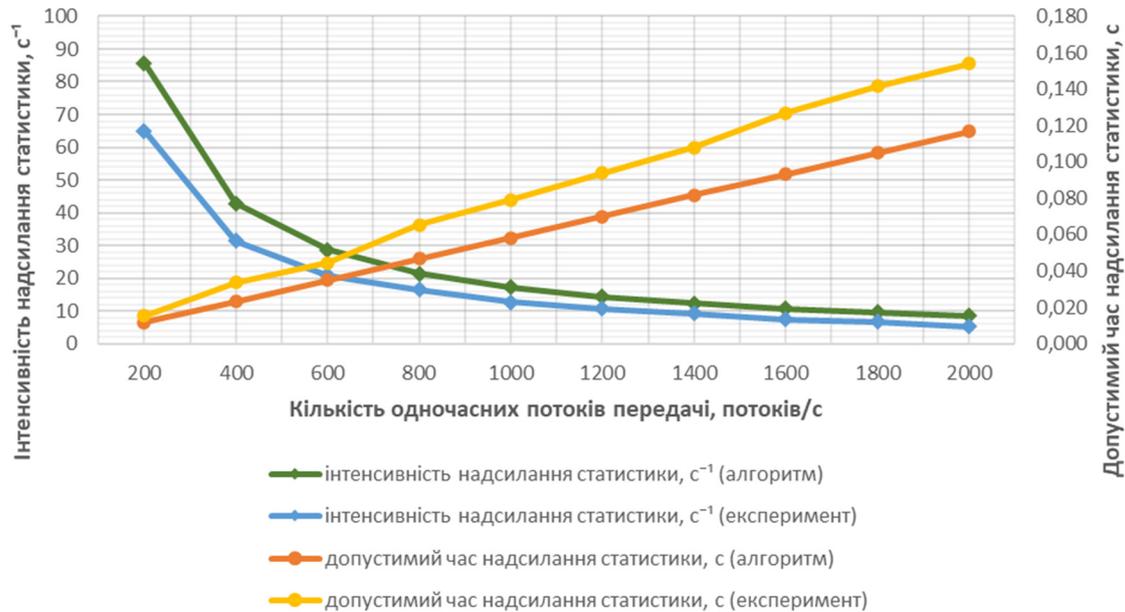


Рис. 7. Залежність часу й інтенсивності надсилання статистики від кількості потоків у мережі

Дискусія і висновки

В результаті дослідження проведено аналіз переваг і недоліків технологій Hedera та DevoFlow. На основі переваг запропоновано модифіковану архітектуру рівня управління з урахуванням підходів DevoFlow. Це дозволило здійснити декомпозицію та визначити функціональні можливості програмних модулів SDN-контролерів дата-центрів для підвищення ефективності оброблення та розподілу вхідних потоків.

Для отримання статистики (службової інформації) з обладнання SDN-мережі дата-центрів запропоновано використовувати push-based метод, який показує кращі результати щодо оперативності збору статистики за колювання навантаження порівняно з методом pull-based.

В подальшому, на основі отриманих статистичних даних від комутаторів пропонується здійснювати класифікацію вхідних потоків та їх розбиття на малі "mice-flows", середні "medium-flows" та великі потоки "elephant-flows". Це дає змогу використовувати різні стратегії оброблення даних потоків для рівномірного завантаження елементів мережі і як наслідок – зменшення ймовірності появи перенавантажень.

У цій статті запропоновано алгоритм визначення допустимого часу надсилання статистики від комутаторів (метод push-based), який враховує динаміку зміни навантаження в лініях в мережі, кількість активних сесій і базується на евристичному правилі обмеженні обсягу службового трафіка.

Для перевірки достовірності запропонованих рішень розроблено імітаційну модель у графічному емуляторі MiniEdit і проведено серію експериментів. Аналіз результатів моделювання підтверджує використання push-based методу як основного для отримання статистики контролером, а також достовірність розроблених аналітичних залежностей для визначення допустимого часу надсилання статистичних даних на контролер.

Подальші дослідження будуть направлені на визначення впливу на мережу різних за обсягом і терміном функціонування потоків ("mice-flows", "medium-flows" "elephant-flows") і розроблення методів (моделей) збалансованого розподілу навантаження.

Внесок авторів. Микола Нестеренко – розроблення архітектури управління на базі модифікованого підходу DevoFlow та аналітичних залежностей для визначення допустимого часу надсилання статистичних даних на контролер, проведення аналізу і порівняння існуючих підходів і технологій, написання висновків; Антон Марінов – огляд літературних джерел, збір теоретичних даних і результатів досліджень, опис балансування навантаженням в мережі модифікованого підходу.

Джерела фінансування. Це дослідження не отримало жодного гранта від фінансової установи в державному, комерційному або некомерційному секторах.

Список використаних джерел

- Costa, L. C., Vieira, A. B., de Brito e Silva, E., Macedo, D. F., Vieira, L. F. M., Vieira, M. A. M., da Rocha Miranda, M., Batista, G. F., Polizer, A. H., Gonçalves, A. V. G. S., Gomes, G., & Correia, L. H. A. (2021). OpenFlow data planes performance evaluation. *Performance Evaluation*, 147. <https://doi.org/10.1016/j.peva.2021.102194>
- Gilliard, E., Liu, J., Aliyu, A. A., Juan, D., Jing, H., & Wang, M. (2024). Intelligent load balancing in data center software-defined networks. *Transactions on Emerging Telecommunications Technologies*, 35(4). <https://doi.org/10.1002/ett.4967>
- Globa, L., Skulysh, M., Romanov, O., & Nesterenko, M. (2019). Quality Control for Mobile Communication Management Services in Hybrid Environment. Y M. Ilchenko, L. Uryvsky, & L. Globa (Eds.), *Advances in Information and Communication Technologies: Vol. 560. Lecture Notes in Electrical Engineering* (pp. 76–100). Springer. https://doi.org/10.1007/978-3-030-16770-7_4
- Romanov, O., & Mankivskyi, V. (2019). Optimal Traffic Distribution Based on the Sectoral Model of Loading Network Elements. In *2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T)* (pp. 683–688). IEEE. <https://doi.org/10.1109/picst47496.2019.9061296>
- Romanov, O., Nesterenko, M., & Mankivskyi, V. (2021). The method of redistributing traffic in mobile network. In D. Ageyev, T. Radivilova, & N. Kryvinska (Eds.), *Data-Centric Business and Applications* (Lecture Notes on Data Engineering and Communications Technologies, Vol. 69, pp. 159–182). Springer. https://doi.org/10.1007/978-3-030-71892-3_7
- Romanov, O., Siemen, E., Nesterenko, M., & Mankivskyi, V. (2021). Mathematical Description of Control Problems in SDN Networks. In *International Conference on Applied Innovations in IT (ICAIIIT)* (pp. 33–40). <http://dx.doi.org/10.25673/36582>.



- Rout, S., Patra, S.S., Sahoo, B. (2017). Performance Evaluation of the Controller in Software-Defined Networking. У H. Behera, & D. Mohapatra (Eds.) *Computational Intelligence in Data Mining*: Vol 556. *Advances in Intelligent Systems and Computing* (c. 543–551). Springer. https://doi.org/10.1007/978-981-10-3874-7_51
- Shirmar, A., & Ghaffari, A. (2020). Performance issues and solutions in SDN-based data center: a survey. *The Journal of Supercomputing*, 76(10), 7545–7593. <https://doi.org/10.1007/s11227-020-03180-7>
- Tennakoon, D., Chowdhury, M., & Luan, T. H. (2018). Cloud-based load balancing using double Q-learning for improved Quality of Service. *Wireless Netw*, 29, 1043–1050. <https://doi.org/10.1007/s11276-018-1888-8>
- Wang, C., Cao, W., Hu, Y., & Liu, J. (2023). Data Center Traffic Scheduling Strategy for Minimization Congestion and Quality of Service Guaranteeing. *Computers, Materials & Continua*, 75(2), 4377–4393. <https://doi.org/10.32604/cmc.2023.037625>
- Yu, H., Qi, H., Li, K., Zhang, J., Xiao, P., & Wang, X. (2018). Openflow Based Dynamic Flow Scheduling with Multipath for Data Center Networks. *Computer Systems Science and Engineering*, 33(4), 251–258. <https://doi.org/10.32604/csse.2018.33.251>

References

- Costa, L. C., Vieira, A. B., de Brito e Silva, E., Macedo, D. F., Vieira, L. F. M., Vieira, M. A. M., da Rocha Miranda, M., Batista, G. F., Polizer, A. H., Gonçalves, A. V. G. S., Gomes, G., & Correia, L. H. A. (2021). OpenFlow data planes performance evaluation. *Performance Evaluation*, 147. <https://doi.org/10.1016/j.peva.2021.102194>
- Gilliard, E., Liu, J., Aliyu, A. A., Juan, D., Jing, H., & Wang, M. (2024). Intelligent load balancing in data center software-defined networks. *Transactions on Emerging Telecommunications Technologies*, 35(4). <https://doi.org/10.1002/ett.4967>
- Globa, L., Skulysh, M., Romanov, O., & Nesterenko, M. (2019). Quality Control for Mobile Communication Management Services in Hybrid Environment. In Ilchenko, M., Uryvsky, L., Globa, L. (Eds.), *Advances in Information and Communication Technologies*: Vol. 560. *Lecture Notes in Electrical Engineering* (pp. 76–100). Springer. https://doi.org/10.1007/978-3-030-16770-7_4
- Romanov, O., & Mankivskiy, V. (2019). Optimal Traffic Distribution Based on the Sectoral Model of Loading Network Elements. In *2019 IEEE International Scientific-Practical Conference Problems of Infocommunications, Science and Technology (PIC S&T)* (pp. 683–688). IEEE. <https://doi.org/10.1109/picst47496.2019.9061296>
- Romanov, O., Nesterenko, M., & Mankivskiy, V. (2021). The method of redistributing traffic in mobile network. In D. Ageyev, T. Radivilova, & N. Kryvinska (Eds.), *Data-Centric Business and Applications* (Lecture Notes on Data Engineering and Communications Technologies, Vol. 69, pp. 159–182). Springer. https://doi.org/10.1007/978-3-030-71892-3_7
- Romanov, O., Siemen, E., Nesterenko, M., & Mankivskiy, V. (2021). Mathematical Description of Control Problems in SDN Networks. In *International Conference on Applied Innovations in IT (ICAIIIT)* (pp. 33–40). <http://dx.doi.org/10.25673/36582>
- Rout, S., Patra, S.S., Sahoo, B. (2017). Performance Evaluation of the Controller in Software-Defined Networking. У H. Behera, & D. Mohapatra (Eds.) *Computational Intelligence in Data Mining*: Vol 556. *Advances in Intelligent Systems and Computing* (c. 543–551). Springer. https://doi.org/10.1007/978-981-10-3874-7_51
- Shirmar, A., & Ghaffari, A. (2020). Performance issues and solutions in SDN-based data center: a survey. *The Journal of Supercomputing*, 76(10), 7545–7593. <https://doi.org/10.1007/s11227-020-03180-7>
- Tennakoon, D., Chowdhury, M., & Luan, T. H. (2018). Cloud-based load balancing using double Q-learning for improved Quality of Service. *Wireless Netw*, 29, 1043–1050. <https://doi.org/10.1007/s11276-018-1888-8>
- Wang, C., Cao, W., Hu, Y., & Liu, J. (2023). Data Center Traffic Scheduling Strategy for Minimization Congestion and Quality of Service Guaranteeing. *Computers, Materials & Continua*, 75(2), 4377–4393. <https://doi.org/10.32604/cmc.2023.037625>
- Yu, H., Qi, H., Li, K., Zhang, J., Xiao, P., & Wang, X. (2018). Openflow Based Dynamic Flow Scheduling with Multipath for Data Center Networks. *Computer Systems Science and Engineering*, 33(4), 251–258. <https://doi.org/10.32604/csse.2018.33.251>

Отримано редакцією журналу / Received: 20.05.25
Прорецензовано / Revised: 02.06.25
Схвалено до друку / Accepted: 18.06.25

Mykola NESTERENKO, DSc (Engin.), Assoc. Prof.

ORCID ID: 0000-0003-0812-2793

e-mail: mykola.nesterenko@viti.edu.ua

Kruty Heroes Military Institute of Telecommunications and Information Technologies, Kyiv, Ukraine

Anton MARINOV, PhD Student

ORCID ID: 0000-0001-8717-2734

e-mail: marinov.anton@ill.kpi.ua

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

MANAGEMENT OF DATA CENTER SDN BASED ON THE MODIFIED DEVOFLOW APPROACH

Background. The paper investigates the features of implementing the SDN network management layer when servicing data center flows of different volumes based on Hedera technology, the OpenFlow protocol, and the DevoFlow approach. The distribution of processing functions, collection of statistical data, and the frequency of updating statistical data affect both the amount of service information and the efficiency and effectiveness of management in terms of distributing incoming traffic for load balancing. The aim of the work is to modify the Hedera approach by redistributing the functions of flow classification and adaptive control depending on the flow volume at the control level, as well as to develop an algorithm for determining the allowable time for sending statistics from switches to the SDN controller.

Methods. The method of system analysis and decomposition is used to study complex systems, methods of collecting data on the state of the network, as well as heuristic rules for determining the volume of service traffic.

Results. The paper analyzes the features of construction and operation of technologies like Hedera and DevoFlow, recommend a modified architecture for building the control plane and identifies the main functions of software modules that can improve the efficiency of SDN controllers. Due to changing the procedure for collecting and processing service information on the equipment of the SDN network of data centers and returning to a centralized type of control.

Approaches to the classification of incoming flows and their division into small "mice-flows", medium "medium-flows" and large "elephant-flows" are determined, which allows further use of multi-path routing for more efficient use of network resources. Also, this article proposes an algorithm for determining the permissible time for sending statistics from switches (push-based method), which takes into account the dynamics of load changes in the lines in the network, the number of active sessions and is based on a heuristic rule for limiting the amount of service traffic.

Conclusions. A modified architecture of the control level based on the DevoFlow approach has been developed and the main functions of the software modules of the SDN controller of the data center network have been determined. Also, some analytical dependencies have been developed to determine the approved time for sending statistics from switches to the SDN controller, taking into account the current state of network congestion. Additionally, specific experiments were conducted to confirm both the reliability of the chosen method for obtaining statistics and the proposed algorithm for determining the permissible time for sending statistics from switches to the SDN controller.

Keywords: Software-defined Networking, data center, OpenFlow, Hedera, DevoFlow, elephant-flows, mice-flows, SDN controller, OpenFlow Switch, flow table.

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.



ОГЛЯДИ Й ОБГОВОРЕННЯ



ЕФЕКТИВНІ МЕТОДИ ВИВЧЕННЯ ПАТЕРНІВ ПРОЄКТУВАННЯ: АНАЛІЗ СУЧАСНИХ ТЕНДЕНЦІЙ І ОСВІТНІХ ПРАКТИК

Вступ. *Визначення ефективних методів вивчення як класичних, так і новітніх патернів проєктування програмного забезпечення, з урахуванням сучасних вимог до його розроблення є актуальним завданням. В основу цього дослідження покладено аналіз новітніх тенденцій щодо практичного застосування різних патернів у процесі розроблення програм, визначення їхньої популярності та пошук методів, що забезпечать ефективніше опанування майбутніми ІТ-фахівцями архітектури застосунків.*

Методи. *У процесі дослідження використано методи аналізу й узагальнення інформації, отриманої з відкритих джерел, зокрема, з наукових публікацій, тез, статей, навчальних планів, робочих програм навчальних дисциплін, силябусів та інших нормативних документів. Проаналізовано робочі програми навчальних дисциплін (силябуси) кількох провідних університетів, що дозволило виокремити ефективні методи для вивчення патернів проєктування програмного забезпечення для майбутніх ІТ-фахівців.*

Результати. *Аналіз навчальних програм показав, що в університетах України увага зосереджена переважно на вивченні класичних патернів проєктування GoF, що дещо обмежує якість підготовки студентів до роботи з новітніми технологіями. Порівняно з традиційними методами навчання, зокрема, читанням лекцій, проведенням лабораторних занять, використанням UML-діаграм для візуалізації програмних кодів, метод проєктів виявився більш ефективним.*

Висновки. *Ефективним методом вивчення патернів проєктування є поєднання класичних лекцій з інтерактивними заняттями та застосуванням методу проєктів. Це сприяє кращому розумінню матеріалу і підготовці студентів до розв'язання реальних задач, які зазвичай є більшими за обсягом і складнішими порівняно з окремою лабораторною роботою, і потребують навичок аналізу і синтезу. Зважаючи на наявні вимоги з боку ринку праці, доцільно оновити навчальні програми, удосконаливши їхній зміст і структуру, зокрема і завдяки впровадженню ефективніших методів навчання.*

Ключові слова: *патерни проєктування, методи навчання, метод проєктів.*

Вступ

В сучасному програмному забезпеченні проєктування є одним із ключових аспектів, що визначає якість, продуктивність і масштабованість розроблюваних програмних систем. Патерни проєктування як узагальнені рішення типових проблем програмування відіграють важливу роль у створенні ефективних, гнучких і підтримуваних програмних рішень (Євланов, 2014). Вивчення патернів проєктування є важливим елементом підготовки фахівців у галузі інформаційних технологій, оскільки дає змогу не лише опанувати кращі практики, а і сформувати навички застосування структурованих підходів до розроблення програмного забезпечення.

