ISSN 2786-5460 (Print)
ISSN 2786-5479 (Online)

МІЖРЕГІОНАЛЬНА АКАДЕМІЯ УПРАВЛІННЯ ПЕРСОНАЛОМ

INTERREGIONAL ACADEMY OF PERSONNEL MANAGEMENT

МАУП

НАУКОВІ ПРАЦІ МІЖРЕГІОНАЛЬНОЇ АКАДЕМІЇ УПРАВЛІННЯ ПЕРСОНАЛОМ

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА СУСПІЛЬСТВО

SCIENTIFIC WORKS OF INTERREGIONAL ACADEMY OF PERSONNEL MANAGEMENT

INFORMATION TECHNOLOGY AND SOCIETY

Випуск 1 (12), 2024
Issue 1 (12), 2024

Видавничий дім
«Гельветика»
2024

Рекомендовано до друку Вченою радою
Міжрегіональної Академії управління персоналом
(протокол № 5 від 24 квітня 2024 року)

Інформаційні технології та суспільство / [головний редактор О. Попов]. – Київ : Міжрегіональна Академія управління персоналом, 2024. – Випуск 1 (12). – 96 с.

Журнал «Інформаційні технології та суспільство» є науковим рецензованим виданням, в якому здійснюється публікація матеріалів науковців різних рівнів у вигляді наукових статей з метою їх поширення як серед вітчизняних дослідників, так і за кордоном.

Редакційна колегія не обов'язково поділяє позицію, висловлену авторами у статтях, та не несе відповідальності за достовірність наведених даних і посилань.

Реєстрація суб'єкта у сфері друкованих медіа:
Рішення Національної ради України з питань телебачення і радіомовлення № 1173 від 11.04.2024 року.

Відповідно до Наказу МОН України № 1290 від 30 листопада 2021 року (додаток 3) журнал включено до Переліку наукових фахових видань України (категорія Б) зі спеціальностей 121 – Інженерія програмного забезпечення, 122 – Комп'ютерні науки, 123 – Комп'ютерна інженерія, 124 – Системний аналіз, 125 – Кібербезпека, 126 – Інформаційні системи та технології.

Усі електронні версії статей журналу оприлюднюються на офіційній сторінці видання
http://journals.maup.com.ua/index.php/it

Статті у виданні перевірені на наявність плагіату за допомогою програмного забезпечення StrikePlagiarism.com від польської компанії Plagiat.pl.

Recommended for publication
by Interregional Academy of Personnel Management
(Minutes No. 5 dated 24 April 2024)

Information Technology and Society / [chief editor Oleksandr Popov]. – Kyiv : Interregional Academy of Personnel Management, 2024. – Issue 1 (12). – 96 p.

Journal «Information Technology and Society» is a peer-reviewed scientific edition, which publishes materials of scientists of various levels in the form of scientific articles for the purpose of their dissemination both among domestic researchers and abroad.

Editorial board do not necessarily reflect the position expressed by the authors of articles, and are not responsible for the accuracy of the data and references.

Registration of Print media entity:
Decision of the National Council of Television and Radio Broadcasting of Ukraine: Decision No. 1173 as of 11.04.2024.

According to the Decree of MES No. 1290 (Annex 3) dated November 30, 2021, the journal was included in the List of scientific professional publications of Ukraine (category B) in specialties 121 – Software engineering, 122 – Computer sciences, 123 – Computer engineering, 124 – Systems analysis, 125 – Cybersecurity, 126 – Information systems and technologies.

All electronic versions of articles in the collection are available on the official website edition
http://journals.maup.com.ua/index.php/it

The articles were checked for plagiarism using the software
StrikePlagiarism.com developed by the Polish company Plagiat.pl.

ЗМІСТ

CONTENTS

УДК 378:004
DOI https://doi.org/10.32689/maup.it.2024.1.1

Василь АНДРУСЯК
студент-магістр, Івано-Франківський національний технічний університет нафти і газу, vasyl.andrusiak-ipm231@nung.edu.ua
ORCID: *0009-0000-9692-6489*

Ліда ГОБИР
асистент кафедри інженерії програмного забезпечення, Івано-Франківський національний технічний університет нафти і газу, lidagobyr@gmail.com
ORCID: *0009-0007-3176-2314*

Тетяна ВАВРИК
асистент кафедри інженерії програмного забезпечення, Івано-Франківський національний технічний університет нафти і газу, vavruk1060@gmail.com
ORCID: *0000-0002-0612-0084*

ОПТИМІЗАЦІЯ НАВЧАЛЬНОГО ПРОЦЕСУ УНІВЕРСИТЕТУ ЗА ДОПОМОГОЮ ЧАТ-БОТА

Анотація. *З появою технологій навчальні заклади постійно шукають інноваційні способи покращити навчальний досвід для студентів. Чат-боти стали цінним інструментом для оптимізації навчального процесу в університетах, надаючи персоналізовану допомогу, швидкі відповіді на запити та безперебійне спілкування між студентами та викладачами.*

Оптимізація навчального процесу університету за допомогою чат-бота може значно полегшити спілкування між студентами та викладачами, а також сприяти швидкому доступу до необхідної інформації. Чат-бот може допомогти студентам отримувати відповіді на питання щодо навчального процесу, матеріали для підготовки до заліків та іспитів, а також надавати загальні поради щодо навчання. Крім цього, за допомогою аналітики можна виявити слабкі місця в навчальному процесі та вдосконалити його. Ця аналітика можлива завдяки використанням зовнішніх інтеграцій з такими інструментами, як: Power BI, Tableau або Qlik Sense.

Додавання чат-бота до системи навчання університету може покращити доступність інформації для студентів, сприяти їх активній участі в навчальному процесі та допомогти вирішувати можливі проблеми швидко та ефективно. Наприклад, чат-бот може відповісти на питання стосовно виконання завдань, допомогти при виникненні труднощів з навчальним матеріалом або навіть вести онлайн-консультації з викладачами.

Висновки. *Чат-бот може стати цифровим помічником студента, який завжди готовий надати необхідну інформацію та підтримку. Це сприятиме покращенню якості навчання, збільшенню мотивації студентів та підвищенню рівня знань.*

Ключові слова: *чат-боти, навчальний процес, університети, оптимізація, персональна допомога*

Vasyl ANDRUSIAK, Lida HOBYR, Tetiana VAVRYK. OPTIMIZING THE EDUCATIONAL PROCESS IN UNIVERSITIES USING CHATBOTS

Abstract. *With the advent of technology, educational institutions are constantly looking for innovative ways to improve the learning experience for students. Chatbots have become a valuable tool for optimizing the educational process in universities, providing personalized assistance, quick responses to queries and seamless communication between students and teachers.*

Optimizing the educational process of the university with the help of a chatbot can greatly facilitate communication between students and teachers, as well as facilitate quick access to the necessary information. The chatbot can help students get answers to questions about the educational process, materials for preparing for tests and exams, and also provide general advice on studying. In addition, with the help of analytics, it is possible to identify weak points in the educational process and improve it. This analytics is possible through the use of external integrations with tools such as: Power BI, Tableau or Qlik Sense.

Adding a chatbot to the university education system can improve the availability of information for students, promote their active participation in the educational process, and help solve possible problems quickly and efficiently. For example, a chatbot can answer questions about completing tasks, help with difficulties with educational material, or even conduct online consultations with teachers.

Conclusions. *A chatbot can become a student's digital assistant who is always ready to provide the necessary information and support. This will help to improve the quality of education, increase the motivation of students and increase the level of knowledge.*

Key words: *chatbots, educational process, universities, optimization, personalized assistance.*

Мета: У цій статті досліджуються переваги впровадження чат-ботів в університетські умови та досліджується, як вони можуть змінити традиційну освітню модель.

Вступ. Сучасні технології впливають на всі сфери життя, в тому числі й на сферу освіти. Розвиток Інтернету та комунікаційних технологій створює нові можливості для вдосконалення процесу навчання,

сприяє зручності та доступності навчання. У традиційній освітній моделі в університетах часто виникає проблема задоволення різноманітних потреб студентів і забезпечення ефективного спілкування між студентами та викладачами. Оскільки технології продовжують розвиватися, навчальні заклади звертаються до чат-ботів для вирішення оптимізації навчального процесу та покращення загального досвіду студентів. Чат-боти на базі штучного інтелекту здатні надавати персоналізовану допомогу, миттєво відповідати на запити та допомагати студентам в їхньому навчанні.

Чат-боти є одним із інноваційних інструментів, які активно використовуються в освітній сфері.

Використання чат-ботів у навчальному процесі вже має свої позитивні результати. Вони надають студентам можливість отримати швидкі та точні відповіді на запитання, підтримують інтерактивність та залучення студентів до навчання, а також сприяють оптимізації роботи викладачів шляхом автоматизації деяких процесів.

Аналіз останніх досліджень та публікацій

Оскільки застосування штучного інтелекту (ШІ) продовжує проникати в різні сектори, освітній простір не є винятком. Ця стаття містить поглиблене дослідження використання та ролі штучного інтелекту в освіті та наукових дослідженнях, зосереджуючись на перевагах (хороших) і потенційних пастках (поганих і потворних), пов'язаних із розгортанням чат-ботів. Розглянуті можливості включають персоналізоване навчання, полегшення адміністративних завдань, розширені дослідницькі можливості та надання платформи для співпраці. Висновки, зроблені на основі цього аналізу, підкреслюють важливість досягнення балансу між можливостями ШІ та людськими елементами в освіті, а також розробку комплексних етичних рамок для розгортання ШІ в освітніх контекстах [1].

У статті [2] розглянуто оптимізацію навчального процесу університету за допомогою чат-бота . Використання чат-бота є одним з сучасних способів покращення якості освіти та забезпечення ефективної взаємодії між студентами та викладачами. Чат-боти можуть бути використані для автоматизації процесів навчання, зручної комунікації інформації та підтримки студентів у режимі реального часу.

Завдяки впровадженню чат-бота у систему навчання університету, можна досягти ефективнішої організації навчального процесу, зменшення адміністративної тяганини та забезпечення доступності інформації для всіх зацікавлених сторін.

Чат-боти — це комп'ютерні програмні системи, які використовують обробку природної мови для допомоги людям у різноманітних видах діяльності. Ці системи зазвичай виконують пошук ключових слів, фраз, прикладів і шаблонів, визначених у їхніх базах знань, і перетворюють їх у запити.

Завдяки останнім технологічним досягненням університети в усьому світі поступово інвестують в освітні програмні системи, які не призначені для заміни вчителів, а постачають корисні інструменти, які дозволяють студентам досягти кращої академічної підготовки. Тому не слід ігнорувати освітній сектор, оскільки надання своєчасного та точного зворотного зв'язку має вирішальне значення для успішної успішності в університеті [2].

У статті [3] описана нова модель розробки чат-ботів. Запропонована модель визначає ключові компоненти, які дозволяють створеному чат-боту діяти як посередник між студентами та персоналом, залученим до навчального процесу. Основні функції, які надає такий чат-бот, включають полегшення спілкування, заповнення часткової інформації, створення нагадувань, консультування студентів, моніторинг подій чи ситуацій, надання корисної інформації та відповіді на запитання студентів [3].

У дослідженні [4] представлені результати дослідження широкого спектру наукової літератури, розглядаються переваги та ризики використання чат-ботів у навчанні, аналізується доцільність та актуальність їхнього застосування. Чат-боти мають безліч плюсів, які удосконалюють навчання та викладання. Але також присутні й мінуси, які потребують коригування, для подальшого залучення новітніх технологій у навчання. У статті представлені результати широкого спектру досліджень наукової літератури, де обговорюються переваги та ризики використання чат-ботів і нейронних мереж в освіті та аналізується доцільність і актуальність їх застосування. Чат-боти та нейронні мережі мають численні переваги, які покращують як навчання, так і викладання. Однак є й недоліки, які потребують коригування для подальшого впровадження цих передових технологій в освіту [4].

Виклад основного матеріалу дослідження

Призначення. Враховуючи вищезазначені фактори, важливо провести дослідження та розробити чат-бот, який відповідатиме потребам університетського середовища та сприятиме покращенню процесу навчання для таких груп користувачів:

- Студенти
- Вчителі
- Абітурієнти
- Незареєстровані користувачі

• Адміністрація університету

Цей чат-бот чудово підходить для студентів, викладачів, абітурієнтів і адміністраторів університетів, оскільки він має кілька основних функцій, які роблять його особливим. а саме:

• Можливість перегляду розкладу
• Можливість створення розкладу
• Можливість видачі кошторису
• Можливість перегляду оцінок
• Можливість залишати та переглядати відгуки
• Отримання повідомлень про зміни в навчальному процесі •
• Отримання інформації про новини університету
• Отримання інформації про вступну кампанію.

Компоненти.

Цей чат-бот складається з наступних компонентів:

• База даних університету ІФНТУНГ
• Чат-бот, що працює в месенджері Telegram (з можливістю додавання інших платформ)
• Проміжне програмне забезпечення (набір методів API) - призначене для взаємодії Чат-бота з базою даних

***Архітектура та результати*.** Архітектура чат-бота показана на рис. 1.

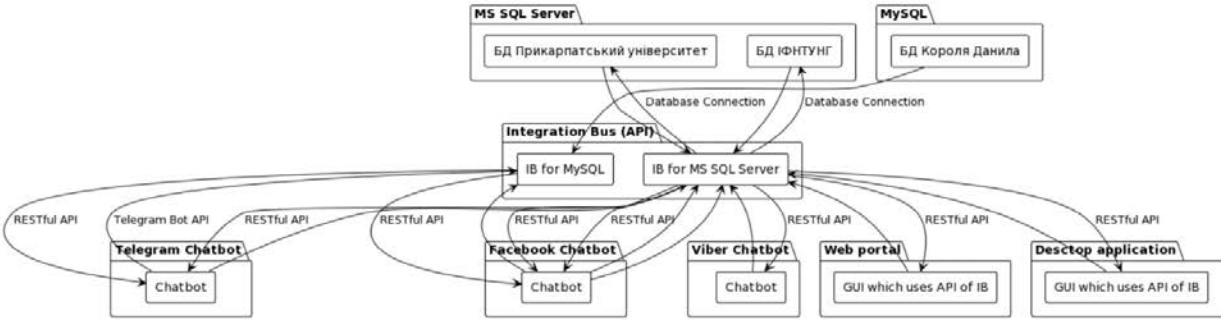


Рис. 1. Архітектура університетського чат-бота

Як ми бачимо на малюнку, з даних, які нам передає чат-бот, ми робимо запити до Middleware. База даних підключена до проміжного програмного забезпечення, і, відповідно, проміжне програмне забезпечення може отримувати або оновлювати дані в цій базі даних. Після цього він повертає відповіді чат-бота.

Головна перевага цієї архітектури полягає в тому, що ми можемо легко масштабувати наше програмне забезпечення.

Окрім масштабування, значною перевагою цієї архітектури також є гнучкість у виборі технологій. Оскільки ці компоненти незалежні один від одного, ми можемо створювати різні чат-боти різними мовами та технологіями. Наприклад, бувають випадки, коли деякі месенджери містять невелику кількість бібліотек для певної мови програмування. У таких випадках ми можемо комбінувати ці мови та фреймворки, наприклад, наш бот Telegram може працювати на C#, Facebook на Python, а Viber на JavaScript. Водночас усі ці боти виконуватимуть ті самі методи API, які надає наша IB для взаємодії з базою даних.

Наприклад, крім бота Telegram, нам потрібно додати інші месенджери, такі як: Facebook і Viber. Більше того, ми навіть можемо додати окремі GUI, наприклад, на веб-портал або в настільну програму, не записуючи нові запити до БД, а повторно використовуючи їх із проміжного програмного забезпечення.

Враховуючи зазначені фактори, важливо провести дослідження та розробити чат-бот [1-5], який задовольнить потреби комунікаційного середовища університету між різними групами користувачів, такими як:

• Учні є учні;
• Студенти – викладачі;
• Вчителі – учні;
• Адміністрація університету – викладачі;
• Адміністрація університету - студенти.

2 показано процес забезпечення комунікаційної безпеки в навчальному процесі.

Щоб забезпечити таку взаємодію спілкування між користувачами, база даних університету має бути підключена до чат-бота, який має містити такі сутності:

– Сутність UserRole – це таблиця бази даних, яка міститиме список користувачів із набором їхніх прав;

– Об'єкт MessageQueue – це таблиця БД, яка міститиме чергу повідомлень. Ця таблиця містить зовнішні ключі SenderID і RecipientID для забезпечення зв'язку між відправником і одержувачем.

Спілкування між користувачами має відбуватися не через прямі повідомлення в чат-боті, а через базу даних. Тобто, коли 100 студентів надсилають повідомлення вчителю в чат-боті, вчитель не отримає всі ці повідомлення в чатах, але ці записи будуть зберігатися в базі. Далі вчитель у чат-боті може переглядати свою папку «Вхідні» з усіма надісланими листами, які він може читати зі своєї скриньки за допомогою попередньо встановлених фільтрів і відповідати на повідомлення. Після цього ці повідомлення автоматично отримають статус «оброблено». Щоб забезпечити ефективне та безпечне спілкування між користувачами, важливо забезпечити безпеку під час авторизації в чат-боті [6-8].

Забезпечення безпеки приватності повідомлень чат-боту здійснюється за допомогою 4 основних сутностей, а саме: User, System, TelegramBot та EmailService. TelegramBot можна замінити на будь-який інший месенджер, який підтримує розробку чат-ботів (рис. 2).

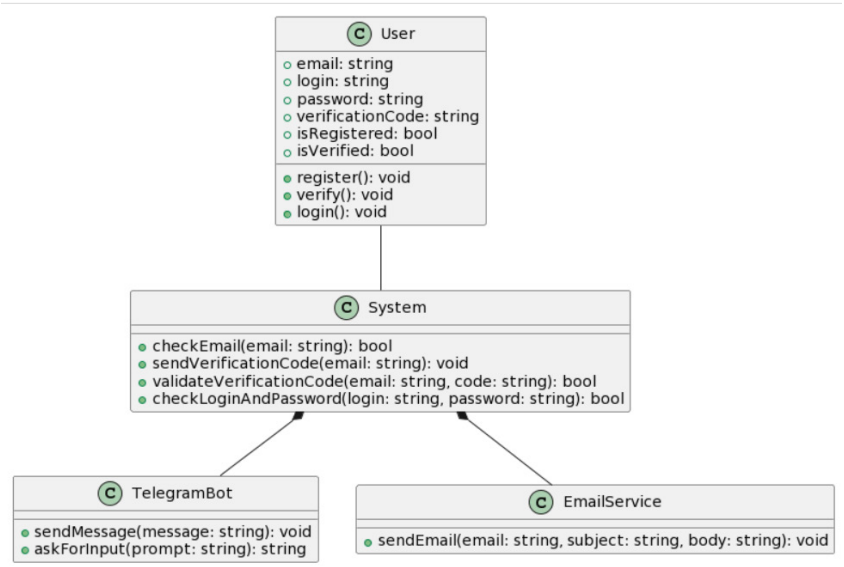


Рис. 2. UML діаграма класів безпеки чат-боту

Сутність User містить поля, які необхідні для визначення прав доступу в чат-боті. Ці поля: email (електронна пошту студента, викладача або представника адміністрації університету), login (логін користувача), password (пароль користувача), verificationCode (код підтвердження, який приходить на електронну пошту користувача для забезпечення двохфакторної авторизації), isRegistered (бітове поле, яке присвоюється під час перевірки системою введеної електронної пошти) та isVerified (бітове поле, яке присвоюється при введенні коду підтвердження, який відправлений на електронну пошту). Також клас User містить 3 методи:

- Register () – виконує логіку реєстрації
- Verify () – виконує логіку верифікації
- Login () – виконує логіку авторизації в систему.

Сутність System – містить набір методів, валідація яких відбувається із бази даних. Вони містить 4 основні методи, а саме:

checkEmail() – метод перевірки введення електронної пошти і на основі цієї пошти визначаються права користувача.

sendVerificationCode() – метод, який призачений для вілправки електронних листів з кодом підтвердження для двохфакторної авторизації.

validateVerificationCode() – метод, який призначений для перевірки введеного коду підтвердження.

checkLoginAndPassword() – метод, який призначений для перевірки введених логіну та паролю користувача.

Сутність TelegramBot (в якості цієї сутності можна використовувати будь-який інший месенджер, наприклад Viber або WhatsAPP де є можливість реалізувати чат-бот) – містить набір методів, які здійснюють

взаємодію з користувачем, а саме можуть відправляти або приймати повідомлення. Також на основі введених даних можуть приймати наступні рішення завдяки використанню штучного інтелекту [9-11].

Сутність EmailService – призначена для відправки електронних листів через mail service.

Візуально процес логіки забезпечення безпеки приватності повідомлень користувачів зображено на рисунку 2.

На рисунках 3, 4 зображено UML діаграму послідовності роботи авторизації/автентифікації чат-боту з методу зробити цю процедуру безпечною та простою у використанні для користувачів.

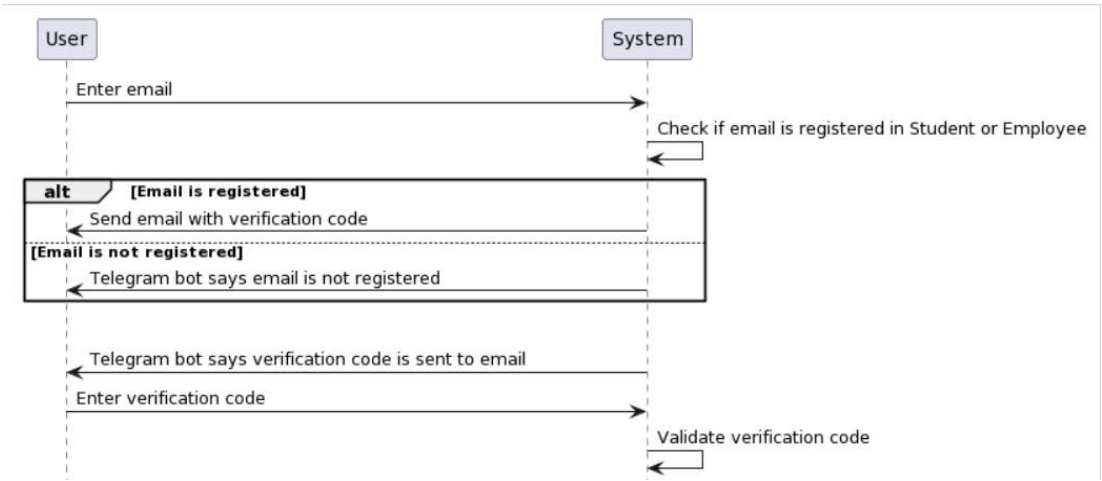


Рис. 3. Забезпечення роботи авторизації/верифікації в освітньому процесі за допомогою чат-бота

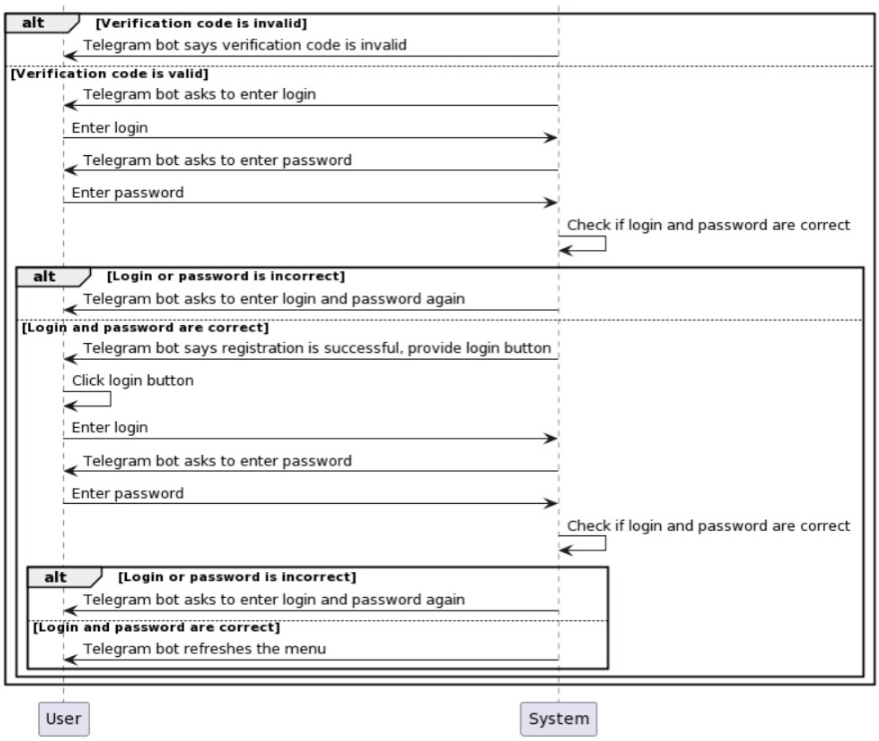


Рис. 4. Забезпечення роботи авторизації/верифікації в освітньому процесі за допомогою чат-бота (продовження)

Під час авторизації роль користувача чат-бота визначається введеними логіном і паролем. Варто враховувати, що кожна роль має свої права доступу до чат-бота.

Детальна інформація про права користувача зображена графічно на UML-схемі варіантів використання (рис. 5).

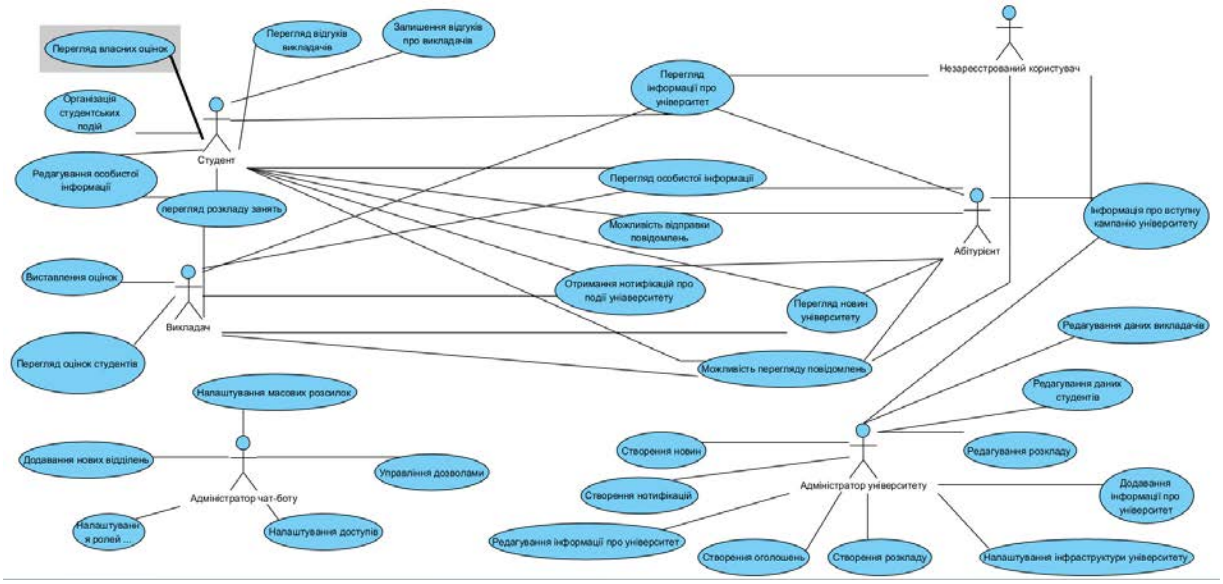


Рис. 5. Права користувачів чат-боту

Спілкування між користувачами має відбуватися не через прямі повідомлення в чат-боті, а через базу даних. Далі вчитель у чат-боті може переглядати свою папку «Вхідні» з усіма надісланими листами, які він може читати зі своєї скриньки за допомогою попередньо встановлених фільтрів і відповідати на повідомлення. Візуальний вигляд чат-боту показано на рис 6.

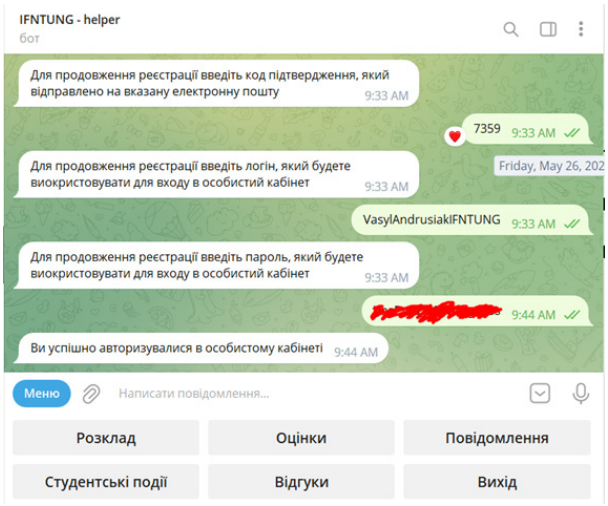


Рис. 6. Процес реєстрації/авторизації в особистому кабінеті чат-бота для забезпечення спілкування

Як показано на малюнку 3, для відправки повідомлення користувач повинен авторизуватися в особистому кабінеті чат-бота [12]. Для цього він (користувач) вводить свою корпоративну електронну адресу. Система чат-ботів на основі введеної електронної пошти ідентифікує користувача та визначає його роль як викладача чи студента. Якщо користувач зареєстрований в базі, йому надсилається код для авторизації в системі, після чого йому необхідно ввести логін і пароль. Візуально, як це відбувається з боку чат-бота, показано на рисунку.

Висновок. Університетське середовище відкриває широкі можливості для використання чат-ботів як основного інструменту взаємодії та підтримки учасників освітнього процесу. Їх реалізація може значно полегшити комунікацію між студентами, викладачами та адміністративним персоналом, сприяючи підвищенню якості навчального процесу. Крім того, важливо враховувати, що для забезпечення безпеки та ефективного зв'язку використання функцій авторизації для безпеки даних.

Переваги використання чат-ботів в університетах:

По- перше: персональна допомога.

Аналізуючи дані та розуміючи поведінку користувача, чат-боти можуть надавати індивідуальні рекомендації та підтримку, щоб допомогти студентам досягти успіху в навчанні.

По-друге: Миттєві відповіді.

Чат-боти можуть надати миттєві відповіді, позбавляючи студентів необхідності чекати відповіді від викладачів чи співробітників.

По-третє: Безперебійне спілкування.

Спілкування відіграє вирішальну роль у навчальному процесі, і чат-боти можуть сприяти безперебійному спілкуванню між студентами, викладачами та адміністративним персоналом. Діючи як централізована платформа для інформації та підтримки, чат-боти можуть забезпечити студентам легкий доступ до ресурсів, необхідних для досягнення успіху.

По четверте: Доступність

Чат-боти доступні 24/7, що робить їх цінним ресурсом для студентів, яким може знадобитися допомога поза звичайним робочим часом. Ця доступність гарантує підтримку, покращення досвіду навчання та загальне задоволення від навчального процесу.

Таким чином, впровадження інноваційних технологій, таких як чат-боти, в університетському середовищі вимагає ретельної розробки та врахування безпеки та ефективності спілкування, але може значно полегшити процес спілкування для всіх учасників навчального процесу, сприяючи покращенню співпраця та доступність інформації. Надаючи персоналізовану допомогу, швидкі відповіді на запити та безперебійне спілкування, чат-боти можуть допомогти університетам задовольнити різноманітні потреби студентів і забезпечити більш ефективне академічне середовище. Оскільки технології продовжують розвиватися, навчальні заклади повинні використовувати інноваційні рішення, такі як чат-боти, щоб залишатися конкурентоспроможними та надавати студентам підтримку, необхідну для досягнення успіху. Така інновація дозволить університетам підвищити конкурентоздатність та стати лідерами у використанні передових технологій у сфері освіти.

Список використаних джерел:

1. Джо Мейо. Програмування Microsoft Bot Framework: багатоплатформний підхід до створення чат-ботів (довідка для розробників), 2-е видання, 2017 р. 400 с.

2. Наливайко О. О. Перспективи використання нейромереж у вищій освіті України. *Інформаційні технології і засоби навчання*. 2023. 97 (5). С. 1–17. https://doi.org/10.33407/itlt.v97i5.5322

3. Офіційна документація Telegram API: https://core.telegram.org/

4. Малиш К., Чабан С., Приходько Я., Наливайко О. Чат-боти у навчанні: перспективи та ризики. Матеріали XXI Всеукраїнської науково-методичної конференції здобувачів вищої освіти та молодих вчених «Наумовські читання», грудень 2023. URL: https://www.researchgate.net/publication/376405897_CATBOTI_U_NAVCANNI_PERESPEKTIVI_TA_RIZIKI. (дата звернення: 11.04.2024).

5. Використання чат-ботів в освіті. [Електронний ресурс]. Режим доступу: https://gerabot.com/article/vikoristannya_chatbotiv_u_osviti.

6. Чат-боти в навчальних закладах: Майбутнє освіти. [Електронний ресурс]. Режим доступу: https://gerabot.com/article/chatboti_v_navchalnih_zakladah_maibutn_osviti

7. Мар`єнко М. В., Коваленко В. М. Штучний інтелект та відкрита наука в освіті. *Фізико-математична освіта*. 2023. Т. 38, № 1. С. 48–53.

8. Smith Amanda. Chatbots: A Valuable Tool for Optimizing the Student Journey. *International Journal of Educational Innovation*, vol. 12, no. 2, 2022.

9. University of California, Irvine. "How Chatbots are Revolutionizing Higher Education." https://www.oit.uci.edu/services/ai/zotgpt/

10. Mendoza, S.; Sánchez-Adame, L.M.; Urquiza-Yllescas, J. F.;González-Beltrán, B.A.; Decouchant,D. A. Model to Develop Chatbots for Assisting the Teaching and Learning Process. *Sensors* 2022, 22, 5532. Published: 25 July 2022 MDPI Journal https://doi.org/10.3390/s22155532

11. Gabriel Babtista Software Architecture with C# 10 and .NET 6: Develop software solutions using microservices, DevOps, EF Core, 2022. 512 c.

12. Srini Janarthanam. Hands-On Chatbots and Conversational UI Development, 2017. 573 c.

УДК 004.056.55:004.384.3:004.738.5
DOI https://doi.org/10.32689/maup.it.2024.1.2

Віктор БОЙКО
кандидат технічних наук, доцент, доцент кафедри кібербезпеки,
Національний університет «Одеська юридична академія», boyko-work@ukr.net
***ORCID:** 0000-0001-5929-657X*

Микола ВАСИЛЕНКО
доктор фізико-математичних наук, доктор юридичних наук, професор,
професор кафедри кібербезпеки,
Національний університет «Одеська юридична академія», vasylenko.it@journals.maup.kiev.ua
***ORCID:** 0000-0002-8555-5712*

Валерія СЛАТВІНСЬКА
доктор філософії в галузі «Право», асистент кафедри кібербезпеки,
Національний університет «Одеська юридична академія», slatvinskaya_valeriya@ukr.net
***ORCID:** 0000-0002-6082-981X*

МОДЕЛЮВАННЯ ЖИВУЧОСТІ ТА ВІДНОВЛЕННЯ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ МЕРЕЖ В УМОВАХ ДІЇ КІБЕРЗАГРОЗ

Анотація. *У статті досліджується актуальна проблема централізації та ієрархізації інформаційно-комунікаційних мереж (ІКМ), яка призвела до розробки методів оцінки та усунення слабких місць у фізичній та функціональній інфраструктурі ІКМ. Розвиток сучасних ІКМ та послуг, що базуються на них, призвів до такого рівня складності та централізації, коли глобальні системні збої та каскадні сценарії недоступності ІКМ у майбутньому стають неодмінними.*

З'ясовано, що існуючі методи моделювання відновлення ІКС можна розділити на дві категорії: аналіз факторів відновлення та аналіз структурної і/або функціональної моделі системи. Доведено, що ІКМ мають розподілену архітектуру, але спостерігається тенденція до централізації та ієрархізації. Приділено увагу специфіці ICN як інфраструктурного об'єкта.

У роботі запропоновано когнітивно-імітаційну модель відновлення ІКМ (CSM ICN), яка базується на загальній моделі CSM CTS та використовується для комплексного прогнозування та сценарного моделювання можливих збоїв та сценаріїв недоступності ІКМ. CSM ICN використовує орієнтований граф (орграф) для моделювання компонентів ICN та зв'язків між ними. Модель може бути «мережевою» (з акцентом на зв'язки) або гібридною (з додаванням «віртуальних вузлів» для причинно-наслідкових зв'язків). Кожен вузол моделює частину ICN, її елемент або робочий компонент. Вузли мають характеристики, що включають час і характер відновлення. Модель використовує критерій Бірнбаума для оцінки впливу виходу з ладу одного елемента на інші. Відновлення вузла моделюється як змінне число кроків дискретного часу. Модель використовує три типи функцій відновлення для різних рівнів підготовленості систем і має 4 рівні оцінки. Модель може оцінювати каскадні сценарії, коли вихід з ладу одного елемента призводить до збоїв в інших частинах ICN.

Висновки. *Отже, моделювання сценаріїв відновлення й перезапуску ІКМ дозволить виявити потенційні вразливості, аналізувати каскадні сценарії недоступності ІКМ у випадку дії надзвичайних подій та природних форс-мажорів. Застосування CTS ICN сприятиме значному підвищенню живучості та стійкості експлуатації ІКМ в умовах зовнішніх атак, помилок персоналу та впливу інших надзвичайних подій та форс-мажорів.*

Ключові слова: *інформаційно-комунікаційні системи, стійкість, живучість, когнітивно-імітаційна модель, моделювання відновлення, кібератаки, інфраструктурні мережі.*

Viktor BOYKO, Nikolai VASILENKO, Valeriia SLATVINSKA. MODELING THE SURVIVABILITY AND RECOVERY OF INFORMATION AND COMMUNICATION NETWORKS IN THE FACE OF CYBER THREATS

Abstract. *The article examines the pressing issue of centralization and hierarchization in information and communication networks (ICNs), which has led to the development of methods for assessing and addressing weaknesses in the physical and functional infrastructure of ICNs. The advancement of modern ICNs and the services based on them has resulted in such complexity and centralization that global systemic failures and cascading scenarios of ICN unavailability in the future become inevitable.*

It has been established that existing methods for modeling ICN recovery can be divided into two categories: recovery factor analysis and analysis of the structural and/or functional model of the system. It is proven that ICNs have a distributed architecture, but there is a tendency towards centralization and hierarchization. Attention is paid to the specificity of ICNs as an infrastructure object.

The paper proposes a cognitive-emulation model for ICN recovery (CSM ICN), which is based on the general CSM CTS model and is used for comprehensive forecasting and scenario modeling of possible failures and unavailability scenarios of ICNs. CSM ICN utilizes a directed graph (digraph) to model ICN components and their connections. The model can be "network-centric" (emphasizing connections) or hybrid (adding "virtual nodes" for cause-and-effect relationships). Each node models a part of

the ICN, its element, or working component. Nodes have characteristics that include recovery time and nature. The model uses the Burnbaum criterion to assess the impact of one element's failure on others. Node recovery is modeled as a variable number of discrete time steps. The model uses three types of recovery functions for different levels of system readiness and has 4 levels of evaluation. The model can assess cascading scenarios where the failure of one element leads to failures in other parts of the ICN.

Conclusions. *Thus, modeling recovery and restart scenarios of ICNs will help identify potential vulnerabilities, analyze cascading scenarios of ICN unavailability in case of emergencies and natural disasters. The application of CTS ICN will significantly increase the resilience and operational stability of ICNs in conditions of external attacks, personnel errors, and the impact of other emergencies and force majeure events.*

Key words: *information and communication systems, resilience, survivability, cognitive simulation model, recovery modeling, cyberattacks, infrastructure networks.*

Актуальність проблеми. Питання стійкості та живучості інформаційних систем нині набуває дедалі більшої актуальності. Більшість наявних інформаційно-комунікаційних систем (ICN) після введення в експлуатацію піддається атакам різного роду і характеру [1] і з великою ймовірністю можуть бути зламані, або виведені з ладу. У разі, якщо зламу або іншим несприятливим впливам піддаються глобальні інформаційні системи, або інформаційні системи, що обслуговують індустріальні та промислові комплекси (industrial control system – ICS) ризики і масштаби збитку багаторазово збільшуються.