Історично основи патернів проєктування були описані у книзі Еріка Гамми, Річарда Гельма, Ральфа Джонсона та Джона Вліссідеса, які 1994 р. у своїй праці "Design Patterns: Elements of Reusable Object-Oriented Software" представили класифікацію 23 класичних патернів проєктування (Gamma et al., 1994). Згодом ці концепції отримали подальший розвиток у зв'язку з еволюцією програмних парадигм і технологій, що призвело до появи нових патернів, зокрема й тих, що стосуються мікросервісної архітектури, хмарних обчислень і розподілених систем.

На сучасному етапі розвитку програмної інженерії вивчення патернів проєктування залишається однією з ключових складових у підготовці фахівців з інженерії програмного забезпечення (Yener, & Theedom, 2014). Зауважимо, що більшість освітніх програм акцентують увагу на класичних патернах GoF, таких як Singleton, Factory, Observer, Proxy та інші, тоді як новітні тенденції, пов'язані з контейнеризацією, подійно-орієнтованою архітектурою (Event-Driven Architecture) та безсерверними (serverless) технологіями, нині меншою мірою інтегровані в навчальні плани (17-695: Design Patterns..., 2022). Це створює певний розрив між академічною підготовкою та реальними вимогами сучасного ринку праці. Інша, не менш важлива невідповідність полягає у переважному застосуванні класичних методів навчання патернам проєктування, що у поєднанні зі складністю власне навчального матеріалу, високим рівнем абстракції не забезпечує їх ефективне опанування (Сілакова, 2017).

Отже, **метою** дослідження є визначення ефективних методів навчання патернам проєктування. Визначення ефективних методів навчання як класичним, так і новітнім патернам дозволить підвищити якість підготовки фахівців у галузі програмної інженерії, сприятиме кращій адаптації випускників до сучасних вимог ІТ-індустрії та забезпечить глибоке розуміння принципів створення гнучких і масштабованих програмних систем.

Аналіз останніх досліджень і публікацій. Вивчення патернів проєктування у провідних університетах світу й України є важливим елементом підготовки майбутніх фахівців у сфері інформаційних технологій. Однак аналіз навчальних програм засвідчує, що зміст курсів із патернів проєктування у більшості університетів зосереджений переважно на класичних підходах, тоді як новітні патерни, що орієнтовані на сучасні архітектурні рішення, представлені недостатньо (Factors Contributing..., 2019).



НТУ України Київський політехнічний інститут імені Ігоря Сікорського (КПІ). У межах опанування освітньої компоненти "Патерни об'єктно-орієнтованого проектування" студенти вивчають основи застосування класичних патернів GoF (Порев, 2022). Зокрема, у навчальному плані передбачено такі теми:

- поняття патерну проектування. Патерн Singleton;
- різновиди патернів Factory;
- патерни Facade, Adapter, Dependency Injection, Bridge;
- патерни Decorator, Observer, Visitor.

Загальний обсяг зазначеного курсу становить 18 навчальних годин, що суттєво обмежує можливості детального опрацювання матеріалу та глибокого розуміння структури коду, враховуючи кількість і різноманіття різних патернів проектування. Крім того, курс орієнтований лише на традиційні патерни, тоді як сучасні архітектурні підходи, зокрема й мікросервісні, безсерверні (serverless) і подійно-орієнтовані (Event-Driven Architecture) патерни, залишаються поза межами програми. Частково вони розглядаються в іншій освітній компоненті "Архітектура програмного забезпечення", проте зміст практичних завдань передбачає лише опрацювання цих шаблонів з погляду системного архітектора. Причому залишається поза увагою практична складова, а саме – розроблення програмного застосування відповідної архітектури.

Національний університет "Києво-Могилянська академія". Навчальна програма передбачає освітню компоненту, що охоплює повний цикл розроблення програмної системи, включаючи методи аналізу та проектування, уніфікований процес розроблення об'єктно-орієнтованих програм і керування програмним проектом. У курсі практикується розроблення навчальних програмних проєктів студентами в невеликих групах із використанням систем управління версіями, веборієнтованих середовищ і рецензування результатів (Освітня програма..., 2018).

Київський національний університет імені Тараса Шевченка. Навчання за спеціальністю "Інженерія програмного забезпечення" проводять на двох факультетах: факультеті комп'ютерних наук і кібернетики (ФКНК) та факультеті інформаційних технологій (ФІТ). Зауважимо, що зміст освітніх компонент, пов'язаних з опануванням об'єктно-орієнтованого програмування, відрізняється суттєво за переліками тем, їх наповненням і розподілом годин, виокремлених на опанування, наприклад, патернів проектування. Вивчення патернів проектування на ФКНК в межах опанування освітньої компоненти "Об'єктно-орієнтоване програмування" концентрується на патернах GoF. Загальний обсяг дисципліни складає 120 год (4 кредити ECTS). На думку авторів дослідження, його не можна вважати достатнім, зважаючи на майбутній фах студентів, адже для вивчення патернів проектування із загального обсягу виокремлено лише три теми, що складає загалом 28 год, з них аудиторних – лише 12 год. Серед методів навчання наявні лише класичні: лекція, лабораторне заняття і самостійна робота (Омельчук, 2022). Проте окремо від основної дисципліни у навчальному плані також присутній "Лабораторний практикум з об'єктно-орієнтованого програмування", що традиційно покращує процес навчання і дозволяє застосовувати той самий метод проєктів, підвищуючи ефективність вивчення матеріалу.

Характерною ознакою організації навчального процесу на ФІТ є запровадження інноваційних технологій навчання, таких як STEM-освіта, в основу якої покладено розвиток критичного мислення, інженерного підходу і вміння застосовувати отримані знання у реальному житті. Щодо освітньої компоненти "Об'єктно-орієнтоване конструювання програм" зауважимо, що вона викладається протягом двох семестрів, є суттєвою за обсягом (становить 330 год, тобто 11 кредитів ECTS), вміщує огляд патернів GoF (загальним обсягом 56 год, з них аудиторних 28 год.) і архітектурних патернів (MVC, MVP, MVVM), подійно-орієнтовані (Event-Driven Architecture) патерни тощо, передбачає застосування крім класичних методів навчання методу проєктів, що, за результатами проведених серед студентів опитувань, позитивно впливає на ефективність вивчення матеріалу (Зубик, 2023).

University of Vienna. На відміну від українських вишів, у Віденському університеті патерни проектування вивчають у межах магістерської програми з інформатики в курсі "Internet Computing and Software Technologies". Цей курс включає детальний розгляд методів проектування, архітектур і патернів, що застосовуються у сучасних розподілених системах. Досліджується підтримка патернів у скриптових мовах програмування та сучасних патернів проектування, таких як *Saga Pattern*, *CQRS*, *Sidecar Pattern*, і це дозволяє студентам отримувати знання не лише про класичні, а й про новітні патерни, орієнтовані на сучасні технології (232 Curriculum..., 2022).

Отже, аналіз навчальних програм засвідчує суттєві відмінності у методах вивчення патернів проектування у різних університетах. У той час як європейські університети, зокрема й Віденський, приділяють увагу сучасним архітектурним рішенням, українські заклади вищої освіти поки що зосереджені переважно на вивченні патернів GoF і застосуванні класичних методів навчання, що може обмежувати готовність випускників до роботи у високотехнологічному середовищі.

Boston University. У Boston University вивчення патернів проектування відбувається у курсі "MET CS665: Software Design and Patterns", де наголос робиться на практичному застосуванні принципів об'єктно-орієнтованого проектування, включаючи GoF-патерни. В курсі інтегровано кейс-стаді з використанням сучасних мов програмування та приклади з проєктів реального бізнес-середовища (Software Design..., 2025). Проєктний метод навчання забезпечує студентам тісний зв'язок між теорією та практикою на сучасному рівні.

Carnegie Mellon University (CMU). Велику увагу CMU приділяє вивченню патернів проектування у контексті системного мислення та гнучких методологій розробки. Курс "17-695: Design Patterns", який викладають у межах програми Master of Software Engineering, включає не лише класичні шаблони, а й аналізує їхню еволюцію, використання в розподілених середовищах і взаємозв'язок із принципами програмної архітектури (17-695: Design Patterns..., 2022). Особливістю підходу в CMU є орієнтація на аналітичне мислення, розв'язання реальних інженерних проблем та адаптацію патернів до складних проєктних сценаріїв. Отже, порівняльний аналіз змісту навчальних програм, виконаний на прикладі шести провідних технічних університетів, свідчить про суттєві відмінності у методах вивчення патернів проектування між вітчизняними та зарубіжними закладами вищої освіти.

У навчальних програмах таких університетів як Carnegie Mellon University (17-695: Design Patterns..., 2022), University of Vienna (232 Curriculum..., 2022) та Boston University (Software Design..., 2025), приділяють більше уваги не лише класичним шаблонам, а й сучасним архітектурним рішенням, що формує здатність студентів швидко адаптуватися до викликів високотехнологічного ринку праці завдяки ефективнішому опануванню шаблонів проектування програмного забезпечення. Вітчизняні програми потребують перегляду їхнього змісту, зокрема й у контексті актуалізації методів навчання, відповідно до глобальних тенденцій у сфері програмного інжинірингу.



Методи

Шаблони проектування є важливим аспектом об'єктно-орієнтованого програмування, що дозволяє створювати гнучкі, масштабовані та підтримувані програмні рішення. Проте навчання цих концепцій часто викликає труднощі серед студентів, оскільки вони мають високий рівень абстракції та потребують глибокого розуміння архітектурних принципів. У цій роботі розглянуто різні методи вивчення шаблонів проектування і запропоновано кращий метод навчання у контексті його ефективності.

Дослідження навчального процесу показує, що студенти, які мають базові знання об'єктно-орієнтованого програмування, часто стикаються з труднощами під час вивчення шаблонів проектування. Основні виклики, з якими вони зустрічаються, містять високу абстрактність шаблонів, що ускладнює розуміння їхніх переваг над традиційними методами розробки, відсутність зв'язку між шаблонами та реальними практичними задачами, а також труднощі у використанні UML-діаграм, які є важливим інструментом для розуміння архітектури програмних систем.

Під час навчального процесу у згаданих вище університетах акцент зазвичай робиться на найпоширеніших шаблонах проектування, які поділяють на три категорії: породжувальні, структурні та поведінкові. До породжувальних шаблонів належать Singleton, який гарантує існування лише одного екземпляра класу, та Factory Method, що забезпечує гнучке створення об'єктів без прив'язки до конкретного класу. Структурні шаблони включають Decorator, який дає змогу динамічно розширювати функціональність класів, не змінюючи їхнього коду. Серед поведінкових шаблонів особливо актуальні такі: Observer, який реалізує механізм спостереження за змінами стану об'єкта, та Strategy, що дозволяє динамічно змінювати алгоритми під час виконання програми.

У процесі навчання зазвичай застосовують різні методи, які можна умовно розділити на традиційні (класичні) й інтерактивні.

Традиційні (класичні) методи, такі як читання лекційних і самостійне опрацювання навчальних матеріалів, за результатами опитувань студентів виявилися менш ефективними (Shatz, 2024), оскільки студенти часто не могли пов'язати отримані знання з практичними завданнями, наближеними до реальних. Лекції, що базувалися на теоретичних поясненнях і UML-діаграмах, демонстрували низьку ефективність через відсутність безпосередньої взаємодії студентів із кодом.

Натомість **інтерактивні методи** виявилися значно ефективнішими (Добровольська, 2018). Живе кодування, під час якого викладач демонструє реалізацію шаблонів у реальному часі, дозволяє студентам краще засвоювати матеріал, проте воно не завжди передбачає активність студентів. Парне програмування, в якому один студент пише код, а інший пояснює логіку дій, сприяє кращому розумінню концепцій, але також має недоліки, адже важко сепарувати компетентність окремого студента, коли вони працюють у парі, і виокремити незрозумілі саме для нього моменти. Найбільш ефективним виявився **метод проєктів** (Петровський, Пасічник, & Скрипник, 2021), у межах застосування якого студенти розробляють власні програмні рішення з використанням шаблонів. Це сприяє насамперед їх самостійності у прийнятті рішень, й ефективному формуванню й закріпленню знань, оскільки вони мають можливість застосовувати шаблони у реальних сценаріях розробки. Також ефективними виявляються вікторини та короткі тести, які виконуються студентами після опанування кожної окремої теми, і які допомагають оцінити індивідуальний рівень засвоєння матеріалу, виявити прогалини у знаннях. Одним із найменш ефективних методів виявляється надмірне використання UML-діаграм без практичних прикладів їхнього застосування, оскільки це ускладнює сприйняття матеріалу студентами, які не мали достатнього досвіду роботи з такими схемами раніше.

Статистику успішності студентів за умови впровадження проєктного методу навчання наведено на рис. 1.

Для оцінювання ефективності вивчення матеріалу за традиційними та сучасними інноваційними методами доцільно використовувати метод порівняння, заснований на аналізі оцінок студентів. Одним із способів кількісного оцінювання є нормалізоване порівняння середніх значень, що дозволяє визначити рівень впливу інноваційного підходу на результати навчання.

Формально ефективність навчання E можна обчислити за такою формулою:

$$E = \frac{S_{new} - S_{old}}{\sigma_{old}}, \quad (1)$$

де S_{new} – середня оцінка студентів, які навчалися за сучасною методикою; S_{old} – середня оцінка студентів, які навчалися за традиційною методикою; σ_{old} – стандартне відхилення оцінок у групі традиційного навчання.

Стандартне відхилення оцінок у групі традиційного навчання можна визначити за такою формулою:

$$\sigma_{old} = \sqrt{\frac{1}{N} \sum_{i=1}^N (S_i - S_{old})^2}, \quad (2)$$

де N – кількість студентів у групі традиційного навчання; S_i – оцінка i -го студента у цій групі; S_{old} – середня оцінка у групі традиційного навчання.

Наведені формули дають змогу оцінити ступінь покращення навчальних результатів шляхом нормалізації різниці між середніми значеннями відносно варіативності оцінок у контрольній групі. Якщо отримане значення E є значущим ($E > 1$), це може свідчити про ефективність проєктного методу засвоєння навчального матеріалу.

У магістерському дослідженні Р. Голубека (Holubek, 2022) здійснено кількісне та якісне оцінювання ефективності вивчення патернів проектування з використанням методу проєктів. Результати дослідження свідчать про підвищення результатів навчання студентів у межах запровадженого методу порівняно з традиційними методами навчання.

Середній бал, отриманий студентами після впровадження методу проєктів, склав 7,8 із 10, при цьому медіанне значення дорівнювало 8, що свідчить про стабільно високі результати більшості учасників. У той самий час, середній бал за результатами іспиту за традиційною методикою становив 6,5, а стандартне відхилення – 1,5. Застосування формули нормалізованого приросту ефективності дозволило кількісно оцінити рівень покращення:

$$E = \frac{S_{new} - S_{old}}{\sigma_{old}} = \frac{7,8 - 6,5}{1,5} \approx 0,87.$$



	I	I	I	I	UML	AE	AE	AD	UML	AD	AE	AD	AE	AD	AE		
	Introduction				Strategy				Decorator				Visitor				
	After break																
	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Q12	Q13	Q14	Q15	%	
Student 1	C	D	B	A	A	A	B	B	A	A	A	C	C	C	C	0,53	
Student 2	A	C	B	C	D	A	B	B	B	B	A	A	A	C	A	0,67	
Student 3	-	-	-	-	-	C	B	B	A	A	A	C	A	C	A	0,8	
Student 4	A	C	B	C	A	A	B	A	C	A	A	C	A	A	A	0,73	
Student 5	A	C	B	C	B	A	-	B	A	C	-	B	A	A	-	0,67	
Student 6	-	-	-	-	-	-	-	-	A	A	A	C	-	-	-	0,75	
Student 7	A	C	B	C	D	C	B	A	B	-	-	-	-	-	-	0,67	
Student 8	A	C	B	C	A	C	B	B	A	C	A	C	A	C	C	0,87	
Student 9	-	-	-	-	D	-	-	-	-	-	-	-	-	-	-	0	
Student 10	-	-	-	-	-	A	B	B	A	A	A	A	A	C	A	0,7	
Student 11	A	C	B	A	-	A	B	-	-	A	A	C	-	-	-	0,78	
Student 12	-	-	-	C	D	A	B	C	-	-	-	-	-	-	-	0,8	
Student 13	A	C	B	C	A	A	B	C	A	C	-	-	-	-	-	1	
Student 14	A	C	B	C	A	C	B	A	A	A	A	C	A	C	A	0,87	
Student 15	A	C	B	-	-	-	B	A	-	-	-	-	-	-	-	0,8	
Student 16	A	C	A	D	A	-	C	A	A	A	A	C	A	C	A	0,64	
Student 17	A	C	B	A	A	A	B	B	A	C	A	C	A	A	-	0,79	
Student 18	A	C	B	C	A	A	B	B	A	B	A	C	B	A	A	0,73	
Student 19	-	-	-	-	A	A	B	B	-	C	A	C	-	-	-	0,86	
Student 20	A	C	B	C	A	A	B	B	A	A	A	C	A	C	C	0,8	
Student 21	A	C	B	B	A	A	B	A	A	C	A	C	A	A	A	0,8	
Student 22	A	C	B	A	D	A	B	B	-	-	-	-	-	-	-	0,63	
Student 23	A	C	C	C	D	C	B	A	A	C	A	C	A	C	A	0,8	
Student 24	-	-	-	C	-	-	-	-	-	-	-	-	-	-	-	1	
Student 25	-	-	-	C	-	-	B	B	-	-	-	-	-	-	-	0,67	
Student 26	-	-	B	C	-	-	-	-	-	-	-	-	-	-	-	1	
Student 27	A	C	B	C	C	A	B	A	B	A	-	A	-	-	-	0,55	
Student 28	A	C	B	C	A	A	B	B	A	C	A	C	A	A	A	0,87	
Student 29	A	D	B	C	-	-	-	-	-	-	-	-	-	-	-	0,75	
Student 30	B	-	C	C	-	-	-	-	-	-	-	-	-	-	-	0,33	
Student 31	-	-	-	-	-	-	-	-	A	C	A	C	-	-	-	1	
Student 32	-	-	-	-	-	-	-	-	-	-	-	-	A	A	B	0,33	
Student 33	B	C	B	C	-	-	-	-	-	-	-	-	-	-	-	0,75	
Student 34	A	C	B	C	D	A	B	C	-	A	A	-	A	-	A	0,83	
Student 35	A	C	C	A	A	C	B	B	A	A	A	C	A	A	A	0,67	
Student 36	A	C	D	C	A	C	C	B	-	C	A	-	-	-	-	0,7	
Student 37	A	D	B	C	D	A	B	C	B	C	A	-	-	-	-	0,73	
Student 38	-	C	B	C	C	A	B	B	B	B	A	C	A	C	B	0,64	
Student 39	A	D	B	C	A	A	C	B	A	A	A	C	A	C	A	0,73	
Student 40	A	C	B	C	-	-	-	-	-	-	-	-	-	-	-	1	
Student 41	A	D	B	A	C	A	B	B	C	C	A	C	A	C	A	0,67	
Student 42	-	-	-	-	-	-	B	A	A	B	A	C	B	-	A	0,63	
Student 43	-	-	-	-	-	-	-	-	A	C	A	C	A	-	-	1	
Student 44	A	C	B	C	A	C	-	-	-	-	-	-	-	-	-	1	
Student 45	A	C	B	C	-	-	-	-	-	-	-	-	-	-	-	1	
Student 46	A	C	B	D	A	A	C	B	A	C	A	C	A	A	A	0,73	
Student 47	C	C	D	A	A	C	B	C	-	-	-	-	-	-	-	0,63	
Student 48	-	-	-	-	B	A	B	B	D	A	A	A	A	C	A	0,55	
Student 49	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	
Student 50	-	-	-	-	A	A	B	C	-	-	-	-	-	C	A	1	
Student 51	A	C	A	A	A	C	B	B	A	C	A	C	-	-	-	0,75	
Student 52	D	C	D	C	D	A	C	A	A	C	A	C	A	A	A	0,6	
Student 53	-	-	-	-	A	A	B	B	A	C	A	C	A	C	A	0,91	
Student 54	A	C	B	C	B	C	B	A	B	A	A	C	B	C	-	0,64	
Student 55	-	-	-	-	-	A	B	C	A	A	A	C	A	A	A	0,8	
Percentage correct	0,9	0,9	0,8	0,7	0,6	1	0,9	0,2	0,7	0,5	1	0,8	0,9	0,6	0,8		
Total answers	36	36	38	40	36	38	40	40	34	37	34	33	29	27	26		
Correct answers	31	31	30	29	21	38	35	7	25	20	34	28	25	16	21		

Рис. 1. Дані успішності студентів у результаті запровадження у навчання методу проєктів



Отримане значення $E \approx 0,87$, хоч і не перевищує умовного порогу значущості ($E > 1$), однак засвідчує помірне покращення результатів навчання. Позитивну динаміку також підтверджують суб'єктивні оцінки студентів, зафіксовані в анкетуванні після проходження курсу: 85 % респондентів схвально оцінили власне лекцію. У межах аналізу практичної діяльності встановлено, що 70 % студентів змогли коректно застосувати патерни у межах контрольного завдання, а 80 % продемонстрували високе розуміння UML-діаграм. Утім, 60 % студентів зазначили труднощі в реалізації функціонального наповнення патернів, що свідчить про потребу у посиленій роботі на етапі практичної імплементації.

Результати

У сукупності зазначені дані вказують на те, що застосування проектного методу до вивчення патернів проектування ефективне і веде до поліпшення успішності студентів, покращення засвоєння ними навчального матеріалу. Незважаючи на те, що отримане числове значення ефективності не є високим у межах статистичної моделі, воно демонструє обґрунтованість застосування методу проектів як доповнення до класичних методів навчання, зокрема в аспекті емоційного залучення студентів до навчального процесу з метою забезпечення кращого розуміння ними навчального матеріалу через призму його майбутнього практичного застосування.

Аналіз методів вивчення шаблонів проектування показав, що метод проектів є ефективним (Тадеуш, 2017). Інтерактивні методи покращують сприйняття навчального матеріалу та дозволяють краще його розуміти. Таким чином, для успішного вивчення шаблонів проектування доцільно поєднувати теоретичні лекції із завданнями, наближеними до практичних, які виконуються поетапно протягом тривалого проміжку часу з поступовим нарощуванням програмного коду, активною взаємодією студентів у межах уявних команд і з викладачем, що дає їм змогу ефективно засвоювати знання та застосовувати їх у реальних сценаріях розробки.

Дискусія і висновки

Підсумовуючи результати дослідження методів вивчення патернів проектування, можна зробити такі висновки. На основі проведеного аналізу навчальних програм і методів навчання в окремих провідних університетах України та деяких зарубіжних закладах вищої освіти виявлено деякі важливі аспекти, що потребують уваги для покращення процесу навчання. Вивчення патернів проектування з використанням традиційних методів навчання, таких як лекції, лабораторні заняття і самостійна робота студентів, виявилось менш ефективним унаслідок обмеженої можливості студентів у застосуванні теоретичних знань у реальних сценаріях розробки. Натомість проектний метод навчання продемонстрував вищу ефективність. Застосування цього методу дозволяє студентам не лише механічно запам'ятовувати теоретичні відомості й відтворювати їх, розв'язуючи задачі, відірвані від реальних проектів, але й підтвердити набуті знання практикою, рухаючись до реалізації задуманого проекту покроково, бути для цього достатньо вмотивованим і орієнтованим на результат. Належна вмотивованість полегшує студентам опанування складних абстракцій.

З огляду на результати дослідження, вважаємо за доцільне зосередити зусилля на оновленні навчальних програм університетів, зокрема і за рахунок коригування їхнього змісту шляхом включення до їхнього змісту сучасних патернів проектування, що відповідають вимогам сьогодення. Інноваційні методи навчання забезпечать студентам кращу підготовку до роботи у сучасному інформаційному середовищі. Метод проектів дозволить студентам досягнути кращих результатів навчання, а також допоможе підготуватися до реальних викликів у їх майбутній діяльності за фахом. Зважаючи на важливість використання патернів проектування у розробці сучасних програмних систем, ефективність їх вивчення є необхідною умовою підготовки висококваліфікованих фахівців у галузі інформаційних технологій.

Внесок авторів. Людмила Зубик – концептуалізація дослідження, розроблення методів дослідження та загальне керівництво; Даниїл Фуркало – дослідження літературних джерел, математичні розрахунки та підготовка графічних матеріалів.

Джерела фінансування. Це дослідження не отримало жодного гранта від фінансової установи в державному, комерційному або некомерційному секторах.

Список використаних джерел

- Євланов, М. В. (2014). Патерни проектування вимог до інформаційної системи. *Вісник Національного університету "Львівська політехніка": Інформаційні системи та мережі*, 783, 429–434. <https://openarchive.nure.ua/server/api/core/bitstreams/cca27a1e-4532-40bc-8a3b-248f662b3d66/content>
- Зубик, Л. В. (2023). Робоча програма навчальної дисципліни "Об'єктно-орієнтоване конструювання програм". КНУ імені Тараса Шевченка. <https://surl.li/djzhuo>
- Омельчук, Л. (2022). Робоча програма навчальної дисципліни "Об'єктно-орієнтоване програмування". КНУ імені Тараса Шевченка. https://drive.google.com/file/d/1_IHCutPL_Om_SeEeXpO1sQn1WmAhAx0/_view
- Освітня програма "Інженерія програмного забезпечення". (2018). *НУ Києво-могилянська академія*. <https://www.ukma.edu.ua/ects/index.php/2011-04-18-08-31-28/163-bakalavr/bpproggin/248-2018-11-01-10-37-52>
- Петровський, С. С., Пасічник, О. А., & Скрипник, Т. К. (2021). Особливості використання методу проектів при викладанні освітніх компонент спеціальності комп'ютерні науки. *Вісник Хмельницького національного університету. Технічні науки*, 4, 30–34. <https://www.doi.org/10.31891/2307-5732-2021-299-4-30-34>
- Порев, В. М. (2022). *Об'єктно-орієнтоване програмування*. НТТУ "КПІ імені Ігоря Сікорського". <https://ela.kpi.ua/server/api/core/bitstreams/854443cb-3375-4e76-961a-1f9a74a15b07/content>
- Сілакова, Т. Т. (2017). Проектні технології підготовки студентів. *Вісник Національного авіаційного університету. Серія: Педагогіка, Психологія*, 11, 153–158. <https://doi.org/10.18372/2411-264X.11.12571>
- Тадеуш, О. М. (2017). Метод проектів як форма продуктивного навчання студентів. *Науковий часопис НПУ імені М. П. Драгоманова. Серія 16: Теорія особистість учителя: проблеми теорії і практики*, 29, 142–146. http://nbuv.gov.ua/UJRN/Nchnpu_016_2017_29_33
- 17-695: Design Patterns. Syllabus (2022). *Carnegie Mellon University*. https://mse.s3d.cmu.edu/0_documents/syllabi/fa2022/17695-design-patterns.pdf
- 232 Curriculum for the master's program in Computer Science (2022). *University of Vienna*. https://informatik.univie.ac.at/fileadmin/user_upload/f_informatik/Studium/Master_Informatik/curricula/MA_Inf_2022_EN.pdf
- Gamma, E., Helm, R., Johnson, R., & Vlissides, J. (1994). *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley.
- Holubek, R. (2022). *Designing the Teaching Pattern: a structured approach to teach design patterns*. Radboud University. https://www.cs.ru.nl/masters-theses/2022/R_Holubek_Designing_the_teaching_pattern_a_structured_approach_to_teach_design_patterns.pdf
- Shatz, I. (2024). Assumption-checking rather than (just) testing: The importance of visualization and effect size in statistical diagnostics. *Behavior Research Methods*, 56, 826–845. <https://doi.org/10.3758/s13428-023-02072-x>
- Software Design and Patterns (MET CS-665). Syllabus. (2025). *Boston University Metropolitan College*. <https://www.bu.edu/csmet/files/2025/01/665-syllabus-2025-spring.docx.pdf>
- Yener, M., & Theedom, A. (2014). *Professional Java EE Design Patterns*. John Wiley & Sons.



References

- 17-695: Design Patterns. Syllabus (2022). *Carnegie Mellon University*. https://mse.s3d.cmu.edu/0_documents/syllabi/fa2022/17695-design-patterns.pdf
- 232 Curriculum for the master's program in Computer Science (2022). *University of Vienna*. https://informatik.univie.ac.at/fileadmin/user_upload/f_informatik/Studium/Master_Informatik/curricula/MA_Inf_2022_EN.pdf
- Educational program "Software Engineering". (2018). *Kyiv-Mohyla National University* [in Ukrainian]. <https://www.ukma.edu.ua/ects/index.php/2011-04-18-08-31-28/163-bakalavr/bpprogin/248-2018-11-01-10-37-52>
- Evlanov, M. V. (2014). Patterns of Designing Requirements for an Information System. *Bulletin of the National University "Lviv Polytechnic": Information Systems and Networks*, 783, 429–434 [in Ukrainian]. <https://openarchive.nure.ua/server/api/core/bitstreams/cca27a1e-4532-40bc-8a3b-248f662b3d66/content>
- Gamma, E., Helm, R., Johnson, R. & Vlissides, J. (1994). *Design Patterns: Elements of Reusable Object-Oriented Software*. Addison-Wesley.
- Holubek, R. (2022). *Designing the Teaching Pattern: a structured approach to teach design patterns*. Radboud University. https://www.cs.ru.nl/masters-theses/2022/R_Holubek_Designing_the_teaching_pattern_a_structured_approach_to_teach_design_patterns.pdf
- Omelchuk, L. (2022). *Work program of the academic discipline "Object-Oriented Programming"*. Taras Shevchenko National University of Kyiv [in Ukrainian]. https://drive.google.com/file/d/1_IHCutPL_0m_SeEeXpO1sQn1WmAhAx0/view
- Petrovsky, S. S., Pasichnyk, O. A., & Skrypnyk, T. K. (2021). Peculiarities of using the project method in teaching educational components of the computer science specialty. *Bulletin of Khmelnytsky National University. Technical Sciences*, 4, 30–34 [in Ukrainian]. <https://www.doi.org/10.31891/2307-5732-2021-299-4-30-34>
- Porev, V. M. (2022). Object-oriented programming. Igor Sikorsky Kyiv Polytechnic Institute [in Ukrainian]. <https://ela.kpi.ua/server/api/core/bitstreams/854443cb-3375-4e76-961a-1f9a74a15b07/content>
- Shatz, I. (2024). Assumption-checking rather than (just) testing: The importance of visualization and effect size in statistical diagnostics. *Behavior Research Methods*, 56, 826–845. <https://doi.org/10.3758/s13428-023-02072-x>
- Silakova, T. T. (2017). Project technologies for student training. *Bulletin of the National Aviation University. Series: Pedagogy, Psychology*, 11, 153–158 [in Ukrainian]. <https://doi.org/10.18372/2411-264X.11.12571>
- Software Design and Patterns (MET CS-665). Syllabus (2025). *Boston University Metropolitan College*. <https://www.bu.edu/csnet/files/2025/01/665-syllabus-/2025-spring.docx.pdf>
- Tadeush, O. M. (2017). The project method as a form of productive student learning. *Scientific Journal of the NPU named after M.P. Dragomanov. Series 16: Creative personality of a teacher: problems of theory and practice*, 29, 142–146 [in Ukrainian]. http://nbuv.gov.ua/UJRN/Nchnpu_016_2017_29_33
- Yener, M., & Theedom, A. (2014). *Professional Java EE Design Patterns*. John Wiley & Sons.
- Zubyk, L. V. (2023). *Work program of the academic discipline "Object-Oriented Program Design"*. Taras Shevchenko National University of Kyiv [in Ukrainian]. <https://surl.li/djzhuo>