Донедавна основним фокусом небезпеки були розробники промислових та інфраструктурних інформаційних систем, які, на відміну від розробників ПЗ «загального призначення», запізнюються з реакцією на нові загрози. Наприклад, під час аналізу безпеки ICS, багато фахівців схильні покладатися на т.зв. «air gap» – «повітряний отвір», що ізолює ICS від глобальних інформаційно-комунікаційних систем. Простіше кажучи – якщо керівна система не підключена до інтернету – звідти не може здійснюватися атака. При цьому один із найперших прикладів атаки – вірус Stuxnet, якраз був побудований на подоланні «повітряного отвору» [2].

Однак останнім часом почастішали збої глобальних інформаційних мереж. На думку авторів це наслідок кількох причин.

По-перше, відбувається вибуховий ріст обсягу та складності ІКТ. Forbes наводить наступні дані щодо мобільних операторів в Україні: 19 мільйонів абонентів Vodafone [3], 8,9 мільйона у Lifecell [4], 24 мільйони у Київстар [5]. За тими ж даними, Lifecell має близько 9000 базових станцій.

По-друге, спостерігається тенденція до централізації та ієрархізації ІКТ. Розвиток технологій та еволюція інформаційно-комунікаційних мереж сприяють все більшій централізації та ієрархізації наявних мереж [6]. Телефонна мережа минулого була відносно децентралізованою системою – при руйнуванні міжміських зв'язків міські телефонні мережі продовжували функціонувати у незалежному режимі, а також залишалася можливість підключень «в обхід» системи. Тепер, як показує практика, «падіння головного офісу» призводить до відключення всієї мережі на рівні країни [7], [8]. У поточному випадку відключення було здійснене вручну, щоб запобігти поширенню кібератаки [8] на ядро системи, однак, враховуючи досвід масштабних збоїв минулого, можна передбачити, що в сучасній централізованій і ієрархічній структурі ІКТ існує ймовірність повного виходу мережі з ладу внаслідок збою, атаки або іншого негативного впливу.

По-третє, досвід безперебійної роботи мобільних та інтернет-мереж призвів до того, що багато різних служб у будь-який спосіб використовують їх як частину своєї інфраструктури. Перш за все це стосується банківської сфери – банкомати, термінали поповнення, POS-термінали в більшості своєму використовують для зв'язку з банком мобільний зв'язок, і при виході з ладу мобільної мережі перестають функціонувати. Сюди також відноситься широко поширена як «просувана технологія безпеки» двофакторна аутентифікація, при якій підтвердження особи користувача відбувається за допомогою SMS-повідомлень. Згідно з джерелами Forbes у ПриватБанка під час відключення «Київстар» не працювало до третини POS-терміналів та близько 5% банкоматів. Це найбільш поширені випадки, однак глобальний збій зв'язку показав, що існують й інші вразливі системи. Зокрема, ЛКП «Львівсвітло» було змушено виконувати відключення ліній вуличного освітлення в ручному режимі [9].

У роботі [10], яка була опублікована ще до початку повномасштабного вторгнення, зроблено такий висновок: «В умовах впливу нових загроз стає важливим не лише управління ліквідацією безпосередніх наслідків загроз, але й боротьба за функціональну живучість з метою запобігання розпаду системи. Внаслідок нових загроз системи можуть безпосередньо виходити з ладу не лише самі системи, а й породжувати «вторинні ефекти», що впливають на суміжні системи. Для загального розпаду інформаційної системи внаслідок нових загроз часто не потрібно, щоб було виведено з ладу 100% її підсистем, а достатньо просто створити зниження працездатності одного або кількох ключових компонентів системи».

Наступні події (чорні вибухи, втрата функціональності систем через бойові дії, нещодавнє відключення мобільного зв'язку «Київстар») підтвердили актуальність викладених у статті положень.

Усе перераховане підкреслює актуальність проблеми забезпечення живучості сучасних інформаційно-комунікаційних мереж та гостро постає питання про надійність і живучість інфраструктурних систем.

Важливою складовою розв'язання проблеми живучості та стійкості ІКМ є розроблення методології комплексного прогнозування і сценарного моделювання можливих збоїв та сценаріїв каскадного виходу зі строю інформаційно-комунікаційних мереж.

Аналіз останніх досліджень і публікацій. У роботі [10] йдеться про те, що сучасні системи «розумний будинок» страждають від фрагментації та прив'язки до постачальника, що ускладнює їх використання, збільшує витрати та знижує рівень безпеки. Натомість авторами запропоноване рішення – модульна система з відкритими протоколами, натхненна глобальними інформаційними системами, що забезпечує взаємодію та незалежну розробку. У роботі [26] запропоновано методику оцінки ризиків та загроз для ICS, розроблено когнітивно-імітаційну модель ICS, запропоновані методи та моделі можуть бути використані для підвищення живучості та стійкості ICS, а також встановлено, що використання даних методів дозволить суттєво підвищити живучість та стійкість ICS. У роботі [27] розглядається питання моделювання кібернетичної складової живучості складних технічних систем. Автор пропонує підхід до моделювання, який ґрунтується на системному аналізі та імітаційному моделюванні.

Метою статті є наукове обґрунтування розробки когнітивно-імітаційної моделі (КІМ) для оцінки стійкості та живучості ІКС шляхом вирішення наступних завдань:

– огляд наявних методів моделювання відновлення інфраструктурних мереж;

– визначення специфіки ІКС як інфраструктурного об'єкта;

– обґрунтування доцільності використання когнітивно-імітаційної моделі (КІМ) для оцінки стійкості та живучості ІКС.

Виклад основного матеріалу.

Існуючі методи моделювання відновлення інфраструктурних мереж

Кількісна оцінка стійкості, надійності та живучості ІКМ є складним питанням і може розглядатися як частковий випадок більш загальної моделі руйнування та відновлення складних технічних систем.

Протягом деякого часу це питання в основному розглядалося в описовому ключі, при цьому об'єктом моделювання слугували процеси руйнування, стійкості та відновлення комунальної інфраструктури (житлового фонду, водопроводу, системи енергопостачання). При цьому як природні катастрофи і психологічні фактори розглядалися стихійні лиха – переважно землетруси. Наприклад, робота [11] описує вплив стихійних лих переважно у якісному вигляді.

Однак у роботах [12] та [13], які вважаються класичними [14], математична модель вже використовується для оцінки часу та процесів відновлення після стихійного лиха. Аналогічно, робота [15] присвячена моделюванню процесів відновлення інфраструктури після землетрусів.

У роботі [16] була запропонована парадигма, в рамках якої виділяються чотири виміри системи – технічний, організаційний, соціальний і економічний – всі з яких можна використовувати для кількісної оцінки показників стійкості різних типів фізичних і організаційних систем. Також автор виділяє живучість в умовах впливу стихійних лих та психологічних факторів, як поєднання двох аспектів – стійкості системи до впливу стихійних лих та психологічних факторів і швидкості відновлення працездатності системи. Аналогічний поділ був зроблений у роботі [17], проте там живучість системи моделювалась також з урахуванням управління системою.

У роботі [1] вводиться досить загальна класифікація того, що можна розуміти під НВ та ПФ. Автори класифікують ці події наступним чином:

– природні небезпеки та стихійні лиха;

– атаки зловмисників, які додатково поділяються на низькочастотні – зломи та проникнення в систему, і високочастотні у вигляді кібератак (найбільш характерні – DDOS, brute-force атаки на пароль і т. д.);

– людські помилки;

– технічні збої (виходить з ладу обладнання);

– комплексні події, що містять одну, кілька або всі перелічені категорії.

У сучасний час до цього списку потрібно додати бойові дії, як результати безпосереднього та опосередкованого виведення з ладу обладнання.

У роботі [18] надано огляд поточного стану концепцій моделювання відновлення після катастроф (Conceptions for Disaster Recovery Modeling – CDRM). Автор відносить до CDRM наступні вісім концепцій:

– моделювання з обмеженими ресурсами,

– машинне навчання,

– моделювання динамічного економічного впливу,

– моделювання системної динаміки,

– агентне моделювання,

– дискретно-подійне моделювання,

– стохастичне моделювання,

– мережеве моделювання.

У роботі [1] була зроблена спроба метааналізу останніх наукових робіт у цій області та загальної категоризації наявних підходів. Основних категорій виділено дві – системні та мережеві.

Системна категорія підходів базується на урахуванні причинно-наслідкових зв'язків і моделюванні їх впливу на досліджувану систему. При цьому, як правило, всі намагаються звести до якихось числових показників, наприклад, коефіцієнтів живучості.

Мережева категорія – це уявлення системи у вигляді якої-небудь топології, використання при цьому статистичних даних про те, як ця система (або системи) себе веде в минулому та дослідження її методами математичної оптимізації.

Також у цій роботі була спроба «остаточно уточнити» поняття живучості, ризику, надійності, вразливості та стійкості.

Ще однією корисною концепцією, детально описаною в роботі, є концепція «цифрових близнюків» (DT) – цифрових подвійок, коли робота реальної інфраструктурної системи дублюється «цифровою системою», при якій кожен реальний компонент або агрегат системи має свого «цифрового подвійника» за термінологією концепції. Таким чином, утворюється віртуальна модель, яку доповнюють даними в режимі реального часу, що відображає стан реальної системи й дозволяє приймати рішення з профілактики, лікування та відновлення системи в умовах впливу НВ та ПФ.

З аналізу вищевказаного випливає, що наявні підходи до оцінки живучості, стійкості та відновлення ICN, як частинний випадок DT, можна розділити на дві основні категорії – аналіз факторів відновлення (системний підхід за термінологією) і аналіз структурної й/або функціональної моделі системи (мережевий підхід за термінологією).

Автор підкреслює, що для кожної конкретної системи не існує єдиного правильного підходу. Якщо порівняти переваги підходів, то найбільші переваги від структурної/функціональної моделі можна отримати у випадках, коли система має розвинену мережеву структуру або функціональні зв'язки. У цьому випадку модель може бути легко побудована на основі реальної схеми структури або функціональних зв'язків, що дозволяє «економити» на розкладанні системи на окремі компоненти (або агрегати, якщо мова йде про складну технічну систему).

Для багатофакторного аналізу підходять випадки, коли є достатня кількість експертів, щоб отримати експертну оцінку, і коли фактори, що впливають на стан (стійкість, живучість, відновлення) системи, визначені та класифіковані.

Специфіка ICN як інфраструктурного об'єкта

З урахуванням усього перерахованого вище, слід виділити специфіку роботи ICN, з одного боку, як інфраструктурної мережі, яка є частковим випадком інфраструктурних мереж, розглянутих вище, а з іншого – має свої власні особливості, характерні саме для ICN.

Хоча до цього часу панує досить поширене уявлення про ICN як мережі, що розподіляється та має розподілену архітектуру, як зазначалося вище, в наш час спостерігається тенденція до централізації та ієрархізації ICN. Здешевлення обладнання та збільшення обсягів зв'язку, разом із постійно зростальним попитом на обчислювальні потужності, канали зв'язку з високою пропускною здатністю та віддалений збір та обробку інформації призвели до того, що апаратна інфраструктура розвивалась непропорційно, без акценту на розподіленість та стійкість. У провайдерів та постачальників послуг склалося обманливе враження про надійність та безперебійність сервісів.

Протягом певного часового інтервалу розвитку ICN мережі справді були досить надійними та стійкими, однак зі зростанням складності та централізації систем їх надійність почала зменшуватись. У цьому розумінні цікаво спостерігати за історією глобальних вибоїв. Зазвичай вибої в ICN відбуваються постійно, однак, до нещодавнього часу, завдяки децентралізованому та розподіленому характеру ICN, їх наслідки та шкоду було місцевими, а самі вибої досить швидко усувалися або обмежувалися їх наслідки.

Один із перших глобальних випадків відмови стався із Skype [19] – система була недоступною протягом двох днів, а саму відмову, за поясненням корпорації, викликав аномально великий обсяг перезапусків після того, як користувачі завантажили оновлення безпеки Windows. Представник корпорації пояснив це так: «Велика кількість перезавантажень позначила на мережеві ресурси Skype. Це призвело до потоку запитів на вхід до системи, що в поєднанні з нестачею ресурсів піринговой мережі викликало ланцюгову реакцію, яка мала критичні наслідки» [20]. Важливо зазначити – відмова не була викликана усвідомленою атакою хакерів або відмовами в критично важливій інфраструктурі (наприклад, перебоями з електропостачанням) – це був наслідок складної взаємодії кількох підсистем єдиної системи. При цьому відновлення функціональності сервісу зайняло тривалий час.

Наступна глобальна відмова сталася 21 вересня 2015 року [21] і тривала понад 15 годин, і знову не була результатом кібератаки або стихійних лих. У блозі корпорації з'явилось наступне пояснення [22]: «Ми випустили зміну конфігурації, яка була більшою, ніж зазвичай, і яку деякі версії Skype не змогли правильно обробити, що призвело до відключення користувачів від мережі. Коли ці користувачі

намагалися знову під'єднатися, виникала велика кількість трафіку, і деякі з вас не могли користуватися безкоштовними послугами Skype, такими як обмін повідомленнями, присутність та керування списком контактів. Інші взагалі не могли увійти в Skype або вийти з нього, а також здійснювати дзвінки на стаціонарні або мобільні телефони».

Слід відзначити, що власники глобальних систем чим далі, тим менш охоче розкривають причини та подробиці відмов, що сталися.

Наступним значним випадком відмов були відмови Facebook – у 2019 [23] та 2021 роках [24]. При цьому відмова призводила до відключення не лише Facebook, але й суміжних проектів – Instagram, WatsApp і т.д. Подібні відмови у більшості сучасних глобальних сервісів сталися протягом останніх кількох років, і кожний раз відмова якої-небудь платформи послужила причиною відмови інших сервісів.

Найбільш помітною була відмова в роботі Amazon[25]. Оскільки сервіс AWS є хостингом для інших сервісів, то відмова в роботі Amazon призвела до відмови в роботі мережі доставлення контенту CDN, крім того, відключилися такі сайти, як Stack Overflow, GitHub, gov.uk, Hulu, HBO Max, Quora, PayPal, Vimeo і Shopify.

Перелічені вище сервіси виходили з ладу в основному через внутрішні помилки – в кожному випадку за повідомленнями їх представників та пресслужб. Однак, варто нагадати, що існують прецеденти цілеспрямованих атак на інфраструктуру. Найбільш відомий – епідемія вірусів WannaCry і Petya. Відомими прикладами подібних відмов є вірус Triton, який був націлений на атаку «останнього рубіжу оборони» – виведення з ладу приладових систем безпеки (SIS) Schneider Triconex, а також епідемії WannaCry, Petya та пов'язані з ними відмови в українській енергосистемі у 2015 році.

Отже, можна стверджувати, що сучасні ІКС та сервіси, що ґрунтуються на них, досягли такого рівня складності та централізації, що подібні збої є неминучими у майбутньому, а їх масштаб буде зростати. У зв'язку з цим набуває максимальної актуальності розробка систем та методів, які дозволяють оцінювати ймовірність збою, визначати слабкі та уразливі місця в фізичній та функціональній інфраструктурі ІКС.

При цьому складність й ієрархізація ІКС дозволяють розглядати їх як частковий випадок складних технічних систем (СТС) та використовувати для аналізу та моделювання їх поведінки когнітивно-імітаційну модель СТС (КІМ СТС), адаптовану під специфіку та особливості роботи ІКС – КІМ ІКС.

Для оцінки можливостей відновлення ІКС після впливу НВ та ПФ суттєво підвищити точність моделі можна було б використовуючи обидва підходи, описані у [1], у поєднаній, гібридній моделі, яка базувалася б на концепції когнітивно-імітаційної моделі складної технічної системи (cognitive-simulation models of complex technical systems CSM CTS), описаній у [17], а також у загальній структурі КІМ, викладеній у [26].

CSM відновлення стійкості ICN

Розглянемо принципи CSM CTS у застосуванні до ICN. Основою такої моделі [26] є орієнтований граф (орграф). Вузли орграфа моделюють компоненти системи ICN, спрямовані ребра (дуги) – зв'язки між компонентами. Залежно від рівня моделювання, така модель з різною точністю і достовірністю відображає взаємодію складових системи, при цьому дуги можуть розглядатися у різних функціональних ролях. Такий підхід дозволяє гнучко перебудовувати CSM CTS згідно з вимогами моделювання.

Якщо в орграфі CSM CTS зв'язки між елементами розглядаються як зв'язок її компонентів за ресурсом, відповідному категоріям стійкості системи («енергія» – «інформація» – «речовина»), то модель має структурно-функціональний акцент («мережевий» за термінологією [1]). Модель допускає використання спеціальних вузлів («віртуальні вузли» за визначенням авторів), які моделюють причинно-наслідкові зв'язки та відносини між компонентами [27] – додавання таких вузлів дозволяє доповнити модель елементами багатофакторного аналізу і отримати гібридну CSM ICN, що використовує «мережеву модель» відновлення ICN, розширену за допомогою віртуальних вузлів, які моделюють причинно-наслідкові зв'язки та залежності від зовнішніх факторів.

Кожен з вузлів системи в рамках моделі відновлення може представляти частину ICN, її елемент або окремий робочий компонент робочої підсистеми. При цьому цей вузол в рамках моделі має характеристики, що включають час і характер відновлення елемента. Запропонована методика оцінки часу відновлення використовує безрозмірні невимірні оцінки «важливості», «уразливості» елемента з погляду загальної задачі мінімізації часу відновлення системи в цілому. Така оцінка ґрунтується на критерії Бірнбаума [28], що визначає частку елементів, на які впливає виходження з ладу розглянутого елемента – чим більше ця частка, тим більше кількість елементів буде заторкнута виходом з ладу даного елемента.

Якщо в моделі використовується дискретний час, то відновлення вузла моделюється як змінне за певним законом число кроків, за які вузол змінює свою функціональність. Окрім зміни функціональності, в моделі передбачено моделювання процесу перезапуску, яке також задається у вигляді послідовності й числа кроків дискретного часу, витрачених на виходження вузла на повне функціонування після його відновлення. Якщо в моделі використовується не дискретний, а неперервний час, то як схеми відновлення вузла може бути прийнята класифікація, наведена у [29], в якій відновлення різних за рівнем підготовленості систем відповідає різним функціям.

$$f_{rec}(t) = ae^{-b\frac{t-t_{0E}}{T_{Я}}} \quad (1)$$

$$f_{rec}(t) = a\left(\frac{t-t_{0E}}{T_{Я}}\right) + b \quad (2)$$

$$f_{rec}(t) = \frac{a}{2}\left(1 + cos\left(\pi b\frac{t-t_{0E}}{T_{Я}}\right)\right) \quad (3)$$

де:

$f_{rec}(t)$ – функція відновлення, що виражає готовність системи до відновлення;

a, b – є постійними значеннями, які обчислюються з використанням кривої, що відповідає доступним даним;

t_{0E} – момент часу, коли відбувається екстремальна подія;

$T_{Я}$ – час відновлення, необхідний для повернення до стану, що передував катастрофі, оцінюваний починаючи з t0E.

При цьому (1) – експоненційна залежність – добре підготовлена система, (2) – лінійна залежність – середньо підготовлена система, (3) – тригонометрична залежність – не дуже добре підготовлена система.

Розподіл за рівнями CSM ICN відповідає розподілу за рівнями у базовій моделі CSM CTS [26] і ґрунтується на розподілі оцінок за різними рівнями.

На загальному рівні оцінюється лише топологія і структура системи в цілому – без врахування реальних характеристик її елементів. Така оцінка є заздалегідь зайвою і моделює «найгірший сценарій» розвитку подій, встановлюючи «потолок» у системній оцінці ролі та важливості елемента.

На рівні оцінки критичності вводиться поняття критичності – елементам на основі експертної оцінки може бути присвоєно значення, що визначає важливість і критичність даного елемента для загальної тривалості відновлення ICN. Залежно від зміни критичності елементів з'являється можливість отримати оцінку тривалості відновлення не за загальною, а за реальною системою – з урахуванням ролі і важливості кожного з її елементів.

На рівні оцінки витрат часу на відновлення враховується не лише положення і роль елемента в системі, а й його власні характеристики відновлення та залежність від ступеня працездатності та часу відновлення інших вузлів системи. Кінцевою інтегральною оцінкою внеску елемента в оперативність відновлення системи є математичне очікування часу відновлення для даного об'єкта.

На реальному рівні моделюється вплив несприятливих факторів на систему, максимально точно відображуючи реальність, що дає можливість оцінити можливі наслідки для швидкості відновлення при виході з ладу тих чи інших елементів ICN і сценарії розвитку негативних наслідків, включаючи каскадні сценарії – коли вихід однієї системи призводить до перерозподілу навантаження на залишені системи і призводить до послідовних збоїв по всьому простору ICN [30].

Висновки. Наразі спостерігається тенденція до централізації та ієрархізації ICN, що призвело до непропорційного розвитку інфраструктури ICN без акценту на розподіленість та стійкість. Сучасні ICN та сервіси, що ґрунтуються на них, досягли такого рівня складності та централізації, що у майбутньому необхідні глобальні системні збої та каскадні сценарії виходу ICN з ладу. Це підтверджує аналіз прецедентів, наведений у роботі. У зв'язку з цим набуває максимальної актуальності розробка систем та методів, які дозволяють оцінювати ймовірність збою, визначати слабкі та вразливі місця в фізичній та функціональній інфраструктурі ICN.

У роботі пропонується когнітивно-імітаційна модель відновлення ICN (CSM ICN), заснована на загальній моделі CSM CTS, в якій поєднуються мережевий та загальний підходи до оцінки часу та сценаріїв відновлення працездатності ICN методами комплексного прогнозування та сценарного моделювання можливих збоїв та каскадних сценаріїв виходу ICN з ладу.

Моделювання сценаріїв відновлення працездатності та перезапуску ICN дозволить виявити можливі вразливості в ICN, проаналізувати каскадні сценарії виходу ICN з ладу в умовах дії НВ та ПФ, що в свою чергу дозволить усунути ці вразливості шляхом прийняття заходів організаційного та технічного характеру (дублювання критично вразливих вузлів, усунення вузьких місць, використання гнучких алгоритмів перерозподілу навантаження тощо). Таким чином, використання CTS ICN дозволить суттєво підвищити живучість та стійкість експлуатації ICS в умовах дії зовнішніх атак, помилок персоналу та впливу інших НВ та ПФ.

Список використаних джерел:

1. Балашова Л., Галкін А., Мельник Т., Шевчук С. Найбільший збій за останні роки: оператор Kyivstar із 24 млн абонентів не працює. Імовірна причина – кібератака. Що відомо. URL: https://forbes.ua/innovations/naybilshiy-zbiy-za-ostanni-roki-operator-kiivstar-iz-24-mlnabonentiv-ne-pratsyue-imovirna-prichina-kiberataka-shcho-vidomo-12122023-17826.

2. Бойко В. Д. Моделювання кібернетичної складової живучості складних технічних систем / Матеріали Міжнародної науково-практичної конференції «Інформаційні управляючі системи та технології» (ІУСТ-Одеса-2014), 23–25 вересня 2014 року. Одеса, 2014. С. 239–241.

3. Бойко В. Д. Оцінка живучості та стійкості компонентів інформаційних систем за допомогою когнітивно-імітаційного моделювання / Європейський вибір України, розвиток науки та національна безпека в реаліях масштабної військової агресії та глобальних викликів XXI століття (до 25-річчя Національного університету «Одеська юридична академія» та 175– річчя Одеської школи права) / ed. by Ківалов С. В. Одеса: Видавничий дім «Гельветика», 2022. С. 746–749.

4. Інформація про Kyivstar – Forbes Ukraine. URL: https://forbes.ua/profile/kiivstar-244

5. Інформація про Lifecell – Forbes Ukraine. URL: https://forbes.ua/profile/lifecell-574

6. Інформація про Vodafone. Forbes Ukraine. URL: https://forbes.ua/profile/vodafone-251

7. Масштабний збій у роботі Київстар – з якими проблемами стикнулись у регіонах – УНІАН. URL: https://www.unian.ua/incidents/masshtabniy-zbiy-u-roboti-kijivstar-z-yakimiproblemami-stiknulis-u-regionah-12481575.html

8. Некряч О. Ієрархічні системи управління конвергентними мережами // Зв'язок. – Міністерство зв'язку України, 2015. – Is. 5. P. 11–13.

9. Через несправність мобільного оператора Київстар відключення вуличного освітлення відбуваються в ручному режимі – Львівська міська рада. URL: https://cityadm.lviv.ua/news/city/housing-and-utilities/299531-cherez-nespravnistiu-mobilnoho-operatorakyivstar-vidkliuchennia-vulychnoho-osvitlennia-vidbuvaiutsia-v-ruchnomu-rezhymi

10. AP Agency by. Skype blackout due to surge of reboots. 2007. URL: https://www.latimes.com/archives/la-xpm-2007-aug-21-fi-skype21-story.html.

11. Arak V. The Microsoft Connection Clarified. 2007. URL: https://web.archive.org/web/20080220105941/http://heartbeat.skype.com/2007/08/ the_microsoft_connection_explained.html.

12. Barrett C., Beckman R., Channakeshava K., Huang F., Kumar V. S. A., Marathe A., Marathe M. V., Pei G. Cascading failures in multiple infrastructures: From transportation to communication network / 2010 5th International Conference on Critical Infrastructure (CRIS). – IEEE, 2010

13. Barzashka I. Are Cyber-Weapons Effective? Assessing Stuxnet's Impact on the Iranian Enrichment Programme // The RUSI Journal. – Taylor & Francis, 2013. – Vol. 158, no. 2. P. 48–56.

14. Bi W., MacAskill K., Schooling J. Old wine in new bottles? Understanding infrastructure resilience: Foundations, assessment, and limitations // Transportation Research Part D: Transport and Environment. – Elsevier BV, 2023. – Vol. 120. P. 103–793.

15. Boyko V., Rudnichenko N., Kramskoy S., Hrechukha Y., Shibaeva N. Concept Implementation of Decision Support Software for the Risk Management of Complex Technical System // Advances in Intelligent Systems and Computing: Selected Papers from the International Conference on Computer Science and Information Technologies, CSIT 2016, September 6-10 Lviv, Ukraine» / ed. by Shakhovska N. – Cham: Springer International Publishing, 2017. P. 255–269.

16. Boyko V., Vasilenko M., Slatvinska V. Survivability and sustainability of smart city information system components // Municipal economy of cities. – O.M. Beketov National University of Urban Economy in Kharkiv, 2021. – Vol. 6, no. 166. P. 20–27.

17. Bruneau M., Chang S. E., Eguchi R. T., Lee G. C., O'Rourke T. D., Reinhorn A. M., Shinozuka M., Tierney K., Wallace W. A., Winterfeldt D. von. A Framework to Quantitatively Assess and Enhance the Seismic Resilience of Communities // Earthquake Spectra. – SAGE Publications, 2003. – Vol. 19, no. 4. P. 733–752.

18. Cimellaro G. P., Reinhorn A. M., Bruneau M. Framework for analytical quantification of disaster resilience // Engineering Structures. – Elsevier BV, 2010. – Vol. 32, no. 11. P. 3639– 3649.

19. Dent S. Amazon, Reddit, Twitter and Twitch Impacted by Huge Network Outage. URL: https://www.engadget.com/amazon-reddit-twitter-twitch-fastly-outage-131112143.html.

20. Harihara J. Skype Outage: An Update, and an Apology. 2015. URL: https://web.archive.org/web/20151204065003/http://heartbeat.skype.com/2015/09/ skype_outage_an_update_and_an.html.

21. Isaac M., Conger K. Facebook, Instagram and WhatsApp Suffer Outages. 2019. URL: https://www.nytimes.com/2019/03/14/technology/facebook-whatsapp-outage.html.

22. Isaac M., Frenkel S. Facebook and Instagram Are Down for Many Users. URL: https://www.nytimes.com/2021/10/04/technology/facebook-down.html.

23. Isumi M., Nomura N., Shibuya T. Simulation of Post-Earthquake Restoration of Lifeline Systems // International journal of mass emergencies and disasters. 1985. – Vol. 3, no. 1. P. 87–105.

24. Kates R. W. Assessing the Assessors: The Art and Ideology of Risk Assessment // Ambio. 1977. – Vol. VI, no. 5. P. 247–252.

25. Kates R. W. Dealing With Disaster // Science Year. 1976. P. 166–179.

26. Kates R. W., Pijawka D. From Rubble to Monument: The Pace of Reconstruction // Reconstruction Following Disaster / ed. by Haas J. E., Kates R. W., Bowden M. J. – Cambridge, MA: MIT Press, 1977. P. 1–23.

27. Miles S. B., Burton H. V., Kang H. Community of Practice for Modeling Disaster Recovery // Natural Hazards Review. – American Society of Civil Engineers (ASCE), 2019. – Vol. 20, no. 1.

28. Rausand M., Høyland A. System Reliability Theory: Models, Statistical Methods, and Applications, Second Edition / 2nd ed. – Wiley-Interscience, 2003.

29. Scott M. Skype Service Disrupted for Some Users Worldwide. 2015. URL: https://www.nytimes.com/2015/09/22/technology/skype-service-disrupted-for-some-usersworldwide.html.

30. Sobhaninia S., Buckman S. T. Revisiting and adapting the Kates-Pijawka disaster recovery model: A reconfigured emphasis on anticipation, equity, and resilience // International Journal of Disaster Risk Reduction. – Elsevier BV, 2022. – Vol. 69. P. 102–738.

УДК 004.9
DOI https://doi.org/10.32689/maup.it.2024.1.3

Надія БОЛЮБАШ
кандидат педагогічних наук, доцент,
доцент кафедри інтелектуальних інформаційних систем,
Чорноморський національний університет імені Петра Могили, Nadiya.Bolubash@chmnu.edu.ua
***ORCID**: 0000-0002-2274-2422*

Олег ЖЕЛТОБРЮХОВ
магістрант, кафедра інтелектуальних інформаційних систем,
Чорноморський національний університет імені Петра Могили, oleg30902228@gmail.com
***ORCID**: 0009-0006-4254-8877*

ЧАТ-БОТ ДЛЯ НАДАННЯ РЕКОМЕНДАЦІЙ ІЗ ПЕРЕГЛЯДУ ВІДЕОФІЛЬМІВ НА ОСНОВІ МАТРИЧНИХ ФАКТОРИЗАЦІЙНИХ МОДЕЛЕЙ

***Анотація**. У статті досліджено основні моделі прогнозування реакцій користувача у рекомендаційних системах, засновані на методах матричної факторизації. Обгрунтовано вибір матричної факторизаційної моделі та розглянуто підходи до забезпечення гнучкості взаємодії рекомендаційної системи і користувача шляхом використання чат-бота, впровадженого у вебзастосунок. **Метою статті** є дослідження ефективності застосування чат-бота, орієнтованого на індивідуальні потреби користувача, у рекомендаційній системі з надання рекомендацій по перегляду відеоконтенту на основі матричних факторизаційних моделей. **Методи дослідження**. Використано методи розробки вебзастосунків та інтелектуальних чат-ботів, методи матричної факторизації з використанням методу декомпозиції сигулярного значення SVD, методи машинного навчання, методи обробки й розпізнавання природної мови та методи оптимізації роботи рекомендаційної системи, що базуються на оцінці точності прогнозу та рівня зодоволеності користувача спілкуванням із чат-ботом. **Наукова новизна дослідження** полягає у виявленні методів і підходів, спрямованих на покращення отримання користувачами персоналізованих рекомендацій по перегляду відеоконтенту відповідно до їх інтересів та уподобань шляхом застосування чат-бота та моделі прогнозування реакцій користувача на основі методів матричної факторизації. **Висновки**. Накопичення великих обсягів цифрової відеоінформації різних форматів вимагає покращення механізмів надання рекомендацій та підвищення точності прогнозу стосовно уподобань користувачів. Дослідження моделей матричної факторизації MF, машини факторизації FM та машини факторизації з урахуванням поля FFM дозволило установити, що кращі показники точності прогнозу має модель машини факторизації з ураховунням поля FFM: MAE=0,86, MSE=1,65, RMSE=1,28. Для забезпечення гнучкості взаємодії користувача з рекомендаційною системою, розробленою на основі моделі FFM, виявлено доцільність її інтеграції з чат-ботом, впровадженим у вебзастосунок. Дослідження якості створеної моделі обробки природної мови показало високу точність розпізнавання намірів користувача при спілкуванні з чат-ботом – 99,17%. Виявлення рівня задоволеності користувачів спілкуванням з чат-ботом та отриманими рекомендаціями дозволило установити, що задоволеність користувачів становила 86,7%. Що свідчить про високий рівень оцінки ефективності взаємодії користувачів з чат-ботом та високу точність системи стосовно прогнозу намірів користувачів з перегляду відеофільмів.*

***Ключові слова:** рекомендаційна система, чат-бот, матрична факторизація, метод декомпозиції сигулярного значення, машинне навчання, машина факторизації з урахуванням поля.*

Nadiia BOLIUBASH, Oleh ZHELTOBRIUKHOV. A CHAT-BOT FOR PROVIDING RECOMMENDATIONS FOR WATCHING VIDEOS BASED ON MATRIX FACTORIZATION MODELS

***Abstract**. The article examines the main models of predicting user reactions in recommendation systems based on matrix factorization methods. The choice of the matrix factorization model is justified and the approaches to ensuring the flexibility of interaction between the recommendation system and the user through the use of a chatbot implemented in web applications are considered. **The purpose** of the article is to study the effectiveness of using a chatbot in providing personalized recommendations for viewing video content on the basis of a matrix factorization model. **Research methods.** General methods of developing web applications and intelligent chatbots are used, methods of matrix factorization using SVD sigular value decomposition method, machine learning methods, natural language processing and recognition methods, and recommendation system optimization methods based on assessment of forecast accuracy, satisfaction level of communication with the chatbot. **The scientific novelty** of the study consists in the identification of methods and approaches aimed at improving users' receipt of personalized recommendations for watching video content in accordance with their interests and preferences by using a chatbot and a model for predicting user reactions based on matrix factorization methods. **Conclusions**. The accumulation of large volumes of digital video information in various formats requires the improvement of mechanisms for providing recommendations and increasing the accuracy of providing predictions regarding user preferences. The research of the matrix factorization models MF, the factorization machine FM, and the field-aware factorization machine FFM made it possible to establish that the model of the field-aware factorization machine FFM had the best indicators of forecast accuracy: MAE=0,86, MSE=1,65, RMSE=1,28. To ensure the flexibility of user interaction with the recommendation system developed on the basis of the FFM model, the expediency of its integration with a chatbot implemented in the web application was found. The research of the quality of the created natural language processing model showed a high accuracy of recognizing the user's*

intentions when communicating with the chatbot - 99.17%. Detection of the level of user satisfaction with communication with the chatbot and received recommendations made it possible to establish that user satisfaction was 86.6%. Which indicates a high level of assessment of the effectiveness of user interaction with the chatbot and the high accuracy of the system in terms of predicting users' intentions to watch videos.

Key words: *recommendation system, chatbot, matrix factorization, sigular value decomposition method, machine learning, field-aware factorization machine.*

Постановка проблеми та аналіз останніх досліджень і публікацій. Накопичення великих обсягів цифрової відеоінформації різних форматів в умовах стрімкого розвитку інформаційного суспільства ускладнює пошук відеоконтенту, який відповідає індивідуальним уподобанням та інтересам користувача [11]. Для вирішення цієї проблеми сервіси, що надають доступ до відео різних типів та жанрів, використовують вбудовані алгоритми фільтрації інформації для надання персоналізованих рекомендацій [4; 12; 17]. Використання рекомендаційних систем дозволяє формувати прогнози стосовно продуктів та послуг, які оптимально будуть відповідати потребам користувачів.

Проте переважна більшість сервісів для перегляду відеоконтенту використовує рекомендаційні системи для вирішення задач комерційного маркетингу і не забезпечує у повній мірі взаємодію між користувачем та системою. Застосування чат-боту дозволяє будувати рекомендації з використанням підходів, які базуються на спілкуванні з користувачем для отримання необхідної інформації про його уподобання та передачі її до рекомендаційної системи для фільтрації відеоконтенту відповідно до його потреб і інтересів [27]. Такий підхід сприяє покращенню формування індивідуальних рекомендацій, дозволяє вирішити проблему холодного старту, спрощує отримання додаткової інформації, забезпечуючи її аналіз у динамічному режимі. Багато застосунків, таких як підтримка клієнтів, інтерактивні платформи обміну повідомленнями, віртуальні помічники та пошук інформації, створені за допомогою чат-ботів [24]. Проведений аналіз досліджень науковців показав, що створення рекомендаційних систем із застосуванням чат-бота у різних предметних сферах дозволяє отримувати результати з високою точністю прогнозу стосовно потреб користувача при наданні рекомендацій та реалізує адаптивні стратегії [7; 9; 20; 27].

У процесі розробки рекомендаційної системи важливим є вибір методів, спрямованих на відбір відеоконтенту з метою забезпечення оптимального прогнозу [15]. Аналіз існуючих підходів до формування прогнозу у рекомендаційних системах дозволив виявити методи колаборативної, контентної та гібридної фільтрації [10; 22]. Контентна фільтрація будує рекомендації, базуючись на інформації про поведінку користувача та його потреби й уподобання. Колаборативна фільтрація будує персональні рекомендації, базуючись на моделі поведінки користувача на основі попередньо зібраної інформації про поведінку інших користувачів із схожими вподобаннями або характеристиками. Серед алгоритмів колаборативної фільтрації виділяють алгоритми, що базуються на даних користувачів та алгоритми, що базуються на даних елементів. Гібридна фільтрація поєднує сильні сторони контентної та колаборативної фільтрації [2].

Підходи, засновані на моделях, при створенні прогнозу рейтингу елементів відеоконтенту використовують накопичену інформацію для формування моделі та її подальшого навчання. Результатом навчання моделі є функція, яка дозволяє отримати рекоменційний прогноз стосовно потреб та інтересів користувача, виходячи із уже наявних у системі оцінок. При отриманні нових даних модель модернізується для підвищення точності рекомендацій.

Останнім часом для прогнозування реакцій користувача набули широкого поширення моделі на основі матричної факторизації, які мають багато модифікацій: базова матрична факторизація (англ. Matrix factorization, MF), машина факторизації (англ. Factorization Machines, FM), нейронна мережа на основі факторизації (англ. Deep Factorization Machines, DeepFM) [29]. Більш ефективними серед них є моделі, які враховують додаткову інформацію про приховані зв'язки між характеристиками користувачів і елементів та пов'язаними з ними полями: матрична факторизація з урахуванням поля (англ. Field-Aware Matrix Factorization, FMF), тензорна факторизація парної взаємодії (англ. Pairwise Interaction Tensor Factorization, PITF), машина факторизації з урахуванням поля (англ. Field-Aware Factorization Machine, FFM) [13; 30; 31].

Проте ефективність методів матричної факторизації при наданні рекомендацій для перегляду відеоконтенту з використанням чат-бота, орієнтованого на індивідуальні потреби користувача, є не достатньо дослідженою й потребує подальшого опрацювання. Це обумовило **мету статті**, яка полягає у дослідженні ефективності застосування чат-бота у рекомендаційній системі з надання рекомендацій по перегляду відеоконтенту на основі матричних факторизаційних моделей.

Виклад основного матеріалу. Рекомендаційна система будує рейтинг елементів відеоконтенту на основі аналізу великого набору даних, що містить інформацію про характеристики елементів і користувачів та уподобання користувачів при виборі елементів.