Отримано редакцією журналу / Received: 03.04.25
Прорецензовано / Revised: 16.04.25
Схвалено до друку / Accepted: 01.05.25

Daniil FURKALO, Student
ORCID ID: 0009-0005-8584-7167
e-mail: daniilf077@knu.ua
Taras Shevchenko National University of Kyiv, Ukraine

Liudmyla ZUBYK, PhD (Ped.), Assoc. Prof.
ORCID ID: 0000-0002-2087-5379
e-mail: zubyk.liudmyla@knu.ua
Taras Shevchenko National University of Kyiv, Ukraine

EFFECTIVE APPROACHED TO TEACHING DESIGN PATTERNS: ANALYSIS OF CURRENT TRENDS AND EDUCATIONAL PRACTICES

Background. *Determining effective methods for studying both classical and modern software design patterns, taking into account modern requirements for its development, is an urgent task. The basis of this study was the analysis of the latest trends in the practical application of various patterns in the process of program development, determining their popularity and searching for methods that will ensure more effective mastering of application architecture by future IT specialists.*

Methods. *The research used methods of analyzing and summarizing information obtained from open sources, in particular, scientific publications, theses, articles, curricula, work programs of educational disciplines, syllabi and other regulatory documents. The work programs of educational disciplines (syllabuses) of several leading universities were analyzed, which allowed us to identify effective methods for studying software design patterns for future IT specialists.*

Results. *The analysis of curricula showed that in Ukrainian universities, attention is focused mainly on the study of classical GoF design patterns, which somewhat limits the quality of students' preparation for working with the latest technologies. Compared to traditional teaching methods, in particular, lectures, laboratory classes, and the use of UML diagrams for visualization of program codes, the project method turned out to be more effective.*

Conclusions. *An effective method of studying design patterns is to combine classical lectures with interactive classes and the use of the project method. This contributes to a better understanding of the material and prepares students to solve real problems, which are usually larger in volume and more complex than separate laboratory work, and require analysis and synthesis skills. Given the existing requirements of the labor market, it is advisable to update curricula by improving their content and structure, including by introducing more effective teaching methods.*

Keywords: *design patterns, learning methods, project method.*

Автори заявляють про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.



AN ANALYTICAL REVIEW OF CONTENT-BASED AND COLLABORATIVE FILTERING IN RECOMMENDER SYSTEMS

Background. *With the rapid growth of digital content, recommender systems are becoming a key tool for providing personalized offers. They contribute to the discovery of new movies, music, and products, maintaining user interest in using platforms. The relevance of researching recommender system algorithms is due to the need to improve their work to satisfy individual user preferences. This paper presents a review and analytical study of recommender system algorithms. The purpose of this paper is to systematize, classify, and critically analyze two main approaches in recommender systems: content-based filtering and collaborative filtering.*

Methods. *A review of existing recommender system methods, a comparative and analytical assessment.*

Results. *The work analyzes recommender system algorithms. A formal definition of the recommendation problem is given, where user preferences are modeled as a functional dependence on object properties. Within the framework of content-based filtering, the use of classification algorithms, such as a Naïve Bayes classifier and decision trees, as well as the Rocchio algorithm, which uses relevant feedback to update the user profile, is considered. The strengths and weaknesses of different similarity measures between vectors are analyzed. In collaborative filtering, the memory-based approach (user-based and item-based methods) and model-based techniques with an emphasis on the k-NN algorithm are investigated. To overcome the shortcomings of individual methods, a hybrid approach is proposed that combines their advantages. Methods for integrating systems into a hybrid model are presented, which allows improving the accuracy of recommendations.*

Conclusions. *The results of the work highlight the features of the specified filtering methods, demonstrate the impact of the implementation of algorithms and input data on the accuracy of recommendations and response time. The analysis of shortcomings emphasizes the importance of the combined use of filtering algorithms to improve the efficiency of recommendation systems, which makes the hybrid approach a promising direction for further research and implementation.*

Keywords: *collaborative filtering, content-based filtering, hybrid filtering, recommender systems, Rocchio algorithm, vector space model.*

Background

The modern digital world is filled with an excess amount of data and information. Recommender systems have become important tools, helping users find what they're looking for among all the options. At their core, these systems aim to predict user preferences and offer personalized recommendations tailored to individual tastes.

Given the increasing reliance on recommender systems in various industries, optimizing their performance and improving their recommendation accuracy is a critical research area. The study of underlying principles, different similarity measures, and hybrid models contributes to enhancing the efficiency and adaptability of these systems. The selection of appropriate models and recommendation algorithms remains an open question in optimizing recommender system performance. As a result, numerous studies have been devoted to the development and evaluation of recommendation algorithms, which makes systematic review and analysis of existing approaches particularly valuable.

Two main methods used in recommender systems are content-based filtering and collaborative filtering. Content-based filtering uses the intrinsic properties of items and user profiles to generate recommendations, while collaborative filtering relies on the collective behavior of users to provide personalized suggestions. Each of these methods has its own advantages and drawbacks.

This paper is an analytical review of recommender system techniques, with a particular focus on content-based and collaborative filtering methods. The primary object of this research is recommender systems, focusing on content-based and collaborative filtering techniques. The aim of the research is to analyze and compare different algorithms used in recommender systems. The objectives include investigating the principles and algorithms behind content-based and collaborative filtering, and evaluating the effectiveness of different similarity measures, such as Euclidean distance, cosine similarity, Pearson correlation etc. The methodology includes an analytical review of existing recommender system techniques, their classification, systematization, and a comparison of their performance based on the related studies, highlighting the advantages and limitations of each approach.

The comparative analysis presented in this study is based on a set of qualitative and performance-oriented criteria. These include the type and characteristics of input data (explicit and implicit feedback, availability of item metadata, data sparsity, and vector dimensionality), recommendation accuracy reported in related studies, and computational aspects such as time performance and algorithmic complexity. Additional comparison dimensions involve scalability with respect to dataset size, interpretability of recommendation results, adaptability to evolving user preferences, and robustness to the cold-start problem for new users and items. Furthermore, the analysis considers the sensitivity of different methods to data quality and feature representation, highlighting how these factors influence both effectiveness and efficiency across content-based, collaborative, and hybrid recommendation approaches.

Methods

General model of recommender systems. A recommender system consists of essential components that facilitate personalized recommendations (Fig. 1). It is built around a set of source data files that can have different structures and be curated from various sources (Xia et al., 2024). The item catalog serves as the repository from which recommendations are drawn, ensuring a diverse selection for users.

User feedback is central to improving recommendation accuracy. Explicit feedback, derived from direct user input such as ratings, likes/dislikes, or comments, provides clear indications of user preferences. However, collecting explicit feedback



can be challenging due to user participation barriers and potential biases (Kumar, 2022). Implicit feedback is inferred from user behavior, such as browsing history, clicks, and purchases, offering an alternative method that does not require direct user engagement. However, it may not fully capture user preferences as it relies on observable actions (Kumar, 2022). Combination of direct user input and observed behavior can provide a more comprehensive view of user interests and preferences (Mandal, & Maiti, 2018).

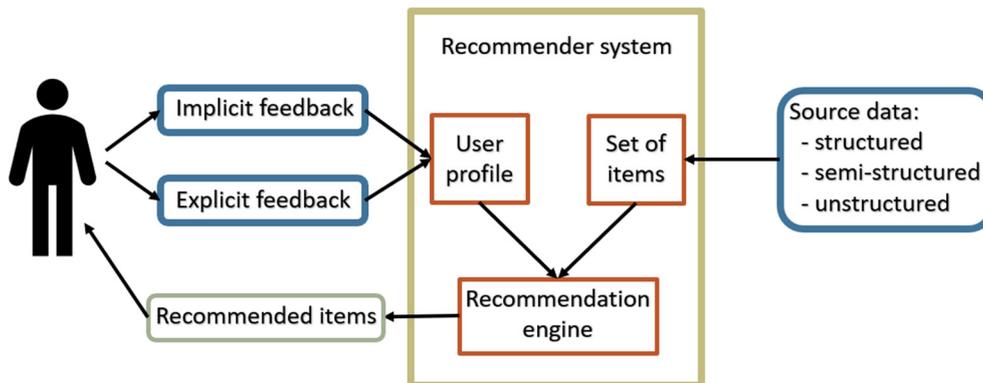


Fig. 1. General design of recommender systems

Leveraging the collected user feedback, the recommender system constructs individual user profiles that encapsulate their preferences. These profiles are dynamic representations that evolve over time based on the user's interactions. User profiles serve as the basis for generating personalized recommendations, allowing the system to match users with items that align closely with their preferences and interests.

In a formal context, recommender systems tackle the following problem (Zisopoulos et al., 2008): given a set of items S , a subset E (estimates) of S , with $|E| \ll |S|$ and the values $f(e)$ for each $e \in E$ of a function $f(x): S \rightarrow \{0, 1\}$, where the exact nature of $f(x)$ is unknown, the aim is to find an estimate $g(x): S \rightarrow \{0, 1\}$, and a subset R (recommendations) of S , with $|R| \ll |S|$ and $E \cap R = \emptyset$, such that the probability $P([g(r)=f(r)]=1)$, for every r in R is maximized. Both $f(x)$ and $g(x)$ can also target numerical sets, employing a threshold mechanism to discern between different classes. In other words, it is assumed that user preferences can be modeled by an unknown target function, and the goal of the recommender system is to approximate this function as accurately as possible.

Content-based filtering. Content-based filtering suggests items to users based on their features and the user's past interactions. The main idea is to find similarities between items and match them with user preferences.

This approach works by creating user profiles and item profiles (Fig. 2). A user profile stores information about a person's interests, preferences, and past interactions. An item profile contains details about an item, such as its description, metadata, or specific features. By comparing these profiles, content-based filtering algorithms find and recommend items that best match a user's interests (Herimanto, Samosir, & Ginting, 2024).

Movie	Adventure	Action	Science-Fiction	Drama	Crime	Thriller	Score	Reviews	User 1	User 2
Star Wars IV	1	1	1	0	0	0	4,54	15845	1	-1
American Beauty	0	0	0	1	0	0	4,13	3484		
City of Gold	0	0	0	1	1	0	3,55	2649	-1	1
Interstellar	0	0	1	1	0	0	4,01	4158	1	
The Matrix	1	1	1	0	0	1	4,87	6533		1

Fig. 2. An example of item and user profiles

Before feeding the data into recommendation algorithms, preprocessing steps are typically applied to clean, normalize, and transform the data. This may involve handling missing values, encoding categorical variables, and scaling numerical features. Additionally, feature extraction techniques may be employed to derive meaningful representations of items and users, such as content features, word embeddings or contextual information, in high-dimensional vectors.

The primary objective of data preprocessing is to transform raw item features into a structured, numerical format suitable for input into mathematical models and machine learning algorithms. These numerical representations form the basis for measuring similarity, learning user preferences, and generating accurate, personalized recommendations.

Since recommender systems aim to estimate the likelihood of user satisfaction with specific items (often categorizing this satisfaction into discrete outcomes such as "like", "dislike", or "neutral"), the recommendation task can naturally be framed as a classification problem. In this formulation, the system predicts the class label that best represents a user's potential reaction to an item, based on historical interactions and item attributes. By training on labeled data, where the labels represent user preferences, classification algorithms build predictive models that can then be used to classify new items for recommendation.

Different classification algorithms can be utilized in content-based filtering. Naive Bayes is a powerful algorithm for content-based filtering in recommender systems due to its simplicity, speed, and effectiveness in handling text data (Ricci, 2002). It operates on the principle of Bayes' theorem, assuming that the presence of a particular feature in a class is unrelated to the presence of any other feature. This "naivety" simplifies the calculation of probabilities, allowing effectively model the content of items and recommend similar items based on the features. However, it's important to note that while Naive Bayes is fast

and easy to implement, it makes strong assumptions about feature independence, which might not hold true in all datasets. Therefore, it's crucial to evaluate its performance on your specific dataset and consider other algorithms if necessary.

Decision trees are powerful and interpretable models used in machine learning to solve classification problems by hierarchically dividing the feature space. In content-based filtering, they provide a structured approach for predicting user preferences based on item attributes (Gershman et al., 2010). A decision tree consists of internal nodes that evaluate feature conditions and branches that represent outcomes, ultimately leading to leaf nodes associated with predicted classes (Fig. 3). During training, the tree splits data recursively, aiming to reduce impurity or maximize information gain, resulting in a set of decision rules that are easy to follow and explain.

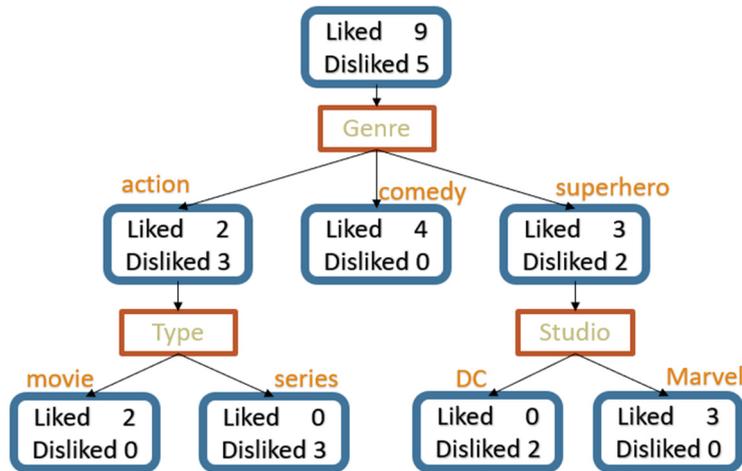


Fig. 3. An example of a decision tree

In recommendation systems, decision trees apply these rules to evaluate item relevance for each user, enabling personalized suggestions. Their visual structure enhances transparency, making them accessible to stakeholders without deep technical expertise. Decision trees support both categorical and numerical features and offer efficient performance suitable for large-scale systems.

However, the effectiveness of decision trees relies heavily on the quality of the input features. Proper feature engineering and preprocessing are critical to accurately model user preferences. Additionally, decision trees can overfit the training data if allowed to grow too deep, capturing noise rather than general trends. To mitigate overfitting, it's crucial to tune the hyperparameters of the decision tree, such as the maximum depth and the minimum number of samples required to split a node, ensuring better generalization to unseen data.

Classic classification algorithms often face challenges in adapting to new data. Traditional models are typically static and require retraining to incorporate additional user interactions. This can be inefficient and difficult to scale, especially with large datasets. Although some modifications to standard algorithms support incremental learning, allowing the model to update with new data without full retraining, implementing such methods can be complex and resource-intensive in practice. As user preferences evolve over time, maintaining model relevance becomes critical. Relying on outdated models risks degrading recommendation quality, while frequent retraining demands substantial computational resources.

To address this challenge, the technique of relevance feedback has emerged as a promising solution. Unlike traditional approaches, relevance feedback empowers systems to dynamically adjust existing models based on user feedback. At the forefront of relevance feedback algorithms stands the Rocchio algorithm, renowned for its effectiveness in refining recommendations in response to user interactions (Meteren, & Someren, 2000). The Rocchio algorithm operates within the vector-space model, necessitating the transformation of unstructured data into a structured format. In this model, each item is represented as a vector in an n-dimensional space (Fig. 4), where n corresponds to the number of characteristics or features.

Central to the Rocchio algorithm is the concept of user profiling, where each user is characterized by a vector that reflects their preferences. This vector is engineered to exhibit high correlation with positively rated items while displaying lower correlation with negatively rated ones. The essence of the Rocchio algorithm can be succinctly captured by the following formula:

$$P' = a \cdot P + b \cdot ST(rel) - c \cdot ST(non - rel). \quad (1)$$

Here, P' represents the updated user profile, P denotes the original user profile vector, $ST(rel)$ signifies the vector sum of positively rated items, and $ST(non - rel)$ represents the vector sum of negatively rated items. Hence, when incorporating a newly rated item into the user's profile, a simple operation suffices. If the item T is deemed relevant, the updated user profile P' is calculated as $P' = a \cdot P + b \cdot T$. Conversely, if the item T is assessed as non-relevant, the updated user profile is determined as $P' = a \cdot P - c \cdot T$. This streamlined process enables the continual refinement of the user's profile, ensuring it remains highly correlated with relevant yet unrated items.