У результаті проведеного дослідження установлено, що рекомендаційна система на основі матричної факторизації базується на створенні та навчанні моделі шляхом аналізу даних і наступному використанні навченої моделі для надання користувачам найкращих пропозицій щодо продуктів і послуг. У рекомендаційних системах відеоконтенту необхідно визначити матрицю рейтингів *R*, кожен рядок якої представляє користувача, а кожен стовпець – елемент відеоконтенту:

$$R = \begin{vmatrix} r_{11} & \dots & r_{1n} \\ \dots & \dots & \dots \\ r_{m1} & \dots & r_{mn} \end{vmatrix}, \tag{1}$$

де r_{ij} – оцінка *j*-го елемента *i*-м користувачем (його рейтинг), *m* – кількість користувачів, *n* – кількість елементів контенту.

Матриця *R* є досить розрідженою, оскільки вона буде містити багато елементів, не оцінених користувачами. Тому у моделі матричної факторизації MF передбачено декомпозицію вихідної матриці на добуток двох матриць меншого рангу (рис. 1).

Для зменшення розмірності матриці рейтингів *R* у матричній факторизації використовують низку методів: метод головних компонент (англ. Principal Component Analysis, PCA), невід'ємне розкладання матриці (англ. Non-Negative Matrix Factorization, NNMF/NMF), метод декомпозиції сигулярного значення (англ. Singular Value Decomposition, SVD) і його модифікації Funk-SVD, SVD++, Asymmetric SVD, timeSVD++ [3; 29]. Аналіз досліджень застосування різних методів матричної факторизації показав, що в області відеоконтенту найкращий час навчання мають методи, реалізовані на основі декомпозиції сингулярного значення SVD [23].

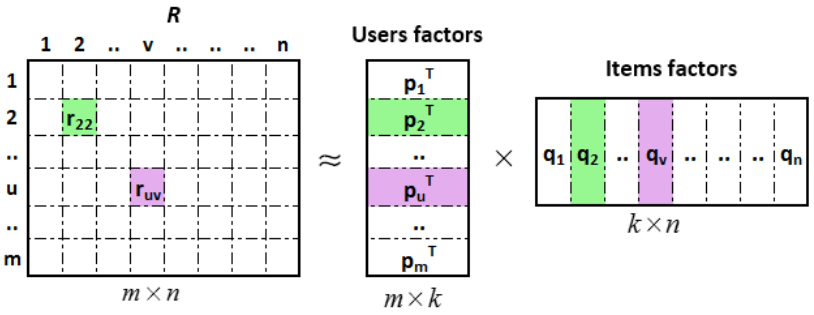


Рис. 1. Схема факторизації матриці рейтингів

Метод SVD передбачає формування матриці $R^{'}$, у якій відсутні значення матриці рейтингів *R* заповнені з використанням глобального середнього або середнього значення користувача чи елемента [29]. Теорема про сингулярний розклад матриці $R^{'}$ розмірністю $m \times n$ дозволяє представити її у вигляді добутку трьох матриць наступним чином:

$$R^{'} = U \cdot S \cdot V^{T}, \tag{2}$$

де *U* та *V* є унітарними матрицями розмірністю $m \times m$ та $n \times n$ відповідно, а матриця *S* є діагональною матрицею розмірністю $m \times n$ із діагональними елементами, рівними сингулярним значенням матриці *R* у порядку їх спадання.

У рекомендаційних системах використовують урізане сингулярне розкладання, коли з усіх діагональних елементів матриці *S* залишають тільки *k* перших елементів, а інші приймають рівними 0. Тоді у матриці *U* залишиться тільки *k* стовпців, у матриці *V* – тільки *k* рядків, а матриця *S* буде квадратною матрицею розмірністю $k \times k$. Щоб виконати декомпозицію, необхідно обрати *k* сингулярних значень для складання діагональної матриці S_k і знайти відповідні рядки та стовпці цих *k* сингулярних значень в матрицях *U* та *V* як U_k і V_k. Нова матриця R_k^* є низькоранговим наближенням матриці R:

$$R_k^* = U_k \cdot S_k \cdot V_k^{T}. \tag{3}$$

Позначивши $P = (U_k \cdot S_k)^T$ та $Q = V_k^{T}$, маємо декомпозицію матриці рейтингів у вигляді добутку двох матриць меншої розмірності:

$$R \approx P^T \cdot Q. \tag{4}$$

Ранг факторизації k є величиною низькорангового наближення R, яку визначають, виходячи з розміру сингулярних значень матриці S. З цією метою можуть бути використані різні методи [29]: евристичні – відношення суми перших k квадратів діагональних елементів матриці S до суми квадратів усіх її діагональних елементів повинно бути більшим за деяке порогове значення (зазвичай рівне 0,95); емпіричні – сума перших k діагональних елементів матриці S повинна бути більшою від суми інших сингулярних значень у декілька раз (наприклад, у 10).

Близькість елементів відеоконтенту, представлених у матриці рейтингів, можна визначати з використанням різних мір близькості: відстані Евкліда, квадрату відстані Евкліда, Манхетенської відстані та інших [1, с. 65-66]. У розрідженій матриці рейтингів R для визначення мір близькості між двома елементами доцільно використовувати косинус подібності:

$$\cos(R_i, R_j) = \frac{R_i \cdot R_j}{\|R_i\| \cdot \|R_j\|}, \tag{5}$$

де R_i та R_j є векторами рейтингів i-го та j-го елеметів, представлених у матриці рейтингів R у i-му та j-му стовпцях, а $\|R_i\|$ та $\|R_j\|$ є нормами цих векторів. Аналогічно може бути розрахована близькість двох користувачів, використовуючи їх вектори, представлені у відповідних рядках матриці рейтингів. Під час розрахунку мір близькості у факторизованій матриці рейтингів R_k^* враховують тільки ті значення, які не є пустими.

Обчислювальну складність традиційного методу SVD можна зменшити, базуючись на навчанні. Такий підхід запропоновано у методах Funk-SVD, SVD++, Asymmetric SVD, timeSVD, які використовують базову ідею методу SVD: матриця рейтингу розкладається на дві матриці нижчого рангу Q і P, добуток яких приблизно рівний матриці рейтингів відповідно до формули 4. Отримана наближена матриця рейтингів для елементів з відомими рейтингами буде містити приблизно рівні їм значення, а для елементів з невідомим рейтингами – розраховані прогнозовані рейтинги. Прогнозований рейтинг i-го елемента відеконтенту для u-го користувача розраховують за формулою:

$$r_{ui} = \sum_f p_{uf}^T q_{if}, \tag{6}$$

де p_{uf} та q_{if} є відповідними елементами матриць Q і P.

Під час навчання моделі матриці Q і P знаходять таким чином, щоб мінімізувати функцію втрат – відхилення між відомими рейтингами та їх прогнозами шляхом оптимізації цільової функції. Ранг факторизації є параметром, при якому навчена модель дає кращий прогноз.

У рекомендаційних системах із матричною факторизацією MF не можуть бути використані додаткові характеристики користувачів і відеоконтенту. Однак ця інформація може бути врахована в моделі машини факторизації FM. Для нового фільму матрична факторизація не може визначити рейтинг, оскільки його ніхто не дивився. Проте знаючи жанр, акторів, режисера та інші атрибути фільму, машина факторизації може надавати рекомендації для його перегляду.

Машина факторизації дозволяє враховувати при наданні рекомендацій додаткову інформацію про елементи шляхом комбінації регресії та матричної факторизації. В моделі FM дані для навчання необхідно структурувати, сформувавши вектори ознак x, перша частина яких кодує користувача, друга – елементи відеоконтенту, наступні – додаткові характеристики користувачів та елементів, поставивши їм у відповідність спостережувані рейтинги y. Формування навчаючої множини відбувається з використанням елементів матриці рейтингів R та додаткових атрибутів елементів і користувачів (рис. 2).

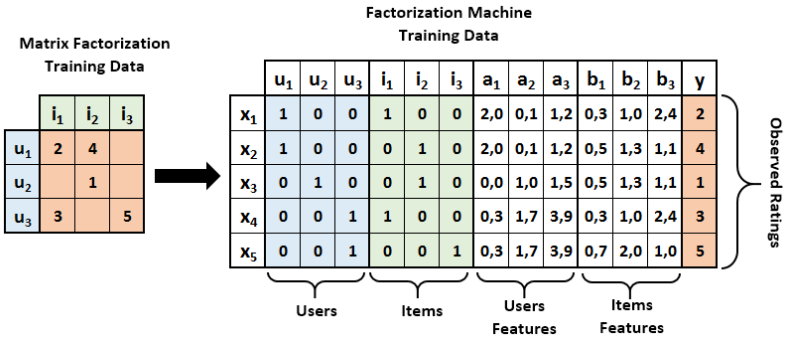


Рис. 2. Структурування даних у моделі машини факторизації FM

Модель FM націлена на моделювання взаємозв'зку між ознаками за допомогою факторизованих параметрів. Для розріджених наборів даних найбільш розповсюдженою є поліноміальна рейгресійна модель другого порядку, яка включає ваги для кожної базової ознаки та для кожної парної комбінації ознак:

$$\hat{y}(x) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} w_{ij} x_i x_j , \quad (7)$$

де w_0 – глобальне зміщення, n – кількість сформованих векторів ознак,
x_i – вектор i-ї ознаки, w_i – вага вектора ознаки x_i,
w_{ij} – вага для комбінації пари векторів i-ї та j-ї ознак,
$\hat{y}(x)$ – вектор прогнозованих рейтингів.

Ранг матриці W розмірністю $n \times n$, яка містить ваги w_{ij}, знижують шляхом застосування матричної факторизації:

$$\hat{y}(x) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle v_i, v_j \rangle x_i x_j , \quad (8)$$

де $w_{ij} \approx \langle v_i, v_j \rangle = \sum_{f=1}^{k} v_{if} v_{jf}$ – скалярний добуток двох векторів розміром $k < n$, k – ранг факторизації.

Тоді рівняння регресії буде мати вигляд:

$$\hat{y}(x) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} x_i x_j \sum_{f=1}^{k} v_{if} v_{jf} . \quad (9)$$

У процесі навчання моделі, яке здійснюють шляхом мінімізації функції втрат, визначають глобальне зміщення w_0, ваги векторів ознак w_i та факторизовані ваги v_{if}. Навчена модель дає можливість отримувати прогнозований рейтинг із врахуванням додаткових характеристик кристувача та елементів відеоконтенту.

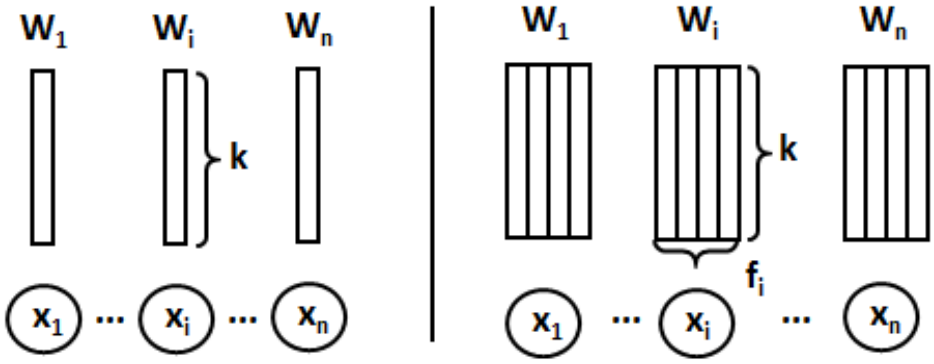
Модель машини факторизації з урахуванням поля FFM є розширенням машини факторизації шляхом додавання при формуванні рекомендацій концепції «поля»: схожі характеристики відносять до одного поля [25; 30]. Це дозволяє вирішити проблему моделі MF, яка полягає у тому, що приховані фактори, спільні для ознак, які інтуїтивно представляють різні категорії інформації, можуть погано узагальнювати кореляцію. Наприклад, може бути прихована взаємодія між атрибутами «жанр фільму» та «вік користувача», яка буде проявлятися у тому, що користувачі, які відносяться до різних вікових груп, надають перевагу різним жанрам фільмів. Для фіксації цієї прихованої взаємодії в моделі необхідно виділити поля в атрибутах «жанр фільму» та «вік користувача».

На відміну від стандартної машини факторизації FM, де кожній ознаці відповідає тільки один прихований вектор v_i, у моделі машини факторизації з урахуванням поля FFM для кожної ознаки створюється стільки векторів, скільки виділено полів. Рівняння поліноміальної регресії другого порядку моделі FFM має вигляд:

$$\hat{y}(x) = w_0 + \sum_{i=1}^{n} w_i x_i + \sum_{i=1}^{n} \sum_{j=i+1}^{n} \langle v_{if(j)}, v_{jf(i)} \rangle x_i x_j , \quad (10)$$

де $v_{if(j)}$ – вага для комбінації ознаки x_i з полем $f(j)$ ознаки x_j,
$v_{jf(i)}$ – вага для комбінації ознаки x_j з полем $f(i)$ ознаки x_i.

Різницю між моделями FM та FFM візуально відображено на рисунку 3.



а) модель FM б) модель FFM

Рис. 3. Порівняння моделей машини факторизації FM та машини факторизації з урахуванням поля FFM

У моделі машини факторизації вектору кожної ознаки x_i відповідає вектор її ваг з усіма іншими ознаками W_i. А в моделі машини факторизації з урахуванням полів вектору кожної ознаки x_i відповідає матриця ваг W_{if_i}, яка має f_i векторів ваг кожного поля ознаки x_i з полями усіх інших ознак.

Навчання моделі FFM здійснюють шляхом мінімізації функції втрат, визначаючи глобальне зміщення w_0, ваги векторів ознак w_i та факторизовані ваги векторів $v_{if(j)}$ із урахуванням виділених полів. Навчена модель дає можливість отримувати прогнозований рейтинг, у якому враховані приховані зв'язки між різними значеннями характеристик користувачів та відеофільмів.

Першим етапом при створенні рекомендаційної системи було формування набору даних та його попердня обробка. Для побудови рекомендаційної системи було досліджено описані вище моделі на наборі даних у вигляді файлів формату CSV із вебсайту MovieLens, який містить оцінку 58 000 фільмів 280 000 користувачами з січня 1995 року по вересень 2021 року [19]. У наборі даних представлено $27 \cdot 10^9$ оцінок фільмів у 5-бальній шкалі, характеристики користувачів – «вік» і «стать» та характеристики фільмів – «жанр» і «назва».

Назва фільму у наборі даних була об'єднана з датою виходу фільму у прокат, тому на етапі попередньої обробки даних ці атрибути було розділено з виділенням окремих характеристик кожного фільму: «назва» та «рік виходу». Попередня обробка даних включала також видалення порожніх рядків і невикористаних стовпців та фільтрування даних на основі критеріїв. Для відсіювання маловідомих фільмів було обрано фільми, які оцінили більше, ніж 15 користувачів. Після цих етапів остаточна кількість рейтингових даних становила 3 0872 062 рядків із 26 897 455 унікальною назвою фільмів і 274 952 унікальним користувачем.

На другому етапі здійснювалося дослідження моделей на основі матричної факторизації та вибір моделі, точність прогнозу якої є найвищою. Створення та навчання моделей матричної факторизації MF і машини факторизації з урахуванням поля FFM здійснювалося з використанням бібліотеки машинного навчання ML.NET (C#) [18]. Модель машини факторизації FM створено та навчено за допомогою бібліотеки Pytorch-Acceleratd [28]. У моделі FM при прогнозуванні рейтингів було враховано додаткову інформацію шляхом включення характеристик користувача: «вік» і «стать» та характеристик відеофільмів: «жанр» і «рік виходу». У моделі FFM досліджено ефективність прогнозування з урахуванням прихованої інформації, пов'язаної з полями, які були виділені у характеристиках «вік» і «стать» користувача та «жанр» фільму: 2 статі, 5 вікових груп, 18 жанрів.

Процес навчання моделей, націлений на мінімізацію відхилення між відомими рейтингами та їх прогнозами, реалізовано з використанням методу стохастичного градієнтного спуску шляхом мінімізації функції втрат [8; 26]. Для моделей MF та FM функція втрат визначалася з використанням методу найменших квадратів як сума квадратів помилок для усіх відомих рейтингів і включала фактори регуляризації для уникнення перенавчання рекомендаційної системи. Для моделі FFM було використано логістичну функцію втрат [6; 16].

Під час навчання моделей до навчаючої множини відбиралося 80% об'єктів набору даних, до тестової – 20%. З метою оцінки точності прогнозу було використано наступні показники: MAE, MSE, RMSE (табл. 1). Результати оцінки навчених моделей наведено у таблиці 2.

Таблиця 1

Показники для оцінювання точності прогнозу рейтингу
(y_i **і** y_i **–** *прогнозоване та фактичне значення рейтингу***)**

Показник	Опис	Формула
MAE	Середня абсолютна похибка (Mean Absolute Error)	$MAE = \frac{1}{n}\sum_{i=1}^{n}\left\|y_i - y_i\right\|$
MSE	Середньоквадратична похибка (Mean Squared Error)	$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$
RMSE	Корінь квадратний із середньоквадратичної похибки (Root Mean Squared Error)	$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$

Проведене дослідження дозволило установити, що кращі показники точності прогнозу мала модель машини факторизації з ураховунням поля FFM: MAE=0,86, MSE=1,65, RMSE=1,28. Тому для створення рекомендаційної системи з надання рекомендацій по перегляду відеофільмів було обрано модель FFM.

Таблиця 2

Результати оцінки точності навчених моделей

Модель	Показники оцінки		
	MAE	MSE	RMSE
MF	1,02	1,74	1,32
FM	0,89	1,71	1,31
FFM	0,86	1,65	1,28

На наступному етапі створення рекомендаційної системи здійснювалася розробка чат-бота та його інтеграція з навченою моделлю. Чат-бот є програмою або моделлю штучного інтелекту, призначеною для імітації людської розмови за допомогою текстової чи голосової взаємодії. Він може інтерпретувати та розуміти дані користувача і надавати необхідні відповіді у розмовній манері [21]. Спілкування користувачів з чат-ботом при наданні рекомендацій з перегляду відеофільмів забезпечує більшу гнучкість у пошуку відеоконтенту, який відповідає їхнім інтересам.

Створення та управління інтелектуальним чат-ботом для надання рекомендацій реалізовано з використанням Microsoft Bot Framework і Azure Bot Service. Таке поєднання технологій дозволяє використовувати бот як інтегрований компонент у різних каналах спілкування: месенджерах, соціальних мережах, вебзастосунках. Обробка природної мови та розпізнавання намірів користувача здійснювалися із використанням хмарних API Azure Cognitive Services та Conversational Language Understanding (CLU). Сервіс Azure Monitor було застосовано з метою моніторингу й аналізу даних у процесі роботи чат-бота. Для зберігання даних використано NoSQL-сховище Azure Table storage, завдяки чому було реалізовано масштабування у відповідності з потребами. У даному дослідженні чат-бот було інтегровано у середовище вебзастосунку, для розробки якого використано мову розмітки HTML та CSS як засіб стилізації.

Розглянемо архітектуру розробленої рекомендаційної системи (рис. 4). На початку роботи здійснюється виявлення намірів користувача на основі його спілкування з чат-ботом. Користувач взаємодіє з ботом у середовищі вебзастосунку шляхом обміну повідомленнями, для обробки яких використано Azure Cognitive Services та CLU. Отримана інформація стосовно уподобань користувача з перегляду фільмів передається до рекомендаційної системи, яка формує рекомендації та надає результат користувачеві через вебзастосунок. Для інтерактивного відображення рекомендацій у середовищі вебзастосунку використано REST API MovieLens.

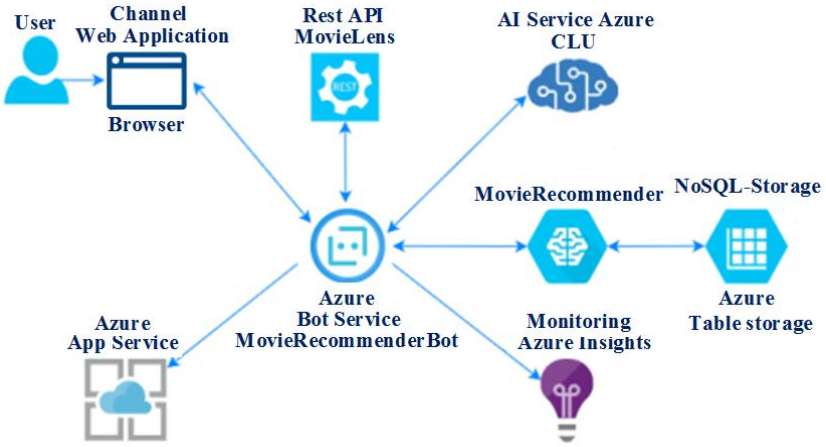


Рис. 4. Архітектура розробленої рекомендаційної системи

Під час спілкування з чат-ботом на основі аналізу обраних користувачем фільмів та їх оцінки рекомендаційна система накопичує інформацію про його уподобання й інтереси. Це дає можливість покращити точність прогнозу при наданні персоналізованих рекомендацій у подальшому.

Для початку спілкування з чат-ботом із метою виявлення намірів користувача необхідно у вікні браузера відкрити сторінку вебзастосунку та натиснути кнопку «Click me!» (рис. 5).

Рис. 5. Вікно вебзастосунку, інтегрованого з чат-ботом

У розробленій системі чат-бот може спілкуватися з користувачем, задаючи йому питання у вигляді текстових повідомлень, отримувати інформацію про уподобання користувача та надавати на основі цього рекомендації по перегляду фільмів. При створенні бота було навчено та протестовано модель розпізнавання мови користувача з використанням сервісу Conversational Language Understanding із визначенням можливих сценаріїв діалогу, намірів і категорій до кожного сценарію та розробки навчаючих висловлювань для кожного наміру. Оцінка якості створеної моделі обробки природної мови показала високу точність розпізнавання намірів користувача при спілкуванні з чат-ботом – 99,17% (рис. 6).

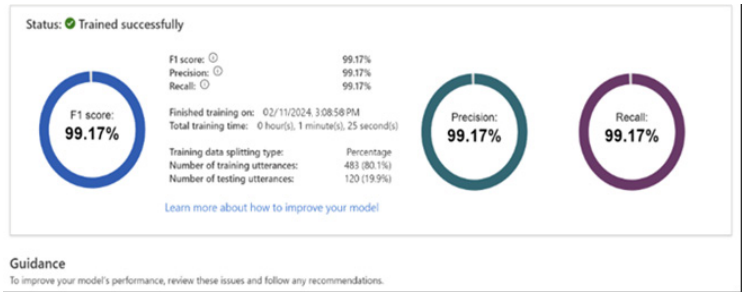


Рис. 6. Оцінка якості та точності моделі розпізнавання мови

При спілкуванні з чат-ботом є дві можливості, які визначають подальший напрямок взаємодії: користувач не вказує назву фільму та користувач вказує назву фільму, який він хоче переглянути. У першому випадку система видає кілька назв популярних фільмів, із яких користувач може обрати один із фільмів як такий, якому він надає перевагу. Передбачена також можливість на основі додаткових питань отримати від користувача бажаний жанр фільму та рік його виходу у прокат. Таким чином реалізується більша гнучкість системи для нових користувачів – обхід проблеми холодного старту. У другому випадку система шукає вказаний фільм та подібні до нього у сховищі даних. Фільми, рекомендавані для перегляду, з коротким описом їх характеристик відображаються у чаті (рис. 7).

Якщо користувач не задоволений наданими результатами, він може повторити запит на отримання інших рекомендацій. Якщо користувач схвально реагує на результат рекомендацій, система припиняє взаємодію з ним, оновлюючи інформацію про рейтингові оцінки користувача. Рекомендовані фільми можна переглянути у будь-якому сервісі, який надає доступ до відеоконтенту.

Рис. 7. Приклад рекомендованих для перегляду фільмів

Цілісна оцінка розробленої рекомендаційної системи повинна включати не тільки оцінку точності реалізованих моделей надання рекомендацій, а й якість реалізації розмовних стратегій чат-бота [14]. Це потребує суб'єктивних методів оцінювання, орієнтованих на сприйняття користувачами характеру взаємодії з чат-ботом.

Для дослідження рівня задоволеності користувачів результатами спілкування з чат-ботом та отриманими рекомендаціями було розроблено і впроваджено у вебзастосунок анкету (табл. 3).

Анкета містить 7 питань, згрупованих за шістьма факторами: інформативність (англ. Informative, INF), легкість у використанні (англ. Easy To Use, ETU), передбачувана якість рекомендацій (англ. Perceived Recommendation Quality, PRQ), легкість розуміння (англ. Ease Of Understanding, EOU), довіра (англ. Trust, TR) і сприймана ефективність (англ. Perceived Efficiency, PE) [5].

У кожному питанні анкети користувач може оцінити рейтинг різних факторів взаємодії з чат-ботом на основі визначених вагових коефіцієнтів у балах шляхом вибору наступних варіантів відповідей:

1) не згоден (англ. *Disagree*, DS) – 1 бал;
2) частково не згоден (англ. *Somewhat Disagree*, SD) – 2 бали;
3) нейтрально (англ. *Neutral*, NT) – 3 бали;
4) частково згоден (англ. *Somewhat Agree*, SA) – 4 бали;
5) абсолютно згоден (англ. *Strongly Agree*, ST) – 5 балів.

Таблиця 3

Питання анкети для оцінювання якості взаємодії з чат-ботом

ID	Фактор	Питання
P1	ETU	Ви можете легко орієнтуватися та взаємодіяти з чат-ботом
P2	ETU	Інструкції та вказівки, які надає система, чіткі та інтуїтивно зрозумілі
P3	EOU	Вам легко зрозуміти рекомендації, які надає система
P4	PE	Ви задоволені швидкістю та чутливістю системи
P5	PRQ	Рекомендовані відео відповідають вашим уподобанням та інтересам
P6	INF	Система надає достатньо інформації про кожен рекомендований фільм
P7	TR	Ви скористаєтеся рекомендаціям системи для майбутнього вибору фільмів

Анкетування проводилося онлайн серед 30 респондентів віком від 17 до 26 років, більшість з яких були студентами. Віковий діапазон містив користувачів, добре знайомих із цифровими технологіями, які часто переглядають відеофільми. За отриманими результатами остаточний бал кожного питання в анкеті K_j розраховувався за формулою:

$$K_j = \frac{\sum_{i=1}^{n} k_i^j}{k_{\max}^j}, \tag{11}$$

де k_i^j – оцінка у балах i–м опитуваним фактору взаємодії, відображеному у j–му питанні анкети, n – кількість опитаних респондентів,

$k_{\max}^j$ – максимальна кількість балів, яку можна отримати за j–те питання.

Враховуючи, що максильний ваговий коефіцієнт для кожного питання рівний 5-ти, а кількість респондентів становила 30, для кожного j–го питання анкети $k_{\max}^j = 5 \cdot n = 5 \cdot 30 = 150$.

Рівень задоволеності фактором взаємодії, відображеному у кожному питанні анкети, розраховують у процентах: $l_j = K_j \cdot 100\%$. Загальний рівень задоволеності користувачів взаємодією з чат-ботом визначають за формулою:

$$L = \frac{\sum_{i=1}^{m} l_j}{m}, \tag{12}$$

де $m = 7$ – кількість питань анкети.

У таблиці 4 показано результати проведеного за допомогою розробленої анкети тестування, які показують, що запропонована система рекомендацій на основі чат-бота може забезпечити задовільні результати для користувачів із кінцевим рівнем їх задоволеності 86,6%.

У питаннях P1 і P2 із фактором ETU – легкість у використанні, отримано позитивні результати від користувачів із рівнем задоволення 85,65%. Результати оцінки фактору EOU – легкості у розумінні отриманих рекомендацій, становлять 90%. Це вказує на те, що потік інструкцій під час спілкування з чат-ботом добре інтерпретують користувачі. Вони також задоволені швидкістю та чутливістю чат-бота, який відповідає на запити користувачів за короткий час: рівнень задоволення фактору PE – 86%.

Оцінка користувачами фактору PRQ – точності рекомендацій відеофільмів, становила 86%. Повнота отриманої інформації та наміри користуватися ситемою у майбутньому оцінені користувачами в 85,3% і 87,3%. У цілому рівень задоволеності користувачів спілкуванням із чат-ботом та отриманими під час спілкування рекомендаціями становив 86,6%.

Таблиця 4

Результати опитування щодо задоволення потреб користувачів

Питання анкети		Кількість відповідей на питання анкети					Рівень задоволеності	
Pj	**Фактор**	**DS (*1 бал*)**	**SD (*2 бали*)**	**NT (*3 бали*)**	**SA (*4 бали*)**	**ST (*5 балів*)**	**K_j, бали**	**l_j, %**
P1	ETU			7	10	13	126	84,0%
P2	ETU			5	9	16	131	87,3%
P3	EOU			2	11	17	135	90,0%
P4	PE		1	2	14	13	129	86,0%
P5	PRQ			4	13	13	129	86,0%
P6	INF		1	2	15	12	128	85,3%
P7	TR			2	15	13	131	87,3%
Остаточний результат, L=								86,6%

Проведене дослідження показало, що реалізація рекомендаційної системи на основі моделі машини факторизації з урахуванням полів та її інтеграція з чат-ботом, впровадженим у вебзастосунок, дозволяє надавати користувачам рекомендації з високим рівнем точності. А їх отримання шляхом спілкування з чат-ботом покращує взаємодію з рекомендаційною системою та робить її адаптованою до індивідуальних потреб користувача. Однак у деяких користувачів можуть виникнути проблеми з отриманням рекомендацій стосовно фільмів, інформація про яких відсутня у системі. Додавання до системи інформації про такі фільми частково допомагає цю проблему вирішити.

Висновки. Надання персоналізованих рекомендацій в умовах високих темпів росту обсягів цифрового відеоконтенту вимагає виявлення методів надання рекомендацій з високою точністю прогнозу стосовно намірів та уподобань користувачів та оптимальною гнучкістю їх взаємодії з рекомендаційною системою.

З метою виявлення ефективної моделі надання рекомендацій було досліджено моделі матричної факторизації MF, машини факторизації FM та машини факторизації з урахуванням поля FFM. Навчання моделей здійснювалося на наборі даних, який містив оцінку 280 000 користувачами 58 000 фільмів із використанням методу стохастичного градієнтного спуску. Отримані результати показали, що найвищі показники точності мала модель машини факторизації з урахуванням поля: MAE=0,86, MSE=1,65, RMSE=1,28.

Установлено, що інтеграція навченої моделі FFM з інтелектуальним чат-ботом, у якому реалізовано модель обробки природної мови, дозволяє реалізувати різні стратегії надання рекомендацій, орієнтовані на конкретного користувача, забезпечує гнучкість взаємодії з рекомендаційною системою, вирішує проблему холодного старту, спрощує отримання додаткової інформації. Доступ до каналу спілкування з чат-ботом, яке здійснюється шляхом обміну повідомленнями, надає розроблений вебзастосунок. Оцінка якості створеної та навченої моделі обробки природної мови показала високу точність розпізнавання намірів користувача при спілкуванні з чат-ботом – 99,17%. Дослідження ефективності взаємодії користувачів із чат-ботом, отримане шляхом обробки результатів анкетування, показало високий рівень задоволеності користувачів результатами спілкування з чат-ботом та отриманими рекомендаціями – 86,6%.

Таким чином, проведене дослідження дозволило виявити, що рекомендаційна система на основі моделі матричної факторизації з урахуванням поля FFM, інтегрована з чат-ботом, націлена на надання персоналізованих пропозицій щодо перегляду відеофільмів з високим рівнем точності та високим рівнем задоволеності користувачів результатами спілкування з чат-ботом. У процесі користування рекомендаційною системою накопичується інформація стосовно інтересів та потреб користувача і точність рекомендацій стає вищою. Проте є проблема, пов'язана з додаванням у систему інформації стосовно нових фільмів, яка у режимі реального часу не є вирішеною, оскільки потребує перенавчання моделі з оновленим набором даних. Що обумовлює необхідність подальших розробок у цьому напрямі.

Список використаних джерел:

1. Болюбаш Н. М. Інтелектуальний аналіз даних. Миколаїв: Вид-во ЧДУ ім. П. Могили, 2023. 320 с. URL: https://dspace.chmnu.edu.ua/jspui/handle/123456789/1461

2. Мелешко Є. В., Семенов С. Г., Хох В. Д. Дослідження методів побудови рекомендаційних систем в мережі Інтернет. *Системи управління, навігації та зв'язку*. Вип. 1(47). 2018. С. 131–136. URL: https://doi.org/10.26906/SUNZ.2018.1.131

3. Мелешко Є. В., Хох В. Д., Босько В. В. Дослідження матричних факторизаційних моделей рекомендаційних систем. *Системи управління, навігації та зв'язку*. Вип. 6(58), 2019. С. 58–62. URL: https://doi.org/10.26906/SUNZ.2019.6.058

4. Abbas M., Riaz M. U., Rauf A., Khan M. T., Khalid S. Context-aware Youtube recommender system. In *International Conference on Information and Communication Technologies* (*ICICT*). Karachi: IEEE, 2017. P. 161–164. URL: https://doi.org/10.1109/ICICT.2017.8320183

5. Attalariq M., Baizal Z.K.A. Chatbot-Based Book Recommender System Using Singular Value Decomposition. *Journal of Information System Research***.** Vol. 4, No 4. 2023. P. 1293–1301. URL: https://doi.org/10.47065/josh.v4i4.3817

6. Chin W. S., Zhuang Y., Juan Y. C., Lin C. J. A Learning-Rate Schedule for Stochastic Gradient Methods to Matrix Factorization. In *Advances in Knowledge Discovery and Data Mining*. *19th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2015)*. *Lecture Notes in Computer Science*. Vol. 9077. 2015. Springer Cham. P. 442–445. URL: https://doi.org/10.1007/978-3-319-18038-0_35

7. Dalton J., Ajayi V. Main R. Vote Goat: Conversational Movie Recommendation. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. Ann Arbor MI USA: ACM, 2018. P. 1285–1288. URL: https://doi.org/10.1145/3209978.3210168

8. Duchi J., Hazan E., Singer Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *Journal of Machine Learning Research*. Vol 12. 2011. P. 2121–2159. URL: https://jmlr.org/papers/volume12/duchi11a/duchi11a.pdf

9. Fajari A. N., Baizal A. Chatbot-based Culinary Tourism Recommender System Using Named Entity Recognition. *Journal Imiah Penelitian dan Pembelajaran Informatika*. Vol. 7, No. 4. 2022. P. 1131–1138. URL: https://doi.org/10.29100/jipi.v7i4.3210

10. Falk K. Practical recommender systems. Shelter Island, NY: Manning, 2019. 432 p.

11. Fayyaz Z., Ebrahimian M., Nawara D., Ibrahim A., Kashef R. Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities. *Applied Sciences*. Vol. 10(21). 2020. URL: https://doi.org/10.3390/app10217748

12. Gomez-Uribe C.A., Hunt N. The netflix recommender system: algorithms, business value, and innovation. *ACM Transactions on Management Information Systems*. Vol. 6, No 4. 2016. P. 1–19. URL: https://dl.acm.org/doi/10.1145/2843948

13. Hong F.X., Zheng X.L., Chen C.C. Latent space regularization for recommender systems. *Information Science*. Vol. 360. 2016. P. 202–216. URL: https://doi.org/10.1016/j.ins.2016.04.042

14. Jannach D. Evaluating conversational recommender systems: A landscape of research. Artificial Intelligence Reviev. Vol. 56, No. 3. 2023. P. 2365–2400. URL: https://doi.org/10.1007/s10462-022-10229-x

15. Jayalakshmi S., Ganesh N., C¨ep R., Senthil Murugan J. (2022). Movie recommender systems: concepts, methods, challenges, and future directions. *Sensors*. Vol. 22(13). URL: https://doi.org/10.3390/s22134904

16. Juan Y., Zhuang Y., Chin W.S. Field-aware Factorization Machines for CTR Prediction. In *Proceedings of the 10th ACM Conference on Recommender Systems*. 2016. P. 43–50. URL: https://doi.org/10.1145/2959100.2959134

17. Ma S., Zha Z., Wu F. Knowing user better: jointly predicting click-through and playtime for micro-video. In *IEEE International Conference on Multimedia and Expo* (*ICME*). 2019. P. 472–477. URL: https://doi.org/10.1109/ICME.2019.00088

18. ML.NET Documentation. URL: https://learn.microsoft.com/en-us/dotnet/machine-learning

19. MovieLens 25M Dataset. URL: https://grouplens.org/datasets/movielens/25m

20. Narducci F., Gemmis M.D., Lops P., Semeraro G. Improving the User Experience with a Conversational Recommender System. In *AI*IA Advances in Artificial Intelligence*. 2018. P. 528–538. URL: https://doi.org/10.1007/978-3-030-03840-3_39

21. Nugraha M., Baizal Z.K.A., Richasdy D. Chatbot-Based Movie Recommender System Using POS Tagging. *Building of Informatics, Technology and Science* (*BITS*). 2022. Vol. 4, No. 2. P. 624–630. URL: https://doi.org/10.47065/bits.v4i2.1908

22. Padti P. G., Hegde K., Kumar P. Hybrid Movie Recommender System. *International Journal of Research in Engineering, Science and Management*. Vol. 4, no. 7, 2021. P. 311–314.

23. Pujahari A., Sisodia D. S. Model-Based Collaborative Filtering for Recommender Systems: An Empirical Survey. In 2020 *First International Conference on Power, Control and Computing Technologies* (*ICPC2T*). Raipur, India: IEEE, 2020. P. 443–447. URL: https://doi.org/10.1109/ICPC2T48082.2020.9071454

24. Singh A., Ramasubramanian K, Shivam S. Building an Enterprise Chatbot: Work with Protected Enterprise Data Using Open Source Frameworks. Berkeley, CA: Apress. 2019. 385 p. URL: https://doi.org/10.1007/978-1-4842-5034-1

25. Sun J., Zhang A., Pan J., Flores A. Field-matrixed Factorization Machines for Recommender Systems. In *WWW'21: Proceedings of the Web Conference*. 2021. P. 2828–2837. URL: https://doi.org/10.1145/3442381.3449930

26. Taddy M. Stochastic Gradient Descent. In *Business Data Science: Combining Machine Learning and Economics to Optimize, Automate, and Accelerate Business Decisions*. New York: McGraw-Hill. 2019. P. 303–307.