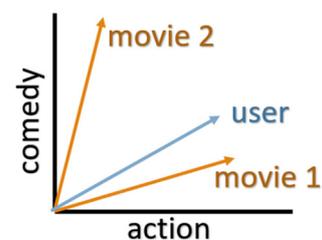


Fig. 4. An example of the user and items profile representation in a vector space model



The coefficients a , b and c modulate the influence of each component on the updated user profile, allowing for fine-grained adjustments based on user feedback. Parameter a manages the balance between past preferences and new interests, ensuring the profile retains useful historical data while adapting to changes. Parameter b determines how quickly the profile updates when encountering relevant items, helping it learn new preferences efficiently. Conversely, parameter c regulates the impact of non-relevant items, preventing incorrect recommendations from distorting the user profile. By carefully tuning these values, recommender systems can better adapt to individual user behavior, improving recommendation accuracy and overall user satisfaction. Continuous updates based on feedback allow the profile to evolve dynamically, making recommendations more responsive and relevant over time.

In content-based recommender systems, various similarity measures, including Euclidean distance, Pearson correlation, and Jaccard similarity, as well as other methods, are used to evaluate the similarity between user profiles and items based on their features (Deutschman, 2023). Each of these methods serves different purposes and is suitable for different types of data and requirements.

Euclidean distance measures the straight-line distance between two points in a multi-dimensional space (2), which can be visualized as the length of the line segment connecting them. It is particularly useful in cases where the magnitude of the vectors is meaningful and should be taken into account. For instance, when the features of the vectors represent absolute values, such as the number of times a user has interacted with certain items or the ratings a user has given, Euclidean distance can effectively capture the overall difference between two users or items:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2}. \quad (2)$$

Cosine similarity quantifies the similarity between two vectors by measuring the cosine of the angle between them in a multi-dimensional space (3). It is particularly effective in scenarios where the magnitude of the vectors is less important than their direction. This characteristic makes cosine similarity ideal for identifying items that are similarly oriented in a high-dimensional space, regardless of their length. For instance, when comparing articles of varying lengths, cosine similarity can effectively find articles that are conceptually similar, even if one article is much longer than another:

$$S_c(x, y) = \cos(\theta) = \frac{x \cdot y}{\|x\| \|y\|} = \frac{\sum_{i=1}^n x_i \cdot y_i}{\sqrt{\sum_{i=1}^n x_i^2} \cdot \sqrt{\sum_{i=1}^n y_i^2}} \quad (3)$$

Pearson correlation measures the linear correlation between two variables (4). It is commonly used in recommender systems to assess the strength and direction of the relationship between two variables. Pearson correlation indicates the degree of similarity between users' preferences. A high positive correlation suggests similar preferences, while a negative correlation suggests divergent preferences. However, Pearson correlation can be problematic with sparse data, as it may not accurately reflect the similarity between items with few common features. In such cases, cosine similarity might be a better choice (Gaurav, 2023):

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4)$$

Jaccard similarity measures the similarity between two sets by dividing the size of the intersection by the size of the union of the two sets (5). It is particularly useful in recommender systems for binary data, such as tag-based recommendations. For instance, in a movie recommender system, Jaccard similarity can be used to determine how similar two movies are based on their top tags:

$$J(X, Y) = \frac{|X \cap Y|}{|X \cup Y|} = \frac{|X \cap Y|}{|X| + |Y| - |X \cap Y|} \quad (5)$$

Semantic similarity and relatedness between words is an important aspect of natural language processing tasks, which can enhance the performance of various applications within recommender systems. By going beyond surface-level comparisons, semantic similarity measures assess the meaning and contextual relationships between words. These methods enable the construction of algorithmic models that enhance context-linguistic analysis, addressing challenges such as word sense disambiguation, named entity recognition, and text analysis. A particularly effective approach involves estimating the semantic distance between words through a weighted modification of the Lesk algorithm (Anisimov, Marchenko, & Kysenko, 2011). By incorporating these refined semantic similarity measures, recommender systems can better understand and utilize the nuanced relationships between words, leading to more precise and contextually relevant recommendations.

These similarity measures have been compared on a real learning platform dataset (Joy, & Renumol, 2020). The study evaluated Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) metrics. It was found that Jaccard similarity and Euclidean distance perform similarly well (MAE 0.66, RMSE 0.91) and better than Pearson correlation (MAE 0.72, RMSE 0.94), while cosine similarity outperforms them all (MAE 0.6, RMSE 0.87). Another study (Wijewickrema, Petras, & Dias, 2019) compared normalized discounted cumulative gain (NDCG). This study showed that BM25 measure can outperform cosine similarity in a journal recommender system – BM25 achieved an average NDCG score of 0.626 for the social science domain and 0.615 for the medicine domain, compared to 0.436 and 0.469 for cosine similarity respectively.

A comparative study has been performed to evaluate the time performance of various distance measures using different sizes of datasets (Marchenko, & Shevchenko, 2024). The study revealed that cosine similarity, inner product, and Euclidean distance produced comparable results in content-based filtering, showing no significant differences in retrieval performance. Also, since feature vectors in content-based filtering have a predefined shape that does not depend on dataset size, the response time for all three measures exhibited a linear growth pattern as the dataset size increased.

These studies are showing that there is no one-size-fits-all answer to which similarity measure performs best in content-based filtering. The optimal choice depends on the characteristics of the data and the specific needs of the recommender system. Experimentation and evaluation using real-world data are crucial to determining the most effective similarity measure for a given application.

Collaborative filtering. Rather than relying on item attributes or content descriptors, collaborative filtering algorithms analyze user-item interactions, such as ratings, purchases, or clicks, represented as a utility matrix (Fig. 5). By identifying users with similar preferences and recommending items favored by their peers, collaborative filtering endeavors to deliver tailored recommendations that resonate with individual tastes. This approach is widely used in various domains, including product recommendations on e-commerce sites, movie and TV show recommendations on streaming platforms and article recommendations on news sites.

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6
User 1	4		5			
User 2	3					4
User 3				2		5
User 4	5	4	5			
User 5		3	5		4	

Fig. 5. An example of a utility matrix

Model-based approaches represent a paradigm in collaborative filtering, where intricate models are crafted to capture underlying patterns and relationships inherent in user-item interactions, and are encompassing various techniques (Grover, 2017). Matrix factorization techniques, such as Singular Value Decomposition (Nguyen, 2016), Principal Component Analysis (Yadav et al., 2021), Nonnegative Matrix Factorization (Anisimov, Marchenko, & Vozniuk, 2014) and Alternating Least Squares (Gosh et al., 2021), as well as the tensor approach (Marchenko, 2016), decompose the user-item interaction matrix into lower-dimensional representations. These latent factors capture user preferences and item characteristics, enabling accurate prediction of user-item interactions. Probabilistic graphical models, such as Bayesian networks (Qazi et al., 2017) or Markov random fields (Steck, 2019), represent the dependencies between users and items probabilistically. Deep learning architectures, such as neural networks or autoencoders (Liu et al., 2018), learn hierarchical representations of user-item interactions. By leveraging multiple layers of abstraction, deep learning models capture intricate patterns and nuances in the interaction data, enabling the generation of highly accurate recommendations.

Unlike model-based approaches that rely on learned parameters, memory-based methods directly compute similarities between users or items to infer preferences and foster recommendation generation. A widely used technique within memory-based collaborative filtering is the k-Nearest Neighbors (k-NN) algorithm (Analytics Vidhya, 2024). The k-NN identifies similar users or items by computing distances in the feature space, based on the assumption that the target user shares preferences with their nearest neighbors. Memory-based collaborative filtering encompasses user-based and item-based approaches (Belhaouari et al., 2023).

User-based approach identifies k users who are similar to the target user based on their past behavior or ratings. The similarity between users is typically measured by the similarity of their ratings or interactions with items. These similar users serve as a reference group whose preferences are aggregated to generate recommendations for the target user. It recommends items that these similar users have liked but the target user has not yet interacted with (Fig. 6). This facilitates the discovery of relevant items that align with the preferences of the target user (Vijay, 2020).

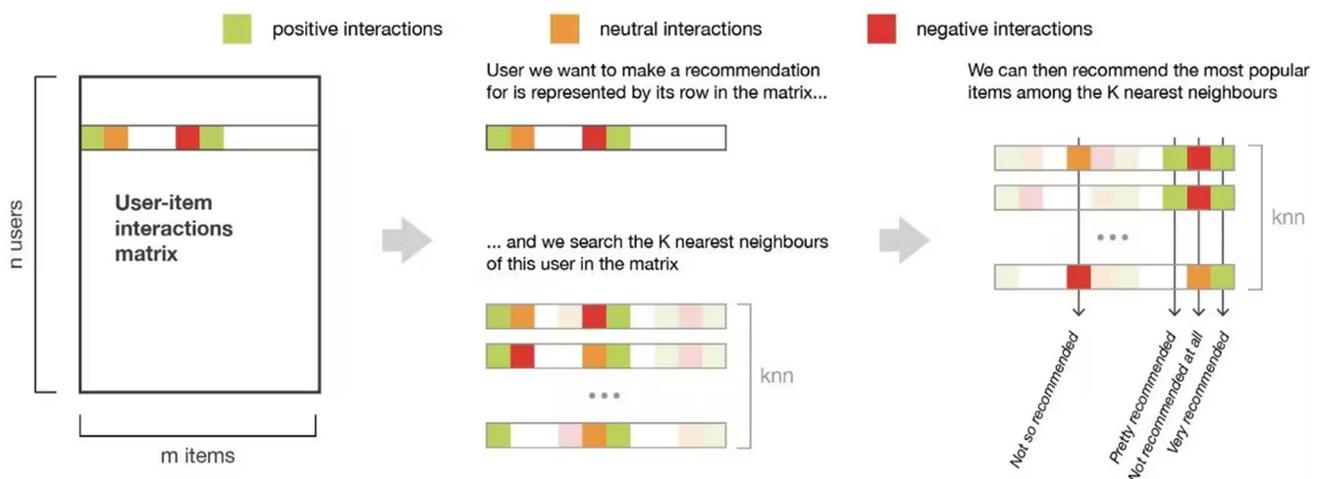


Fig. 6. An illustration of the user-based collaborative filtering algorithm (Vijay, 2020)

Item-based method focuses on the items themselves rather than the users. K-NN operates by identifying items that are similar to those a user has interacted with or rated highly (Fig. 7). The similarity between items is determined by the similarity of the ratings of those items by the users who have rated both items. The algorithm finds the k items that are most similar to the items the user has interacted with and recommends these items to the user.

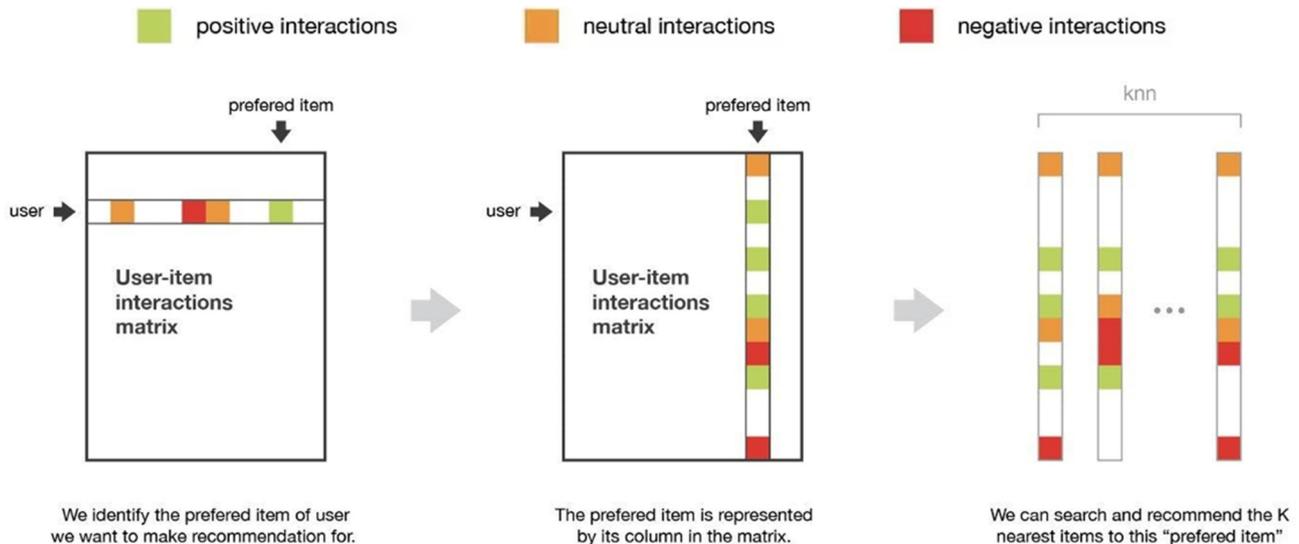


Fig. 7. An illustration of the item-based collaborative filtering algorithm (Vijay, 2020)

Collaborative filtering and content-based filtering may differ in their underlying principles, but they often share common ground when it comes to similarity measures. The same similarity measures utilized in content-based filtering can also be seamlessly integrated into collaborative filtering frameworks, enriching the recommendation process with insights derived from user-item interactions (Kumar, 2022).

The experimental study (Fkih, 2022) demonstrates that no single similarity measure is universally effective for both user-based and item-based approaches, with some measures performing better in one approach than the other. Additionally, the study highlights that the effectiveness of a similarity measure can vary based on dataset density, even within the same filtering approach. The study (Saranya, Sadasivam, & Chandralekha, 2016) proposed a new similarity measure that incorporates Pearson Correlation Coefficient and Jaccard Coefficient. The proposed combined similarity measure outperformed the Bhattacharyya Coefficient and the Jaccard Uniform Operator Distance, and achieved an average F-score measure of 0.77 compared to 0.72 and 0.75 in the competitors respectively. Another study (Sun et al., 2017) proposed integrating Triangle and Jaccard similarities for recommendation to enhance recommender system performance. This combination allows the system to leverage the strengths of different measures, addressing their individual limitations and providing a more robust and adaptable recommendation process. The proposed measure achieved MAE of 0.614 and RSME of 0.816 for the FilmTrust dataset, compared to MAE of 0.617–0.852 and RSME of 0.817–1.043 for other similarity measures.

Another research investigated the impact of similarity measures on recommendation time (Baxla, 2014). The findings revealed that correlation-based measures require significantly more time for execution. One more comparative study (Marchenko, & Shevchenko, 2024) demonstrated that cosine similarity has significantly worse time performance on high-dimensional datasets compared to inner product and Euclidean distance. Additionally, the experiment revealed that due to data sparsity, the response time on a high-dimensional collaborative filtering dataset can be lower than on a lower-dimensional content-based filtering dataset. Therefore, when selecting an appropriate similarity measure, it is crucial to consider not only the accuracy but also the computational efficiency.

Results

Pros and cons. Comparative analysis. Content-based filtering relies on structured representations of item attributes and user profiles derived from explicit or implicit individual feedback. Because recommendations are generated by matching item features to learned user preferences, the underlying reasoning is typically transparent and easy to interpret, especially in rule-based models, decision trees, or relevance feedback approaches such as the Rocchio algorithm (Ricci, 2002). From a data perspective, these methods are particularly well suited to environments with rich and expressive metadata, where item characteristics can be encoded as fixed-dimensional vectors obtained from metadata, textual descriptions, or semantic representations. Under such conditions, comparative studies report stable and reliable recommendation accuracy, provided that similarity measures are appropriately selected and features adequately capture user interests (Joy, & Renumol, 2020).

From a computational standpoint, most content-based algorithms are efficient and scalable. Similarity calculations between user profiles and items typically exhibit linear time complexity with respect to the number of items, while fixed-dimensional feature spaces enable predictable performance even in large and frequently updated catalogs (Marchenko, & Shevchenko, 2024). Another notable advantage is robustness to the item cold-start problem: new items can be recommended immediately based solely on their attributes, without requiring historical user-item interaction data or information about other users' preferences (Herimanto, Samosir, & Ginting, 2024). This self-sufficiency also allows content-based systems to effectively serve users with niche or specialized interests, including items that may be unpopular in the broader population.