27. Theosaksomo D., Widyantoro D. H. Conversational Recommender System Chatbot Based on Functional Requirement. In *IEEE 13th International Conference on Telecommunication Systems, Services, and Applications* (*TSSA*). 2019. P. 154–159. URL: https://doi.org/10.1109/TSSA48701.2019.8985467

28. Welcome to pytorch-accelerated's documentation. URL: https://pytorch-accelerated.readthedocs.io/en/latest

29. Zhang Y. An Introduction to Matrix factorization and Factorization Machines in Recommendation System, and Beyond. 2022. URL: https://doi.org/10.48550/arXiv.2203.11026

30. Zhang Z., Lui Y., Zhang Zh. Field-Aware Matrix Factorization for Recommender Systems. *IEEE Access*. Vol. 6. 2018. P. 45690–45698. URL: https://doi.org/10.1109/ACCESS.2017.2787741

31. Zhao X., Li X., Liao L., Song D., Cheung W.K. Crafting a time-aware point-of-interest recommendation via pairwise interaction tensor factorization. In *Knowledge Science, Engineering and Management. KSEM*. 2015. *Lecture Notes in Computer Science*. Vol. 9403. Cham, Switzerland: Springer. P. 458–470. URL: https://doi.org/10.1007/978-3-319-25159-2_41

УДК 004.4
DOI https://doi.org/10.32689/maup.it.2024.1.4

Дмитро БУХАЛЕНКОВ
магістрант, НТУУ «КПІ імені Ігоря Сікорського», 3a43mka@gmail.com
ORCID: *0009-0001-0224-8873*

Тетяна ЗАБОЛОТНЯ
кандидат технічних наук, доцент,
доцент кафедри програмного забезпечення комп'ютерних систем,
НТУУ «КПІ імені Ігоря Сікорського», zabolotnia@pzks.fpm.kpi.ua
ORCID: *0000-0001-8570-7571*

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ МОДИФІКОВАНОГО МЕТОДУ АВТОМАТИЗОВАНОГО ПОШУКУ КЛЮЧОВИХ СЛІВ У ТЕКСТІ

Анотація. *В умовах невпинного зростання обсягу текстових даних, які доводиться обробляти людині майже в усіх сферах її діяльності, непересічної важливості набуває задача забезпечення швидкого доступу до необхідної інформації. Для вирішення цієї задачі наявні пошукові системи, як правило, проводять індексацію даних: спеціальні боти сканують ресурси і намагаються відшукати пов'язані з ними ключові слова. Від коректності знайдених ключових слів напряму залежить релевантність результатів пошуку, що будуть видані користувачу пошукової системи.*

В даній статті розглянуто модифікований метод автоматизованого пошуку ключових слів у природномовних текстових даних. Він ґрунтується на аналізі складних синтаксичних зв'язків між словами в реченнях тексту та здатний шукати ключові терміни, що складаються з кількох слів.

Метою дослідження *є програмна реалізація та експериментальне дослідження ефективності модифікованого методу автоматизованого пошуку ключових слів у тексті.*

Методика реалізації*. Для випробувань модифікований метод було реалізовано на платформі Python NLTK. У якості тестового масиву даних було обрано два набори текстів: тексти невеликого обсягу (до 400 слів) та тексти більшого обсягу (до 2500 слів). Порівняння проводилися з трьома популярними аналогами, кожен з яких реалізовано на основі різних підходів (машинне навчання, аналіз N-грам, статистичний аналіз). Для кількісного вимірювання ефективності та порівняння з існуючими аналогами запропоновано використовувати метрики абсолютної точності та повноти за Жаккаром.*

Висновки*. Результати випробувань продемонстрували перевагу запропонованого методу над аналогами в точності пошуку ключових слів. Відмічено, що зі збільшенням обсягу текстів абсолютна точність зростає майже в усіх випадках, втім повнота за Жаккаром зменшується. На основі результатів випробувань сформульовано подальші напрямки роботи над покращенням запропонованого методу.*

Ключові слова: *ключові слова, аналіз ефективності, оброблення текстових даних, Python NLTK, стенфордська класифікація.*

Dmytro BUKHALENKOV, Tetiana ZABOLOTNIA. STUDY OF THE EFFECTIVENESS OF THE MODIFIED METHOD OF AUTOMATED SEARCH FOR KEYWORDS IN TEXT

Abstract*. In the conditions of constant growth of the volume of text data, which a person has to process in almost all spheres of his activity, the task of ensuring quick access to the necessary information becomes extremely important. To solve this problem, existing search engines, as a rule, perform data indexing: special bots scan resources and try to find keywords related to them. The relevance of the search results that will be issued to the user of the search engine directly depends on the correctness of the keywords found.*

This article discusses a modified method of automated search for keywords in natural language text data. It is based on the analysis of complex syntactic relationships between words in the sentences of the text and is able to search for key terms consisting of several words.

The research objective *is the programmatic implementation and experimental study of the effectiveness of the modified method of automated search for keywords in text data.*

Methodology of implementation*. For testing, the modified method was implemented on the Python NLTK platform. Two sets of texts were chosen as a test dataset: texts of a small volume (up to 400 words) and texts of a larger volume (up to 2500 words). Comparisons were made with three popular analogues, each of which is implemented on the basis of different approaches (machine learning, N-gram analysis, statistical analysis). For quantitative measurement of efficiency and comparison with existing analogues, it is proposed to use absolute accuracy and completeness metrics according to Jaccard.*

Conclusions*. The results of the tests demonstrated the superiority of the proposed method over analogues in the accuracy of searching for keywords. It was noted that with an increase in the volume of texts, the absolute accuracy increases in almost all cases, but the completeness according to Jaccard decreases. Based on the test results, further directions of work on improving the proposed method are formulated.*

Key words: *keywords, performance analysis, text data processing, Python NLTK, Stanford classification.*

Вступ. Задача пошуку ключових слів виникає у багатьох сферах роботи з текстовими даними. Інформація про ключові слова використовується при інформаційному текстовому пошуку, класифікації,

кластеризації даних тощо. За багато років досліджень спеціалістами було запропоновано методи, різні за точністю, ефективністю та можливістю застосування. Але на сьогоднішній день досі не існує універсального способу визначити перелік ключових термінів для довільного тексту будь-якої тематики. Кожен текст має свою структуру, стиль викладення, стилістичні особливості написання. Тож тривають пошуки нових шляхів вирішення задачі автоматизованого визначення ключових слів в текстових даних, а також спроби підвищити ефективність уже існуючих методів.

Аналіз останніх досліджень і публікацій. Дана стаття присвячена дослідженню ефективності модифікованого методу автоматизованого пошуку ключових слів у природномовних текстових даних [1]. В основу даного методу покладено сучасний гібридний метод пошуку ключових слів в англомовних текстах, що був запропонований українським фахівцем О.В. Яхимовичем [2] в 2021 році. Він застосовує можливості сучасних програмних лінгвістичних пакетів для побудови розмітки тексту і аналізу слів. Головною з особливостей цього методу, окрім фільтрації вербального шуму, можна назвати використання даних залежностей між парами слів та даних про частини мови, що отримуються за допомогою синтаксичного аналізатора. Але цей гібридний метод має суттєвий недолік: він здатен шукати лише поодинокі ключові слова, що погіршить точність видачі результатів пошуку. Запропонована модифікація усуває даний недолік і дозволяє знаходити ключові терміни, що складаються з кількох слів.

Постановка завдання. Метою даної роботи є експериментальне дослідження ефективності модифікованого методу автоматизованого пошуку ключових слів у текстових даних, сформульованого та теоретично обґрунтованого у [1].

У відповідності до поставленої мети задачами дослідження є:

– програмна реалізація модифікованого методу автоматизованого пошуку ключових слів;

– дослідження ефективності роботи розробленої програмної реалізації запропонованого методу автоматизованого пошуку ключових слів за критеріями абсолютної точності та повноти пошуку ключових слів за Жаккаром;

– визначення подальших кроків щодо підвищення ефективності модифікованого методу автоматизованого пошуку ключових слів.

Виклад основного матеріалу дослідження. Ключовими словами називають такі слова або вирази, якими можна описати основний зміст деякого тексту. Іноді ключовими словами, що відображають суть тексту, називають цілі словосполучення. В більшості випадків для одного тексту наводять близько десяти ключових слів [3].

Задача визначення ключових слів в тексті є складною і нетривіальною задачею, адже знайдені ключові слова повинні найбільш точно передавати тематику тексту. Таким чином, бажано не обирати в якості ключових загальновживані слова, або такі, що не несуть змістового навантаження. Точно визначеного алгоритму пошуку ключових слів людиною, нажаль, не існує, що робить цей процес складним для автоматизації [4].

Методи пошуку ключових слів

Більшість відомих методів автоматизованого пошуку ключових слів поділяється на кілька типів:

1. Статистичні методи – такі, що ґрунтуються на законах статистики [5].

2. Словникові методи – використовують наперед зібрані словникові дані, або тезауруси з деяких тематик [6].

3. Гібридні методи – поєднання особливостей статистичних та словникових методів для найбільш ефективного пошуку ключових слів [7].

Статистичні та словникові методи мають свої переваги і недоліки, тож сучасні дослідження проводяться в напрямках розроблення і покращення гібридних методів.

Модифікований метод автоматизованого пошуку ключових слів у тексті

Запропонований у [1] метод створений на базі гібридного методу, сформульованого у [2] і використовує інструменти сучасних програмних синтаксичних аналізаторів для оброблення текстів і отримання необхідних даних для подальшого зважування слів-кандидатів у ключові слова.

В загальному вигляді запропонований авторами модифікований метод є таким:

1. Синтаксичний аналіз тексту і отримання даних про зв’язки між парами слів і частини мови, до яких належать слова тексту.

2. Отримання з тексту набору всіх виразів з типами зв’язків flat та compound.

3. Фільтрування пар слів, зв’язки між якими належать до переліку неінформативних.

4. Заміна займенників в парах слів відповідними іменниками.

5. Відсіювання слів, які при синтаксичному аналізі було віднесено до неінформативних частин мови.

6. Фільтрування стоп-слів.

7. Визначення кількості зв’язків для кожного слова з пари.

8. Прийняття перших n слів з найбільшою кількістю зв’язків як ключові (де n – бажана кількість шуканих ключових слів).

9. Фільтрація отриманих багатослівних виразів за допомогою попередньо отриманих ключових слів.

Загальну схему запропонованого модифікованого методу наведено на рис. 1.

Для отримання пар слів використовується стенфордська класифікація [8] зв'язків між лексичними одиницями речень тексту. Для фільтрації слів, що відносяться до неінформативних частин мови, автори гібридного методу використовують класифікацію Пенна [9].

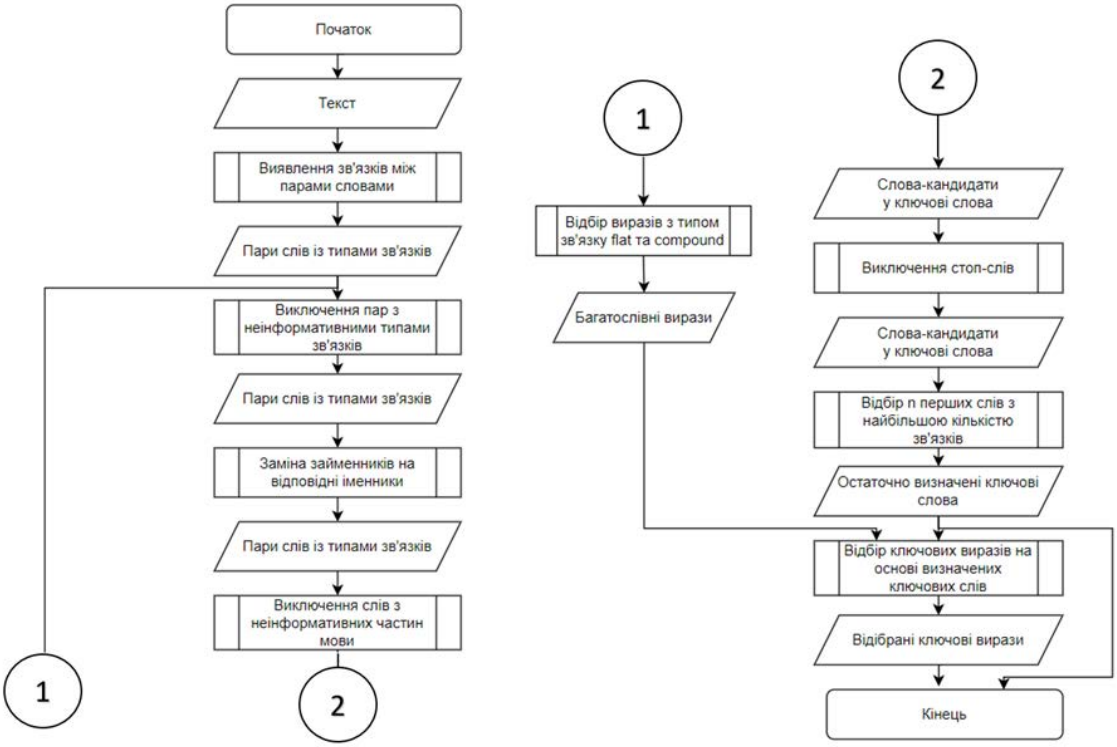


Рис. 1. Загальна схема запропонованого модифікованого методу [1]

Зазначимо, що, на відміну від гібридного методу [2], запропонована модифікація здатна знаходити ключові терміни з кількох слів. З точки зору використання таких методів в якості складових комплексної пошукової системи, однослівні ключові терміни сприяють більш загальному пошуку, але недостатньо добре покривають специфічні і конкретні запити. Таким чином, модифікований метод може покращити релевантність результатів пошуку, виданих пошуковою системою.

Кількісне оцінювання ефективності модифікованого методу автоматизованого пошуку ключових слів

Перш, ніж проводити практичне дослідження ефективності модифікованого методу автоматизованого пошуку ключових слів в тексті, необхідним є обрання способу кількісного оцінювання якості результатів роботи методу. У даній статті дослідження ефективності модифікованого методу проводиться із застосуванням двох метрик: абсолютної точності та повноти пошуку за Жаккаром. З огляду на постановку задачі, час знаходження ключових слів для вибраного тексту не має такого вирішального значення, як якість знайдених слів. Окрім того, більшість існуючих інструментів для пошуку ключових слів є онлайн сервісами, обчислювальні потужності яких можуть складатися із мережі серверів. В свою чергу, розроблене програмне забезпечення з реалізації запропонованого методу можливо протестувати лише на одному комп’ютері, тож порівняння часу виконання з хмарними серверами не є доречним.

Абсолютна точність визначається як відношення кількості правильно знайдених ключових слів за допомогою використання програмної реалізації методу до кількості ключових слів, визначених автором тексту. Якщо взяти множину еталонних ключових слів до деякого тексту як A, а множину ключових слів, що було знайдено програмою як B, тоді абсолютну точність а пошуку ключових слів можна обчислити за формулою:

$$a = \frac{n(A \cap B)}{n(A)}\# \tag{1}$$

де $n(A \cap B)$ – кількість правильно знайдених ключових слів; $n(A)$ – кількість еталонних ключових слів.

Повнота за Жаккаром визначається як відношення кількості правильно знайдених ключових слів до загальної кількості еталонних ключових слів і знайдених ключових слів мінус кількість правильно знайдених ключових слів. Повнота за Жаккаром J обчислюється за формулою:

$$J = \frac{n(A \cap B)}{n(A) + n(B) - n(A \cap B)} = \frac{n(A \cap B)}{n(A \cup B)}\# \tag{2}$$

де $n(B)$ – кількість програмно знайдених ключових слів; $n(A \cup B)$ – кількість елементів об'єднання обох множин [10].

Для застосування вищенаведених метрик до термінів, що складаються з кількох слів, у [1] пропонується використовувати метрику Word Accuracy (WAcc) з пороговим значенням 66,66%, що є оберненою до метрики Word Error Rate (WER) [11]. Значення метрики WER може бути обчислене за наступною формулою:

$$WER = \frac{S + D + I}{N}\# \tag{3}$$

де S – кількість замін; D – кількість видалень; I – кількість вставлень; N – кількість слів в "еталонному", або довідковому варіанті.

Значення метрики WAcc є оберненим до WER і обчислюється за формулою:

$$WAcc = 1 - WER\# \tag{4}$$

Для перевірки ефективності модифікованого методу було розроблено програмну реалізацію у середовищі Python за допомогою платформи Python NLTK та допоміжних пакетів AllenNLP та JiWER, особливості якої наведено в [1].

Аналіз результатів експериментальних досліджень ефективності модифікованого методу автоматизованого пошуку ключових слів в текстових даних

Порівняння ефективності роботи розробленого програмного забезпечення проводилися з наступними існуючими сервісами, що надають подібні можливості з пошуку ключових слів у тексті:

1. MonkeyLearn на основі машинного навчання [12].
2. WordCount, що використовує аналіз N-грам [13].
3. Komprehend, що побудований на статистичному аналізі [14].

Для експериментальної апробації ефективності модифікованого методу автоматизованого пошуку ключових слів в тексті було взято 200 довільних текстів тез до статей з наукового технічного журналу [15]. Це невеликі тексти обсягом 150-400 слів, із наперед зазначеним переліком ключових слів, тож їх зручно використовувати для швидкого тестування. Спочатку наведемо порівняння значень абсолютної точності для власної розробки та аналогів.

Рис. 2. Гістограма порівняння значень абсолютної точності пошуку ключових слів для текстів обсягом 150-400 слів

Результати випробувань (рис. 2) демонструють, що середнє значення абсолютної точності пошуку ключових слів для програмної реалізації модифікованого методу становить 0,402, для сервісу MonkeyLearn – 0,266, WordCount – 0.34, Komprehend – 0.294. Отже, власна розробка збільшує абсолютну точність пошуку ключових слів у межах від 6.2% до 13,6% у порівнянні з аналогами в текстах обсягом 150-400 слів.

Додатково були проведені випробування на текстах більших розмірів, близько 1500-2500 слів. Для цього було взято 100 довільних текстів статей з того ж наукового журналу.

Результати випробувань (рис. 3) демонструють, що середнє значення абсолютної точності пошуку ключових слів для власної розробки становить 0,592, для сервісу MonkeyLearn – 0,296, WordCount – 0.251, Komprehend – 0.575. Отже, власна розробка збільшує абсолютну точність пошуку ключових слів у межах від 1.7% до 34,1% у порівнянні з аналогами в текстах обсягом 1500-2500 слів.

Рис. 3. Гістограма порівняння значень абсолютної точності пошуку ключових слів для текстів обсягом 1500-2500 слів

За результатами отриманих значень абсолютної точності пошуку ключових слів можна відмітити наступне:

1. На текстах невеликого обсягу (150-400 слів) запропонований модифікований метод має явну перевагу над іншими аналогами.

2. На текстах більшого обсягу (1500-2500 слів) запропонований метод все ще має найкращі результати, однак сервіс Komprehend, що базується на методах статистики, майже його наздоганяє.

3. Збільшення середньої точності майже всіх 4 інструментів зі збільшенням обсягу тексту можна пояснити більш явним проявом статистичних закономірностей. Таким чином, за достатньо великого обсягу тексту, ключові вирази будуть неодноразово повторюватися, що збільшить їх вагу при зважуванні кандидатів у ключові терміни.

На основі отриманих результатів порівнянь можна стверджувати, що запропонований модифікований метод є кращим за критерієм абсолютної точності. Гіпотеза про використання інформації, отриманої з синтаксичного аналізатора, для пошуку багатослівних виразів в тексті надає модифікованому методу можливість шукати ключові терміни, які складаються з кількох слів, що і було очікуваним. Причому зі збільшенням обсягу тексту збільшується і кількість правильно знайдених ключових термінів, що можна пояснити більш частим входженням ключового терміну в текст. Виявлення ключових термінів з кількох слів було неможливим для методу, взятого за основу модифікації.

Для порівняння значень повноти пошуку ключових слів за Жаккаром випробування були проведені на тих самих 200 довільно обраних текстах тез статей з наукового технічного журналу [15].

Результати отриманих значень повноти пошуку ключових слів за Жаккаром для запропонованої модифікації та існуючих аналогів (рис. 4) демонструють, що середнє значення для власної розробки становить 0,088, для сервісу MonkeyLearn – 0,089, WordCount – 0.087, Komprehend – 0.048. Отже маємо зменшення повноти пошуку ключових слів за Жаккаром на 0,1% у порівнянні з найкращим результатом іншого метода при застосуванні на текстах обсягом 150-400 слів.

Аналогічно були проведені випробування на текстах обсягом 1500-2500 слів.

Рис. 4. Гістограма порівняння значень повноти пошуку ключових слів за Жаккаром для текстів обсягом 150-400 слів

Рис. 5. Гістограма порівняння значень повноти пошуку ключових слів за Жаккаром для текстів обсягом 1500-2500 слів

Результати отриманих значень повноти пошуку ключових слів за Жаккаром (рис. 5) демонструють, що середнє значення для власної розробки становить 0,058, для сервісу MonkeyLearn – 0,108, WordCount – 0.065, Komprehend – 0.012. Отже маємо зменшення повноти пошуку ключових слів за Жаккаром на 5% у порівнянні з найкращим результатом іншого метода при застосуванні на текстах обсягом 1500-2500 слів.

За результатами отриманих значень повноти пошуку ключових слів за Жаккаром можна відмітити наступне:

1. На текстах невеликого обсягу (150-400 слів) значення повноти майже однакові, окрім значень сервісу Komprehend, що базується на використанні методів статистики. Це можна пояснити тим, що даний сервіс видавав помітно більше ключових слів у результатах, чим значно понижував повноту через збільшення вербального шуму.

2. На текстах більшого обсягу (1500-2500 слів) запропонована модифікація має третій результат, що можна пояснити більшою кількістю ключових слів на виході, у порівнянні з такими аналогами як MonkeyLearn та WordCount. MonkeyLearn має найкращий результат, адже сервіс завжди обмежує кількість ключових слів до 10. Найгірший результат – у сервіса Komprehend, який отримує в результаті занадто багато вербального шуму і має повноту пошуку близько 1%.

На основі отриманих результатів порівнянь можна стверджувати, що запропонований модифікований метод має меншу повноту пошуку ключових слів за Жаккаром у порівнянні з аналогами, однак не найменшу. На етапі пошуку ключових термінів, що складаються з кількох слів, дещо збільшується вербальний шум, що зменшує значення повноти пошуку, причому кількість вербального шуму збільшується зі збільшенням обсягу тексту.

Виходячи з вищенаведених результатів дослідження ефективності модифікованого методу автоматизованого пошуку ключових слів у тексті, можна визначити такі подальші напрямки роботи над підвищенням ефективності роботи запропонованого методу та розширенням його можливостей: проведення додаткового тренування моделей для синтаксичних парсерів; використання синонімічних та тематичних словників для пошуку ключових термінів, які не зустрічаються в тексті; зменшення кількості вербального шуму в результатах пошуку; проведення більшої кількості досліджень на текстах різних тематик та стилю; додавання підтримки пошуку ключових слів в текстах інших природних мов.

Висновки. В рамках даного дослідження виконано програмну реалізацію модифікованого методу автоматизованого пошуку ключових слів в тексті. Проведені порівняння ефективності розробленої програми та існуючих сервісів MonkeyLearn (машинне навчання і статистичний аналіз), WordCount (аналіз N-грам) та Komprehend (статистичний аналіз) за критеріями абсолютної точності та повноти пошуку за Жаккаром показали, що запропонований модифікований метод має кращі показники абсолютної точності, але повнота може бути нижчою, ніж в інших аналогів. На основі отриманих результатів дослідження ефективності модифікованого методу для автоматизованого пошуку ключових слів запропоновано подальші кроки щодо підвищення його ефективності та розширення можливостей його застосування.

Список використаних джерел:

1. Бухаленков Д.О., Заболотня Т.М. Модифікований метод пошуку ключових слів та термінів у текстових даних. *Проблеми програмування* № 1 (2024). С. 12–22. Київ, 2024.

2. Яхимович О.В. Інформаційна технологія пошуку ключових слів на основі парсингу англомовних текстів. Вінниця, 2021.

3. Shibamouli Lahiri, Sagnik Ray Choudhury, Cornelia Caragea. Keyword and Keyphrase Extraction Using Centrality Measures on Collocation Networks, 2014.

4. C. Zhang, H. Wang, Y. Liu, D. Wu, Y. Liao, and B. Wang, «Automatic keyword extraction from documents using conditional random fields», *Journal of Computational Information Systems* №4, pp. 1169–1180, 2008.

5. Rafael Geraldeli Rossi, Ricardo Marcondes Marcacini, Solange Oliveira Rezende. Analysis of Statistical Key-word Extraction Methods for Incremental Clustering. Proceedings of the 10th of the Encontro Nacional de Intelig^encia Artificial e Computacional (ENIAC), Fortaleza, Brazil, 2013, 1–12.

6. Takashi Yamauchi, Dongshik Kang, Hayao Miyagi. The Keyword Search Using Thesaurus Concept, 2002 [Електронний ресурс] URL: https://koreascience.kr/article/CFKO200211921321260.pdf (дата звернення 27.03.2024).

7. K. S. Sampada, N Kavya. Machine Learning Methods for Keyword extraction and Indexing, 2019.

8. Marie-Catherine de Marneffe, Christopher D. Manning (2008). Stanford typed dependencies manual [Електронний ресурс] URL: https://downloads.cs.stanford.edu/nlp/software/dependencies_manual.pdf (дата звернення 27.03.2024).

9. Beatrice Santorini (1990). Part-of-Speech Tagging Guidelines for the Penn Treebank Project [Електронний ресурс] URL: https://www.cis.upenn.edu/~bies/manuals/tagguide.pdf (дата звернення 27.03.2024).

10. NC Chung, B. Miasojedow, M. Startek, A. Gambin (2019). «Jaccard/Tanimoto similarity test and estimation methods for biological presence-absence data». BMC Bioinformatics.

11. Klakow, Dietrich; Jochen Peters (September 2002). «Testing the correlation of word error rate and perplexity». Speech Communication. 38 (1–2): 19–28. doi:10.1016/S0167-6393(01)00041-3. ISSN 0167-6393

12. Keyword Extractor – MonkeyLearn [Електронний ресурс] URL: https://monkeylearn.com/keyword-extractor-online/ (дата звернення 27.03.2024).

13. Keyword Extractor – WordCount [Електронний ресурс] URL: https://wordcount.com/keyword-extractor (дата звернення 27.03.2024).

14. Keyword Extractor – Komprehend [Електронний ресурс] URL: https://komprehend.io/keyword-extractor (дата звернення 27.03.2024).

15. Journal of Aerospace Technology and Management [Електронний ресурс] URL: https://jatm.com.br/jatm/issue/archive (дата звернення 27.03.2024).

УДК 004.056
DOI https://doi.org/10.32689/maup.it.2024.1.5

Андрій ГЛАЗУНОВ
аспірант спеціальності 122 «Комп'ютерні науки»,
Національний університет біоресурсів і природокористування України,
glasgarick2013@gmail.com
***ORCID:** 0009-0003-8631-8430*

ОГЛЯД ТА АНАЛІЗ ДОСЛІДЖЕНЬ З ПРОБЛЕМАТИКИ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ ХМАРНИХ ІНФРАСТРУКТУР

Анотація. *Хмарні обчислення є моделлю забезпечення доступу до мережевих ресурсів, таким як сховища даних і обчислювальні потужності на вимогу, без прямого управління з боку користувачів. В даний час хмарні обчислення включають як публічні, так і приватні центри обробки даних, що надають клієнтам єдину платформу через інтернет. Периферійні обчислення (edge computing) – це стратегія, спрямована на наближення виконання обчислень та збереження інформації до кінцевих користувачів, скорочення часу відгуку та оптимізації пропускної спроможності хмарних сервісів. Мобільні хмарні обчислення використовують розподілені обчислення для передачі програм на мобільні пристрої, такі як телефони та планшети. Численні дослідження показують, що хмарні обчислення і мобільні хмарні обчислення стикаються з проблемами інформаційної безпеки (ІБ), загрозами та вразливістю для клієнтів, і одним із перспективних методів боротьби з цими загрозами є використання методів машинного навчання (МН). У цій статті проведено аналіз загроз та проблем з ІБ, а також виконано огляд, запропонованих різними авторами рішень, щодо забезпечення ІБ хмарних обчислень та хмарних сервісів. Насамперед, розглянуто дослідження, що ґрунтуються на застосуванні алгоритмів МН для забезпечення безпеки хмарних обчислень та хмарних сервісів.*

Ключові слова: *хмарні обчислення; інформаційна безпека; машинне навчання; кібератаки, аномалії.*

Andrii HLAZUNOV. REVIEW AND ANALYSIS OF RESEARCH ON THE ISSUES OF INFORMATION SECURITY OF CLOUD INFRASTRUCTURES

Abstract. *Cloud computing is an access to network resources, such as data storage and computing power, on demand, without direct control by users. Currently, cloud computing includes both public and private data centers that provide customers with a single platform over the Internet. Peripheral computing (edge computing) is a strategy aimed at bringing computing and information storage closer to end users, reducing response time and optimizing the bandwidth of cloud services. Mobile cloud computing uses distributed computing to deliver applications to mobile devices such as phones and tablets. Numerous studies show that cloud computing and mobile cloud computing face information security (IS) challenges, threats, and vulnerabilities for customers, and one of the promising methods to combat these threats is the use of machine learning (ML) techniques. In this article, an analysis of IS threats and problems is carried out, as well as a review of the solutions proposed by various authors for the IS provision of cloud computing and cloud services. First of all, research based on the application of MN algorithms to ensure the security of cloud computing and cloud services is considered.*

Key words: *cloud computing; informational security; machine learning; cyber attacks, anomalies.*

Вступ. Хмарні обчислення з'явилися відносно нещодавно, як нова структура для спрощення та надання послуг через Інтернет [22]. Організація хмарних обчислень включає розміщення одного або декількох центрів обробки даних (ЦОД), які взаємопов'язані між собою. Ця система спроектована таким чином, що для користувача немає різниці між фізичними компонентами системи та їх віртуальними уявленнями. Завдяки цьому користувач хмарних обчислень може взаємодіяти з обчислювальними ресурсами, не турбуючись про технічні деталі та організацію процесу, оскільки ці завдання повністю покладаються на оператора хмарного сервісу. Фінансові обмеження, пов'язані з оптимізацією витрат приватних та державних структур (у загальному випадку об'єктів інформаційної діяльності – ОІД), а також зростаючі потреби в обчислювальних ресурсах, вимагають зростання обсягів сховищ даних із паралельним збільшенням потреб в аналізі. Ці та інші чинники сприяли розширенню попиту на різні хмарні моделі [10, 13]. Однак, як було показано в роботах [24, 30] хмарні обчислення та хмарні сервіси, мають низку проблем із забезпеченням інформаційної безпеки (ІБ). Зауважимо, що з точки зору ІБ є певна різниця між забезпеченням ІБ хмарних обчислень і хмарних сервісів. Ці відмінності можна звести до таких категорій:

1. Рівень контролю ІБ для хмарних обчислень та хмарних сервісів. У хмарних обчислень клієнти мають великий контроль за безпекою своїх даних і додатків, тоді як у хмарних сервісів більше контролю за безпекою, має провайдер. У хмарних сервісів, клієнт може мати досить обмежені можливості для налаштування та управління політиками безпеки.

2. Поверхня атаки для хмарних обчислень та хмарних сервісів. У хмарних обчислень клієнтська інфраструктура та програми можуть бути вразливими перед атаками, зокрема мережевими. У той же час

для хмарних сервісів поверхня атаки, як правило, набагато менша, оскільки провайдер відповідає за ІБ своєї інфраструктури.

3. Відповідність хмарних обчислень та хмарних сервісів. У хмарних обчисленнях клієнти безпосередньо повинні відповідати вимогам безпеки. У хмарних сервісів тільки провайдер несе відповідальність за дотримання вимог ІБ.

Таким чином, хмарні обчислення та хмарні сервіси мають дещо відмінні моделі безпеки. Вибір конкретної моделі залежатиме від потреб клієнта, його технічної експертизи та вимог до безпеки. Все сказане вище і мотивувало виконати аналіз наукових публікацій, присвячених виключно проблематиці забезпечення ІБ хмарних обчислень і хмарних сервісів.

2. Огляд попередніх досліджень.

У [17] автори розглядають загальний алгоритм вирішення проблем безпеки підвищення продуктивності хмарної системи. Автори використовували штучні нейронні мережі (ШНМ) для аналізу захищеності хмарного середовища.

У [19] розглядається організація системи безпеки хмарних обчислень, заснована на довірі, у хмарних моделях. Тобто провайдер забезпечує надійність, безпеку та конфіденційність даних у хмарній системі. Авторами запропоновано модель управління доступом на основі довіри як ефективний метод забезпечення ІБ у розподілених обчислювальних інфраструктурах. Як клієнтські, так і хмарні активи клієнтів у хмарної системі, у цьому дослідженні оцінюються з урахуванням аналізу їх довіри.

У [31] аналізуються моделі забезпечення ІБ хмарної інфраструктури. Автори розглянули у своїй роботі відмінні проблеми ІБ розподілених обчислень, що виникають у результаті використання об'єктами інформаційної діяльності різних моделей хмарних обчислень. Як показано авторами, приватні хмари зазвичай використовуються організаціями для своїх внутрішніх потреб. Вони вимагають суворого контролю доступу та управління. Громадські хмари надаються сторонніми провайдерами і можуть бути менш прозорими щодо безпеки. У громадських хмарах ресурси можуть розподілятися між різними клієнтами. Це потребує додаткових заходів безпеки.

У [9] автори розглядають моделі машинного навчання (далі МН) підвищення безпеки даних у хмарних системах. Концепція забезпечення ІБ розподілених обчислень обговорюється авторами в контексті віртуалізації серверних ферм як практичного середовища розгортання бізнес-додатків. На думку авторів, віртуалізація серверних ферм допомагає забезпечити безпеку хмарних обчислень шляхом ізоляції, оскільки віртуальні сервери у фермі можуть бути ізольовані один від одного, запобігаючи несанкціонованому доступу (НСД). Крім того, ферма може динамічно масштабуватись в залежності від навантаження, забезпечуючи гнучкість та ефективність. Віртуальні сервери можуть мати різні рівні доступу, що сприяє безпеці.

У [37] автори наводять модель класифікації загроз для хмарних обчислень. Особливість класифікації полягає в тому, що вона заснована на можливості алгоритмів МН для виявлення та вирішення проблем безпеки. Крім того, авторами пропонується модель угруповання ризиків для хмарних обчислень. Модель ґрунтується на алгоритмах МН.

Аналогічні дослідження були проведені в роботі [13]. Дослідження присвячене аналізу сучасних підходів, вкладених у забезпечення ІБ хмарних сервісів. З огляду на те, що хмарні обчислення є однією з найбільш зростаючих областей у сфері інформаційних технологій, забезпечення безпеки та надійності процесів, що відбуваються у хмарах, а також захист механізмів взаємодії між клієнтами та постачальниками хмарних сервісів, становлять вкрай важливе наукове та прикладне завдання. Побоювання щодо втрати даних та їх компрометації стоять біля витоків небажання деяких компаній переміщати свої обчислення до хмар. Автор аналізує різноманітність хмарних сервісів, що надаються різними провайдерами, та порівнює існуючі підходи до забезпечення ІБ у цій сфері. Крім того, пропонується новий підхід, що базується на принципі диверсифікації. На думку автора, застосування диверсифікації необхідне забезпечення надійності та безпеки критичних компонентів хмарних систем. Цей принцип полягає у використанні унікальної версії кожного ресурсу завдяки особливій комбінації провайдерів хмарних обчислень, географічного розміщення центрів обробки даних, моделей надання хмарних сервісів та моделей розгортання хмарної інфраструктури.

У [29] автори досліджують алгоритми МН, які можуть бути використані для усунення загроз ІБ, пов'язаних із поширенням шкідливого програмного забезпечення (ПЗ) у хмарних системах. Авторами запропоновано бар'єрну структуру, яка використовує три алгоритми МН та призначена для виявлення шкідливого ПЗ.

Як було показано в [19, 35] хмарні обчислення мають значний потенціал для зростання і стає все більш популярними. Однак, незважаючи на свої унікальні характеристики, хмарні обчислення пов'язані з різними загрозами безпеці. Категоризація загроз була виконана багатьма авторами, зокрема у роботах [13, 19, 35].

Загрози конфіденційності включають інсайдерські загрози для клієнтської інформації, а також ризики зовнішніх атак [14]. [14] показано, що, по-перше, інсайдерський ризик для клієнтської інформації, пов'язаний з несанкціонованим або незаконним доступом до інформації про клієнта з боку інсайдера постачальника хмарних послуг. Це серйозна проблема безпеки [19]. По-друге, ризик зовнішніх атак стає дедалі актуальнішим для хмарних обчислень. Цей ризик включає віддалені програмні або апаратні атаки, спрямовані на клієнтів та хмарні програми [14]. По-третє, витік інформації є необмеженим ризиком для хмарних даних через навмисні та/або ненавмисні людські помилки.

Загрози цілісності інформації з організацією хмарних обчислень розглянуті у роботах [16, 31]. По-перше, це ризик ізоляції інформації, яка неточно поєднує значення параметрів безпеки, необачне проектування віртуальних машин (ВМ) та зовнішні клієнтські гіпервізори. По-друге, це погане управління доступом клієнтів, яке через неефективний контроль доступу може зіткнутися з різними проблемами та загрозами ІБ. Що дозволить потенційним зловмисникам завдати шкоди інформаційним активам, розміщеним у хмарі [5, 20].

Як показано в [12, 20] загрози доступності включають, наприклад, фізичне переривання роботи хмарних обчислень та/або хмарних сервісів, а також пов'язані з неефективними стратегіями відновлення після атак на хмарні системи.

У роботах [1, 2, 4, 18, 23, 34] аналізуються різні види атак на хмарні системи. У [18] обговорюються сценарії мережевих атак. Зокрема, як зазначають автори, сканування портів становить для хакерів значний інтерес, оскільки дає інформацію про запуск успішної атаки. У [34] розглядаються спуфінг-атаки при яких хакер чи шкідливе ПЗ діють від імені іншого користувача (або системи), видаючи себе за дані.

У [18] також розглянуті особливості організації атак на основі ВМ. У роботі показано, що різні ВМ, що використовуються у хмарних платформах, можуть викликати різні проблеми з ІБ. Наприклад, у разі, коли шкідливий код, розміщений в середині образу ВМ, буде реплікований під час створення ВМ. Також автори аналізують атаки з урахуванням додатків, які у хмарі. Такі атаки можуть вплинути на його продуктивність хмарних додатків і спричинити витік інформації в зловмисних цілях.

Як самостійний напрямок досліджень із проблематики ІБ хмарних інфраструктур можна вважати роботи, пов'язані із застосуванням методів МН для забезпечення ІБ хмарних обчислень.

Згідно [11] машинне навчання – це логічна перевірка розрахунків та вимірних моделей, які комп'ютерні системи використовують для реалізації конкретного завдання.

З погляду ІБ методи МН [27] настільки значущі у хмарі, що у найближчому майбутньому кожна хмарна система використовуватиме методи МН. Зі збільшенням затребуваності хмарних обчислень та зростанням навантаження на систему, а також обсягу трафіку стає необхідним активне моніторингове втручання в роботу ЦОД. Це дозволяє забезпечити безперебійну роботу інфраструктури, оскільки оперативне реагування на загрози ІБ та несправності сприяє стабільності та безпеці системи. Моніторинг, включаючи відстеження стану компонентів та управління хмарною інфраструктурою, відіграє ключову роль у забезпеченні високого рівня послуг, оптимізації розподілу ресурсів та забезпеченні надійності та ІБ. Це однаково важливо як для клієнтів, так і для провайдерів хмарних послуг. Моніторинг хмарного середовища включає кілька підтипів, кожен з яких виконує свої функції [18]. Зокрема, моніторинг ІБ – виявлення потенційно небезпечних алгоритмів та запобігання порушенням безпеки хмарних систем.

Для ефективної побудови системи моніторингу необхідно розробити математичну модель, що ґрунтується на оцінці параметрів системи в різних станах та часі, а також на основі застосування методів МН. Такий підхід дозволить створити формальний апарат із вхідними, проміжними та вихідними станами, що сприятиме забезпеченню ІБ як хмарних обчислень загалом, так і хмарних сервісів, зокрема.

У роботах [26, 32, 36] розглядаються інстанси хмарної інфраструктури у контексті забезпечення ІБ хмарних обчислень. Інстанси є віртуальними або фізичними обчислювальними ресурсами, що надаються хмарним провайдером для виконання різних завдань і додатків. Ці ресурси можуть включати ВМ, контейнери, сервери, бази даних (БД) та інші обчислювальні ресурси, які можуть бути масштабовані та налаштовані відповідно до потреб клієнта. На думку авторів [26], інстанси (ВМ або контейнер) хмарної інфраструктури відіграють важливу роль у забезпеченні ІБ хмарних сервісів, див. рис. 1.

По-перше, провайдери часто використовують віртуалізацію для створення та управління інстансами хмарної інфраструктури. Це дозволяє забезпечити ізоляцію та сегментацію ресурсів між різними клієнтами, що допомагає запобігти НСД до даних та додатків. По-друге, інстанси хмарної інфраструктури можуть використовуватися для виявлення аномалій, потенційних загроз та несанкціонованих дій. Це дозволяє операторам хмарних сервісів оперативно реагувати на можливі інциденти безпеки та запобігати їм. По-третє, використання інстансів також дозволяє налаштовувати права доступу та політики ІБ для різних користувачів та програм. Це забезпечує контроль над доступом до даних та ресурсів та допомагає запобігти НСД. По-четверте, інстанси можуть бути налаштовані за допомогою

спеціалізованих засобів захисту від DDoS-атак, що допомагає забезпечити безперервність роботи сервісів навіть при масованих мережевих атаках.

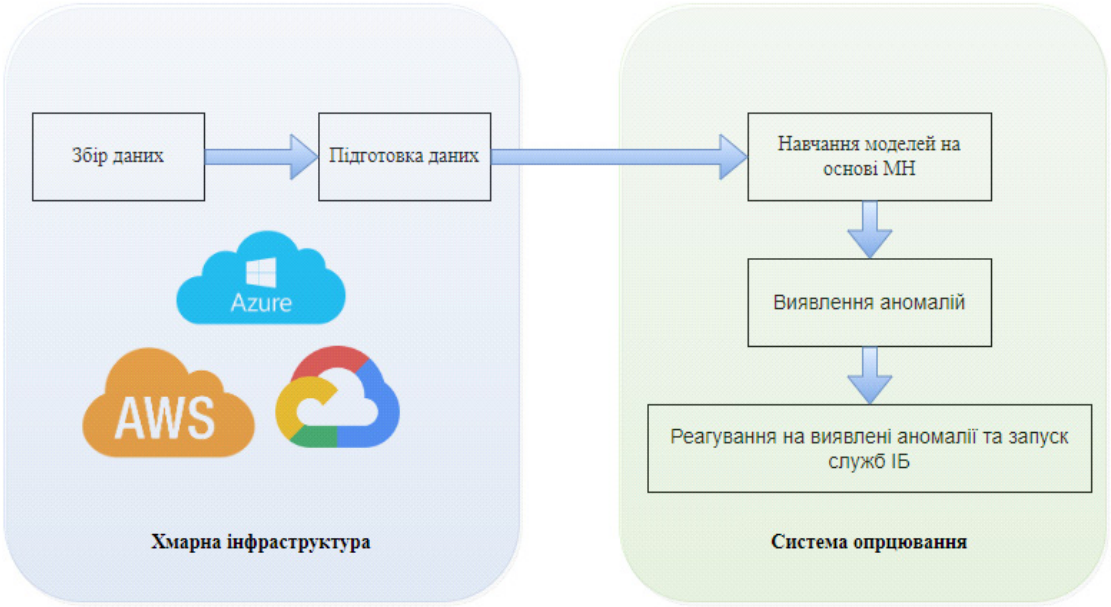


Рис. 1. Схема взаємодії хмарної інфраструктури та інстансів у системі виявлення аномалій ІБ на основі застосування методів МН

На рисунку 1 схематично показано взаємодію хмарної інфраструктури та інстансів у системі виявлення аномалій ІБ на основі застосування методів МН. Така взаємодія може бути реалізована у кілька етапів.

Етап 1. Збір даних.

Хмарна інфраструктура надає середовище для розгортання та виконання інстансів, які можуть бути використані для виконання різних завдань та програм. У процесі роботи інстанси генерують дані про поведінку, такі як використання ресурсів, мережевий трафік, журнали подій тощо.

Етап 2. Підготовка даних.

Дані, зібрані з інстансів, обробляються та готуються до аналізу. Це може включати очистку даних, масштабування ознак, перетворення форматів і т.д.

Етап 3. Навчання моделей з урахуванням МН.

На основі підготовлених даних будуються моделі МН виявлення аномалій. Ці моделі можуть включати алгоритми класифікації, кластеризації, дерева рішень, та ін, здатні виявляти незвичайні або підозрілі патерни в даних.

Етап 4. Виявлення аномалій.

Навчені моделі застосовуються до нових даних, що надходять від інстансів хмарних обчислень, виявлення аномалій. Це може включати аналіз аномальних шаблонів використання ресурсів, незвичайних мережевих пакетів, незапланованих подій у журналах та інших незвичайних сценаріїв.

Етап 5. Реагування на виявлені аномалії.

Після виявлення аномалій система ІБ може вживати різних заходів реагування, таких як надсилання повідомлень адміністраторам, блокування доступу до ресурсів, запуск додаткових заходів безпеки тощо.

У даній структурній схемі, взаємодія між хмарною інфраструктурою, та інстансами в системі виявлення аномалій ІБ, засноване на зборі та аналізі даних від інстансів, з подальшим навчанням моделей МН для виявлення аномалій та вжиття відповідних заходів щодо забезпечення ІБ. Таким чином, інстанси хмарної інфраструктури відіграють ключову роль у забезпеченні ІБ хмарних сервісів, забезпечуючи ізоляцію, моніторинг, керування доступом та захист від мережевих атак. Використання методів МН для виявлення аномалій в інстансах хмарної інфраструктури, може бути ефективним способом виявлення незвичайних або шкідливих дій, які можуть загрожувати безпеці системи. Однак, зазначимо, що для ІБ хмарних сервісів у такій схемі можуть мати місце деякі відмінності. Що пов'язано з такими чинниками:

- Рівень абстракції. Хмарні послуги надають абстрактніший рівень доступу до обчислювальних ресурсів, ніж просто хмарна інфраструктура. Користувачі хмарних сервісів часто мають справу з більш високорівневими сервісами, такими як платформи як сервіс (PaaS), програмне забезпечення як сервіс

(SaaS) і т.д. Це може вплинути на способи збирання та аналізу даних для виявлення аномалій, оскільки доступ до низькорівневих деталей інфраструктури може бути обмежений.

• Типи даних. У хмарних сервісів можуть генеруватися різні типи даних, наприклад, дані про взаємодію користувачів з додатками або обробку транзакцій. Відповідно методи виявлення аномалій можуть бути спрямовані на аналіз цих специфічних типів даних.

• Інтеграція з API. Багато хмарних сервісів надають API для взаємодії з ними. Використання таких API може полегшити збирання даних для виявлення аномалій та інтеграцію із системами ІБ.

• Керування доступом та ІБ. Багато хмарних сервісів мають власні механізми керування доступом і заходи безпеки, які можуть впливати на способи виявлення аномалій. Наприклад, наявність механізмів автентифікації та авторизації може сприяти ідентифікації підозрілих активностей.

Відповідно, взаємодія з хмарними сервісами може вимагати врахування специфічних особливостей цих сервісів та адаптацію методів виявлення аномалій відповідно до їх характеристик.

Як було показано в [15, 17, 21, 28, 36] методи МН, такі як алгоритми кластеризації або методи спостереження без вчителя (наприклад, метод головних компонентів), можуть використовуватися для аналізу поведінки інстансів хмарної інфраструктури та виявлення аномалій. Наприклад, якщо виявляється незвичайна активність у використанні ресурсів або мережному трафіку, це може вказувати на можливу атаку або порушення ІБ.

Відповідно до [15] методи МН можуть використовуватися для аналізу системних параметрів інстансів хмарної інфраструктури, таких як завантаження процесора, використання пам'яті, дискова активність тощо.

У [28] наголошується, що методи навчання з вчителем можуть бути застосовані для аналізу журналів подій (логів) інстансів хмарної інфраструктури з метою виявлення аномальних чи підозрілих дій. Наприклад, можна навчити модель класифікації з урахуванням ретроспективних даних про події визначення, які події є нормальними а які – аномальними.

Методи МН можуть бути застосовані для аналізу мережевого трафіку інстансів хмарної інфраструктури з метою виявлення аномальних патернів або атак. Наприклад, можна використовувати алгоритми виявлення викидів для виявлення незвичайної мережевої поведінки, яка може вказувати на атаку або наявність шкідливого програмного забезпечення [17, 21].

Як показав аналіз попередніх досліджень, у зв'язку з міграцією все більшого обсягу даних та додатків у хмарні сервіси, ІБ стикається з низкою нових та унікальних викликів. Таблиця 1 містить систематизований огляд основних загроз, із якими зіткнулися організації під час використання хмарних сервісів.

Таблиця 1

Систематизація основних загроз, з якими зіткнулися організації під час використання хмарних сервісів (складено автором за результатами аналізу літературних джерел, наведених у цьому дослідженні)

Загроза для ІБ хмарного сервісу	Пріоритетні заходи щодо нівелювання загрози
Недостатній контроль над обліковими записами, правами, доступом та паролями у хмарних сервісах	Дискретна ізоляція користувачів та додатків. Ефективні інструменти управління правами доступу. Багатофакторна автентифікація (MFA). Керування доступом на основі ролей (RBAC). Аудит доступу та моніторинг з метою виявлення підозрілих активностей та НСД чи спроби вторгнення в реальному часі.
Інтерфейси та API з недостатнім захистом	Проведення регулярного аудиту та оцінки безпеки інтерфейсів та API допоможе виявити потенційні вразливості та недоліки в їх реалізації. Використання стандартів безпеки, таких як OAuth, OpenID Connect, SSL/TLS, а також відповідність принципам RESTful API допомагає забезпечити захист при роботі з інтерфейсами та API хмарних сервісів. Реалізація суворої системи авторизації для доступу до API допоможе запобігти НСД до хмарного сервісу та захистить дані від витоків.
Некоректна конфігурація та недостатнє керування змінами у хмарному сервісі.	Використання засобів автоматизації для налаштування та керування конфігурацією хмарними сервісами допоможе запобігти людським помилкам та забезпечити стандартизацію налаштувань ІБ. Проведення регулярних аудитів конфігурації хмарного сервісу допоможе виявляти та виправляти потенційні вразливості та помилки конфігурації. Впровадження систем моніторингу змін дозволить відстежувати та аналізувати всі зміни, що вносяться до хмарного сервісу, що сприятиме оперативному виявленню несанкціонованих дій та запобігатиме загрозам ІБ хмарного сервісу. Розробка та впровадження суворих політик ІБ, включаючи правила конфігурації та процедури керування змінами, допоможе мінімізувати ризики місконфігурації та несанкціонованих змін.

Продовження таблиці 1

Загроза для ІБ хмарного сервісу	Пріоритетні заходи щодо нівелювання загрози
Проблеми безпеки, пов'язані з архітектурою хмарних систем	При розгляді бізнес-цілей, ризиків, загроз ІБ та відповідності законодавству в контексті хмарногих сервісів, а також особливостям їх інфраструктури, об'єктам інформаційної діяльності слід врахувати високу динаміку змін та обмежений централізований контроль у хмарних сховищах. Необхідно акцентувати увагу на розвитку та адаптації інфраструктурної стратегії хмарних сервісів. При адаптації рішень необхідно враховувати основні практики оцінки ІБ, що надаються вендором.
Загрози та ризики, пов'язані з розробкою додатків для хмарних сервісів	Забезпечення навчання та сертифікації розробників з безпечної розробки додатків для хмарних сервісів допоможе підвищити обізнаність з ІБ та знизити ризик помилок у коді. При розробці програм для хмарних сервісів слід використовувати перевірені фреймворки та бібліотеки, які мають вбудовані механізми безпеки та пройшли перевірку на вразливості. Проведення статичного та динамічного аналізу коду допоможе виявляти потенційні вразливості та помилки у додатках ще на стадії розробки. Налаштування програм з урахуванням принципів захисту за промовчанням, таких як мінімізація привілеїв та обмеження доступу до ресурсів, допоможе знизити поверхню атаки та зменшити ризик компрометації системи.
Загрози та вразливості, що виникають під час роботи з хмарними сервісами, що надаються зовнішніми компаніями	Перед використанням хмарних сервісів необхідно провести ретельний аналіз безпеки постачальника, включаючи його репутацію, стандарти безпеки, сертифікації та рейтинги надійності. Важливо укласти SLA (Service Level Agreement), в якому мають бути чітко визначені зобов'язання постачальника в галузі безпеки, включаючи процедури реагування на інциденти, резервне копіювання даних та доступ до аудиту. Здійснення регулярного моніторингу та аудиту ІБ дозволить виявляти потенційні вразливості та недоліки у безпеці хмарних сервісів, а також контролювати їхню відповідність стандартам безпеки. Розробка та регулярне оновлення плану реагування на інциденти дозволить оперативно та ефективно реагувати на можливі загрози ІБ в хмарних сервісах.
Загрози, пов'язані з системними вразливостями у хмарних сервісах	Необхідно регулярно оновлювати та патчити всі компоненти хмарної інфраструктури, включаючи операційні системи, програми та сервіси, щоб виправити відомі вразливості. Проведення регулярного сканування та моніторингу вразливостей у хмарній інфраструктурі допоможе оперативно виявляти та усувати потенційні загрози ІБ. Обмеження доступу та привілеїв до системних ресурсів та даних у хмарних сервісах допоможе знизити ризик експлуатації вразливостей.
Загрози, пов'язані з ненавмисним витіканням інформації з хмарного сховища	Необхідно провести перевірку баз даних (БД) PaaS, сховищ та БД, розміщених на хостингу, включаючи ВМ, контейнери (інстанси) та встановлене на них програмне забезпечення. Слід вибирати пошукові машини, які повністю інтегровані у хмарне середовище, для того, щоб своєчасно виявити будь-які кореневі або мережеві сервіси, що роблять трафік видимим ззовні. Ці заходи також включають балансувальники навантаження, мережі доставки контенту, мережевий піринг (network peering) та хмарні фаєрволи. Пошукова машина повинна враховувати безліч мережевих компонентів, таких як кластерні IP, сервіси Kubernetes та правила доступу.
Загрози, пов'язані з некоректною конфігурацією та застосуванням безсерверних та контейнерних рішень	Використання засобів автоматизації конфігурації та деплойменту, таких як Ansible, Terraform або Kubernetes, допоможе запобігти людським помилкам при налаштуванні та розгортанні контейнерів та безсерверних додатків. Впровадження систем моніторингу та аудиту конфігурації дозволить своєчасно виявляти та виправляти місконфігурації в реальному часі, а також відстежувати зміни у конфігурації для виявлення потенційних уразливостей в ІБ хмарних сервісів. Також ефективним може бути застосування принципів least privilege, оскільки налаштування прав доступу та привілеїв для контейнерів та безсерверних функцій згідно з принципом "найменших привілеїв" допоможе знизити ризик експлуатації вразливостей.

Закінчення таблиці 1

Загроза для ІБ хмарного сервісу	Пріоритетні заходи щодо нівелювання загрози
Загрози, пов'язані з діями організованих злочинних груп та/або хакерських угруповань	Встановлення засобів захисту мережі, firewalls, системи виявлення вторгнень (IDS) та системи запобігання вторгненням (IPS), а також регулярний моніторинг мережного трафіку для виявлення аномальної активності. Впровадження багаторівневої автентифікації та суворого контролю доступу до хмарних ресурсів, включаючи механізми двохфакторної автентифікації, обмеження доступу за ролями та привілеями, а також регулярне оновлення паролів. Застосування шифрування даних у спокої та під час їх передачі між клієнтами та хмарними сервісами допоможе захистити конфіденційну інформацію від НСД. Проведення регулярних аудитів ІБ та моніторингу подій для виявлення та реагування на потенційні загрози та інциденти ІБ у реальному часі. Регулярне створення резервних копій даних та розробка планів відновлення після інцидентів допоможе мінімізувати втрату даних у разі атаки або інциденту ІБ.
Загрози, пов'язані з ексфільтрацією даних хмарних сховищ	Налаштування прав доступу до хмарних сховищ відповідно до принципу "необхідності" (least privilege), коли тільки необхідним користувачам та групам має надаватися доступ до даних. Впровадження систем моніторингу активності та виявлення загроз дозволить виявляти аномальну активність у хмарних сховищах даних, таку як незвичайні спроби доступу або завантаження великих обсягів даних. Впровадження систем запобігання витоку даних (DLP), які можуть автоматично виявляти та блокувати спроби несанкціонованого експорту або завантаження конфіденційної інформації з хмарних сховищ. Проведення навчання та тренінгів для співробітників за правилами безпечного поводження з даними у хмарних сховищах, а також щодо розпізнавання та запобігання соціальній інженерії та фішингових атак.

Майбутні напрями досліджень.

У роботах, які були проаналізовані у даній статті, переважно обговорюється коло питань, присвячених проблемі комплексного підходу до забезпечення ІБ хмарних інфраструктур. Як показано в більшості робіт, хмарні обчислення стрімко набирають популярності, витісняючи традиційні моделі ведення бізнес-процесів для об'єктів інформатизації. Проте, проаналізовані роботи недостатньо охоплюють питання реалізації системи моніторингу розподілу ресурсів. Також було виявлено, що значною мірою розглянуті роботи не торкаються такого аспекту забезпечення інформаційної безпеки хмарних інфраструктур, як застосування методів МН та для створення прогнозної моделі завантаження та методики збору метрик ІБ інстансів. Серед основних питань, які потребують вирішення у рамках майбутніх досліджень, слід виділити:

– розробку нових моделей, що описують аномальну поведінку інстансів, внаслідок порушення політики ІБ;

– розвиток методів МН з вчителем для аналізу журналів подій (логів) інстансів хмарної інфраструктури з метою виявлення аномальних чи підозрілих дій.

Висновки.

Показано, що хмарні обчислення забезпечують доступ до мережевих ресурсів, таких як сховища даних та обчислювальні потужності, на запит, без прямого управління з боку користувачів. В даний час хмарні обчислення включають як публічні, так і приватні центри обробки даних (ЦОД), що надають клієнтам єдину платформу через інтернет. Мобільні хмарні обчислення використовують розподілені обчислення для передачі програм на мобільні пристрої, такі як телефони та планшети.

Встановлено, що численні дослідження вказують на проблеми інформаційної безпеки (ІБ), загрози та вразливості для клієнтів, з якими стикаються хмарні обчислення та мобільні хмарні обчислення, та одним із перспективних методів боротьби з цими загрозами є використання методів машинного навчання (далі МН).

Проведено аналіз загроз та проблем з ІБ хмарних структур, а також огляд рішень, запропонованих різними авторами, щодо забезпечення безпеки хмарних обчислень та хмарних сервісів. Насамперед, розглянуто дослідження, що ґрунтуються на застосуванні алгоритмів МН для забезпечення безпеки хмарних обчислень та хмарних сервісів.

Список використаних джерел:

1. Горбань О., Браіловський М. Захист хмарної інфраструктури від DDoS атак. Проблеми кібербезпеки інформаційно-телекомунікаційних систем: Збірник матеріалів доповідей та тез; м. Київ, 27-28 жовтня 2022 року; Київський національний університет імені Тараса Шевченка / Редкол.: В.В. Ільченко. (голова) та ін. – К.: ВПЦ «Київський університет», 2022. 159 с.

2. Маковоз К. О. Методи виявлення вторгнень у хмарних системах відеоспостереження. Хмарні технології в освіті: матеріали Всеукраїнського науково-методичного Інтернет-семінару (Кривий Ріг–Київ–Черкаси–Харків, 21 грудня 2012 р.). – Кривий Ріг: Видавничий відділ КМІ, 2012. 173 с.

3. Фролов В. В. Analysis of approaches providing security of cloud sevices. Radioelectronic and Computer Systems, 2020. (1), 70-82.
4. Шимчук Г., Голотенко О., Золотий Р. З. Основні проблеми та загрози хмарної безпеки. Матеріали науково-технічної конференції «Інформаційні моделі, системи та технології» Тернопільського національного технічного університету імені Івана Пулюя, 2022. 59–60.
5. Aawadallah N. Security Threats of Cloud Computing. Int. J. Recent Innov. Trends Comput. Commun. 2015, 3, 2393–2397.
6. Al-Janabi S., Shehab A. Edge Computing: Review and Future Directions. REVISTA AUS J. 2019, 26, 368–380.
7. Alsolami E. Security threats and legal issues related to Cloud based solutions. Int. J. Comput. Sci. Netw. Secur. 2018, 18, 156–163.
8. Barona R., Anita M. A survey on data breach challenges in cloud computing security: Issues and threats. In Proceedings of the International Conference on Circuit, Power and Computing Technologies (ICCPCT), Paris, France, 17–18 September 2017; pp. 1–8.
9. Bhamare D., Salman T., Samaka M., Erbad A., Jain, R. Feasibility of Supervised Machine Learning for Cloud Security. In Proceedings of the International Conference on Information Science and Security, Jaipur, India, 16–20 December 2016; pp. 1–5.
10. Borylo P., Tornatore M., Jaglarz P., Shahriar N., Cholda P., Boutaba R. Latency and energy-aware provisioning of network slices in cloud networks. Comput. Commun. 2020, 157, 1–19.
11. Butt U. A., Mehmood M., Shah S. B. H., Amin R., Shaukat,M. W., Raza S. M., Piran M. J. A review of machine learning algorithms for cloud computing security. Electronics, 2020. 9(9), 1379.
12. Callara M., Wira P. User Behavior Analysis with Machine Learning Techniques in Cloud Computing Architectures. In Proceedings of the 2018 International Conference on Applied Smart Systems, Médéa, Algeria, 24–25 November 2018; pp. 1–6.
13. Carmo M., Dantas Silva F. S., Neto A.V., Corujo D., Aguiar, R. Network-Cloud Slicing Definitions for Wi-Fi Sharing Systems to Enhance 5G Ultra-Dense Network Capabilities. Wirel. Commun. Mob. Comput. 2019, 2019, 8015274.
14. Deshpande P., Sharma S. C., Peddoju S. K. Security threats in cloud computing. In Proceedings of the International Conference on Computing, Communication and Automation, Greater Noida, India, 11–14 December 2011; pp. 632–636.
15. Elzamly A., Hussin B., Basari, A. S. Classification of Critical Cloud Computing Security Issues for Banking Organizations: A Cloud Delphi Study. Int. J. Grid Distrib. Comput. 2016, 9, 137–158.
16. Kazim M., Zhu S. Y. A survey on top security threats in cloud computing. Int. J. Adv. Comput. Sci. Appl. 2015, 6.
17. Khan A. N., Fan M. Y., Malik A., Memon R. A. Learning from Privacy Preserved Encrypted Data on Cloud Through Supervised and Unsupervised Machine Learning. In Proceedings of the International Conference on Computing, Mathematics and Engineering Technologies, Sindh, Pakistan, 29–30 January 2019; pp. 1–5.
18. Khan M. A survey of security issues for cloud computing. J. Netw. Comput. Appl. 2016, 71, 11–29.
19. Khilar P., Vijay C., Rakesh S. Trust-Based Access Control in Cloud Computing Using Machine Learning. In Cloud Computing for Geospatial Big Data Analytics; Das, H., Barik, R., Dubey, H., Roy, D., Eds.; Springer: Cham, Switzerland, 2019; Volume 49, pp. 55–79.
20. Le Duc T., Leiva R. G., Casari P., Östberg, P. O. Machine Learning Methods for Reliable Resource Provisioning in Edge-Cloud Computing: A Survey. ACM Comput. Surv. 2019, 52, 1–39.
21. Lee Y., Yongjoon P., Kim, D. Security Threats Analysis and Considerations for Internet of Things. In Proceedings of the International Conference on Security Technology (SecTech), Jeju Island, Korea, 25–28 November 2015; pp. 28–30.
22. Lim S. Y., Kiah M. M., Ang T. F. Security Issues and Future Challenges of Cloud Service Authentication. Polytech. Hung. 2017, 14, 69–89.
23. Lin C., Lu H. Response to Co-resident Threats in Cloud Computing Using Machine Learning. In Proceedings of the International Conference on Advanced Information Networking and Applications, Caserta, Italy, 15–17 April 2020; Volume 926, pp. 904–913.
24. Mathkunti N. Cloud Computing: Security Issues. Int. J. Comput. Commun. Eng. 2014, 3, 259–263.
25. Nadeem M. Cloud Computing: Security Issues and Challenges. J. Wirel. Commun. 2016, 1, 10–15.
26. Salah K., Hammoud M., Zeadally, S. Teaching cybersecurity using the cloud. IEEE Transactions on Learning Technologies, 2015. 8(4), 383–392.
27. Sarma M., Srinivas Y., Ramesh N., Abhiram, M. Improving the Performance of Secure Cloud Infrastructure with Machine Learning Techniques. In Proceedings of the International Conference on Cloud Computing in Emerging Markets (CCEM), Bangalore, India, 19–21 October 2016; pp. 78–83.
28. Sayantan G., Stephen Y., Arun-Balaji B. Attack Detection in Cloud Infrastructures Using Artificial Neural Network with Genetic Feature Selection. In Proceedings of the IEEE 14th International Conference on Dependable, Autonomic and Secure Computing, Athens, Greece, 12–15 August 2016; pp. 414–419.
29. Selamat N., Ali F. Comparison of malware detection techniques using machine learning algorithm. Indones. J. Electr. Eng. Comput. Sci. 2019, 16, 435.
30. Stefan H., Liakat M. Cloud Computing Security Threats And Solutions. J. Cloud Comput. 2015, 4, 1.
31. Subashini S., Kavitha V. A Survey on Security Issues in Service Delivery Models of Cloud Computing. J. Netw. Comput. Appl. 2011, 35, 1–11.
32. Sulistio A., Reich C., Doelitzscher, F. Cloud infrastructure & applications–CloudIA. In Cloud Computing: First International Conference, CloudCom 2009, Beijing, China, December 1-4, 2009. Proceedings 1 (pp. 583–588). Springer Berlin Heidelberg.
33. Varun K. A., Rajkumar N., Kumar N. K. Survey on security threats in cloud computing. Int. J. Appl. Eng. Res. 2014, 9, 10495–10500.
34. Venkatraman S., Mamoun A. Use of data visualisation for zero-day malware detection. Secur. Commun. Netw. 2018, 1–13.
35. Xue M., Yuan C., Wu H., Zhang Y., Liu W. Machine Learning Security: Threats, Countermeasures, and Evaluations. IEEE Access 2020, 8, 74720–74742.
36. Yau S. S., Buduru A. B., Nagaraja, V. Protecting critical cloud infrastructures with predictive capability. In 2015 IEEE 8th International Conference on Cloud Computing (2015, June). (pp. 1119-1124). IEEE.
37. Yuhong L., Yan S., Jungwoo R., Syed R., Athanasios V. A Survey of Security and Privacy Challenges in Cloud Computing: Solutions and Future Directions. J. Comput. Sci. Eng. 2015, 9, 119–133.

УДК 004.4'24.056.52:336.764./768(045)
DOI https://doi.org/10.32689/maup.it.2024.1.6

Наталія ГУЛАК
кандидат технічних наук,
доцент кафедри комп'ютеризованих систем захисту інформації,
Національний авіаційний університет, gulak_n@ukr.net
***ORCID:** 0009-0000-9584-7113*

Андрій МАЙСТРЕНКО
магістр, асистент кафедри комп'ютеризованих систем захисту інформації,
Національний авіаційний університет, hondor553@gmail.com
***ORCID:** 0009-0002-1612-9178*

АВТОМАТИЗАЦІЯ МОДУЛЯ ІНФОРМАЦІЙНИХ АКТИВІВ

Анотація. *Стаття докладно розглядає процес автоматизації модуля інформаційних активів як ключовий етап у забезпеченні ефективного управління інформаційною безпекою в сучасних організаціях. Основний акцент робиться на використанні ORM фреймворків, зокрема Hibernate, як засобу спрощення доступу до бази даних та оптимізації управління інформаційними ресурсами.* ***Мета роботи*** *цієї статті полягає у вивченні можливостей підвищення рівня захищеності інформаційних активів через автоматизацію модуля управління ними. Для досягнення цієї мети використано методологію розробки та імплементації програмного модуля на основі Hibernate ORM фреймворку та Java Persistence API. В статті детально описано етапи розробки та впровадження модуля, включаючи інвентаризацію, категоризацію та автоматизацію управління активами. Надано огляд основних переваг використання Hibernate, таких як підвищення продуктивності та надійності системи. Детально проаналізовано критерії ефективності розробленого рішення, включаючи надійність, захищеність, швидкість роботи та доступність.* ***Наукова новизна*** *полягає в застосуванні сучасних технологій для підвищення ефективності та надійності управління інформаційною безпекою. Зроблено висновок про значний внесок автоматизації модуля у підвищення ефективності управління інформаційною безпекою та полегшення користування системою.* ***Висновки*** *статті підтверджують успішність використання ORM фреймворків для автоматизації управління інформаційними активами, що призводить до покращення ефективності та забезпечення високого рівня захисту даних.* ***Ключові слова:*** *інформаційні активи, управління інформаційною безпекою, ORM фреймворки, оцінка ризиків, захист інформаційних активів, система менеджменту інформаційної безпеки.*

Nataliia GULAK, Andrii MAISTRENKO. AUTOMATION OF THE INFORMATION ASSETS MODULE

Abstract. *The article examines in detail the process of automating the information assets module as a key stage in ensuring effective management of information security in modern organizations. The main emphasis is on the use of ORM frameworks, in particular Hibernate, as a means of simplifying access to the database and optimizing the management of information resources.* ***The purpose*** *of this article is to study the possibilities of increasing the level of security of information assets through the automation of their management module. To achieve this goal, the methodology of developing and implementing a software module based on the Hibernate ORM framework and the Java Persistence API was used. The article describes in detail the stages of development and implementation of the module, including inventory, categorization and automation of asset management. An overview of the main benefits of using Hibernate, such as improved system performance and reliability, is provided. The criteria for the effectiveness of the developed solution were analyzed in detail, including reliability, security, speed of operation and availability.* ***Scientific novelty*** *consists in the application of modern technologies to increase the efficiency and reliability of information security management. It was concluded that the automation of the module made a significant contribution to improving the efficiency of information security management and facilitating the use of the system.* ***The conclusions*** *of the article confirm the success of using ORM frameworks for automating the management of information assets, which leads to improved efficiency and ensuring a high level of data protection.* ***Key words:*** *information assets, information security management, ORM (Object Relational Mapping) frameworks, system integration, risk assessment, protection of information assets.*

Вступ. Широке впровадження сучасних інформаційних технологій пов'язане, поряд з раціональним використанням ресурсів розподіленої комп'ютерної мережі, з організацією ефективної протидії загрозам атак на її інфраструктуру. Постійні зміни в конфігурації системи, її параметрах і складі програмного забезпечення вимагають постійного аналізу стану безпеки системи, передбачення та виявлення нових загроз безпеці та застосування превентивних заходів. Автоматизація модуля «Інформаційні активи» є лише першим кроком у напрямку покращення системи управління інформаційною безпекою компанії. Подальші дослідження та розвиток дозволять розширити можливості системи і підвищити її ефективність у забезпеченні безпеки та ефективності використання інформаційних ресурсів.

Головна частина. Модуль обліку інформаційних активів використовується для автоматизації діяльності власників інформаційних активів. Головною метою автоматизації каталогу інформаційних активів є оптимізація та вдосконалення системи управління інформаційної безпеки (ІБ).

До інформаційних активів компанії зазвичай відносять інформацію, апаратне, програмне забезпечення та інші засоби, необхідні для отримання, обробки та зберігання даних, що використовуються у певних бізнес-процесах. Це може бути сховища даних, бази даних, бази клієнтів, виробничі показники (звіти), фінансові звіти, інформаційні системи тощо.

Система повинна забезпечувати єдину інформаційну інфраструктуру у системі менеджменту інформаційної безпеки (СМІБ), що створюється через інтеграцію максимальної кількості інформації в модулі і, тим самим, що дозволяє удосконалювати інформаційну проникність і синергію між структурними підрозділами.

За класифікацією автоматизованих комплексів система, яка розглядається, відноситься до багатофункціональних програмно-технічних комплексів для автоматизації управління організаційними процесами в умовах розподіленого використання інформації різними фахівцями.

Для ефективної роботи таких систем необхідно щоб були виконані наступні вимоги: забезпечення необхідного обсягу інформації; механізм своєчасної актуалізації змісту та базовий набір сервісів роботи з інформацією; створення інтуїтивно зрозумілих інтерфейсів для користувачів; надання послуг, що мають очевидну цінність; можливість вибору типового профілю для користувачів.

Основними принципами створення СМІБ є використання програмного і апаратного забезпечення, здатного працювати з мережею Інтернет; відповідність міжнародним стандартам у сфері ІБ; уніфікація форматів і протоколів інформаційного обміну; використання ефективних методів захисту від несанкціонованого доступу [2].

Особливу увагу потрібно звернути на людський фактор, який відіграє не останню роль як джерело загрозі інформації. Для цього передбачено три основні категорії користувачів: авторизований користувач, адміністратор і супер-адміністратор. Кожен користувач має права і обов'язки відповідно до свого рівня доступу. Коректний розподіл певних прав користувачів надає можливість виконувати конкретні дії з різними типами документів, що знижує ризики незаконного розкриття інформації [1].

Виходячи з політики системи управління інформаційної безпеки (СУІБ) задачі модернізації і розвитку автоматизованих комплексів полягають у наступному: розширенню функціональності для більш ефективного управління інформаційними активами; інтеграції із сучасними технологіями та стандартами безпеки; підтримка масштабованості та гнучкості системи для відповіді на зростаючі потреби користувачів; створення гнучкої архітектури системи, що дозволяє легко додавати нові функції та модулі [3].

«Автоматизація модуля інформаційних активів» має на меті оптимізацію та удосконалення управління інформаційною безпекою. Заснована на застосуванні ORM Фреймворку, вона спрощує роботу з базами даних, забезпечуючи швидкість та масштабованість. Головною метою є забезпечення централізованого зберігання та оцінка ризиків для кожного окремого інформаційного активу відповідно до міжнародних стандартів. Рішення включає функції з розробки плану обробки ризиків, розмежування прав доступу користувачів, відправлення повідомлень власникам активів та побудову звітів. Забезпечується оперативна підтримка інформації, що дозволяє використовувати дані для управління ризиками та інцидентами. Автоматизація модуля сприяє підвищенню продуктивності управління інформаційною безпекою, що є важливим умовою при обмеженому числі персоналу та великій кількості облікових об'єктів [4].

Об'єктом автоматизації є модуль управління інформаційними активами, спрямований на оптимізацію та покращення системи управління інформаційною безпекою в організації. Цей модуль містить інструменти для централізованого зберігання інформації про всі об'єкти оцінки ризиків, а також проведення класифікації та оцінки ризиків для кожного інформаційного активу відповідно до міжнародних стандартів. Об'єкт автоматизації надає можливість розробки планів обробки ризиків, налаштування прав доступу користувачів, сповіщення власників активів та створення звітів. Основною метою автоматизації є забезпечення ефективного управління інформаційною безпекою організації та підвищення його ефективності в умовах обмежених ресурсів та великої кількості облікових об'єктів.[5]

Рішення для автоматизації модуля інформаційних активів реалізоване на основі програмного забезпечення з закритим вихідним кодом і є готовим інструментом для оцінки та обробки ризиків інформаційних активів компанії. Для користування системою не потрібна установка додаткових клієнтських програм, адже достатньо наявності будь-якого сучасного веб-браузера.

Впровадження даного рішення сприяє підвищенню ефективності управління інформаційною безпекою, спрощує процес оцінки та обробки ризиків і забезпечує зручний доступ до необхідної інформації для зацікавлених сторін.

Процес автоматизації спрямований на оптимізацію та покращення системи управління інформаційною безпекою, а також на створення реєстру інформаційних активів і керування записами. Це є центральною точкою, де зберігаються дані про всі активи компанії, визначається їхнє значення, вплив на

бізнес-процеси та подальше використання цих даних у модулі [6].

- Створення довідників активів компанії.
- Оцінка критичності об'єктів захисту.
- Класифікація активів та їх передача в модуль.
- Інтеграція з системами інвентаризації.
- Створення єдиного інформаційного середовища для спільної роботи з суміжними системами.

Проект автоматизації модуля інформаційних активів включає наступні етапи:

1. Інвентаризація інформаційних активів: Визначення всіх інформаційних активів компанії та їх поточний стан.

2. Категоризація інформаційних активів: Визначення ключових активів, які впливають на функціонування бізнес-процесів компанії.

3. Логічне проектування процесу управління каталогом інформаційних активів: Структурування, опис та проектування процесів для подальшої автоматизації.

4. Автоматизація процесу управління каталогом інформаційних активів: Реалізація автоматизації управління каталогом для забезпечення ефективності та точності процесів.

5. Проектування розроблених процесів на систему автоматизації. Імплементація та налагодження розроблених процесів на систему автоматизації з метою їх подальшого використання.[9]

В результаті автоматизації модуля інформаційних активів формується перелік активів, визначається їхня критичність з урахуванням впливу на бізнес-процеси компанії, а також створюються матриці доступу для розподілу прав до інформаційних активів.

Код модуля «Інформаційні активи» використовує ORM (Object Relational Mapping) фреймворк для забезпечення зручного доступу до бази даних. ORM дозволяє розробникам працювати з об'єктами даних, не звертаючись безпосередньо до SQL запитів (рис. 1).

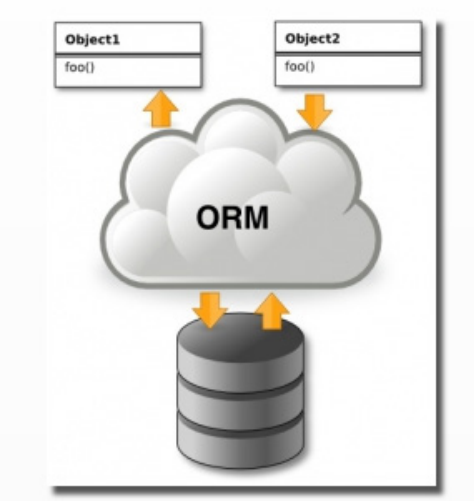


Рис. 1. Структурна схема роботи ORM Фреймворку

В реалізації модуля використовується Hibernate ORM, який є реалізацією специфікації JPA (Java Persistence API). Hibernate спрощує взаємодію з базою даних, адже він автоматично генерує SQL запити на основі об'єктів Java.[10]

Основні переваги використання Hibernate для цього модуля:

- Спрощення роботи з базою даних за рахунок автоматичної генерації SQL запитів.
- Зручний доступ до даних у вигляді об'єктів Java, що полегшує розробку та збереження коду.
- Підтримка високої продуктивності та швидкодії завдяки оптимізації запитів та кешуванню даних.

Крім того, важливою перевагою Hibernate є можливість легкої масштабованості та розширення функціональності за рахунок його гнучкої архітектури.

Для поліпшення роботи СМІБ необхідно проведення робіт з оцінки та обґрунтування різних заходів які можуть забезпечити необхідний рівень захищеності інформаційних активів.[7]

Для оцінки ефективності розробленого програмного продукту необхідно обрати критерії за якими вони будуть порівнюватися.[8]

Перша оцінка була проведена за критеріями (рис. 2): надійність, захищеність, швидкість роботи, доступність.

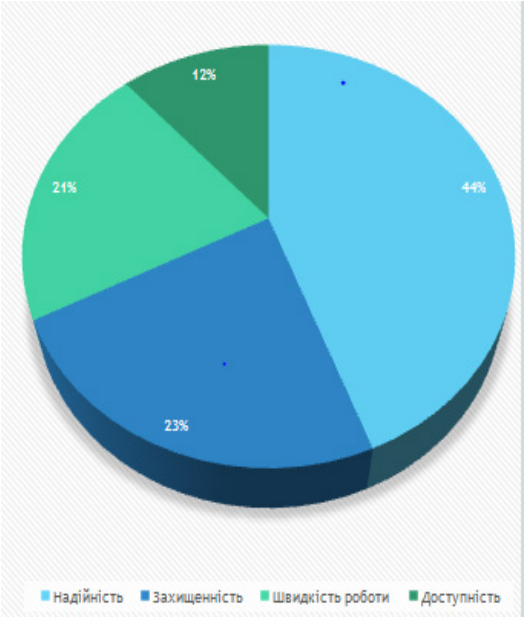


Рис. 2. Кругова діаграма оцінки ефективності за першими критеріями

Друга оцінка ефективності розробленого продукту була проведена за наступними критеріями (рис.3): переносимість, зручність супроводу, ефективність, зручність використання, функціональність.