Despite these strengths, content-based filtering suffers from several limitations. Its effectiveness is highly sensitive to feature engineering quality: incomplete, noisy, or poorly designed item representations can substantially degrade recommendation performance and narrow the recommendation space (Zisopoulos et al., 2008). Typical implementation pitfalls include inadequate preprocessing, inappropriate similarity measure selection for the data type, and lack of proper feature normalization, all of which may bias similarity computations. Moreover, content-based filtering systems are prone to over-specialization, repeatedly recommending items that are very similar to those previously consumed and thereby limiting diversity and serendipity (Ricci, 2002).



Finally, content-based filtering may exhibit unstable or unsatisfactory behavior in dynamic or subjective domains. Because recommendations are constrained to known item characteristics, these methods struggle to introduce entirely novel or conceptually different items and often fail to capture rapidly evolving user interests without frequent retraining. As a result, delayed adaptation and increased computational overhead can occur when user preferences change quickly. Such weaknesses are especially evident in scenarios with sparse or poorly defined metadata, context-dependent or subjective tastes, and highly dynamic user behavior, where static content representations are insufficient to model the true drivers of user satisfaction (Herimanto, Samosir, & Ginting, 2024).

Unlike content-based approaches, collaborative filtering does not require explicit item features or rich metadata, allowing it to recommend items even when descriptive attributes are incomplete or unavailable (Vijay, 2020). By exploiting patterns in collective user behavior, collaborative filtering can uncover latent preferences and recommend items that users might not have discovered on their own, thereby introducing serendipity and promoting exposure to diverse content. When sufficient interaction data are available, both memory-based and model-based methods demonstrate high recommendation accuracy by capturing shared interests and implicit behavioral signals (Fkih, 2022).

From an accuracy and adaptability perspective, collaborative filtering is particularly effective in domains where subjective preferences dominate and are difficult to encode through explicit features. Model-based approaches, such as matrix factorization and probabilistic models, are especially successful at learning latent structures that generalize beyond observed interactions, often outperforming simpler methods in dense data settings (Nguyen, 2016; Gosh et al., 2021). However, this advantage diminishes in sparse environments, where unreliable similarity estimates and limited historical data reduce recommendation quality (Belhaouari et al., 2023). Adaptability to evolving user interests varies by method: memory-based approaches naturally incorporate recent interactions, whereas model-based techniques require periodic retraining or incremental updates to reflect preference shifts.

Computational complexity and scalability represent central challenges for collaborative filtering. Memory-based methods, including user-based and item-based k-NN, incur substantial computational costs due to pairwise similarity calculations over high-dimensional and sparse matrices, which can lead to unstable response times without proper optimization (Marchenko, & Shevchenko, 2024). Model-based methods mitigate online computation by shifting complexity to an offline training phase, enabling faster recommendation at runtime but at the cost of increased training overhead and resource consumption. As a result, large-scale collaborative filtering-based systems often rely on dimensionality reduction, approximation techniques, or distributed computing frameworks to achieve acceptable scalability (Nguyen, 2016).

In terms of interpretability, collaborative filtering generally offers limited transparency compared to content-based methods. Recommendations are derived from aggregated user behavior or latent factors rather than explicit item characteristics, making it difficult to provide clear explanations to end users. While this opacity may be acceptable in some applications, it can reduce user trust and hinder system debugging or evaluation, particularly in high-stakes domains (Fkih, 2022).

Despite its strengths, collaborative filtering is subject to several fundamental limitations and common pitfalls. The cold-start problem remains a major challenge, as new users and new items lack sufficient interaction data to generate reliable recommendations (Belhaouari et al., 2023). Data sparsity further exacerbates this issue, leading to noisy or biased similarity estimates and degraded accuracy, especially in user-based approaches. Popularity bias is another frequent concern, whereby frequently interacted items are over-recommended, limiting diversity and marginalizing niche content (Vijay, 2020). Moreover, inappropriate choices of similarity measures, neighborhood size, or update strategies can significantly harm both accuracy and efficiency (Baxla, 2014).

Hybrid approach. Hybrid recommender systems combine the strengths of both content-based and collaborative filtering approaches to enhance the effectiveness of recommendations (Wayesa et al., 2023). This approach is particularly useful when there is insufficient data to effectively apply either method alone. By leveraging both types of data, hybrid systems can offer more personalized and accurate recommendations.

Hybrid recommender systems have been extensively studied as a means to improve recommendation quality by combining multiple filtering techniques. The foundational work (Burke, 2007) introduced a systematic framework for hybridization, focusing on integrating collaborative filtering and content-based methods. Based on the interaction patterns between constituent models, hybrid recommender systems can be classified into several major types listed below, each addressing specific challenges of recommendation tasks to a different extent.

The weighted hybrid approach (Fig. 8) combines the outputs of multiple recommendation models using predefined or learned weights. Typically, predictions generated by content-based and collaborative filtering models are linearly aggregated to produce the final recommendation score. This approach effectively improves recommendation accuracy by balancing complementary information sources and partially alleviates data sparsity issues. However, weighted hybrids do not fully resolve the cold-start problem for new users, as collaborative components still require sufficient interaction history. Additionally, static weighting schemes may limit adaptability to dynamic user behavior and may lead to suboptimal performance if weights are not carefully tuned.

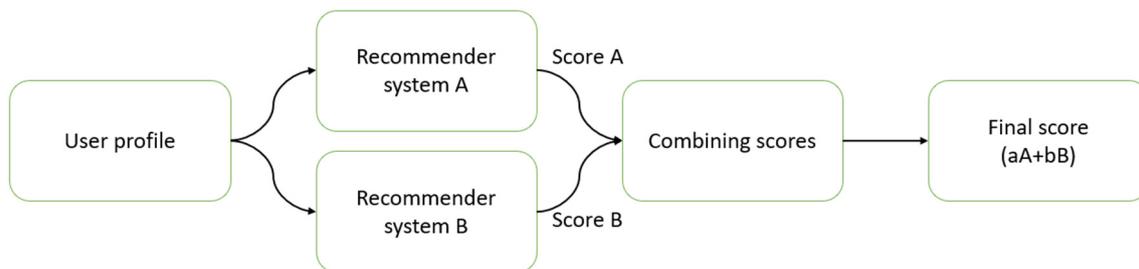


Fig. 8. Weighted hybrid recommender system



In switching hybrid systems (Fig. 9), the recommendation process dynamically selects one of several models based on contextual conditions, such as user profile completeness, interaction history, or item availability. This approach explicitly addresses the cold-start problem by activating content-based methods for new users or items and switching to collaborative filtering when sufficient data becomes available. Switching hybrids offer improved robustness across heterogeneous usage scenarios but do not directly combine model outputs, which may limit accuracy gains compared to weighted or mixed hybrids. Furthermore, incorrect switching criteria can lead to unstable or inconsistent recommendations.

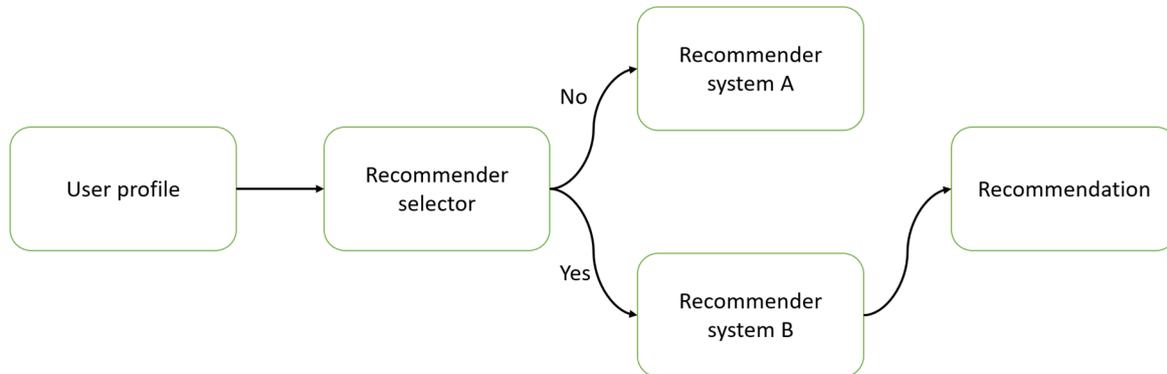


Fig. 9. Switching hybrid recommender system

The mixed hybrid (Fig. 10) approach generates recommendations from multiple models simultaneously and merges the resulting candidate lists into a unified output. This strategy enables the system to exploit diverse recommendation signals in parallel and is particularly effective for increasing recommendation coverage and diversity. Mixed hybrids can handle sparse and partial datasets more effectively by matching subsets of users or items to appropriate models. However, this approach increases computational complexity and may introduce redundancy or conflicting recommendations if result aggregation is not properly managed.

Cascaded hybrid systems (Fig. 11) employ a hierarchical structure in which a primary recommendation model generates an initial ranked list, while secondary models refine the output by resolving ambiguities, such as tie-breaking or missing values. This approach is well suited for improving precision and ranking quality without significantly increasing computational overhead. Cascading hybrids, however, depend heavily on the quality of the primary model and do not inherently address cold-start or data sparsity problems unless explicitly incorporated into the cascade structure.

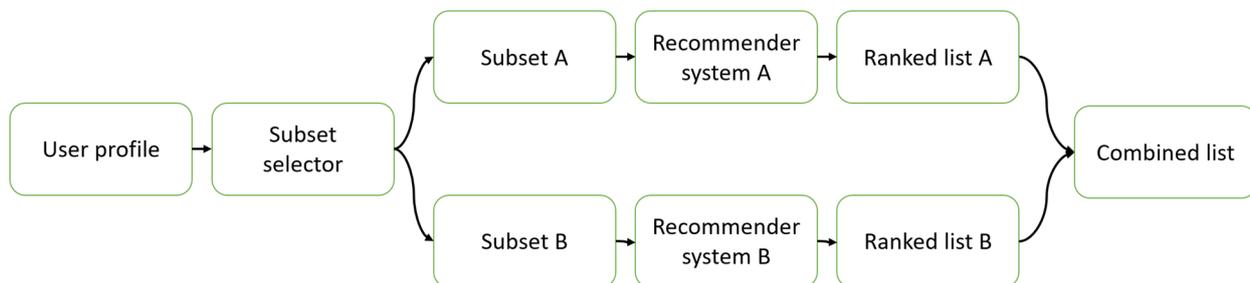


Fig. 10. Mixed hybrid recommender system

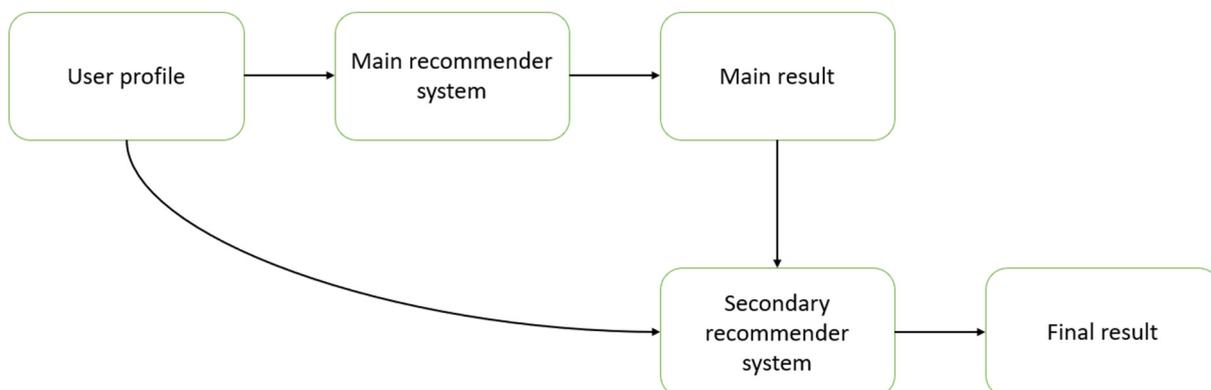


Fig. 11. Cascaded hybrid recommender system

Feature augmentation (Fig. 12) integrates a contributing recommendation model by using its output, such as predicted ratings or class labels, as input features for the main recommendation system. This strategy enhances the representational



power of the core model and improves recommendation accuracy without altering its fundamental structure. Feature augmentation is effective for incorporating collaborative signals into content-based models and partially mitigating data sparsity. Nevertheless, it does not fully eliminate cold-start limitations and may propagate errors from the auxiliary model into the main system.

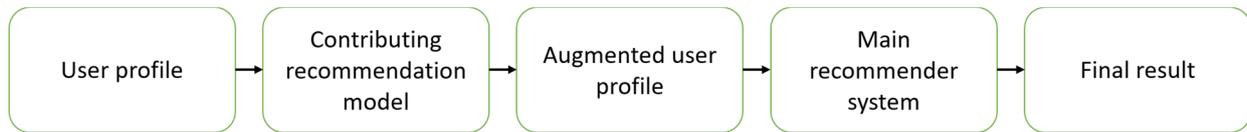


Fig. 12. Feature augmentation hybrid recommender system

The feature combination approach (Fig. 13) merges features derived from different recommendation paradigms into a unified feature space. For example, collaborative latent factors may be injected into content-based user profiles, enabling the model to consider both item attributes and user interaction patterns. This method provides strong flexibility and improves personalization by jointly modeling heterogeneous data. However, feature combination increases model complexity and may suffer from scalability issues in high-dimensional feature spaces. Additionally, interpretability can be reduced due to the integration of latent and explicit features.

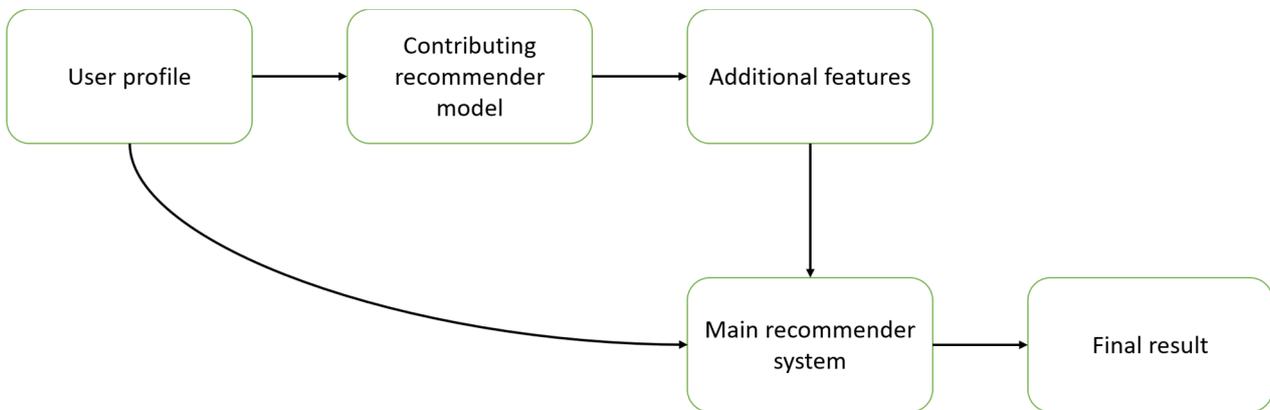


Fig. 13. Feature combination hybrid recommender system

The study (Burke, 2007) highlighted the trade-offs of hybrid systems, including increased computational complexity and the necessity for mindful parameter tuning. Recent systematic reviews on hybrid recommender systems highlight their growing significance in addressing key challenges such as cold-start behavior, data sparsity, and recommendation accuracy. The study (Çano, 2017) found that most hybrid models combine collaborative filtering with other strategies, often using weighted hybridization due to its simplicity and flexibility. The review emphasized that cold-start and data sparsity remain the most frequently addressed issues. Moreover, accuracy metrics dominate the assessment of hybrid recommenders, but user satisfaction evaluations remain a challenge. Another study (Sabiri et al, 2025) provided a broader perspective by exploring hybridization techniques within big data environments. This review underscored the increasing adoption of machine learning-based hybridization methods and the necessity for scalability improvements. The review also highlighted the importance of balancing accuracy with computational efficiency, suggesting the use of hybrid parallel algorithms and large-scale datasets to address scalability concerns. Ethical considerations and user engagement metrics were also noted as emerging research priorities.