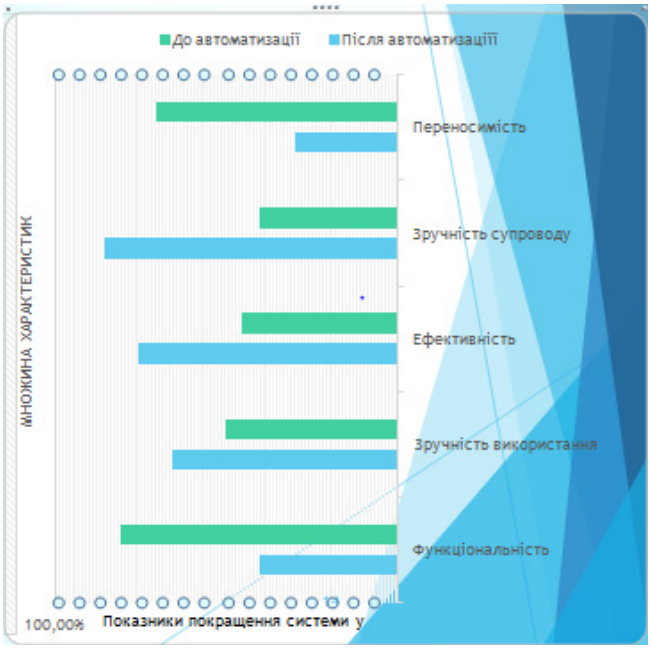


Рис. 3. Показники покращення системи

На основі оцінки ефективності проведеної автоматизації модуля удалося підвисити такі показники: зручність супроводу, ефективність, зручність використання.

Висновки:

1. У процесі розробки модуля "Інформаційні активи" було виявлено, що використання ORM фреймворків, зокрема Hibernate, значно спрощує доступ до бази даних і полегшує управління інформаційними активами компанії. Це дозволяє швидко реалізувати функціонал модуля і забезпечити його високу продуктивність та надійність.

2. Додавання нових можливостей до модуля, таких як інтеграція з іншими системами управління, підтримка розширених звітів та аналітичних інструментів, зробить систему ще більш універсальною та корисною для користувачів, тобто розширить її функціональність.

Список використаних джерел:

1. ДСТУ ISO/IEC 27007:2018: Настанова щодо аудиту систем керування інформаційною безпекою.
2. ДСТУ ISO/IEC 27001: Система менеджменту інформаційної безпеки.
3. Електронний ресурс Деякі питання проведення незалежного аудиту інформаційної безпеки на об'єктах критичної інфраструктури: http://www.dsszzi.gov.ua/dsszzi/control/ uk/publish/ Режим доступу: вільний.
4. Жилін А. В., Шаповал О. М., Успенський О. А. Технології захисту інформації в інформаційно-телекомунікаційних системах. Навчальний посібник. – Київ: КПІ ім. Ігоря Сікорського, Вид-во «Політехніка», 2021. 213 с.
5. Інформаційна та кібербезпека: соціотехнічний аспект / В.Бурячок, В. Толубко, В. Хорошко, С. Толюпа. – Київ: ДУТ, 2015. 288 с.
6. Когут Ю. Кібербезпека та ризики цифрової трансформації компанії /Ю. Когут. –Київ: вид-во консалтингова компанія Сидкон, 2021. 364 с.
7. Корченко О.Г. Менеджмент інформаційної безпеки: навчальний посібник/ О.Г. Корченко, М.Є. Шелест, С.В. Казмірчук, Ю.М. Ткач, Є.В. Іванченко. Ніжин: ФОП Лук'яненко В.В. ТПК «Орхідея», 2019. 408 с.
8. НД ТЗІ 1.1-005-07 Захист інформації на об'єктах інформаційної діяльності. Створення комплексу технічного захисту інформації. Основні положення.
9. Остапов С. Е. Технології захисту інформації : навчальний посібник / С.Е. Остапов, С. П. Євсеєв, О. Г. Король. – Х. : Вид. ХНЕУ, 2013. 476 с.
10. Bauer C. Java Persistence with Hibernate / C. Bauer, G. King, G. Gregory. – New York, USA: Manning Publications, 2016. 960 с. – (2nd edition).

UDC 004.4`6:004.4`4
DOI https://doi.org/10.32689/maup.it.2024.1.7

Oleksandr DEINEHA
PhD student, Department of Theoretical and Applied Computer Science, School of Mathematics and Computer Sciences, V. N. Karazin Kharkiv National University, oleksandr.deineha@karazin.ua
ORCID: *0000-0001-8024-8812*

LAMBDA CALCULUS TERM REDUCTION: EVALUATING LLMS' PREDICTIVE CAPABILITIES

Abstract. *This study is part of a research series of optimizing compilers and interpreters of functional programming languages. Lambda Calculus was chosen as the most straightforward functional programming language, which can process any operation available to other functional programming languages but with the simplest syntax. Using machine learning methods allows for uncovering relations inside lambda terms, which might indicate which reduction strategy better suits their reduction. Finding those techniques for lambda terms allows optimizing not only lambda term reduction but also interpreters and compilers of functional programming languages.*

This research aims *to scrutinize LLMs' understanding of Lambda term reduction to predict reduction steps and evaluate prediction accuracy. Artificially generated Lambda terms were employed Utilizing OpenAI's GPT-4 and GPT-3.5 models. However, due to model constraints and cost considerations, experiments were limited to terms with specific token counts.*

Despite its larger size, results revealed that the GPT-4 model did not significantly outperform GPT-3.5 in understanding reduction procedures. Moreover, while the GPT-3.5 model exhibited improved accuracy with reduced token counts, its performance with more complex prompts was suboptimal. This underscores the LLMs' limitations in grasping Lambda terms and reduction strategies, especially with larger and more intricate terms.

Conclusions. *The research concludes that general-purpose LLMs like GPT-3.5 and GPT-4 are inadequate for accurately predicting Lambda term reductions and distinguishing between strategies, particularly with larger terms. While fine-tuning may enhance model performance, the current findings highlight the need for further exploration and alternative approaches to achieve a deeper understanding of lambda term reduction using LLMs.*

Key words: *Lambda Calculus, Large Language Model, reduction process, prompt engineering.*

Олександр ДЕЙНЕГА. РЕДУКЦІЯ ТЕРМІВ ЛЯМБДА-ЧИСЛЕННЯ: ОЦІНКА ПРОГНОЗИВНИХ ЗДАТНОСТЕЙ LLM

Анотація. *Це дослідження є частиною серії досліджень оптимізації компіляторів та інтерпретаторів функціональних мов програмування. Лямбда-числення було обрано як найпростішу мову функціонального програмування, яка може обробляти будь-які операції, доступні іншим мовам функціонального програмування, але з найпростішим синтаксисом. Використання методів машинного навчання дозволяє виявити зв'язки всередині лямбда-термів, які можуть вказати, яка стратегія редукції краще підходить для їх нормалізації. Пошук цих методів для лямбда-термів дозволяє оптимізувати не тільки редукцію лямбда-термів, але й інтерпретатори та компілятори функціональних мов програмування.*

Мета. *Це дослідження має на меті вивчити як LLM розуміє лямбда-терми, для цього передбачити кроки редукції та оцінити точність передбачень. Використовувалися штучно створені лямбда-терми з використанням моделей OpenAI GPT-3.5 і GPT-4. Однак через обмеження моделей та міркування щодо вартості експериментів були обмежені термами з певною кількістю токенів.*

Незважаючи на більший розмір, результати показали, що модель GPT-4 незначно перевершила GPT-3.5 у розумінні процесу редукції. Крім того, у той час як модель GPT-3.5 продемонструвала підвищену точність із зменшеною кількістю токенів, її продуктивність із більш складними термами була неоптимальною. Це підкреслює обмеження LLM у розумінні лямбда-термів і стратегій скорочення, особливо з більшими та складнішими термами.

Висновки. *Дослідження показує, що LLM загального призначення, такі як GPT-3.5 і GPT-4, недостатні для точного прогнозування скорочень лямбда-термів і розрізнення стратегій, особливо з більшими термами. Хоча точне налаштування може підвищити продуктивність моделі, поточні результати підкреслюють необхідність подальшого дослідження та альтернативних підходів для досягнення глибшого розуміння редукції лямбда-терму за допомогою LLM.*

Ключові слова: *Лямбда-числення, Велика Мовна Модель, процес редукції, інженерія промпту.*

Introduction. Our research aimed at optimizing functional programming compilers and interpreters. For this purpose, we considered Lambda Calculus the most straightforward possible representation of functional programming languages [1]. Lambda Calculus allows the execution of its programs called terms as the expression reduction process. The lambda terms can be divided into Applications, Abstractions, and Variables. The term reduction is possible using redexes, special combinations of Abstract, and any other term inside an Application. Some terms may contain more than one redex, and choosing a specific redex by some rule defines a reduction strategy. The most famous reduction strategies are the normal order or the leftmost outermost (LO) strategy and the applicative order or the rightmost innermost (RI) strategy. An example of the Y term is shown in Figure 1. The example describes all lambda term elements and the RI and LO strategies.

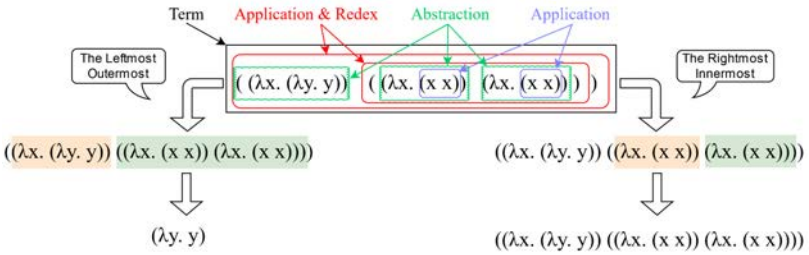


Fig. 1. Example of the Y term, applying the rightmost innermost and the leftmost outermost strategies and showing term types

Research in the Lambda Calculus reduction process may help optimize the interpretation and compilation of other function programming languages using similar optimization techniques [2]. Understanding the Lambda terms reduction process may help uncover essential features and methods of discovering strategy priorities.

Analysis of recent research and publications. Usually, reduction steps are estimated as equal [3], which allows comparison of reduction strategies via a simple number of reductions. In the research [4], we considered another approach to estimating reduction steps via their computational efficiency, which allows us to develop a greedy strategy that minimizes computational resources required for reduction. Although this approach enables estimating computational resources needed for single-step normalization, it does not allow us to understand the relation between specific terms and strategies that better suit its normalization.

The problem of measuring term complexity was considered in the studies [5, 6], where memory consumption. Also, cost models were proposed in the article [7] to solve the same problem. All works show that it is possible to define the complexity of reduction steps via different approaches.

The research [8] considers using the Transformer models for sequential analysis of lambda terms for predicting the type of term in Typed Lambda Calculus, which simply extends Lambda Calculus with defining types via specific terms without modifying the expressions lexicon. This usage highlights the idea that extracting particular term features that indicate its type is possible. That idea can be extended to the strategy priority, and in the research, [9] was considered to estimate the reduction steps number for the RI and the LO strategies. However, research [9], due to computational limitations, considers only simplified term representation, which does not count variable information. The study [9] results show that this approach allows accurate identification of reductions if the expected number is less than 10, but for bigger expected numbers, the accuracy drops.

Studies [2, 10] solved the problem of losing variable information with more complex pretrained machine learning models, namely Microsoft CodeBERT [11] and OpenAI [12] embedding models. Those studies used vector representations of lambda terms created with Large Language models (LLM) and uninformed Machine Learning techniques to find the relation between collected vectors and the most suitable reduction strategy. The main issue with the studies [2, 10] is using pretrained models on programming languages or for general problems. It does not provide information about LLMs' understanding of lambda terms.

Also, recent research has shown promising results with the implementation of LLMs for solving mathematical problems [13], code execution [14], and compilation optimization [15]. All such works show that it is possible to use LLMs as a discovery tool for code-related tasks, but no one has checked how good general tasks LLMs are for understanding lambda terms.

The research objective is to investigate LLM's understanding of the lambda term reduction process. Achieving this objective can be highlighted in the following tasks:

1. Prepare a lambda terms dataset containing the following reduction step term for the selected strategies.
2. Predict the following reduction step using the selected strategy, general LLM, and prepared terms.
3. Calculate the accuracy of such predictions and conclude that LLMs can be used to understand lambda term reduction and strategy differences.

Scientific novelty. For the first time, the ability of LLM to understand lambda calculus was investigated.

Research methodology. The research is considered one of the biggest publicly available general task LLM models, developed and trained by OpenAI: the GPT-4 and GPT-3.5 models [16]. Table 1 shows the considered models, their weights number, and price per million tokens [16].

Table 1

Comparison of the GPT-3.5 and GPT-4 models

Model name	Weights Number	Price of input per million tokens	Price of output per million tokens
GPT-3.5	~20 Billion	0.50$	1.50$
GPT-4	~220 Billion	30.00$	60.00$

In this research, artificially generated lambda terms were used. The procedure for generating those terms was described in the study [9]. Accordingly to the procedure, with some probability, choose the next term element (Application, Abstraction, or Variable from a set of available variables), which recursively builds a term. This procedure allows consider the maximum available terms in the selected bound of variables and probabilities of elements. This research uses the same terms dataset used in previous studies [2, 10]. However, considering price limitations, the number of terms used for testing decreased, considering the number of input and expected output tokens. The results of such data preparing shown in Table 2. Although the number of terms used in experiments has significantly shrunk, there are still enough terms to check the proposed LLM's ability to understand the reduction process with differing strategies. Also, shown LLMs require special text descriptions, called prompts, for problem statements, which a LLM must solve. The prompt size also increases the number of required input tokens for each term, and depending on the prompt, it can increase the number of expected output tokens

Table 2

Results of term dataset preparing

	Original dataset	**Cropped to 77 max tokens per term**	**Cropped to 40 max tokens per term**
Terms number	4282	1019	305
Input tokens	523k	52k	8k
Expected output LO tokens	503k	45k	6.4k
Expected output RI tokens	521k	45k	6.4k

Considering the analysis of available GPT models' price and weight numbers and the concluded size of datasets, it is possible to define the methodology of experiments:

1. Prepare prompts for predicting the LO / RI steps using the GPT-3.5 / GPT-4 models.
2. Using prepared prompts, predict the following reduction step using the selected model (use for prediction cropped dataset to 77 tokens per term with GPT-3.5 and cropped to 40 tokens – GPT-4 model).
3. Postprocces predictions to formulate actual term answers.
4. Using the Lambda Calculus interpreter, the expected following terms are compared with actual predictions.

Results of research. The first stage of the study requires preparing a prompt. There are a few prompt types [17]: some require detailed descriptions of solved tasks, some require examples of solving, and others require simply asking about the task. Due to the high price of using the GPT-4 model, the simplest approach was chosen. On another site, the GPT-3.5 model was considered a few approaches.

```
f"""Given the lambda term, apply the leftmost-outermost (LO) strategy to perform the next step of reduction.
The LO strategy, also known as normal order reduction, prioritizes the reduction of the leftmost-outermost redex first.
This means that if there's a choice between reducing an expression inside a lambda abstraction or an application outside,
the application takes precedence unless there's no other redex outside the abstraction.

Lambda Calculus Reduction Rules:

1. Alpha Conversion (α-conversion): Rename bound variables, ensuring no variable name conflicts.
This step is essential for avoiding collisions between variables.

2. Beta Reduction (β-reduction): Apply the function to its argument.
The formal rule is ((λx.M) N) → M[x:=N], where M[x:=N] denotes substituting N for x in M.

3. Eta Conversion (η-conversion): Simplify functions with unnecessary abstractions. The rule is λx.(M x) → M if x does not appear in M.

Prioritization in LO Strategy:

- Outermost First: Reduce the outermost redex before any inner redexes, even if the inner one is to the left of an outer one.
- Leftmost First: When faced with multiple outermost redexes, choose the leftmost one.

Examples:

- Given (λx.x x) ((λy.y) z), the LO strategy first reduces the outermost leftmost redex, resulting in (λx.x x) z.
- For ((λx.λy.x y) (λa.a)) b, the first step of reduction under LO strategy would yield (λy.(λa.a) y) b.

Procedure:

- Identify the leftmost-outermost redex in the term.
- Apply the appropriate reduction rule based on the structure of this redex.
- If multiple steps are available, choose the one that aligns with the LO strategy's prioritization.

Take your time to analyze the term <<<{str_term}>>. Consider each part of the term carefully and apply the reduction rules as described.
Remember to use α-conversion to avoid variable naming conflicts, especially when dealing with nested lambda expressions.
In the end provide the next reduction term in format: Result: next step term
"""
```

Fig. 2. Using the description prompt, the GPT-3.5 model generates the following term according to the LO strategy

```
f"""Example of performing task #1:
Given term: (λx.((λy.((λz.z) x)) (λa.a))). Provide the next step of term reduction

Your output:
1. Identify the leftmost-outermost redex in the given term: ((λy.((λz.z) x)) (λa.a))
1.1. Where object of the redex is (λy.((λz.z) x)) (λa.a)
1.2. And subject of the redex is
2. Apply β-reduction: ((λy.((λz.z) x)) (λa.a)) [x:= (λa.a)]
3. Result: (λy.((λz.z) (λa.a)))

Example of performing task #2:
Given term: (((λx.x) (λy.(y (λz.z)))) (λa.a)). Provide the next step of term reduction.

Your output:
1. Identify the leftmost-outermost redex in the given term: (((λx.x) (λy.(y (λz.z)))) (λa.a))
2. Apply β-reduction: ((λy.(y (λz.z))) (λa.a)) [x:= (λy.(y (λz.z)))]
3. Result: ((λy.(y (λz.z))) (λa.a))

Given term: {str_term} Provide the next step of term reduction using example.

Your output:
"""
```

Fig. 3. Using the detailed step prompt, the GPT-3.5 model generates the following term according to the LO strategy

```
f"""
Please generate the next step of reduction a Lambda Calculus term. Provide only term expression.

Lambda term: '''{str_term}'''
"""
```

Fig. 4. Using the simplest command prompt, the GPT-4 model generates the following term according to the LO strategy

Figure 2 shows an example of the description prompt used for the GPT-3.5 model, Figure 3 shows an example of the detailed step prompt used for the GPT-3.5 model, and Figure 4 shows an example of the simplest command prompt used for the GPT-4 model. All shown prompts are used for the LO strategy, but the prompt for the RI strategy differs only in the description of the RI strategy, but the logic is kept the same.

Table 3

Accuracy of the following step predictions with the GPT-3.5 and GPT-4 models

	GPT-3.5 to LO (77 tokens)	GPT-3.5 to LO (40 tokens)	GPT-3.5 to RI (77 tokens)	GPT-3.5 to RI (40 tokens)	GPT-4 to LO (40 tokens)	GPT-4 to RI (40 tokens)
Description prompt	9.5%	23.6%	6.47%	17.04%	–	–
Detailed step prompt	4.0%	10.82%	3.0%	9.83%	–	–
Simplest command prompt	10.0%	27.21%	3.23%	9.83%	41.3%	36.39%

Table 3 shows all the experiments. Due to the high price of GPT-4 model generation, only the simplest command prompt and terms with a maximum of 40 tokens were considered, which showed the best results with the GPT-3.5 model. Experiments with the GPT-3.5 model were considered terms with a maximum of 77 tokens; 40 token results were extracted from the collected results.

Discussion. Low accuracy on more complex prompts (description and detailed step) might indicate overloading the model with redundant details. Also, increasing accuracy with a decreasing maximum number of tokens shows that the GPT-3.5 model cannot profoundly analyze and understand lambda terms. Real programs can contain hundreds of variables, which is a big problem to analyze. A drop in accuracy of 5-7% on prediction following terms for the RI strategy compared to the LO can be explained by the fact that GPT-3.5 and GPT-4 models do not wholly understand the redex and reduction strategy concept. However, the most significant accuracy for predicting the following RI term was achieved using the description prompt, indicating that LLMs can improve their redex understanding, but it depends on prompt construction.

The 10% difference between the best results achieved on GPT-3.5 and GPT-4 indicates that increasing the number of model weights doesn't significantly improve understanding of the reduction procedure. However, the GPT-4 model was the closest to accurate results.

This research benefits from showing that general task LLMs are unsuitable for predicting the following term step. Fine-tuning techniques can improve such models but with more affordable ones.

The research disadvantages are not uncovering all possible experiments on the GPT-4 model with different prompts due to the high generation price, using a limited number of tokens in experiment terms, and considering only OpenAI models. Considering these disadvantages, the following research step could fine-tune some LLM for more accurate results.

Conclusions. The results of the research were solved in the following tasks:

1. A lambda terms dataset has been prepared considering the limitations of selected LLM models. The prepared dataset allowed to check how selected models understand lambda calculus reduction and the difference in strategies by selecting two reduction strategies (LO and RI).

2. The following reduction steps were predicted using GPT-3.5 and GPT-4 models. The predictions were cleaned to check their reliability.

3. Using the Lambda Calculus interpreter, the predictions' results were compared to expected terms, which allowed the predictions to be calculated accurately. The applicability of GPT-3.5 and GPT-4 was examined, and it was concluded that selected LLMs are insufficient to understand lambda term reduction and strategy differences, especially on larger terms.

Bibliography:

1. Cummins, C., Seeker, V., Grubisic, D., Elhoushi, M., Liang, Y., Roziere, B., Gehring, J., Gloeckle, F., Hazelwood, K., Synnaeve, G., Leather, H. Large Language Models for Compiler Optimization. *ArXiv*, 2023. URL: https://arxiv.org/abs/2309.07062.

2. Dal Lago, U., and Martini, S. On Constructor Rewrite Systems and the Lambda Calculus. *Logical Methods in Computer Science*, 2012, Volume 8. DOI: 10.2168/LMCS-8(3:12)2012.

3. Deineha, O. Supervised data extraction from Transformer representation of lambda-terms. *Radioelectronic And Computer Systems*, 2024. In press.

4. Deineha, O., Donets, V., & Zholtkevych, G. Deep Learning Models for Estimating Number of Lambda-Term Reduction Steps. *ProfIT AI 2023: 3rd International Workshop of IT-professionals on Artificial Intelligence (ProfIT AI 2023)*, 2023, vol. 3624, pp. 147-156. URL: https://ceur-ws.org/Vol-3641/paper12.pdf.

5. Deineha, O., Donets, V., & Zholtkevych, G. Estimating Lambda-Term Reduction Complexity with Regression Methods. *International Conference "Information Technology and Interactions"*, 2023, no. 3624, pp. 147–156. URL: https://ceur-ws.org/Vol-3624/Paper_13.pdf.

6. Deineha, O., Donets, V., & Zholtkevych, G. Unsupervised Data Extraction from Transformer Representation of Lambda-Terms. *Eastern European Journal of Enterprise Technology*, 2024. In press.

7. Deliyannis, E.P., Paul, N., Patel, P.U., & Papanikolaou, M. A comparative performance analysis of Chat GPT3.5, Chat GTP4.0 and Bard in answering common patient questions on melanoma. *Clinical and experimental dermatology*, 2023. DOI: 10.1093/ced%2Fllad409.

8. Dezani-Ciancaglini, M., Ronchi Della Rocca, S., and Saitta, L. Complexity of lambda-term reductions. *RAIRO Theor. Informatics*, 1979, Appl. 13: 257-287. DOI: 10.1051/ita/1979130302571.

9. Feng, Z., Guo, D., Tang, D., Duan, N., Feng, X., Gong, M., Shou, L., Qin, B., Liu, T., Jiang, D., & Zhou, M. CodeBERT: A Pre-Trained Model for Programming and Natural Languages. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 2020, pp 1536–1547. DOI: 10.18653/v1/2020.findings-emnlp.139.

10. Grabmayer, C. Linear Depth Increase of Lambda Terms along Leftmost-Outermost Beta-Reduction. *ArXiv,* 2019. URL: https://doi.org/10.48550/arXiv.1604.07030.

11. Liu1, C., Lu, S., Chen, W., Jiang, D., Svyatkovskiy, A., Fu, S., Sundaresan, N., Duan, N. Code Execution with Pre-trained Language Models. *Accepted to the Findings of ACL 2023*, 2023. URL: https://arxiv.org/abs/2305.05383.

12. Miranda, Brando, Shinnar, Avi, Pestun, Vasily, Trager, Barry. Transformer Models for Type Inference in the Simply Typed Lambda Calculus: A Case Study in Deep Learning for Code. *Computer Science*, 2023. URL: https://arxiv.org/abs/2304.10500.

13. Pollak, D., Layka, V., & Sacco, A. Functional Programming. Beginning Scala 3. 2020. DOI:10.1007/978-1-4842-7422-4_4.

14. Qi, Xiaochu. Reduction Strategies in Lambda Term Normalization and their Effects on Heap Usage. *ArXiv*, 2004. URL: https://arxiv.org/abs/cs/0405075.

15. White, J., Hays, S., Fu, Q., Spencer-Smith, J., & Schmidt, D.C. ChatGPT Prompt Patterns for Improving Code Quality, Refactoring, Requirements Elicitation, and Software Design. *ArXiv*, 2023. DOI: 10.48550/arXiv.2303.07839.

16. Yang, Zhen, Ding, Ming, Lv, Qingsong, Jiang, Zhihuan, He, Zehai, Guo, Yuyi, Bai, Jinfeng, Tang, Jie. GPT Can Solve Mathematical Problems Without a Calculator. Machine Learning. *ArXiv*, 2023. URL: https://arxiv.org/abs/2309.03241.

17. New embedding models and API updates. *Blog OpenAI*, 2024. URL: https://openai.com/blog/new-embedding-models-and-api-updates (accessed 24.04.2024).

УДК 004
DOI https://doi.org/10.32689/maup.it.2024.1.8

Леся ЛЮШЕНКО
кандидат технічних наук, доцент кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», lyushenkol@gmail.com
***ORCID:** 0000-0003-4319-5955*

Ярослав ПЕРЕГУДА
аспірант кафедри програмного забезпечення комп'ютерних систем, Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», findershtein@gmail.com
***ORCID:** 0009-0002-7292-7887*

СПОСІБ ПОБУДОВИ ПРОГРАМНИХ ДЕТЕКТОРІВ ДЛЯ ВИЯВЛЕННЯ ПРОГРАМНИХ БОТІВ В СОЦІАЛЬНИХ МЕРЕЖАХ

***Анотація. Мета даної роботи** полягає у детальному дослідженні ефективності використання великих мовних моделей (Large Language Models), або LLM, для виявлення програмних ботів в соціальних мережах. Робота зосереджується на аналізуванні ефективності різних методів виявлення та визначення потенціалу LLM як засобу для підвищення точності та ефективності процесу ідентифікації ботів.*

Дослідження охоплює аналіз трьох основних підходів до виявлення програмних ботів: аналіз метаданих, текстовий аналіз та аналіз графів. Аналізуються як традиційні методи машинного навчання, так і новітні LLM, які використовуються для аналізу великих даних з соціальних мереж. Основною методикою є порівняльний аналіз, який включає використання розширених наборів даних, таких як TwiBot20 і TwiBot-22, для оцінки продуктивності кожного методу з використанням метрик, таких як точність та F1-міра, що дозволяє отримати об'єктивне уявлення про ефективність різних підходів до виявлення ботів.

***Наукова новизна** даної роботи полягає у використанні LLM для аналізу різноманітних видів даних з соціальних мереж для виявлення програмних ботів. Автори розглядають інтеграцію LLM у традиційні методи виявлення, що дозволяє адаптувати процеси виявлення до складної поведінки програмних ботів, забезпечуючи високу точність і ефективність.*

***Висновки.** LLM демонструють високу ефективність у виявленні програмних ботів, проте мають високу обчислювальну вимогливість. Тому актуальним є застосування гібридних підходів, які поєднують LLM з традиційними методами. Така гібридизація дозволить зменшити використання ресурсів і забезпечити більш стійку та адаптовану систему виявлення ботів. Такий підхід може сприяти поліпшенню загальної продуктивності систем виявлення ботів, зменшенню витрат на обчислювальні ресурси та забезпеченню більш точного і ефективного виявлення шкідливих програм у соціальних мережах. Рекомендується подальше дослідження для вдосконалення інтеграції LLM у системи виявлення ботів, особливо в контексті динамічної поведінки соціальних мереж та еволюції програмних ботів.*

***Ключові слова:** великі мовні моделі, нейронні мережі, аналіз метаданих, програмні боти, соціальні мережі.*

Lesya LYUSHENKO, Yaroslav PEREHUDA. METHOD OF BUILDING SOFTWARE DETECTORS FOR DETECTING SOFTWARE BOTS IN SOCIAL NETWORKS

***Abstract. The purpose** of this work is to study in detail the effectiveness of using large language models (LLM) to detect software bots in social networks. The work focuses on analyzing the effectiveness of different detection methods and determining the potential of LLM as a means to improve the accuracy and efficiency of the bot identification process.*

The study covers the analysis of three main approaches to bot detection: metadata analysis, text analysis, and graph analysis. Both traditional machine learning methods and the latest LLM are analyzed for their ability to analyze big data from social networks. The main technique is benchmarking, which involves the use of extended datasets such as TwiBot20 and TwiBot-22 to evaluate the performance of each method using metrics such as accuracy and F1-measure. It provides an objective view of the performance of different approaches to bot detection.

***The scientific novelty** of this work is the use of LLM to analyze various types of data from social networks to detect software bots. The authors consider the integration of LLM into traditional detection methods, which allows adapting detection processes to the complex behavior of software bots, ensuring high accuracy and efficiency.*

***Conclusions.** LLMs demonstrate high efficiency in detecting software bots, outperforming traditional methods by some indicators. However, given the computational demands of LLM, the authors recommend considering hybrid approaches that combine the advantages of LLM with the efficiency of traditional methods to optimize resource usage and provide a more robust and adaptive bot detection system. This approach can improve the overall performance of bot detection systems, reduce computing resource costs, and provide more accurate and effective detection of malicious actors in social networks. Further research is recommended to improve the integration of LLM into bot detection systems, especially in the context of the dynamic behavior of social networks and the evolution of software bots.*

***Key words:** large language models (LLM), neural networks, metadata analysis, software bots, social networks.*

Соціальні мережі змінили принципи комунікації в суспільстві та стали невід'ємною частиною повсякденного життя. Висока популярність таких мереж призвела до створення різних соціальних онлайн-платформ, кожна з яких надає унікальний досвід та забезпечує спілкування людям з однаковими інтересами у режимі реального часу, без географічних та часових обмежень.

Однак, такі платформи страждають від наявності програмних ботів, які використовуються для маніпулювання та поширення неправдивої інформації. За даними дослідницького співтовариства

маніпулювання за допомогою програмних ботів фіксується під час широкого спектру різноманітних тематичних дискусій. Аналіз облікових записів програмних ботів у соціальних мережах показує, що дії більшості таких ботів призводять до різнопланових онлайн-загроз, а саме: дезінформація [24], втручання у вибори [11], екстремістські кампанії [17], теорії змови [10] тощо. Додатково програмні боти використовуються для поширення пропаганди, нав'язливої реклами та фейкових новин. Так вплив ботів зафіксовано під час дебатів щодо вакцинації [30], реклами електронних сигарет [1] та дебатів щодо пандемії COVID-19 [23]. Така висока активність програмних ботів викликає занепокоєння дослідницького співтовариства та користувачів соціальних мереж щодо цілісності і правдивості інформації.

Дослідження щодо розуміння та виявлення програмних ботів в соціальних мережах завжди були «гонкою озброєнь» [26]. Перші методи були зосереджені на аналізуванні метаданих користувачів за допомогою класифікаторів машинного навчання [22], тоді як оператори програмних ботів маніпулюють поведінковими характеристиками та метаданими облікового запису, щоб уникнути виявлення [6]. Пізніше з'являються мовні моделі для аналізу текстів всередині облікового запису та в постах [31], тоді як оператори програмних ботів періодично публікують скопійовані у справжніх людей пости, щоб заплутати дані моделі і видавати себе за живих користувачів [6]. Новітні моделі збирають мережеві дані про взаємодію користувачів і аналізують її з використанням нейронних мереж на основі графів [2], тоді як сучасні комплексні програмні боти стежать за користувачами, з якими вони пов'язані тим чи іншим чином та стратегічно обривають зв'язки з ними, щоб бути більш непомітними для таких нейронних мереж [16].

Виникнення великих мовних моделей (Large Language Model) або LLM, призвело до значних проривів у сфері обробки великих масивів даних. Завдяки здатності швидко аналізувати, узагальнювати та витягувати інформацію з різнопланових джерел, LLM змінили підходи до виявлення взаємозв'язків. Це особливо важливо у галузях, де необхідно обробляти великі набори неструктурованих даних для розпізнавання тенденцій, аналізу патернів поведінки та розробки нових теорій тощо. Такі моделі відмінно справляються з різноманітними завданнями, здатні слідувати інструкціям [27], але їх використання має певні ризики [25].

Виявлення програмних ботів є надзвичайно складним завданням, головним чином через зростаючу складність та постійну модифікацію цих ботів. Тому актуальним є дослідження можливості застосування гібридних підходів, які поєднують переваги LLM з ефективністю традиційних методів, щоб оптимізувати використання обчислювальних ресурсів і забезпечити більш стійку та адаптовану систему виявлення програмних ботів в соціальних мережах.

1 Існуючі методи виявлення програмних ботів

Існуючі системи виявлення програмних ботів можна класифікувати за об'єктами аналізу: метадані, тексти, графи.

1.1 Виявлення програмних ботів на основі аналізу метаданих

Методи, які виявляють програмних ботів на основі аналізу метаданих, базуються на обробці даних, отриманих з облікових записів та шаблонів активності користувачів. Як правило, для виявлення програмних ботів використовується машинне навчання або нейронні мережі з алгоритмами класифікації. Даними для аналізування є різноманітні характеристики облікового запису користувача [12].

Розглянемо модель **SGBot** (Scalable and Generalizable Bot), яка використовує технології машинного навчання за парадигмою навчання під наглядом (supervised learning) [22]. Методом класифікації в даній моделі є метод «випадковий ліс» (Random forest) [4]. Для ефективного тренування та роботи даної моделі достатньо мати набір даних, який складається лише з невеликої кількості метаданих та їх похідних даних облікового запису користувача (табл. 1).

Таблиця 1

Характеристики облікового запису потрібні для роботи SGBot

Метадані	Похідні дані
Кількість постів	Частота постів
Кількість підписників	Швидкість зросту кількості підписників
Кількість друзів	Швидкість зросту кількості друзів
Кількість підписок	Швидкість зростання кількості підписок
Кількість груп	Швидкість зросту кількості груп
Чи наявна картинка облікового запису	Співвідношення кількості послідовників до друзів
Чи наявна картинка фону (якщо є така можливість)	Довжина псевдоніму, кількість чисел в псевдонімі, довжина імені, кількість чисел в імені, довжина опису профілю, вірогідність вибору псевдоніму
Чи підтверджений користувач	

Дані характеристики облікового запису аналізуються моделлю і результат аналізу повертається у вигляді дробового числа від 0 до 1. Чим ближче це число до 1, тим більша вірогідність, що даний обліковий запис управляється програмним ботом. Такий формат результату є досить поширеним серед моделей, які використовуються як детектори ботів.

Модель **SGBot** має значну масштабованість завдяки зосередженню лише на даних облікового запису користувача, до якої зазвичай можна легко отримати доступ без значних затрат на ресурси. Використання меншої кількості характеристик призводить до незначного зниження в точності визначення, проте отримується можливість аналізувати потік облікових записів у режимі реального часу. Проте, слід зазначити, що в певних випадках доречно аналізувати як можна більшу кількість характеристик. Так, аналогічна модель **Botometer** аналізує більше 1000 різноманітних характеристик (метадані, статистичні дані, похідні дані, шаблони поведінки тощо) [32].

Розповсюдження моделей, які працюють на основі аналізу характеристик облікових записів, призвело до нових реалізацій програмних ботів, які є набагато успішнішими в «обмані» таких методів аналізу [6]. Дана програмні боти успішно маніпулюють профілями облікових записів, а їх шаблони поведінки є досить непередбачуваними. Отже, існуючі методи, засновані на аналізі характеристик облікових записів, стикаються з проблемами в точному виявленні цих нових облікових записів програмних ботів [21].

1.2 Виявлення програмних ботів на основі аналізу тексту

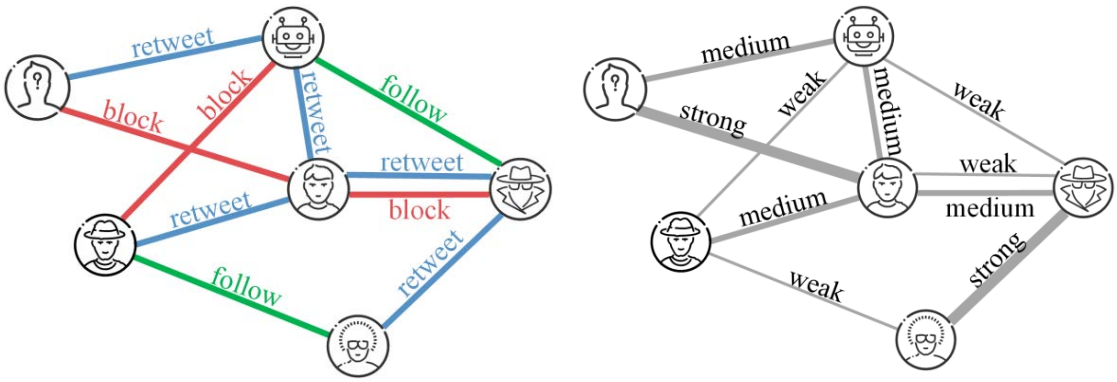
Методи на основі аналізу тексту для виявлення програмних ботів в основному покладаються на методи обробки природньої мови (Natural Language Processing), або NLP. За допомогою NLP зазвичай аналізуються опублікований вміст (пости) та описи облікових записів користувачів. Підходи в цій категорії включають вкладання слів (word embedding) [31], рекурентні нейронні мережі [12], механізми уваги (attention mechanisms) [21] і використання попередньо навчених мовних моделей для кодування постів [9]. Поєднання аналізу постів та аналізу характеристик облікового запису користувача, дозволяє застосували моделі машинного навчання без нагляду [21]

Однією з моделей, яка використовує даний метод є модель для виявлення програмних ботів **RoBERTa** (**R**obustly **O**ptimized **BERT A**pproach) [18]. Дана мовна модель використовується для кодування постів та описів користувачів, закодовані данні яких потім передаються у класифікатор «бот - людина», що заснований на роботі багатошарового перцептрона Румельхарта [19].

Незважаючи на численні існуючі дослідження та вражаючу ефективність методів на основі аналізу тексту, нові облікові записи програмних ботів все ще можуть уникнути виявлення, ділячись викраденим вмістом від справжніх користувачів [6]. Крім того, нещодавні дослідження показали, що покладатися виключно на текстові дані недостатньо для надійного та точного виявлення програмних ботів [5].

1.3 Виявлення програмних ботів на основі аналізу графів

Користувачі та програмні боти в соціальних мережах взаємодіють по-різному та мають різний вплив на інших, що призводить до неоднорідності відношень і впливу (рис. 1).



а) граф відносин користувачів б) граф сили впливів користувачів

Рис. 1. Взаємодія користувачів та програмних ботів в соціальних мережах

Виявлення програмних ботів на основі аналізу графів передбачає аналіз зав'язків та відношень між користувачем та його підписниками, друзями, підписками, спільними групами, тощо. Із таких «об'єктів», зв'язків та відношень можна утворити відповідні структурні графи.

Для виявлення програмних ботів структурні графи аналізуються методами: показників центральності [8], навчання за поданими вузлами (node representation learning) [3], графових нейронних мереж

(Graph Neural Network), або GNN [7]. Комбінування різних методів аналізу графів і текстів [32], а також створення покращених архітектур GNN для аналізу неоднорідних мереж [21], мають значні перспективи для виявлення програмних ботів.

Однією з моделей виявлення програмних ботів на основі аналізу графів, яка використовує неоднорідності зв'язків в соціальних мережах, є RGT (Relational Graph Transformers) [21]. Дана модель використовує топологічну структуру соціальної мережі і будує граф з неоднорідними відношеннями і впливами. У такому графі користувачі виступають у ролі вершин, а неоднорідні відношення у ролі ребер. Даний граф оброблюється шаром перетворення неоднорідного графа, який використовує методи на основі механізмів уваги. В ході обробки формуються графи інтенсивності впливів вершин, мережа семантичної уваги агрегує графи впливів між користувачами і в результаті шар перетворення неоднорідного графа видає результат аналізу.

Проте існуючі методи виявлення програмних ботів мають свої обмеження, які вимагають значних обчислювальних ресурсів і, найважливіше, великих наборів даних для свого тренування. Нещодавнє оголошення монетизації Twitter API (Twitter 2023), який використовувався для отримання тренувальних даних для моделей, робить зазначені вище методи дорогими в обслуговуванні, підтримці та адаптації до нововведень.

1.4 LLM як детектори програмних ботів

Особливістю LLM є використання нейронних мереж, які навчені на великому обсязі немаркованого тексту. Великі мовні моделі, як правило, працюють краще, ніж звичайні мовні моделі, у широкому діапазоні завдань природної мови, особливо коли вони мають доступ до більшої кількості даних і тонкої настройки, що означає, що вони можуть адаптуватися до різних сфер роботи і сценаріїв.

Існує кілька відмінностей між звичайними мовними моделями та великими мовними моделями. LLM є більш комплексними, ніж звичайні мовні моделі і для їх навчання використовувалася більша кількість різноманітних даних, ніж для звичайної мовної моделі. Це означає, що вони можуть фіксувати більш загальні та різноманітні лінгвістичні знання, але також більше інформаційного шуму та помилок.

Поглибимо розуміння особливостей використання LLM в подальшому дослідженні.

2 Порівняльний аналіз можливостей різних моделей щодо виявлення програмних ботів

Метою даного дослідження є порівняльний аналіз існуючих моделей, які використовуються для виявлення програмних ботів. Результатом дослідження є визначення моделей, які найбільш достовірно розрізняють реальних користувачів від програмних ботів. Особлива увага приділена порівнянню ефективності LLM з іншими моделями та визначення потенціалу LLM як засобу для підвищення точності та ефективності процесу ідентифікації ботів.

Об'єктами дослідження є наступні моделі: SGBot, RoBERTa, RGT, LOBO, BotBuster, Botometer, BotPercent, і LMBot, та три великі мовні моделі: Mistral-7B [15], LLaMA2-70b [13] і ChatGPT. Умовно, їх можна поділити на три групи за видами даних:

1. Моделі, які аналізують тільки один вид даних: SGBot аналізує метадані, RoBERTa – текст, і RGT – графи.
2. Моделі які аналізують два виду даних: BotBuster та LOBO, обидва з яких аналізують метадані та текст.
3. Моделі які аналізують три види даних: метадані, текст та графи. До них відносяться моделі Botometer, BotPercent, LMBot та три моделі на основі LLM: Mistral-7B, LLaMA2-70b і ChatGPT.

Порівняння моделей проводитиметься шляхом порівняння результатів тестування моделей. В результаті тестування визначається точність та F1-міри (F-score, F-measure). Під точністю мається на увазі відсоток достовірно визначених реальних користувачів та відсоток програмних ботів. Під F-мірою мається на увазі одна з мір точності тесту, яка обчислюється через влучність та повноту тесту, де влучність є числом правильно визначених позитивних результатів, поділеним на число всіх позитивних результатів, включно з визначеними неправильно, а повнота є числом правильно визначених позитивних результатів, поділеним на число всіх зразків, які повинно було бути визначено як позитивні [20].

Навчання моделей та їх тестування проводилися на двох наборах даних: TwiBot20 [28] і набагато більш великому та різноплановому наборі TwiBot-22 [29]. Зазначені набори даних містять: метадані, статистичні дані, похідні дані, шаблони поведінки, взаємодію та зв'язки з іншими користувачами, облікові записи реальних користувачів та програмних ботів. Дані збиралися із соціальної мережі Twitter, в час, коли Twitter API ще був доступний для дослідницької роботи. Кожен набір даних розділявся на дві нерівні частини: перший і більший набір даних використовувався для навчання моделей, менший набір – для безпосереднього тестування моделей.

2.1 Навчання моделей LLM для виявлення програмних ботів

Оскільки LLM навчені на більшій кількості різноманітних даних, ніж звичайні мовні моделі і є більш комплексними, для їх подальшого навчання для ефективного виявлення програмних ботів достатньо використовувати так звані запити (prompts). Так, використовуючи спеціально сформовані до LLM

запити, їх можна навчити на прикладах, надавши при цьому відповідний контекст. Для більш ефективного навчання, було вирішено розділити запити відповідно до виду даних, які ці запити будуть оброблювати: метадані, текст та графи (рис. 2).

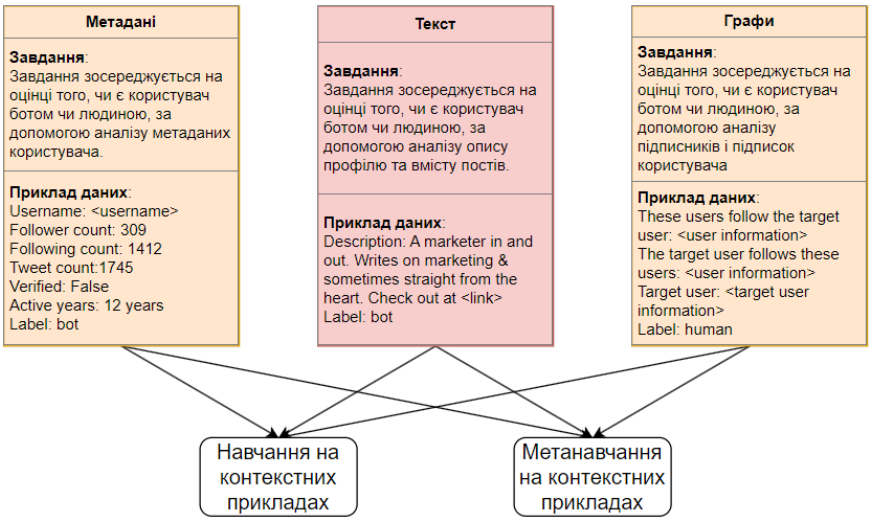


Рис. 2. Схема запитів до LLM для її навчання у виявленні облікових записів програмних ботів

2.1.1 Навчання за допомогою аналізу метаданих

З набору даних випадково вибирається набір із *n* користувачів, метадані яких використовуються для навчання LLM. Метадані облікового запису було послідовно об'єднано у лінійну форму, схожу на природню мову. До кожного запису були надані відповідні мітки про те, чи є користувач програмним ботом чи людиною. Далі, навченій на контекстних прикладах LLM дають запит самій сформувати мітку для наступного користувача (рис. 3).

Для аналізу метаданих облікового запису використовувалася наступні дані: кількість підписників, кількість підписок, кількість постів, чи є обліковий запис підтвердженим та кількість активних років, оскільки саме ці дані найбільше допомагають ідентифікувати соціальних програмних ботів.

2.1.2 Навчання за допомогою аналізу тексту

З набору даних випадково вибирається набір із *n* користувачів, описи облікових записів та зміст постів, яких використовують для навчання LLM.

Evaluate whether a user is a bot or human with the help of several labeled examples. Output the label first and explanation after.

Username: <redacted>. Follower count: 352. Following count: 1432. Tweet count: 1641. Verified: False. Active years: 12 years.

Label: bot

Username: <redacted>. Follower count: 4712064. Following count: 41. Tweet count: 6226. Verified: True. Active years: 14.

Label: human

Username: <redacted>. Follower count: 16491. Following count: 19928. Tweet count: 49652. Verified: False. Active years: 5

Label: ?

Рис. 3. Приклад запиту для навчання LLM на основі аналізу метаданих

До даних постів/описів додається мітка про те, чи є користувач програмним ботом чи людиною, і отриманий набір використовується для навчання LLM на прикладах.

Далі, навченій на контекстних прикладах LLM дають запит самій сформувати мітку для наступного користувача. Для цього їй надається опис цільового облікового запису та вміст його постів. LLM аналізує опис пости і ставить до кожного мітку. Загальним результатом є переважна кількість міток (рис. 4).

Evaluate whether a user is a bot or human with the help of the user's self-written description.
Output the label first and explanation after.
Description: sc/ shenallemoorr ig/ shenallemoore
Label: bot
Description: A marketer in and out. Writes on marketing straight from heart. Check out at <link>
Label: bot
Description: Day 1 Trump supporter. I rode the escalator!
Label: ?

Рис. 4. Приклад запиту для навчання LLM на основі аналізу тексту

2.1.3 Навчання за допомогою аналізу графів

З набору даних випадково вибирається набір із *n* користувачів. Для того, щоб навчити LLM на контекстних прикладах, їй надаються списки із підписників та підписок, кожен з яких має мітку про те, чи є користувач програмним ботом чи людиною. Далі навчена на контекстних прикладах LLM використовує списки підписників та підписок цільового облікового запису, щоб самій сформувати його мітку (рис. 5).

Evaluate whether a user is a bot or human with the help of the user's followers and followings and their labels. Output the label first and explanation after.
These users follow the target user:
<user metadata and description>
Label: bot
The target user follows these users:
<user metadata and description>
Label: human

Target user:
<target user metadata and description>
Label: ?

Рис. 5. Приклад запиту для навчання LLM на основі аналізу графів

2.1.4 Метанавчання на контекстних прикладах

Слід зазначити, що звичайне навчання LLM на контекстних прикладах не є ідеальним способом навчання LLM. Враховуючи специфіку дослідження, було визначено, що більш ефективним методом навчання буде *метанавчання* на контекстних прикладах [14]. Метанавчання на контекстних прикладах має на меті покращити здатність LLM виконувати інструкції шляхом тонкого налаштування LLM до запитів типу {інструкція, вхідні дані, вихідні дані} [27].

Ще однією особливістю є навчання LLM на відносно невеликому наборі даних. Незважаючи на те, на якому наборі даних проводився результат тестування Twibot-20 чи Twibot-22, для навчання або метанавчання на контекстних прикладах з кожного набору відповідно бралися дані лише однієї тисячі облікових записів. Це відносно мало, якщо порівнювати з іншими моделями, для навчання яких потрібно декілька тисяч, або навіть сотень тисяч облікових записів. Незважаючи на, здавалось, малий набір навчальних даних, наданих LLM, вони показали достовірні результати при тестуванні, ознайомитися з якими можна далі.

2.2 Результати порівняльного аналізу моделей

Результати тестування та порівняльного аналізу моделей були розділені на три групи відповідно до кількості видів даних які аналізує та чи інша модель.

2.2.1 SGBot, RoBERTa та RGT

До першої групи моделей належать SGBot, RoBERTa та RGT. Результати їх тестування наведені в таблиці 2.

Таблиця 2

Результати тестування моделей SGBot, RoBERTa та RGT

Назва моделі	Набір даних Twibot-20		Набір даних Twibot-22	
	Точність, %	F1-міра	Точність, %	F1-міра
SGBot	81.6	0.847	62.3	0.394
RoBERTa	75.5	0.732	63.3	0.431
RGT	86.6	0.880	50.9	0.509

За результатами тестування модель RGT показала себе краще, ніж дві інші моделі на наборі даних Twibot-20, і гірше на наборі даних Twibot-22. На другому наборі даних у даної моделі все ще найкращий показник F1-міри, проте вона значно відстає від інших моделей по критерію точності. Це обумовлено тим, що набір даних Twibot-22 є більш великим і різноплановим, а отже надає більш широкий спектр даних для аналізу даних облікового профілю. Таким чином, модель RGT краще працює на обмежених наборах даних, в той час як модель RoBERTa має кращі результати на більш широких наборах даних. Схожу тенденцію також можна побачити в результатах тестування інших моделей.

2.2.2 BotBuster та LOBO

До другої групи моделей належать BotBuster та LOBO. Результати їх тестування наведені в таблиці 3.

Таблиця 3

Результати тестування моделей BotBuster та LOBO

Назва моделі	Набір даних Twibot-20		Набір даних Twibot-22	
	Точність, %	F1-міра	Точність, %	F1-міра
BotBuster	77.2	0.811	62.7	0.439
LOBO	76.2	0.806	55.2	0.197

За результатами тестування зазначимо, що модель BotBuster перевершує модель LOBO на обох наборах даних: Twibot-20 та Twibot-22.

2.2.3 LMBot, BotPercent, Botometer, Mistral-7B, LLaMA2-70b та ChatGPT

До третьої та найбільшої групи моделей належать LMBot, BotPercent, Botometer та три моделі на основі LLM: Mistral-7B, LLaMA2-70b та ChatGPT. Слід зазначити, що для моделі ChatGPT представлені результати тестування як при звичайному навчанні на контекстних прикладах, так і при метанавчанні.

Результати тестування моделей показують, що ChatGPT, навчений за допомогою метанавчання на контекстних прикладах, має кращі показники, ніж інші моделі на обох наборах даних Twibot-20 та Twibot-22 (таблиця 4).

Хоч результати тестування мають значні відмінності в залежності від набору даних, як було сказано раніше, це обумовлюється різницею між цими наборами даних. Отже ChatGPT, навчений на наборі даних Twibot-20 за допомогою простого навчання на контекстних прикладах, має значно гірші результати, ніж при метанавчанні. Аналогічна ситуація і при тестування на наборі даних Twibot-22, хоча в даному випадку різниця порівняно не така велика. Це, в свою чергу, ще раз показує перевагу метанавчання перед звичайним навчання для тренування LLM.

Таблиця 4

Результати тестування моделей LMBot, BotPercent, Botometer, Mistral-7B, LLaMA2-70b та ChatGPT

Назва моделі	Набір даних Twibot-20		Набір даних Twibot-22	
	Точність, %	F1-міра	Точність, %	F1-міра
LMBot	85.6	0.876	-	-
BotPercent	84.5	0.864	73.1	0.726
Botometer	53.1	0.531	75.5	0.585
Mistral-7B	60.9	0.573	58.2	0.534
LLaMA2-70B	66.2	0.658	66.8	0.685
ChatGPT	63.2	0.557	73.5	0.705
ChatGPT (метанавчання)	89.9	0.914	76.9	0.792

2.2.4 RoBERTa, RGT, BotBuster та ChatGPT

Об'єднаємо результати таблиць 6-7, виокремивши моделі, які показали себе найкраще в кожній із трьох груп. Отримаємо результати тестування моделей RoBERTa, RGT, BotBuster та ChatGPT в таблиці 5.

Таблиця 5

Результати тестування моделей RoBERTa, RGT, BotBuster та ChatGPT

Назва моделі	Набір даних Twibot-20		Набір даних Twibot-22	
	Точність, %	F1-міра	Точність, %	F1-міра
RoBERTa	75.5	0.732	63.3	0.431
RGT	86.6	0.880	50.9	0.509
BotBuster	77.2	0.811	62.7	0.439
ChatGPT (метанавчання)	89.9	0.914	76.9	0.792

В результаті велика мовна модель ChatGPT, навчена за допомогою метанавчання на контекстних прикладах, показала найкращі результати серед усіх трьох груп моделей. При правильному методі навчання вона має найкращі результати точності та F1-міри незалежно від набору даних, який використовувався для навчання та тестування моделі.

Отже, результати порівняльного аналізу показали що великі мовні моделі мають високий потенціал щодо виявлення програмних ботів. Для свого навчання вони не потребують відносно широкого набору даних порівняно з іншими моделями, і при правильному методі навчання вони можуть досягти високої ефективності роботи, вираженої в точності визначення програмних ботів та людей серед користувачів.

3 Покращення системи виявлення програмних ботів

Моделі LLM є потужними сучасними інструментами, які можуть досягти чудових результатів у багатьох сферах та задачах, без потреби у обширних наборах навчальних даних. З іншої сторони моделі LLM вимагають більше обчислювальних ресурсів, таких як пам'ять, час та обчислювальних потужностей для навчання та роботи. Інші моделі є менш вимогливими до ресурсів, ніж великі мовні моделі. Це означає, що вони можуть працювати швидше й дешевше. Тому незважаючи на позитивні результати, використання моделей на основі LLM в тому вигляді, в якому вони існують на даний момент, не є оптимальним.

З урахуванням вищезазначеного, має сенс комбінований підхід в використанні досліджених методів. На (рис. 6) запропонована принципова схема системи детектора ботів, що використовує моделі LLM разом з іншими моделями для виявлення програмних ботів. Основа даної системи полягає у використанні моделей LLM лише у випадках, коли інші моделі мають складності у ідентифікації користувача. Розглянемо даний метод на прикладі моделі SGBot. Результат роботи даної моделі видається у вигляді дробового числа від 0 до 1. Чим ближче це число до 0,5 тим більше у моделі виникало складностей при ідентифікації користувача. У таких випадках можна використати LLM модель для уточнення результату моделі SGBot**.** Діапазон, при якому використовується модель LLM слід підібрати так, щоб збільшити точність результату ідентифікації і при цьому не потребувати значних обчислювальних ресурсів постійно. В результаті, сформована система буде мати підвищену точність з незначним збільшенням потреб в обчислювальних ресурсах.

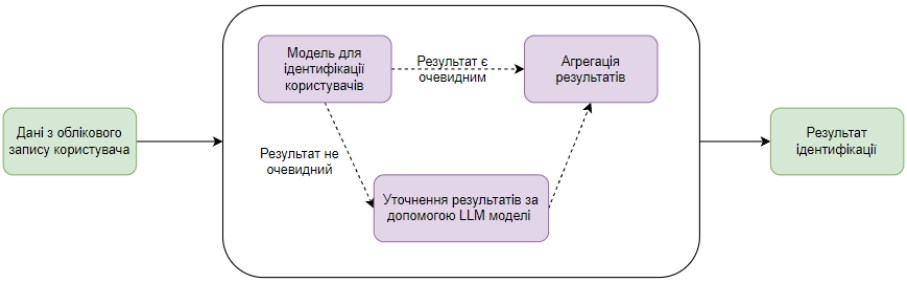


Рис. 6. Принципова схема запропонованої системи визначення програмних ботів

Висновки

Наразі існують різноманітні методи виявлення програмних ботів на основі машинного навчання та нейронних мереж. Кожен із методів підходить для вирішення конкретних сценаріїв виявлення програмних ботів, що забезпечує оптимальні рішення для конкретних випадків використання. Точність цих методів значно знижується в загальніших сценаріях виявлення програмних ботів, що охоплюють, зокрема, різні часові періоди, теми обговорення та мови.

Експерименти на двох широко поширених наборах даних демонструють, що виявлення програмних ботів на основі LLM може досягти високої точності, незважаючи на низьку кількість навчальних даних. Проте, для своєї роботи LLM моделі вимагають значних обчислювальних ресурсів, і, беручи до уваги

кількість користувачів і відповідну величину потоків даних у різноманітних соціальних мережах, використання таких моделей не є оптимальним.

Для вирішення цієї проблеми була запропонована схема системи виявлення програмних ботів. Основа запропонованої системи полягає у використанні ресурсоємних моделей LLM лише у випадках, коли інші моделі мають складності щодо ідентифікації користувача.

Список використаних джерел:

1. Allem J.-P., Ferrara E. The importance of debiasing social media data to better understand e-cigarette related attitudes and behaviors. *Journal of medical Internet research*. 2016. Vol. 18. № 8.
2. BIC: Twitter bot detection with text-graph interaction and semantic consistency. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics / Lei Z. et al. Canada, 2023. Vol. 1. P. 10326–10340.
3. Bot2Vec: A general approach of intra-community oriented representation learning for bot detection in different types of social networks. Information Systems / Pham P. et al. 2022. Vol. 103. DOI: 10.1016/j.is.2021.101771.
4. BotOrNot: A system to evaluate social bots. Proceedings of the 25th International Conference Companion on World Wide Web / Davis C.A. et al. 2016. P. 273–274.
5. BotRGCN: Twitter bot detection with relational graph convolutional networks. Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining / Feng S. et al. IEEE, 2021. P. 236–239.
6. Cresci S. A decade of social bot detection. *Communications of the ACM*. 2020. Vol. 63. № 10. P. 72–83.
7. Detect me if you can: Spam bot detection using inductive representation learning. Companion proceedings of the 2019 World Wide Web conference / Ali A.S. et al. 2019. P. 148–153.
8. Detecting bots in social-networks using node and structural embeddings. *Journal of Big Data* / Dehghan A. et al. 2023. Vol. 10. № 1. P. 1–37.
9. Dukic D., Keca D., Stipic D. Are you human? detecting bots on twitter using BERT. 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA). IEEE, 2020. P. 631– 636.
10. Ferrara E. What types of covid-19 conspiracies are populated by twitter bots? First Monday, 25(6), 2020. DOI: 10.5210/fm.v25i6.10633.
11. Howard P.N., Kollanyi B., Woolley S. Bots and automation over twitter during the US election. *Computational propaganda project : working paper series*. 2016. № 21(8).
12. Kudugunta S., Ferrara E. Deep neural networks for bot detection. *Information Sciences*. 2018. № 467. P. 312–322.
13. Llama 2: Open foundation and fine-tuned chat models / Touvron H. et al. 2023. (Preprint arXiv:2307.09288).
14. MetaICL: Learning to learn in context. Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies / Min S. et al. 2022. P. 2791–2809.
15. Mistral 7b / Jiang A.Q. et al. 2023. (Preprint arXiv:2310.06825).
16. Multi-modal social bot detection: Learning homophilic and heterophilic connections adaptively. Proceedings of the 31st ACM International Conference on Multimedia / Li S. et al. ACM, 2023 P. 3908–3916.
17. Predicting online extremism, content adopters, and interaction reciprocity. Social Informatics / Ferrara E. et al. Bellevue, 2016. P. 22–39.
18. Roberta: A robustly optimized BERT pretraining approach / Liu Y. et al. 2019. (Preprint arXiv:1907.11692).
19. Rumelhart D.E., Hinton G.E., Williams R.J. Learning Internal Representations by Error Propagation, Parallel Distributed Processing, Explorations in the Microstructure of Cognition. *Biometrika* / ed. Rumelhart D.E., McClelland J. 1986. Vol. 1. № 71. P. 599-607.
20. Sasaki Y. The truth of the F-measure. *Teach tutor mater*. 2007. Vol. 1. № 5. P. 1-5.
21. Satar: A self-supervised approach to twitter account representation learning and its application in bot detection. Proceedings of the 30th ACM International Conference on Information & Knowledge Management / Feng S. et al. ACM, 2021. P. 3808–3817.
22. Scalable and generalizable social bot detection through data selection. *Proceedings of the AAAI conference on artificial intelligence* / Yang K.-C. et al. 2020. Vol. 34. P. 1096–1103.
23. Shahi G.K., Dirkson A., Majchrzak T.A. An exploratory study of COVID-19 misinformation on Twitter. *Online social networks and media*. 2021. № 22.
24. Social bot-aware graph neural network for early rumor detection. Proceedings of the 29th International Conference on Computational Linguistics / Huang Z. et al. Gyeongju, 2022. P. 6680– 6690.
25. Taxonomy of risks posed by language models. Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency / Weidinger L. et al. ACM, 2022. P. 214–229.
26. The paradigm-shift of social spambots: Evidence, theories, and tools for the arms race. Proceedings of the 26th international conference on world wide web companion / Cresci S. et al. 2017. P. 963–972.
27. Training language models to follow instructions with human feedback. Advances in Neural Information Processing Systems / Ouyang L. et al. 2022. № 35.
28. Twibot-20: A comprehensive twitter bot detection benchmark. Proceedings of the 30th ACM International Conference on Information & Knowledge Management / Feng S. et al. 2021. P. 4485–4494.
29. Twibot-22: Towards graph-based twitter bot detection. Advances in Neural Information Processing Systems / Feng S. et al. 2022. Vol. 35. P. 35254–35269.
30. Weaponized health communication: Twitter bots and Russian trolls amplify the vaccine debate. *American journal of public health* / Broniatowski D.A. et al. 2018. Vol. 108. № 10. P. 1378–1384.
31. Wei F., Nguyen U.T. Twitter bot detection using bidirectional long short-term memory neural networks and word embeddings. 2019 First IEEE International conference on trust, privacy and security in intelligent systems and applications (TPSISA). IEEE, 2019. P. 101–109.
32. Yang K.-C., Ferrara E., Menczer F. Botometer 101: Social bot practicum for computational social scientists. *Journal of Computational Social Science*. 2022. Vol. 5. № 2. P. 1511–1528.

УДК 004.51
DOI https://doi.org/10.32689/maup.it.2024.1.9

Володимир МАТУЗКО
аспірант кафедри програмної інженерії,
Запорізький національний університет, matuzkovd@ukr.net
***ORCID:** 0000-0002-3005-6051*

АЛГОРИТМИ В ПРОГРАМНІЙ РЕАЛІЗАЦІЇ АВТОМАТИЗОВАНОГО ПЕРЕКЛАДУ ІНТЕРФЕЙСУ ПРОГРАМ

Анотація. *Велика кількість повсякденних дій вже давно виконується за допомогою мобільних додатків та міжнародних ресурсів у мережі Інтернет. Через це у користувачів постає проблема знання мови для можливості користування цими програмами. Не всі розробники в світі мають доступ до професійних послуг перекладу або можливість створити належний переклад власноруч. Важливим та зазвичай вирішальним фактором є вартість програмного забезпечення для створення перекладу. Великі професійно-спрямовані програмні пакети мають вартості ліцензій, що вимірюються в сотнях доларів США. Також більшість з існуючих засобів спрямовані на переклад довільних текстів у загальній формі. Альтернатиним підходом є залучення команд перекладачів, але в цьому випадку треба враховувати додатковий час та обсяг перекладу, а також обмежений доступ до перекладачів на менш відомих мовах світу.*

Зважившиці фактори можна зазначити, що існує потреба в зручній та доступній програмі, що спеціалізована для створення та забезпечення якості перекладу саме інтерфейсів програмного забезпечення. Після проведеного аналізу визначені вимоги до запропонованої програми, а саме зручний інтерфейс та набір функцій, що спеціалізовані для роботи з програмними інтерфейсами та файлами вихідного коду.

***Мета роботи** – розробка програми з необхідним функціоналом для автоматизації перекладу інтерфейсів користувача.*

***Методологія.** Програмне забезпечення розроблене з використанням мови C# в середі Microsoft Visual Studio. Інтерфейс програми створено з урахуванням вимог розробників програмного забезпечення.*

***Наукова новизна.** Виявлено напрямки вдосконалення існуючих методів та засобів та розробка програмного забезпечення для надання нового засобу для перекладу. В даній роботі описано та розроблено комп'ютерну утиліту у версії для ОС Microsoft Windows, яка реалізує всі вимоги до базового функціоналу. Програма спеціалізована для перекладу інтерфейсів користувача та дозволяє автоматизувати цей процес. Оптимізовано та спрощено процес створення перекладу інтерфейсів для власних програмних розробок невеликих команд розробників шляхом повної прив'язки існуючого вихідного коду незалежно від використаної мови програмування.*

***Висновки.** Розроблену програму можна використовувати для створення та перекладу інтерфейсів користувача. Функція автоматизованого перекладу потребує володіння особистого ключа для користування Google Translate API.*

***Ключові слова:** машинний переклад, розробка програмного забезпечення, інтерфейс користувача, Google Translate.*

Volodymyr MATUZKO. ALGORITHMS IN SOFTWARE SOLUTION FOR AUTOMATED USER INTERFACE TRANSLATION

Abstract. *Numerous daily activities are long since accomplished using mobile applications and international resources available through the Internet. This raises the issue of the need for end users to know the languages required to operate and use these programs. Not every developer in the world has access to professional translation services, or the ability to create such translations on their own. An important and usually determining factor is the cost of translation software. Large professionally-targeted software packages have license fees measured in hundreds of US dollars. Also, the majority of existing solutions are meant for various kinds of general freeform text. An alternative approach would be hiring teams of translators, but this needs accounting for the extra time and size of translation, as well as limited availability of translators for the lesser known world languages.*

Evaluating these factors shows an existing need of a convenient and accessible program specialized in creating and ensuring quality translation of software user interfaces specifically. Analysis leads to confirmed requirements for the program, those being a comfortable user interface and a set of functions specific to working with user interfaces and source code files.

***The purpose of this work** is to develop software that meets the functionality requirements for automated user interface translation.*

***Methodology.** Software is developed using the C# language and Microsoft Visual Studio environment. The program's interface was designed according to needs of software developers.*

***Scientific novelty.** Determine ways and approaches to improve existing methods and tools, and the development of software to provide a new tool for translation. This work describes the development of such software tool for Microsoft Windows that implements every listed requirement for base functionality. The program is specialized for translation of user interfaces and enables automation of the process. The process of providing user interface translation for software created by small teams of developers were optimized and simplified via a full link to existing source code regardless of programming language used.*

***Conclusions.** The developed program can be used to create and translate user interfaces. Machine translation functionality requires possession of a personal key to utilize Google Translate API.*

***Key words:** machine translation, software development, user interface, Google Translate.*

Постановка проблеми. У сучасних умовах надзвичайно зростає роль синхронного перекладу як засобу, який обслуговує економічні, суспільно-політичні, наукові, культурно-естетичні та інші відносини народів світу. Тому проблема якісного перекладу є актуальною на даний час і вимагає розробки все нових методів та засобів синхронного перекладу.

Науково-технічний прогрес, який охоплює все нові сфери життя і пов'язані з нею міжнародне співробітництво наук, очікуваний демографічний вибух і інші найважливіші явища розвитку цивілізації призводять до небувалого розвитку різного роду контактів між державами та іншими різномовними товариствами людей. Глобалізація політики, економіки, виробництва й досліджень, а також об'єднання зусиль для подолання наслідків кризових явищ та катастроф планетарного масштабу жорстко поставили на порядок денний проблему забезпечення комунікації в умовах сучасної полілінгвокультурної світової спільноти. У таких умовах наріжним каменем взаємодії є оптимізація процесу комунікації, основним засобом якої завжди була й залишається природна мова. Ці та інші виклики, які з часів промислової революції постали перед людством, стимулювали створення новітніх напрямків лінгвістики, спрямованих на вивчення функціональних аспектів природної мови на кшталт лінгвістики фахових мов, лінгвістики тексту тощо, а також виникнення суміжних дисциплін, що сформувались в результаті спеціалізації й інтеграції наукових досліджень з метою оптимізації процесу комунікації: термінознавство, психолінгвістика, когнітологія, комп'ютерна лінгвістика, корпусна лінгвістика, штучний інтелект тощо. Поступова імплементація доробків цих наук та накопичення серйозного масиву лінгвістичних ресурсів результували виокремленням комплексного наукового напрямку — автоматизованого опрацювання природної мови. До пошуку нових можливостей вирішення проблеми глобальної комунікації стимулювала усіх зацікавлених і еволюція інформаційних та мережевих технологій [1].

Синхронний переклад також присутній у світі комп'ютерних інтерфейсів. Особливо гостро питання стоїть серед користувачей мобільних додатків, тому що кількість смартфонів та інших подібних пристроїв вимірюється мільярдами [5]. Велика кількість цих додатків має версії лише на одній мові, або використовує машинний переклад без перевірки коректності тексту та урахування контексту в інтерфейсі. Також якість машинного перекладу також залежить від конкретної мовної пари. Як приклад, Китай є одним з найбільших постачальників сучасних мобільних додатків. До того ж, машинний переклад з китайської історично відрізнявся своєю складністю у порівнянні з перекладом між європейськими мовними групами. Дослідження та розробки у цьому напрямку відбуваються постійно [2]. Щодо інтерфейсів програм для комп'ютерів ситуація дуже схожа, але зазвичай розробники користуються допомогою команд перекладачів, тим самим уникаючи ситуацій з низькою якістю перекладу. Однак ці перекладачі мають виконувати переклад та знаходити методи і засоби для цього власноруч.

Одним з можливих варіантів реалізації такого функціоналу є використання існуючого формату .PO, який використовується в системі локалізації GNU gettext [3]. При використанні цього алгоритму розробник генерує PO-файл, який буде містити набір рядків, відповідних тексту інтерфейсу його програми. Для цього розробник має форматувати рядки з текстом згідно до вимог алгоритму gettext – кожний рядок має бути у вигляді функції, для якої вхідними даними є сам текст інтерфейсу на деякій базовій мові, а як вихіні дані повертає відповідний текст з доступного набору даних для перекладу. Використання цього функціоналу також потребує від розробника залучення та налаштування бібліотеки GNU. Gettext, як у випадку з мовою програмування C# [4].

```
white-space
#  translator-comments
#. extracted-comments
#: reference…
#, flag…
#| msgid previous-untranslated-string
msgid untranslated-string
msgstr translated-string
```

Рис. 1. Формат запису одного рядку в файлі PO

Отриманий таким чином файл PO потім можна відкрити в утиліті для перекладу. Прикладом вже існуючої такої програми-утиліти є Poedit [6]. Poedit дозволяє користувачу редагувати ці файли

в інтерфейсі, схожому на інші існуючі програми для перекладу. Також Poedit надає вбудовану можливість використання онлайн-сервісів для машинного перекладу, таких, як Microsoft Translator та Google Translate, але цей функціонал наявний лише в платних версіях програми [7]. Poedit націлений на перекладачів, які працюють виключно з файлами перекладу, тому в ньому відсутня можливість переглянути вихідний код для отримання повного розуміння, де і як використовується текст.

Після завершення роботи над перекладом розробнику потрібно конвертувати отриманий PO-файл в коректний формат для обраної мови програмування за допомогою інструментів GNU gettext. Також потрібно зазначити, що автоматизація за допомогою gettext реалізована тільки для конкретного переліку популярних мов програмування, тому цей алгоритм важко назвати універсальним рішенням для всіх розробників.

Аналіз останніх досліджень і публікацій. В [1, 2, 5] розглянуто досягнутий прогрес в сфері машинного перекладу та визначенно актуальні проблеми. Доцільність розробки альтернативних алгоритмів доведена на основі існуючих рішень, що описано в ресурсах [3, 4, 6, 7].

Постановка завдання. Розробити алгоритм та відповідну програмну реалізацію автоматизації перекладу інтерфейсів комп'ютерного програмного забезпечення, спрямовані на використання невеликими командами розробників та незалежно від обраної ними мови програмування. Для досягнення цієї мети поставлено завдання визначити сильні і слабкі сторони існуючих сучасних методів та засобів перекладу, а також їх варіанти застосування; сформувати перелік технічних вимог.

Виклад основного матеріалу дослідження. Однією з головних вимог до розробленої програми є можливість завантаження та взаємодії з вихідним кодом для забезпечення коректного перекладу елементів інтерфейсу. Для цього розробнику потрібно наперед підготувати код згідно вимог утиліти, а саме зберігати всі рядки зі змістом тексту інтерфейсу в належній формі. Головною перевагою такого підходу є майже повна незалежність від обраної мови програмування.

Вимоги поточної версії перекладача до форматування:

- двовимірний масив рядків розміром [кількість_мов, кількість_рядків];
- локалізація рядків інтерфейсу у вигляді = UITranslator_language[0,##], де 0 – порядковий номер першої мови локалізації, а ## – порядковий номер рядку в файлі локалізації.

Після завантаження вихідний код можна переглянути в правому вікні утиліти. Коректно оформлені рядки локалізації виділені синім кольором в тексті. При створенні нового проекту на основі файлу вихідного коду додатково відбувається створення порожніх рядків для заповнення в вікні для перекладу в кількості, що відповідає кількості коректно оформлених рядків в завантаженному коді.

Цей алгоритм створений для зручності пошуку та надання контексту в вихідному коді при перекладі тексту інтерфейсу програми. Для цього лише потрібно двічі натиснути на порядковий номер в списку – курсор в вікні коду буде переміщено до відповідного за номером рядку локалізації.

Інтерфейс програми забезпечує створення перекладу за допомогою таблиці з двома стовпчиками, які відповідають обраній мовній парі. Перший стовпчик містить текст для редагування, другий стовпчик містить оригінал тексту на іншій мові. Кожен рядок таблиці відповідає окремому рядку в масиві для локалізації, що знаходиться в завантаженному файлі вихідного коду програми.

Для початку роботи та створення перекладу потрібно ініціювати файл перекладу одним з трьох шляхів:

- створити новий проект на базі файлу вихідного коду;
- створити новий проект на базі існуючого файлу локалізації;
- відкрити два існуючі файли локалізації для редагування.

Після завантаження коректно оформлених файлів розробник може створювати та редагувати текст інтерфейсу для своєї програми. В перших двох випадках надаються порожні рядки для створення нового тексту, який потім зберігається у вигляді нового файлу локалізації.

Іншою важливою функцією утиліти є надання автоматизованого перекладу. Для цього потрібен завантажений існуючий файл локалізації в правому стовпчику таблиці. Цей файл використовується як джерело рядків на мові оригіналу в обраній мовній парі. Для початку автоматизованого перекладу розробник надає свій особистий ключ API для доступу до обраного сервісу перекладу (на даний момент реалізовано доступ до Google Translate), та обирає мовну пару в переліку доступних. Після цього треба лише обрати потрібні рядки для перекладу та натиснути відповідну кнопку.

Машинний переклад виконується за допомогою HTTP-запитів до онлайн-сервісу. Обрані рядки форматуються в необхідній формі та відсилаються у запиті. Для створення запиту до Google Translate необхідно використати нотацію JSON. Приклад інформації в запиті на переклад для N обраних рядків:

{“q”: [
"текст-рядку1", "текст-рядку2", "текст-рядку3", ... , “текст-рядкуN”
] }

Також в HTTP-запиті вказано індивідуальний API-ключ розробника, початкову та цільову мову перекладу, а також вказівка, що текст надається в базовому текстовому форматі. Таким чином можна перекласти до 128 рядків одразу (Google, 2024). При успішності запиту від онлайн-сервісу надходить HTTP-відповідь також у форматі JSON:

{ “data”: {
“translations”: [
{ “translatedText”: “переклад-тексту-рядка1” },
{ “translatedText”: “переклад-тексту-рядка2” },
{ “translatedText”: “переклад-тексту-рядка3” },
...
{ “translatedText”: “переклад-тексту-рядкаN” }]
} }

На рис. 2 показано як отриманий результат декодується і вставляється у відповідні запиту рядки лівого стовпчика таблиці. Також програма відслідковує кількість перекладених символів за поточну сесію.