These studies emphasize that while hybrid recommender systems provide substantial benefits, their successful implementation requires significant computational resources and careful design considerations. In many cases it also requires distributed or parallel processing to maintain scalability and acceptable response times (Çano, 2017). Interpretability varies widely across hybridization strategies: weighted and cascaded models tend to preserve a moderate level of transparency, whereas deeply integrated or machine learning-based hybrids are less interpretable due to latent representations and multi-stage transformations (Liu et al., 2018).

From a performance and robustness perspective, hybrid recommender systems demonstrate higher accuracy and greater adaptability to evolving user preferences. The content-based component enables recommendations for new users or items with limited interaction history, while the collaborative component contributes serendipity and discovery of novel items beyond a user's immediate profile. This balance makes hybrid systems especially suitable for large-scale, real-world applications characterized by heterogeneous data and dynamic user behavior (Sabiri et al., 2025).

Despite their overall robustness, hybrid recommender systems are not immune to instability or suboptimal performance. Poorly chosen integration strategies, imbalanced weighting between components, or misalignment between algorithmic assumptions and data characteristics can lead to degraded accuracy or inconsistent behavior. Additional risks include overfitting due to model complexity, inefficient handling of high-dimensional data, and unstable recommendations in extremely sparse, noisy, or rapidly changing environments. These limitations highlight the importance of context-aware hybridization.

The continued evolution of hybridization methods, particularly through machine learning and contextual adaptation, remains a key focus in recommender system research. Despite the fact that designing and implementing hybrid recommender systems can be complex, they can offer more comprehensive and effective recommendations, addressing the limitations of using content-based or collaborative filtering alone.



Modern challenges in recommender systems. Several fundamental challenges of recommender system such as the cold start problem, data sparsity, and scalability have been discussed in earlier sections of this paper. While these issues are central to recommendation accuracy and system performance, they do not fully capture the complexity of deploying recommender systems in real-world applications. In practice, a number of additional challenges must be considered to ensure robustness, efficiency, fairness, and user trust in operational systems.

From a system-level perspective, latency remains a critical concern, especially in interactive and time-sensitive applications such as e-commerce and online advertising. Even when scalability challenges are addressed through distributed architectures, complex recommendation models may still introduce unacceptable response delays. High latency negatively affects user experience and can lead to reduced engagement or abandonment. Common mitigation approaches include offline precomputation of recommendations, model simplification for real-time inference, and approximate similarity search techniques that trade minimal accuracy loss for substantial speed improvements (Roy, & Dutta, 2022).

Beyond technical performance, bias and fairness have emerged as key ethical challenges in modern recommender systems. Recommendation algorithms may unintentionally reinforce existing biases present in training data, resulting in popularity bias, exposure bias, or demographic bias. Such effects can limit content diversity, disadvantage minority groups, and raise concerns about transparency and accountability. Addressing these issues requires fairness-aware modeling, diversity-promoting re-ranking strategies, and regular auditing of recommendation outputs to balance accuracy with equitable exposure (Mehrabi et al., 2021).

One important yet often overlooked challenge is the vulnerability of recommender systems to shilling attacks, also referred to as profile injection attacks. In such attacks, malicious users or automated agents intentionally introduce biased or fraudulent interactions to manipulate recommendation outcomes, for example by artificially promoting specific items or suppressing competitors. These attacks can significantly distort ranking mechanisms and undermine the credibility of recommendation platforms. Mitigation strategies include anomaly detection techniques, trust-aware recommendation models, and machine learning approaches that identify abnormal rating patterns and suspicious user behavior (Gunes et al., 2014).

Another challenge arises from semantic ambiguity and synonymy in item descriptions and user queries. In content-based and knowledge-aware recommender systems, different terms may refer to the same concept, leading to fragmented representations and reduced retrieval effectiveness. This issue is particularly prevalent in systems that rely on textual metadata or user-generated content. Advances in natural language processing, such as latent semantic analysis and neural word embedding models, have been shown to alleviate synonymy by capturing semantic similarity in latent representation spaces (Roy, & Dutta, 2022).

Privacy concerns are increasingly important as recommender systems rely on extensive collection and processing of personal user data. The use of behavioral signals and detailed user profiles raises risks related to data leakage, unauthorized inference, and regulatory non-compliance. Privacy-preserving techniques, including differential privacy, federated learning, and secure computation, have been proposed to reduce these risks while maintaining recommendation effectiveness. Moreover, transparent data usage policies and user-controlled privacy settings play a crucial role in fostering trust and long-term adoption of recommendation technologies (Yang et al., 2019).

Modern recommender systems must operate under complex and often conflicting constraints, balancing accuracy, efficiency, robustness, fairness, and privacy. Addressing these challenges requires integrated solutions that combine advances in machine learning, distributed systems, and ethical AI, moving beyond purely performance-driven optimization toward socially responsible recommendation technologies.

Discussion and conclusions

The analytical comparison conducted in this study demonstrates that the effectiveness of recommender systems is determined not by the choice of a single algorithmic paradigm, but by the alignment between model assumptions, data characteristics, and operational constraints. As summarized in Table 1, content-based and collaborative filtering exhibit complementary strengths that manifest across multiple evaluation dimensions, including data requirements, interpretability, scalability, robustness, and adaptability.

The analysis confirms that content-based filtering is most reliable in environments with well-structured and expressive item metadata, where interpretability and predictable time performance are critical. Its stable computational behavior and robustness to item cold-start scenarios make it particularly suitable for domains with rapidly evolving catalogs or limited interaction data. However, empirical evidence reviewed in this work indicates that recommendation quality in content-based systems is highly sensitive to feature engineering choices and similarity measures, and that over-specialization remains a persistent limitation in practice.

Collaborative filtering, in contrast, demonstrates superior capability in uncovering latent user preferences and promoting serendipity when interaction data are sufficiently dense. Both memory-based and model-based approaches achieve high accuracy under favorable data conditions, yet the analysis highlights substantial trade-offs in terms of scalability, interpretability, and robustness to sparsity. Experimental findings reviewed in the paper further indicate that similarity measure selection and computational cost play a critical role in large-scale deployments, often requiring approximation techniques or offline training pipelines to maintain acceptable response times.

Hybrid recommender systems emerge from the analysis as a practically validated strategy for mitigating the structural limitations of single-method approaches. Different hybridization schemes address different problem dimensions: switching hybrids effectively handle cold-start scenarios, weighted and feature-level hybrids improve accuracy and robustness, while cascading approaches refine ranking quality with limited overhead. At the same time, the review makes clear that hybrid systems introduce increased model complexity, higher maintenance costs, and sensitivity to integration design choices, which can lead to unstable behavior if not carefully managed.

Several open challenges remain evident from the analytical perspective of this study. First, there is no universally optimal hybridization strategy. Performance strongly depends on domain characteristics, data quality, and system constraints. Second, evaluation practices remain fragmented, with accuracy metrics dominating experimental studies, while user-centric measures such as satisfaction, trust, and long-term engagement are less consistently addressed. Finally, scalability, fairness,



and privacy considerations increasingly shape real-world recommender system design but are not yet systematically integrated into comparative evaluation frameworks.

Table 1

Comparative analysis of recommendation approaches, summary

Criterion	Content-based filtering	Collaborative filtering	Hybrid filtering
Core principle	Recommends items similar to those a user liked in the past based on item features	Recommends items based on preferences of similar users or items	Combines content-based and collaborative signals using different hybridization strategies
Required data	Item metadata and user profiles	User-item interaction matrix	Item metadata, interaction data, and possibly contextual features
Personalization level	High (user-specific models)	High (community-driven patterns)	High, with increased robustness due to multi-source signals
Recommendation accuracy	Moderate; limited by feature quality	High when sufficient interaction data is available	Generally higher than single methods if properly tuned
Ability to recommend novel items	Strong (does not rely on popularity)	Weak (new items lack interactions)	Improved compared to collaborative filtering; depends on hybrid type
Cold-start (user)	Handles reasonably well	Performs poorly	Typically mitigated, but not fully eliminated
Cold-start (item)	Handles well if metadata is available	Performs poorly	Largely alleviated through content signals
Diversity and serendipity	Low to moderate; tends to over-specialization	Higher diversity due to social effects	Higher than content-based filtering; controllable via hybrid design
Explainability	High (feature-based explanations)	Low to moderate	Medium to high, depending on dominant component
Scalability	Scales well with users; depends on feature dimensionality	Scalability issues for large user-item matrices	Often more computationally expensive than single methods
Robustness to noise	Sensitive to poor or sparse metadata	Sensitive to noisy or biased interaction data	More robust due to redundancy of information sources
Adaptability to context	Limited unless explicitly modeled	Limited	High when contextual or feature-level hybrids are used
Model complexity	Low to moderate	Moderate to high	High, varies by hybridization strategy
Implementation complexity	Relatively simple	More complex (similarity learning, factorization)	Most complex; requires integration and tuning
Maintenance cost	Low	Moderate	High (multiple models, parameters, and pipelines)
Typical failure cases	Overspecialization, feature misrepresentation	Data sparsity, popularity bias	Error propagation, overfitting, misbalanced model dominance
Best-suited scenarios	Domains with rich item descriptions and stable user interests	Large-scale platforms with dense interaction data	Production systems requiring robustness and higher accuracy

Future research should therefore focus on developing standardized, multi-criteria evaluation methodologies that jointly assess accuracy, efficiency, interpretability, and ethical considerations. Promising directions include context-aware and adaptive hybridization strategies, integration of lightweight machine learning models to balance performance and complexity, and exploration of privacy-preserving and fairness-aware recommendation techniques. From a practical standpoint, the findings of this study suggest that robust recommender systems should be designed as adaptive architectures rather than static algorithmic choices, with hybrid models serving as a flexible foundation for addressing evolving data and application requirements.

Sources of funding. This study did not receive any grant from a funding institution in the public, commercial, or non-commercial sectors. The funding is covered by the authors' own expense.

References

Anisimov, A. V., Marchenko, O. O., & Kysenko, V. K. (2011). A method for the computation of the semantic similarity and relatedness between natural language words. *Cybernetics and Systems Analysis*, 47, 515–522. <https://doi.org/10.1007/s10559-011-9334-2>

Anisimov, A. V., Marchenko, O. O., & Vozniuk, T. G. (2014). Determining Semantic Valences of Ontology Concepts by Means of Nonnegative Factorization of Tensors of Large Text Corpora. *Cybernetics and Systems Analysis*, 50, 327–337. <https://doi.org/10.1007/s10559-014-9621-9>

Baxla, M. A. (2014). *Comparative study of similarity measures for item based top n recommendation* [Unpublished thesis, National Institute of Technology Rourkela]. CORE. <https://files.core.ac.uk/download/53190130.pdf>

Belhaouari, S. B., Fareed, A., Hassan, S., & Halim, Z. (2023). A collaborative filtering recommendation framework utilizing social networks. *Machine Learning with Applications*, 14, 1–20. <https://doi.org/10.1016/j.mlwa.2023.100495>



- Burke, R. (2007). Hybrid Web Recommender Systems. In P. Brusilovsky, A. Kobsa, & W. Nejdl (Eds.), *Lecture Notes in Computer Science*, 4321. *The Adaptive Web* (pp. 377–408). Springer. https://doi.org/10.1007/978-3-540-72079-9_12
- Çano, E. (2017). Hybrid Recommender Systems: A Systematic Literature Review. *Intelligent Data Analysis*, 21, 1487–1524. <https://doi.org/10.3233/IDA-163209>
- Deutschman, Z. (2023, August 7). *Recommender Systems: Machine Learning Metrics and Business Metrics*. Neptune AI. <https://neptune.ai/blog/recommender-systems-metrics>
- Fkih, F. (2022). Similarity measures for Collaborative Filtering-based Recommender Systems: Review and experimental comparison. *Journal of King Saud University - Computer and Information Sciences*, 34(9), 7645–7669. <https://doi.org/10.1016/j.jksuci.2021.09.014>
- Gaurav, P. (2023, February 14). *Step by Step Content-Based Recommender system*. Medium. <https://medium.com/@prateekgaurav/step-by-step-content-based-recommendation-system-823bbfd0541c>
- Gershman, A., Meisels, A., Luke, K.-H., Rokach, L., Schlar, A., & Sturm, A. (2010). A Decision Tree Based Recommender System. In G. Eichler, P. Kropf, U. Lechner, P. Meesad, & H. Unger (Eds.), *10th International Conference on Innovative Internet Community Services: Vol. 165. Lecture Notes in Informatics* (pp. 170–179). Gesellschaft für Informatik. <https://dl.gi.de/server/api/core/bitstreams/ca0e5035-3a82-48a1-8eb8-8f49ee374161/content>
- Gosh, S., Nahar, N., Wahab, M. A., Biswas, M., Hossain, M. S., & Andersson, K. (2021). Recommender system for E-commerce Using Alternating Least Squares (ALS) on Apache Spark. In P. Vasant, I. Zelinka, & G. W. Weber (Eds.), *Intelligent Computing and Optimization: Vol. 1324. Advances in Intelligent Systems and Computing* (pp. 880–893). Springer. https://doi.org/10.1007/978-3-030-68154-8_75
- Grover, P. (2017, December 28). *Various Implementations of Collaborative Filtering*. Medium. <https://towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0>
- Gunes, I., Kaleli, C., Bilge, A., & Polat, H. (2014). Shilling attacks against recommender systems: a comprehensive survey. *Artificial Intelligence Review*, 42, 767–799. <https://doi.org/10.1007/s10462-012-9364-9>
- Herimanto, H., Samosir, K., & Ginting, F. (2024). A Comparative Analysis of Content-Based Filtering and TF-IDF Approaches for Enhancing Sports Recommendation Systems. *Innovation in research of informatics*, 6(2), 90–97. <https://doi.org/10.37058/innovatics.v6i2.12404>
- Joy, J., & Renumol, V. G. (2020). Comparison of Generic Similarity Measures in E-learning Content Recommender System in Cold-Start Condition. In *IEEE Bombay Section Signature Conference* (pp. 175–179). Institute of Electrical and Electronics Engineers. <https://doi.org/10.1109/IBSSC51096.2020.9332162>
- Kumar, S. (2022, September 25). *Collaborative Filtering based Recommender Systems for Implicit Feedback Data*. Sumit's Diary. <https://blog.reachsumit.com/posts/2022/09/explicit-implicit-cf/>
- Liu, Y., Wang, S., Khan, M. S., & He, J. (2018). A novel deep hybrid recommender system based on auto-encoder with neural collaborative filtering. *Big Data Mining and Analytics*, 1(3), 211–221. <https://doi.org/10.26599/BDMA.2018.9020019>
- Mandal, S., & Maiti, A. (2018). Explicit Feedbacks Meet with Implicit Feedbacks: A Combined Approach for Recommender system. In L.M. Aiello, H. Cherifi, P. Lió, L.M. Rocha, C. Cherifi, R. Lambiotte (Eds.), *7th International Conference on Complex Networks and their Applications: Vol. 813. Studies in Computational Intelligence* (pp. 169–181). Springer. https://doi.org/10.1007/978-3-030-05414-4_14
- Marchenko, O. O. (2016). A Method for Automatic Construction of Ontological Knowledge Bases. Development of a Semantic-Syntactic Model of Natural Language. *Cybernetics and Systems Analysis*, 52, 20–29. <https://doi.org/10.1007/s10559-016-9795-4>
- Marchenko, O., & Shevchenko, M. (2024). Influence of distance measures and data characteristics on time performance in content-based and collaborative filtering datasets. In A. Anisimov, V. Snytyuk, A. Chris, A. Pester, F. Mallet, I. Krak, N. Cogan, O. Chertov, O. Marchenko, S. Bozókí, T. Needham, V. Tsyganok, & V. Vovk (Eds.), *Information Technology and Implementation: Vol. 3909. Central Europe University Repository Workshop Proceedings* (pp. 99–108). CEUR-WS. https://ceur-ws.org/Vol-3909/Paper_8.pdf
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A Survey on Bias and Fairness in Machine Learning. *ACM Computing Surveys*, 54(6), 115. <https://doi.org/10.1145/3457607>
- Meteren, R., & Someren, M. (2000). Using Content-Based Filtering for Recommendation. In *Proceedings of the machine learning in the new information age: Vol. 30. MLnet/ECML2000 workshop* (pp. 47–56). ICS. https://users.ics.forth.gr/~potamias/mlnia/paper_6.pdf
- Analytics Vidhya. (2024, October 14). *Movie Recommendation and Rating Prediction using K-Nearest Neighbors*. <https://www.analyticsvidhya.com/blog/2020/08/recommendation-system-k-nearest-neighbors/>
- Nguyen, A. (2016). *Singular Value Decomposition in Recommender Systems* [Honors project, Texas Christian University]. TCU Digital Repository. <https://repository.tcu.edu/server/api/core/bitstreams/7483e691-6fc0-4a82-9185-3adeb00cde44/content>
- Qazi, M., Fun, G. M., Meissner, K. J., & Fontes, E. R. (2017). An insurance recommendation system using Bayesian networks. In *11th ACM Conference on Recommender Systems* (pp. 274–278). Association for Computing Machinery. <https://doi.org/10.1145/3109859.3109907>
- Ricci, F. (2002). *Content-Based Filtering and Hybrid Methods*. EIA. <http://eia.udg.es/arl/Agentsoftware/3-ContentBasedHybrid.pdf>
- Roy, D., & Dutta, M. (2022). A systematic review and research perspective on recommender systems. *Journal of Big Data*, 9, 59. <https://doi.org/10.1186/s40537-022-00592-5>
- Sabiri, B., Khtira, A., El Asri, B., & Rhanoui, M. (2025). Hybrid Quality-Based Recommender Systems: A Systematic Literature Review. *Journal of Imaging*, 11(1), 12. <https://doi.org/10.3390/jimaging11010012>
- Saranya, K. G., Sadasivam, G. S., & Chandralekha, M. (2016). Performance Comparison of Different Similarity Measures for Collaborative Filtering Technique. *Indian Journal of Science and Technology*, 9(29), 1–8. <https://doi.org/10.17485/ijst/2016/v9i29/91060>
- Steck, H. (2019). Markov Random Fields for Collaborative Filtering. *Advances in Neural Information Processing Systems*, 32, 5473–5484. <https://doi.org/10.48550/arXiv.1910.09645>
- Sun, S.-B., Zhang, Z.-H., Dong, X.-L., Zhang, H.-R., Li, T.-J., Zhang, L., & Min, F. (2017). Integrating Triangle and Jaccard similarities for recommendation. *PLoS ONE*, 12(8), e0183570. <https://doi.org/10.1371/journal.pone.0183570>
- Vijay, H. (2020, April 11). *Recommendation System using kNN*. Auriga. <https://aurigait.com/blog/recommendation-system-using-knn/>
- Wayesa, F., Betalo, M. L., Asefa, G., & Kedir, A. (2023). Pattern-based hybrid book recommender system using semantic relationships. *Scientific Report*, 13, 3693. <https://doi.org/10.1038/s41598-023-30987-0>
- Wijewickrema, M., Petras, V., & Dias, N. (2019). Selecting a text similarity measure for a content-based recommender system: A comparison in two corpora. *The Electronic Library*, 37(3), 506–527. <https://doi.org/10.1108/EL-08-2018-0165>
- Xia, Z., Sun, A., Xu, J., Peng, Y., Ma, R., & Cheng, M. (2024). Contemporary Recommendation Systems on Big Data and Their Applications: A Survey. *IEEE Access*, 12, 196914–196928. <https://doi.org/10.1109/ACCESS.2024.3517492>
- Yadav, V., Shukla, R., Tripathi, A., & Maurya, A. (2021). A New Approach for Movie Recommender System using K-means Clustering and PCA. *Journal of Scientific & Industrial Research*, 80(2), 159–165. <https://doi.org/10.56042/JSIR.V80I02.40102>
- Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019). Federated Machine Learning: Concept and Applications. *ACM Transactions on Intelligent Systems and Technology*, 10(2), 12. <https://doi.org/10.48550/arXiv.1902.04885>
- Zisopoulos, Z., Karagiannidis, S., Demirtsoğlu, G., & Antaris, S. (2008, October). *Content-Based Recommender systems*. ResearchGate. https://www.researchgate.net/publication/236895069_Content-Based_Recommendation_Systems

Отримано редакцією журналу / Received: 21.07.25

Прорецензовано / Revised: 20.09.25

Схвалено до друку / Accepted: 20.09.25



Максим ШЕВЧЕНКО, асп.
ORCID ID: 0009-0008-5104-9767
e-mail: maksym_shevchenko@knu.ua
Київський національний університет імені Тараса Шевченка, Київ, Україна

АНАЛІТИЧНИЙ ОГЛЯД КОНТЕНТНОЇ ТА КОЛАБОРАТИВНОЇ ФІЛЬТРАЦІЙ У РЕКОМЕНДАЦІЙНИХ СИСТЕМАХ

Вступ. За стрімкого зростання обсягу цифрового контенту рекомендаційні системи стають ключовим інструментом для надання персоналізованих пропозицій. Вони сприяють відкриттю нових фільмів, музики та товарів, підтримуючи зацікавленість користувачів у використанні платформ. Актуальність дослідження алгоритмів рекомендаційних систем зумовлена необхідністю вдосконалення їхньої роботи для задоволення індивідуальних уподобань користувачів. Ця робота являє собою огляд та аналітичне дослідження алгоритмів рекомендаційних систем. Метою цієї роботи є систематизація, класифікація та критичний аналіз двох основних підходів у рекомендаційних системах: фільтрації на основі вмісту (контентної) та колаборативної фільтрації.

Методи. Огляд існуючих методів рекомендаційних систем, порівняльне й аналітичне дослідження.

Результати. Проаналізовано алгоритми рекомендаційних систем. Дано формальне визначення задачі рекомендацій, де вподобання користувачів моделюються як функціональна залежність від властивостей об'єктів. У межах фільтрації на основі вмісту розглянуто використання класифікаційних алгоритмів, таких як найвигідніший басейн класифікатор, і дерев рішень, а також алгоритму Роккіо, який застосовує релевантний зворотний зв'язок для оновлення профілю користувача. Проведено аналіз сильних і слабких сторін різних мір подібності між векторами. У колаборативній фільтрації досліджено *memory-based* підхід (*user-based* та *item-based* методи) і *model-based* техніки з акцентом на алгоритми *k-NN*. Для подолання недоліків окремих методів запропоновано гібридний підхід, який об'єднує їхні переваги. Представлено способи інтеграції систем у гібридну модель, що дає змогу покращити точність рекомендацій.

Висновки. Результати роботи виокремлюють особливості зазначених методів фільтрації, демонструють вплив реалізації алгоритмів і вхідних даних на точність рекомендацій і час відповіді. Аналіз недоліків підкреслює значення комбінованого використання алгоритмів фільтрації для підвищення ефективності рекомендаційних систем, що робить гібридний підхід перспективним напрямом для подальших досліджень і впровадження.

Ключові слова: алгоритм Роккіо, векторна модель, гібридна фільтрація, колаборативна фільтрація, контентна фільтрація, рекомендаційні системи.

Автор заявляє про відсутність конфлікту інтересів. Спонсори не брали участі в розробленні дослідження; у зборі, аналізі чи інтерпретації даних; у написанні рукопису; в рішенні про публікацію результатів.

The author declares no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses or interpretation of data; in the writing of the manuscript; in the decision to publish the results.



ІНФОРМАЦІЯ ПРО АВТОРІВ / INFORMATION ABOUT AUTHORS



Олексій Батрак, аспірант кафедри комп'ютерної інженерії Відкритого міжнародного університету розвитку людини "Україна".
Наукові інтереси: алгоритми автоматизації систем, автономні роботизовані системи, дрони.

Oleksiy Batrak, PhD Student at the Department of Computer Engineering, Open International University of Human Development "Ukraine".
Research interests: systems automation algorithms, autonomous robotic systems, drones.



Сергій Виговський, аспірант кафедри інженерії програмного забезпечення Приватного акціонерного товариства "Вищий навчальний заклад «Міжрегіональна академія управління персоналом»".

Наукові інтереси: алгоритми автоматизації систем, автономні роботизовані системи, дрони.

Sergiy Vyhovskyy, PhD Student at the Department of Software Engineering, Interregional Academy of Personnel Management.
Research interests: algorithms for system automation, autonomous robotic systems, drones.



Андрій Дуднік, д-р техн. наук, доц., професор кафедри мережевих та інтернет-технологій факультету інформаційних технологій Київського національного університету імені Тараса Шевченка.
Наукові інтереси: комп'ютерні мережі, інтернет речей, вимірювання механічних величин.

Andrii Dudnik, DSc (Engin.), Assoc. Prof., Professor at the Department of Networking and Internet Technologies, Faculty of Information Technologies, Taras Shevchenko National University of Kyiv.
Research interests: computer networks, Internet of Things, measuring mechanical quantities.



Людмила Зубик, канд. пед. наук, доц., доцент кафедри програмних систем і технологій факультету інформаційних технологій Київського національного університету імені Тараса Шевченка.
Наукові інтереси: педагогіка вищої школи, вебтехнології, системи штучного інтелекту.

Liudmyla Zubyk, PhD (Ped.), Assoc. Prof., Associate Professor at the Department of Software Systems and Technologies, Faculty of Information Technologies, Taras Shevchenko National University of Kyiv.
Research interests: higher education pedagogy, web technologies, artificial intelligence.



Андрій Лемешко, д-р філософії, доцент кафедри прикладних інформаційних систем факультету інформаційних технологій Київського національного університету імені Тараса Шевченка.
Наукові інтереси: комп'ютерні мережі та кібербезпека, дослідження у сфері безпроводних комп'ютерних мереж.

Andriy Lemeshko, PhD, Associate Professor at the Department of Applied Information Systems, Faculty of Information Technology, Taras Shevchenko National University of Kyiv.
Research interests: computer networks and cybersecurity; studies in wireless computer networks.



Антон Марінов, аспірант кафедри телекомунікацій Навчально-наукового Інституту телекомунікаційних систем Національного технічного університету України "Київський політехнічний інститут імені Ігоря Сікорського".
Наукові інтереси: керування дата-центрами, емуляція у Mininet, програмно-конфігуровані мережі, інжиніринг трафіка.

Anton Marinov, PhD Student at the Department of Telecommunications of Educational and Research Institute of Telecommunication Systems, National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
Research interests: data-center management, Mininet emulation, software-defined networking, traffic engineering.



Віра Миколайчук, д-р філософії, асистент кафедри мережевих та інтернет-технологій факультету інформаційних технологій Київського національного університету імені Тараса Шевченка.
Наукові інтереси: роботизовані системи, Інтернет речей (IoT), штучний інтелект, машинне навчання.

Vira Mykolaichuk, PhD, Assist. at the Department of Networking and Internet Technologies, Faculty of Information Technologies, Taras Shevchenko National University of Kyiv.

Research interests: robotic systems, Internet of Things (IoT), artificial intelligence, machine learning.



Микола Нестеренко, д-р техн. наук, доц., заступник начальника кафедри комп'ютерних інформаційних технологій факультету інформаційних технологій Військового інституту телекомунікацій та інформатизації імені Героїв Крут.

Наукові інтереси: управління електронними комунікаційними мережами, програмно-конфігуровані мережі, інжиніринг трафіка.

Mykola Nesterenko, DSc. (Engin.), Assoc. Prof., Deputy Head of the Department of Computer Information Technologies, Faculty of Information Technology, Kruty Heroes Military Institute of Telecommunications and Information Technology.

Research interests: management communication networks, software-defined networking, traffic engineering.



Олена Павлюк, канд. техн. наук, доцент кафедри автоматизованих систем управління інституту комп'ютерних наук та інформаційних технологій Національного університету "Львівська політехніка".

Наукові інтереси: ШІ, смарт-індустрія, надійність систем.

Olena Pavliuk, PhD (Engin.), Assoc. Prof. at the Department of Automated Control Systems, Lviv Polytechnic National University.

Research interests: AI, smart industry, system reliability.



Тарас Семенченко, аспірант кафедри програмних систем і технологій факультету інформаційних технологій Київського національного університету імені Тараса Шевченка.

Наукові інтереси: комп'ютерний зір, машинне навчання, штучний інтелект, обробка природної мови.

Taras Semenchenko, PhD Student at the Department of Software Systems and Technologies, Faculty of Information Technologies, Taras Shevchenko National University of Kyiv.

Research interests: computer vision, machine learning, artificial intelligence, natural language preprocessing.



Ольга Ткаченко, д-р техн. наук, проф., професор кафедри програмних систем і технологій факультету інформаційних технологій Київського національного університету імені Тараса Шевченка.

Наукові інтереси: оптимізація та моделювання систем і мереж, інформаційні технології, розподілені інформаційні системи.

Olga Tkachenko, DSc (Engin.), Prof., Professor at the Department of Software Systems and Technologies, Faculty of Information Technologies, Taras Shevchenko National University of Kyiv.

Research interests: optimization and modeling of systems and networks, information technologies, distributed information systems.



Олександр Торошанко, канд. техн. наук, асистент кафедри кібербезпеки та захисту інформації факультету інформаційних технологій Київського національного університету імені Тараса Шевченка.

Наукові інтереси: кібербезпека, телекомунікаційні системи та мережі, інформаційна безпека.

Oleksandr Toroshanko, PhD (Engin.), Assist. at the Department of Cybersecurity and Information Protection, Faculty of Information Technologies, Taras Shevchenko National University of Kyiv.

Research interests: cybersecurity, telecommunications systems and networks, information security.



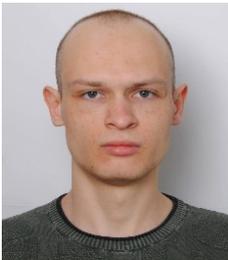
Андрій Фесенко, канд. техн. наук, доцент, декан факультету комп'ютерних наук та технологій, Державне некомерційне підприємство "Державний університет «Київський авіаційний інститут»".
Наукові інтереси: захист інформації, квантові обчислення, біометрія, криптографія.

Andrii Fesenko, PhD (Engin.), Assoc. Prof., Dean of the Faculty of Computer Science and Technology, State Non-Commercial Company "State University "Kyiv Aviation Institute".
Research interests: information protection, quantum computing, biometrics, cryptography.



Даниїл Фуркало, студент 3-го курсу кафедри програмних систем і технологій, факультет інформаційних технологій, Київський національний університет імені Тараса Шевченка.
Наукові інтереси: програмування, нейронні мережі, системне програмування.

Daniil Furkalo, 3rd-year student at the Department of Software Systems and Technologies, Faculty of Information Technologies, Taras Shevchenko National University of Kyiv.
Research interests: programming, neural networks, system programming.



Максим Шевченко, аспірант кафедри математичної інформатики факультету комп'ютерних наук та кібернетики Київського національного університету імені Тараса Шевченка.
Наукові інтереси: програмування на Python, Big Data, штучний інтелект.

Maksym Shevchenko, PhD Student at the Department of Mathematical Informatics, Faculty of Computer Science and Cybernetics, Taras Shevchenko National University of Kyiv.
Research interests: computer programming in Python, Big Data, artificial intelligence.



Ірина Юрчук, канд. фіз.-мат. наук, доцент кафедри програмних систем і технологій факультету інформаційних технологій Київського національного університету імені Тараса Шевченка.
Наукові інтереси: спеціальні функції, топологічний аналіз даних, методи машинного навчання.

Iryna Yurchuk, PhD (Phys. & Math.), Assoc. Prof. at the Department of Software Systems and Technologies, Faculty of Information Technology, Taras Shevchenko National University of Kyiv.
Research interests: special functions, topological data analysis, machine learning methods.

Наукове видання



СУЧАСНІ ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

№ 1(4)/2025

Редактор *Л. Магда*

Оригінал-макет виготовлено Видавничо-поліграфічним центром "Київський університет"

Автори опублікованих матеріалів несуть повну відповідальність за підбір, точність наведених фактів, цитат, економіко-статистичних даних, власних імен та інших відомостей. Редколегія залишає за собою право скорочувати та редагувати подані матеріали.



Формат 60x84^{1/16}. Обл.-вид. арк. 8,5. Ум. друк. арк. 8,4. Наклад 100. Зам. № 225-11618.
Гарнітура Arial. Папір офсетний. Друк офсетний. Вид. № Іт4.
Підписано до друку 17.11.25

Видавець і виготовлювач
ВПЦ "Київський університет",
Б-р Тараса Шевченка, 14, м. Київ, 01601, Україна
☎ (38044) 239 32 22; (38044) 239 31 58; (38044) 239 31 28
e-mail: vpc@knu.ua
<http://vpc.knu.ua>

Свідоцтво суб'єкта видавничої справи ДК № 1103 від 31.10.02