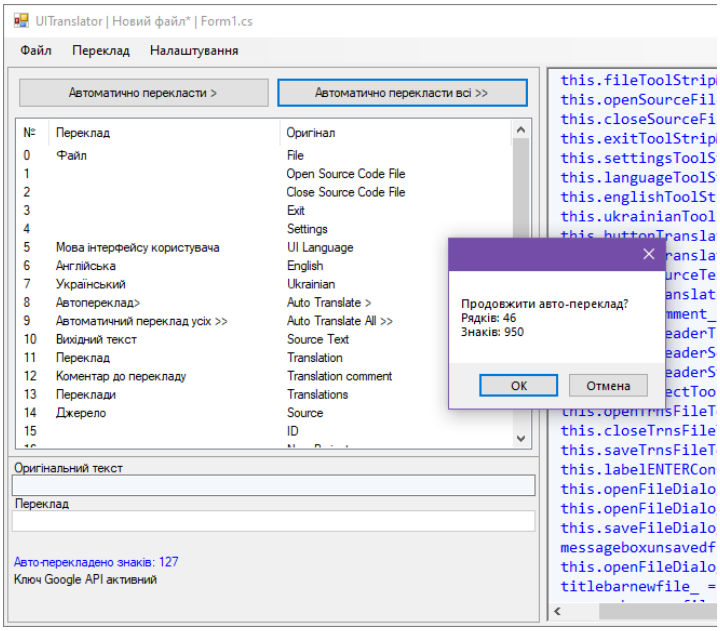


Рис. 2. Вікно програми під час використання машинного перекладу

Результатом роботи утиліти для перекладу є файл локалізації. Цей файл має розширення .uit та містить в собі рядки тексту та їх порядковий номер (починаючи з 0) в звичайному текстовому вигляді для легкої взаємодії з будь-якою програмою:

str0 “текст-рядку1”
str1 “текст-рядку2”
str2 “текст-рядку3”
...
strN “текст-рядкуN+1”

Один файл локалізації відповідає одній мові інтерфейсу користувача. Розробник має лише на свій розсуд налаштувати свою програму для коректного завантаження файлів локалізації в відповідні масиви рядків, які потім можна використовувати для зміни мови інтерфейсу в тому числі і під час роботи програми. Зразок оформлення коду в раніше наведеному форматі:

```
UIElement1.Text = UITranslator_language[0,0];
UIElement2.Text = UITranslator_language[0,1];
UIElement3.Text = UITranslator_language[0,2];
...
UIElementN+1.Text = UITranslator_language[0,N];
```

В цьому прикладі UITranslator_language є двовимірним масивом, що містить в собі всі доступні файли локалізації. Перший номер відповідає номеру мови, другий номер є порядковим номером рядку в файлі локалізації, N – загальна кількість рядків.

Як зазначено вище, результатом роботи розробленої утиліти є файли локалізації, що повністю готові до використання розробником в програмному проекті. До того ж, у порівнянні з алгоритмом gettext відсутня необхідність так би мовити початкової мови інтерфейсу, що незмінно наявна в вихідному коді програми. При використанні розробленої програми зменшується кількість необхідних етапів для отримання якісного інтерфейсу користувача на багатьох мовах світу, та незалежно від обраної мови програмування.

Висновки з даного дослідження та перспективи подальшого розвитку в даному напрямі. В даній статті описано алгоритми роботи першої початкової версію програми для відносно швидкого створення тексту інтерфейсів користувача перекладу. Головним завданням було спроектувати та створити програму, яка продемонструє весь необхідний для цього функціонал на належному рівні зручності.

Описана версія програми має простір для додавання нового фукнціоналу в подальшому. Запропонований формат масивів є лише одним з можливих варіантів оформлення. В подальших версіях утиліти розробник зможе сам обирати зручний синтаксис коду, за умови використання нумерованого переліку рядків. Також є потенціал для підтримки роботи з іншими сервісами машинного перекладу.

Список використаних джерел:

1. Міщенко А. Л. Лінгвістика фахових мов та сучасна модель науково-технічного перекладу : монографія. Вінниця : Нова Книга, 2013. 448 с

2. Hany Hassan, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, et al. Achieving Human Parity on Automatic Chinese to English News Translation. arXiv:1803.05567 [cs.CL] URL: https://arxiv.org/abs/1803.05567 (дата звернення: 30.03.2024).

3. gettext GNU Project Free Software Foundation (FSF). URL: https://www.gnu.org/software/gettext/ (дата звернення: 03.04.2024).

4. C# (GNU gettext utilities). URL: https://www.gnu.org/software/gettext/manual/html_node/C_0023.html (дата звернення: 03.04.2024).

5. How Many People Have Smartphones Worldwide. URL: https://www.bankmycell.com/blog/how-many-phones-are-in-the-world (дата звернення: 30.03.2024).

6. Poedit Translation Editor. URL: https://poedit.net/ (дата звернення: 03.04.2024).

7. Purchase Poedit Pro – Poedit. URL: https://poedit.net/purchase (дата звернення: 03.04.2024).

UDC 004.738.5:004.8:005.334:330.8:519.2
DOI https://doi.org/10.32689/maup.it.2024.1.10

Vasyl NESTEROV
Data Analyst, Itel inc, USA, Florida, Jacksonville, vasil.nesterov@gmail.com
ORCID: *0009-0000-3204-1382*

EXPLORING THE IMPACT OF BIG DATA ANALYTICS ON BUSINESS PERFORMANCE IN THE DIGITAL ERA

Abstract. *The corporate world is benefiting from the trends of BIG DATA (BD) and business modeling and analysis. Previous studies have demonstrated the enormous and exponential growth of data created in the modern world. These consist of the everyday inundation of unstructured and structured information in companies.*

Problem statement. *The main research gap addressed by previous literature studies is the lack of a comprehensive analysis of BD's application for digital transformation.*

Purpose of study. *This is filled by looking at the strategic benefits, opportunities, and challenges that BD presents to companies as they digitally transform their IT platforms. Therefore, the purpose of this study is to draw attention to the numerous uses and advantages of the technology of BD among researchers and companies. Methodology. Qualitative Research Methods, Utilizes qualitative research methods for a broad perspective. Emphasizes exploratory research to advance knowledge in the field. Uses an epistemological approach to find relevant literature sources from reputable databases like Google Scholar and Science Direct. Scientific novelty. Based on the research that is currently accessible, the article evaluates and discusses the latest trends, possibilities, and dangers of BD and how it has helped firms stay competitive by enabling them to develop successful business strategies. The assessment also covers the several uses for BD and analytics in business, as well as the data sources that are produced and their salient features.* ***Conclusion:*** *Lastly, the paper not only describes the difficulties in putting BD projects into practice successfully but also points up open research paths in BD analytics that need further attention. According to the BD topics under evaluation, effective administration and manipulation of massive data sets utilizing BD techniques and technologies may produce valuable business insights.*

Key words: *Big data, Digital transformation, Machine Learning, Artificial intelligence, Customer insight, Market trend, AI tools.*

Василь НЕСТЕРОВ. ДОСЛІДЖЕННЯ ВПЛИВУ АНАЛІТИКИ ВЕЛИКИХ ДАНИХ НА ЕФЕКТИВНІСТЬ БІЗНЕСУ В ЦИФРОВУ ЕПОХУ

Анотація. *Корпоративний світ отримує вигоду від тенденцій BIG DATA (BD) та бізнес-моделювання і аналізу. Попередні дослідження продемонстрували величезний і експоненціальний ріст даних, що створюються в сучасному світі. Вони складаються з щоденного потоку неструктурованої та структурованої інформації в компаніях. Постановка проблеми. Основною прогалиною в попередніх дослідженнях є відсутність комплексного аналізу застосування BD для цифрової трансформації.* ***Мета дослідження****. Заповнити цю прогалину шляхом аналізу стратегічних переваг, можливостей та викликів, які BD надає компаніям у процесі цифрової трансформації їхніх IT-платформ. Тому метою цього дослідження є привернення уваги дослідників та компаній до численних застосувань та переваг технологій BD.* ***Методологія.*** *Якісні методи дослідження, використовує якісні методи дослідження для широкої перспективи. Наголошує на пошукових дослідженнях для поглиблення знань у цій галузі. Використовується епістемологічний підхід для пошуку відповідних літературних джерел з авторитетних баз даних, таких як Google Scholar та Science Direct. Наукова новизна: На основі доступних на даний момент досліджень у статті оцінюються та обговорюються останні тенденції, можливості та небезпеки BD, а також те, як він допомагає фірмам залишатися конкурентоспроможними, дозволяючи їм розробляти успішні бізнес-стратегії. Оцінка також охоплює кілька сфер застосування бізнес-аналітики в бізнесі, а також джерела даних, які створюються, та їхні основні характеристики.* ***Висновок:*** *Насамкінець, стаття не лише описує труднощі в успішному впровадженні BD-проектів на практиці, але й вказує на відкриті дослідницькі шляхи в BD-аналітиці, які потребують подальшої уваги. Відповідно до розглянутих тем BD, ефективне адміністрування та маніпулювання великими масивами даних з використанням методів і технологій BD може дати цінні бізнес-інсайти.*

Ключові слова: *Великі дані, цифрова трансформація, машинне навчання, штучний інтелект, розуміння клієнтів, ринкові тенденції, інструменти штучного інтелекту.*

Introduction. Data processing, Artifical intelligent (AI) technologies are evolving quickly these days, and social media is becoming more and more significant. Data transmission security methods are becoming more and more important as technologies advance. Businesses are undergoing significant changes as a result of the extensive usage of information technology in many areas of life. It is important to note the increasing significance of information technology in corporate management. Businesses both create their solutions and apply outside ones. This is to guarantee the safety of data storage and transfer [19]. The growth of new fields of knowledge and the integration of technology for communication and information constitute the foundation of an enterprise's organizational process. The lines between the many industries and businesses nowadays are unclear. This is a result of the thinning of the lines separating the competencies of different organizations. New, very adaptable, and effective organizational solutions are built on the foundation of business networks, contemporary IT tools and databases, and, most importantly, creative individuals [18].

When large data sets are computationally examined to identify patterns, trends, and connections–particularly those pertaining to human behavior and interaction–they are referred to as "Big data (BD)". These massive information sets require state-of-the-art computational techniques for analysis. The phrase "BD analytics" has become more common in academic and professional contexts, including papers, journals, and conferences. In essence, it describes the enormous volume, variety, and speed with which data is produced and made accessible in the modern world. Large quantities, high-velocity, high-variety, and high-value information resources are referred to as BD, and to provide insights and support wise decision-making, new and economical information processing techniques must be used [21]. Despite the impressive achievements in BD analytics, many businesses find it difficult to implement these technologies due to high costs and other obstacles. Furthermore, there is scant experimental proof of the overall positive effects. Therefore, the study issue addressed in this article is if information-driven decision-making using BD Analytics leads to higher performance and a competitive edge in Pakistan's industrial industry. Effective business tales highlight how important it is to make deliberate decisions based on trustworthy information. High-quality data is required by the industrial sector to improve productivity and the effectiveness of business operations. Reliable data is essential to organizational decision-making processes because it converts inputs into knowledge that can be put into practice. This is how data-driven choices are made [19].

Problem statement. In an attempt to fully explore the potential of BD analytics, many companies have begun to invest large sums of money in the field. Nevertheless, the majority of research publications are general and do not provide industry-specific guidance on how businesses should change to take full advantage of these technological advancements.

BD is still a relatively new idea, and most of the studies that have been done on it have focused on its theoretical characteristics rather than its use in the age of digital transformation. The significant gaps hamper BD's strategic and commercial potential in our understanding of how it creates corporate value, despite the widespread excitement surrounding this technology.

The majority of research has advanced our knowledge of BD's supporting infrastructure, tools, and other resources, but it has seldom addressed BD's function in an organization's digital transformation.

Recent research and study. Managing bigger datasets becomes more challenging. The phrase "BD" describes databases that have grown to such an extent that they are difficult for traditional database management systems to handle (Almeida, Brás, Sargento, & Pinto, 2023). Moreover, the scope of BD surpasses the capacity of current data management, storage, and processing techniques. Three primary attributes of BD are volume, velocity, and diversity [2]

Three factors–volume, velocity, and variety–determine an organization's capacity for making well-informed judgments. Variety refers to the range of forms and types of data, volume represents the amount of the data, and velocity defines the pace at which it is changing. IBM presented Veracity, its fourth [19], Furthermore, according to some scholars, data significance is a fifth in the process of decision-making [27].

BDA uses sophisticated techniques to analyze BD sets. Larger datasets, however, also come with greater obstacles and problems. Improved decision-making, risk mitigation, and the finding of insightful information may all be facilitated by advanced analytics. Many academics have studied management decision-making extensively throughout the years, and it is very important. Simon's four stages of decision-making–intelligence, design, choice, and execution–are widely applied by decision-makers in many situations. Furthermore, there are several stages involved in the BD analysis pipeline, each with unique decision-making needs and obstacles [22, p.15].

These options include how to collect data, which data to gather, how to portray data once it has been extracted for analysis, and how to make decisions using the data that has been acquired. The report further highlights that adopting a data-driven method of decision-making necessitates modifying the organizational environment, management, HRM, and other methods of management. By putting these changes into practice, a business may improve its competitive position by fostering stronger customer interactions, reducing management risks, and increasing operational efficiency [13].

In strategic management, BD has come to be seen as an essential corporate asset for the success of organizations. To establish a framework that facilitates decision-making, boosts organizational effectiveness, and provides a sustained competitive advantage, BD must be combined with other assets and skills. An examination of the literature yielded the conclusion that, although some studies use comparable methodologies, other research investigations use a variety of both theoretical and practical methods for data collecting and refining. All organizations nowadays rely on information-driven decision-making. BD analytics provides helpful resources and insights to enhance traditional data mining techniques and decision-making algorithms [21].

The basic goal of every firm is to improve performance. Company strategic management is regarded to have one ultimate goal: to improve organizational performance. As a result, organizations have shifted their attention to this area. The variety of definitions, views, and measuring indicators provided suggests that experts are divided on what an organization's performance entails and how it will be judged. As a result, it can be challenging for

businesses to define, analyze, and assess performance [3]. This study presents a model for investigating the effects of business process adoption (BPA) and the role of mediator that business process performance (BPER) plays in the relationship between business process adoption and firm performance. It accomplishes this by building on principles from the resource-based viewpoint (RBV). The empirical study's findings, which are based on statistics collected from 204 moderately to high-level company executives throughout a variety of industries, show that the use of BA has a favorable influence on BPER. Furthermore, firm performance (FP) and BPER get together well. Furthermore, the results show that BPER mediates the connection between FP and BA adoption [6].

Purpose of the study. This study's primary goal is to investigate BD's importance in the age of digital transformation. While a number of academics have examined the BD idea from a technological standpoint, there hasn't been much research done on the topic from a management one. Furthermore, the use of BD in an organization's digital transformation to meet changing business needs has received little attention from scholars and practitioners.

This study aims to explore BD as a concept in general, with a focus on its strategic benefits, possibilities, and problems as well as the impact of Machine Learning (ML) and market trends as companies digitally modernize their infrastructures. Additionally, because BD is still a relatively new field, this research aims to close the current gap by identifying BD as a facilitator of organizational digital transformation. Encouraging managers who are either new to BD or looking to extend their horizons to have a deeper comprehension of it is another crucial goal of this research.

Research design. Given the novelty of BD, this study will use qualitative research methods to provide a broad perspective of the subject. Additionally, emphasizing exploratory research methods, this project will advance knowledge in the field. Most significantly, based on the abstract and introduction parts provided in each journal article, this research will consider secondary resources and understand crucial words and theoretical frameworks. Among these sources are news periodicals, corporate reports, and press releases. In addition, depending on the validity and dependability of the chosen papers, this study will assess the present status of the field's research.

The search parameters have been created using an epistemological approach, given the multidisciplinary nature of the study issue. Therefore, to find relevant literature sources from reputable databases like Google Scholar and Science Direct, a variety of keywords like “Big Data”, “digital transformation”, “importance of Big Data”, “impact of Big Data”. Therefore, to find relevant literature sources from reputable databases like Google Scholar and Science Direct, a variety of keywords like “BD” “ML business strategy”, “digital transformation”, “organizational efficiency”, “impact of Big Data”, “customer insights” and “Big Data and digital transformation” have been used.

Main analysis of study. BD’s effect on business operations BD's great potential has prompted major organizational reforms, particularly in the field of operations optimization. The following is a discussion of some of these [16]:

1. Improving production and operational efficiency:

Consider an efficient system that manufactures goods or services with pinpoint accuracy. BD assists businesses with this. Businesses get invaluable insights into their operations by leveraging massive amounts of structured and unstructured data. Regular monitoring and analysis immediately identify and address obstacles and inefficiencies.

As a result, it's reasonable to say that it empowers businesses to make based on evidence selections, allowing them to optimize operations and manage assets more accurately. Businesses can now focus on what matters: innovation and development, thanks to optimized manufacturing lines and automating boring labor.

2. Improving Data-Driven Management of Supply Chains

Every business's logistics network is its basis; it is a fragile system that can experience serious effects from even minor disruptions. BD analytics is a phenomenon responsible for such comprehensive perception of the whole supply chain, beginning from the purchase of raw materials till delivering finished products. Companies will be able to project disruptions and build flexible plans by observing supplier information, stock levels, transportation, and consumer trend on a regular basis. This enables retailers to balance the market and ensure a continuous flow of items; hence, the consumer is satisfied thus increasing their loyalty.

3. Better Cost Management should include listing of resources and their allocation in management.

For business people, managing funds is a problematic tightrope. BD relies on high-precision analysis as well as constant data input to avoid uncertainty. This thus gives organizations an insight into their habit of spending and allows them to detect where they have merely wasted their money.

Implementing resource allocation optimization enables an organization to reduce expenses and result in higher returns on investments at the same time. Firms with those new capabilities can choose to direct their money to areas that stimulate creativity, technology, and customers’ joy as opposed to less functional areas.

The fundamentals of management in an age of emerging technologies. For production companies, for instance, implementing BD Analytics projects entails several steps, including defining the business problem, determining the extent of the data, assembling a cross-functional team, creating a schedule for each task, gathering and choosing data, analyzing and modeling it, visualizing it, producing a report, integrating it into information systems, and providing specialized training [4].

BD management, or BDA, offers a chance to change workforce and business methods. By analogy, it is feasible to analyze the method of automation and its impact on the manufacturing workforce thanks to the insights obtained from BD Analytics, way it was before the production operations were mechanized, many years ago. Using predictive analytics, machine learning, or methods similar to MapReduce [1], BD analytics provides fast and reliable insights to help improve production decisions. Public databases facilitate the creation of new uses for data resources by linking businesses or specific systems inside an organization. Similar to how the Internet of Things is creating data, employees are also becoming "data generators" as they may generate data both internally and outside through the use of IP addresses and different kinds of sensors. Machine interpretation of data is made possible by more complex software, which allows for a more thorough integration of applications based on BD with conventional value-creation processes and largely autonomous decision-making. Many sectors are facing challenges to their business models due to digitalization and BD Analytics [6].

Some businesses, even those with a dominant market position, may struggle to modify their operations in response to the changing circumstances and fail to fully capitalize on the potential presented by the process of digitization and BD Analytics. The continuous digitization trend lowers transaction costs related to control, communication, and information collection by a large margin. Companies can analyze the interrelated nature of purchasing behavior to better suit advertising material, for example, thanks to easier access to an updated pool of data and powerful BD Analytics. This might reflect in greater total customer demand [5].

As a result, over time, less effective business models may be replaced by gradually improving current company models through greater digitization and data analytics. Deployed and standardized BD solutions could not be sufficient to provide a long-term competitive edge, nevertheless, as standards rise. Organizations may obtain valuable insights not just from publicly accessible online datasets but also from privately acquired data by utilizing analytical tools that examine both structured and unstructured data. Connecting data from websites, product rating sites, and social network data with consumer choices and product features gives businesses a wealth of opportunities to comprehend customer needs, anticipate their requirements, and, most importantly, maximize resource utilization [6].

BD ML Applications' Challenges. The following are general ML challenges [12, 13]: (i) creating flexible and scalable computational architectures; (ii) comprehending data properties before utilizing ML tools and algorithms; and (iii) being able to build, learn, and predict as you increase sample size, dimension, and label categories. Many significant specialized subfields of large-scale machine learning, including large-scale recommendation systems, natural language processing, rule-based association learning, and ensemble learning, continue to struggle with scaling issues despite the availability of numerous large-scale ML algorithms [4].

A crucial component of ML is absent from the fundamental MapReduce architecture that is frequently offered by first-generation “BD analytics” platforms like Hadoop. Iteration, recursion, and other essential properties needed to effectively iterate “around” a MapReduce program are not supported by MapReduce. On these platforms, programmers creating MLmodels must implement looping in non-standard ways that are not part of the standard MapReduce architecture. The recent creation of several specific techniques or libraries to enable iterative programming on big clusters has been spurred by this lack of support. In the meantime, an iteration failure in MapReduce is the direct target of newer MapReduce extensions like HaLoop, Twister, and PrItr [14].

The following are the main reasons ML approaches are not appropriate for handling BD classification problems [20]: (i) An ML method trained on a specifically labeled dataset may not be appropriate for another dataset; (ii) an ML method is typically trained using a certain number of class types, which means a large variety of class kinds discovered in a dynamically growing dataset will lead to insufficient classification results; and (iii) an ML method develops based on a single learning task, making them unsuitable for the multiple training tasks and knowledge transfer requirements of BD analysis that are present today.

Technological Development of BD ML Applications. The majority of scalable ML advancements (such Madlib, Apache Mahout, etc.) take place in the field of massively parallel database processing. ML algorithms with scalable predictive functions may be designed and implemented to enable better work in the BD age. The following techniques have been investigated and assessed [9].

(i) Progressive enhancement neural networks in associative memory architectures that can easily adapt to new datasets and sources;

(ii) facets developing that can learn a hierarchical arrangement in the data;

(iv) multi-task learning that can learn multiple predictive functions in parallel;

(iii) deep learning techniques that automate the method of feature engineering by learning to generate and sift through data-driven features. BD's vast and expanding data domain necessitates the employment of the multi-domain representation-learning (MDRL) approach for categorization.

The distance-metric learning, feature extraction, and feature variable learning components of the MDRL approach are all included. Many representation-learning techniques have been put forth in machine learning.

In addition to the recommended network model, the cross-domain, representation-learning (CDRL) approach may be appropriate for BD categorization [15]. Deep learning is a particularly helpful tool for BD analytics because it can analyze and learn from vast volumes of unstructured data, which is one of its main advantages. It was investigated how deep learning may be used to BD analytics, specifically in relation to simplifying discriminative tasks, quick information retrieval, semantic indexing, data tagging, and the extraction of complicated patterns from large amounts of data. Additional research was conducted on deep learning in BD, encompassing data streaming, high-dimensional data, distributed computing, and the scalability of Deep Learning models [27].

The Bayesian Network (BN) is a prominent ML approach commonly used to describe probabilistic correlations between variables. A novel weight-based ensemble technique was presented to train a BN architecture from an ensemble of local outcomes; an intelligent BD initial processing technique and a data quality score have been suggested to test and assure the data quality and data fidelity. The whole learning process was built using the Kepler scientific workflow, which made it simple to integrate the algorithm with data-parallelism distributed (DDP) engines like Hadoop. It was also shown how Kepler may help with the development and operation of the BD BN learning application [28].

Machine learning, cloud computing, and workflow methodologies were combined to create a Scalable Bayesian Network Learning (SBNL) workflow. The technique makes use of distributed computing models and ensemble learning to enable efficient BN learning from BD, as well as intelligent pre-processing of BD [26].

Through HBase and the Hadoop Distributed File System (HDFS), the architecture offers dependable permanent data storage. The modules for batch and stream processing make up the architecture's core. It offers MLtools and algorithms that developers may use with ease to do tasks like classification, recommendation, clustering, and prediction, among others [15].

Scalable Advanced Massive Online Analysis (SAMOA), an open-source platform for large data stream mining, has the technique accessible. Adaptive Model Rules (AMRules) are distributed throughout a cluster using a combination of vertical and horizontal parallelism. AMRules creates understandable representations of decision rules. Developing novel distributed ML algorithms and implementing them on top of cutting-edge distributed stream processing engines (DSPEs) is made easier by SAMOA. Additionally, it provides a library of distributed ML algorithms that anyone may use or alter [2]

This study [19] paper looks at how Russian aggression has affected Ukraine's cyberspace and suggests ways to uninstall infected malware from digital equipment. Taking into account the hostile acts of the aggressor state, the research attempts to identify sensitive sectors and specify the vector of development of digital technologies that may be utilized securely in Ukraine. The study makes use of information research, statistical research, and analytic definition to pinpoint areas of Ukraine's cyberspace that urgently need assistance to liberate digital tools from corrupted software belonging to the aggressor state.

The Risks and Concerns of Russian Influence. The report emphasizes how vital it is to rid Ukraine's cyberspace of the parasitic effects of digital tools created by Russian businesses. It highlights issues that pose a danger to state security and offers workable suggestions for guaranteeing state digital security as well as the future growth of Ukraine's digital sector. The urgent necessity to replace Russian software in the organizational and managerial domains of Ukraine's cyber-digital infrastructure is also emphasized in the study. The growth of the Ukrainian IT cluster and Ukraine's potential as a digital state with high levels of digital means integration are also covered.

Recommendations for Protecting Cyberspace in Ukraine. According to the study the Ukrainian authorities still depend on Russian soft-ware so it will be necessary to support national organic development of the digital tools if the country will be able to clear the reliance on the third party. The document issues its staging specifications based on the declared study goals, such as the immediate isolation and elimination of Russian software and services, and taking decisive efforts for fueling domestic IT industry growth. Also, it advises to form Government assisting programs and to let domestic computer engineering researches to take place for the purpose to provide financial support and to increase adequacy of the cyber-digital area in Ukraine.

However, the main idea of the final study is how we can derive profit from AI in the digital age. Companies must be ready to question the common approach and take an adventure into non-explored zones to position itself ahead of the trend and adjust to the dynamic environment as AI provides new opportunities. Consequently, businesses are capable to get a notable competitive advantage which can hardly be copied and offer a wide array of opportunities to create value [7].

Conclusion. AI has been promoted as the game-changing tech issue that could transform the functions and management of the organization. This research will examine the ways in which AI is especially adapted into the IT & business strategies of an organization that makes the company's objectives and plan to thrive in the digital world. The research indicated that creative and routine AI integrations collude with each other to surpass the amount of efficaciousness of solo operations. The study also pointed out that strategic business/IT integrations are crucial in bridging the digital transformation gaps.

The key finding of the study is that AI functions as an engine that can enforce serious transformations within organizations regardless of whether it should be seen as a tool or not. The organizational realm has to be ready for challenges of AI and understand the ways in which it could be implemented to overcome hurdles and develop entirely new value. The ability to apply knowledge and explore – meaning more than theoretical thinking – is thus also required**.**

This is the reason, that employing AI by companies into their business strategy is exactly like changing lead into gold, since in this case companies turn technology and data into new forms of competitive advantage and value. The process of AI adoption requires a firm grasp of its principles but also the ability to experiment zealously and in an agile manner.

Bibliography:

1. A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects / A. E. Ezugwu et al. *Engineering Applications of Artificial Intelligence*. 2022. Vol. 110. P. 104743. URL: https://doi.org/10.1016/j.engappai.2022.104743 (date of access: 28.03.2024).

2. Alghanmi N., Alotaibi R., Buhari S. M. HLMCC: A Hybrid Learning Anomaly Detection Model for Unlabeled Data in Internet of Things. *IEEE Access*. 2019. Vol. 7. P. 179492–179504. URL: https://doi.org/10.1109/access.2019.2959739 (date of access: 28.03.2024).

3. Time series big data: a survey on data stream frameworks, analysis and algorithms / A. Almeida et al. *Journal of Big Data*. 2023. Vol. 10, no. 1. URL: https://doi.org/10.1186/s40537-023-00760-1 (date of access: 26.04.2024).

4. The mediating role of supply chain management on the relationship between big data and supply chain performance using SCOR model / R. O. K. Alshawabkeh et al. *Uncertain Supply Chain Management*. 2022. Vol. 10, no. 3. P. 729–736. URL: https://doi.org/10.5267/j.uscm.2022.5.002 (date of access: 26.04.2024).

5. Aydiner A. S., Bayraktar E. Business Analytics and Firm Performance: The Mediating Role of Business Process Performance. *Academy of Management Proceedings*. 2018. Vol. 2018, no. 1. P. 18111. URL: https://doi.org/10.5465/ambpp.2018.18111abstract (date of access: 28.03.2024).

6. Bannikov V., Zalialetdzinau K., Siasiev A., Ivanenko R., Saveliev D. Computer science trends and innovations in computer engineering against the backdrop of Russian armed aggression. International Journal of Computer Science and Network Security.2022. Vol.22, no.9. P. 465-470. URL: https://doi.org/10.22937/IJCSNS.2022.22.9.60 (date of access: 28.03.2024)

7. Exploring the Impact of Big Data Analytics on Organizational Decision-Making and Performance: Insights from Pakistan's Industrial Sector / A. Latif et al. *Pakistan Journal of Humanities and Social Sciences*. 2023. Vol. 11, no. 2. URL: https://doi.org/10.52131/pjhss.2023.1102.0475 (date of access: 28.03.2024).

8. Ezugwu, A. E., Ikotun, A. M., Oyelade, O. O., Abualigah, L., Agushaka, J. O., Eke, C. I., & Akinyelu, A. A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects. *Engineering Applications of Artificial Intelligence.* 2022. Vol. 110. P. 104743. URL: https://doi.org/10.1007/s10462-022-10325-y (date of access: 28.03.2024)

9. How Big Data Analytics Boosts Organizational Performance: The Mediating Role of the Sustainable Product Development / S. Ali et al. *Journal of Open Innovation: Technology, Market, and Complexity*. 2020. Vol. 6, no. 4. P. 190. URL: https://doi.org/10.3390/joitmc6040190 (date of access: 26.04.2024).

10. How Big Data Analytics Boosts Organizational Performance: The Mediating Role of the Sustainable Product Development / S. Ali et al. *Journal of Open Innovation: Technology, Market, and Complexity*. 2020. Vol. 6, no. 4. P. 190. URL: https://doi.org/10.3390/joitmc6040190 (date of access: 28.03.2024).

11. Injadat, M., Moubayed, A., Nassif, A.B. and Shami, A. Machine learning towards intelligent systems: applications, challenges, and opportunities. *Artificial Intelligence Review*, 2021. Vol. 54, no. 5. P.3299–3348. URL: https://doi.org/10.48550/arXiv.2101.03655 (date of access: 28.04.2024)

12. Internet of Things in arable farming: Implementation, applications, challenges and potential / A. Villa-Henriksen et al. *Biosystems Engineering*. 2020. Vol. 191. P. 60–84. URL: https://doi.org/10.1016/j.biosystemseng.2019.12.013 (date of access: 28.03.2024).

13. Exploring the Impact of Big Data Analytics on Organizational Decision-Making and Performance: Insights from Pakistan's Industrial Sector / A. Latif et al. *Pakistan Journal of Humanities and Social Sciences*. 2023. Vol. 11, no. 2. URL: https://doi.org/10.52131/pjhss.2023.1102.0475 (date of access: 26.04.2024).

14. Lee I., Shin Y. J. Machine learning for enterprises: Applications, algorithm selection, and challenges. *Business Horizons*. 2020. Vol. 63, no. 2. P. 157–170. URL: https://doi.org/10.1016/j.bushor.2019.10.005 (date of access: 28.03.2024).

15. Machine learning and data analytics for the IoT / E. Adi et al. *Neural Computing and Applications*. 2020. Vol. 32, no. 20. P. 16205–16233. URL: https://doi.org/10.1007/s00521-020-04874-y (date of access: 28.03.2024).

16. Organizational business intelligence and decision making using big data analytics / Y. Niu et al. *Information Processing & Management*. 2021. Vol. 58, no. 6. P. 102725. URL: https://doi.org/10.1016/j.ipm.2021.102725 (date of access: 26.04.2024).

17. Organizational business intelligence and decision making using big data analytics / Y. Niu et al. *Information Processing & Management*. 2021. Vol. 58, no. 6. P. 102725. URL: https://doi.org/10.1016/j.ipm.2021.102725 (date of access: 28.03.2024).

18. Pizło W., Parzonko A. Virtual Organizations and Trust. *Trust, Organizations and the Digital Economy*. New York, 2021. P. 61–78. URL: https://doi.org/10.4324/9781003165965-6 (date of access: 28.03.2024).

19. Prabhakaran V., Kulandasamy A. Integration of recurrent convolutional neural network and optimal encryption scheme for intrusion detection with secure data storage in the cloud. *Computational Intelligence*. 2020. URL: https://doi.org/10.1111/coin.12408 (date of access: 28.03.2024).

20. The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities / M. M. Rathore et al. *IEEE Access*. 2021. Vol. 9. P. 32030–32052. URL: https://doi.org/10.1109/access.2021.3060863 (date of access: 26.04.2024).

21. Sbai I., Krichen S. A real-time Decision Support System for Big Data Analytic: A case of Dynamic Vehicle Routing Problems. *Procedia Computer Science*. 2020. Vol. 176. P. 938–947. URL: https://doi.org/10.1016/j.procs.2020.09.089 (date of access: 28.03.2024).

22. Schmidt S., von der Oelsnitz D. Innovative business development: identifying and supporting future radical innovators. *Leadership, Education, Personality: An Interdisciplinary Journal*. 2020. Vol. 2, no. 1. P. 9–21. URL: https://doi.org/10.1365/s42681-020-00008-z (date of access: 28.03.2024).

23. The mediating role of supply chain management on the relationship between big data and supply chain performance using SCOR model / R. O. K. Alshawabkeh et al. *Uncertain Supply Chain Management*. 2022. Vol. 10, no. 3. P. 729–736. URL: https://doi.org/10.5267/j.uscm.2022.5.002 (date of access: 28.03.2024).

24. The Role of AI, Machine Learning, and Big Data in Digital Twinning: A Systematic Literature Review, Challenges, and Opportunities / M. M. Rathore et al. *IEEE Access*. 2021. Vol. 9. P. 32030–32052. URL: https://doi.org/10.1109/access.2021.3060863 (date of access: 28.03.2024).

25. Time series big data: a survey on data stream frameworks, analysis and algorithms / A. Almeida et al. *Journal of Big Data*. 2023. Vol. 10, no. 1. URL: https://doi.org/10.1186/s40537-023-00760-1 (date of access: 28.03.2024).

26. Internet of Things in arable farming: Implementation, applications, challenges and potential / A. Villa-Henriksen et al. *Biosystems Engineering*. 2020. Vol. 191. P. 60–84. URL: https://doi.org/10.1016/j.biosystemseng.2019.12.013 (date of access: 26.04.2024).

27. Visvizi A., Troisi O., Grimaldi M. Mapping and Conceptualizing Big Data and Its Value Across Issues and Domains. *Big Data and Decision-Making: Applications and Uses in the Public and Private Sector*. 2023. P. 15–25. URL: https://doi.org/10.1108/978-1-80382-551-920231002 (date of access: 28.03.2024).

28. Wadoux A. M. J. C., Minasny B., McBratney A. B. Machine learning for digital soil mapping: Applications, challenges and suggested solutions. *Earth-Science Reviews*. 2020. Vol. 210. P. 103359. URL: https://doi.org/10.1016/j.earscirev.2020.103359 (date of access: 28.03.2024).

УДК 004.75
DOI https://doi.org/10.32689/maup.it.2024.1.11

Юлія ПАРФЕНЕНКО
кандидат технічних наук, доцент, доцент кафедри інформаційних технологій, Сумський державний університет, yuliya_p@cs.sumdu.edu.ua
***ORCID:** 0000-0003-4377-5132*

Володимир НАГОРНИЙ
кандидат технічних наук, доцент, доцент кафедри інформаційних технологій, Сумський державний університет, v.nahornyi@cs.sumdu.edu.ua
***ORCID:** 0000-0001-5223-7219*

Роман ДАНИЛЕНКО
магістр комп'ютерних наук, Сумський державний університет, danylenko@outlook.com
***ORCID:** 0009-0007-0683-5833*

РОЗРОБЛЕННЯ МОБІЛЬНОГО ДОДАТКУ ПІДТРИМКИ НАДАННЯ ПОСЛУГ ВІД ЕНЕРГЕТИЧНИХ МІКРОМЕРЕЖ

***Анотація. Метою** роботи є розроблення мобільного додатку для інформування клієнтів енергетичних мереж про стан роботи мереж, а також про прогнозовані обсяги генерації електричної енергії від відновлюваних джерел. **Методологія.** Для вивчення актуальності використання мобільних додатків для отримання доступу до інформації про поточний та прогнозований стан енергетичних мереж застосовано аналітичний метод, що включає в себе пошук та аналіз відповідної наукової літератури, дослідження існуючих мобільних додатків, а також встановлення функціональних вимог до мобільного додатку. Методи структурно-функціонального моделювання використано для представлення системи у вигляді функцій, які пов'язані між собою та функцій, що перетворюють вхідні дані у вихідні. Метод проектування використано при розробленні UML-діаграми варіантів використання. Для практичної реалізації мобільного додатку обрано методику з використанням принципу Clean Architecture.*

***Результати.** Проведено огляд сучасних тенденцій розроблення мобільних додатків, у тому числі для моніторингу та управління енергетичними мережами. Виконано моделювання варіантів використання мобільного додатку, а також моделювання послідовності дій основних акторів по роботі з мобільним додатком засобами мови UML. Мобільний додаток розроблений під операційну систему Android для роботи з даними енергетичних мереж, які зберігаються в системі керування базами даних SQLite. Спроектовано та реалізовано екрани мобільного додатку – авторизації, вибору локацій та головний екран, на якому відображається інформація про енергетичну мікромережу на обраній локації. Протестовано роботу мобільного додатку для підтримки надання послуг від електричних мікромереж клієнту з двома мікромережами, встановленими на різних локаціях. **Наукова новизна** роботи полягає у тому, що розроблений мобільний додаток має таку архітектуру, яка дозволяє інтегруватися в систему підтримки прийняття рішень при управлінні енергетичними мікромережами і через API інтерфейс відображати дані з єдиної бази даних інформаційної системи.*

***Висновок.** У роботі представлено розроблення мобільного додатку для моніторингу користувачем поточного стану енергетичної мікромережі, а також інформування про прогнозні обсяги генерації електричної енергії.*

***Ключові слова:** мобільний додаток, енергетичні мікромережі, інтерфейс, функціональний стан, прогнозування.*

Yuliia PARFENENKO, Volodymyr NAHORNYI, Roman DANYLENKO. DEVELOPMENT OF A MOBILE APPLICATION TO SUPPORT THE PROVISION OF ENERGY MICROGRID SERVICES

***Absctract. The purpose of the work** is to develop a mobile application for informing customers of energy networks about the state of operation of the networks, as well as about the forecasted volumes of electricity generation from renewable sources. **Methodology.** To study the relevance of using mobile applications to gain access to information about the current and projected state of energy networks, an analytical method was applied, which includes the search and analysis of relevant scientific literature, the study of existing mobile applications, as well as the establishment of functional requirements for the mobile application. The methods of structural-functional modeling are used to represent the system in the form of functions that are related to each other and functions that transform input data into output. The design method was used in the development of UML use case diagrams. For the practical implementation of the mobile application, a methodology using the principle of Clean Architecture was chosen.*

***The results.** An overview of modern trends in the development of mobile applications, including for monitoring and managing energy networks, was conducted. Modeling of the mobile application use case was carried out, as well as modeling of the sequence of actions of the main actors in working with the mobile application using the UML language. The mobile application is developed for the Android operating system to work with energy network data stored in the SQLite database management system. The screens of the mobile application - authorization, location selection, and the main screen, which displays information about the energy microgrid at the selected location, were designed and implemented. The operation of the mobile application to support the provision of services from electric microgrids to a client with two microgrids installed in different locations has been tested. **The scientific novelty** lies in the fact that the developed mobile application has such an architecture that allows integration into the decision support system for managing energy microgrids and displaying data from a single database of the information system through the API interface.*

***Conclusion.** The work presents the development of a mobile application for user monitoring of the current state of the energy microgrid, as well as informing about the forecast volumes of electricity generation.*

***Key words:** mobile application, energy microgrids, interface, functional state, forecasting.*

Постановка проблеми. Зростання споживання електричної енергії, підвищення її вартості та потреба обмеження використання вуглеводневих видів палива спонукають до переходу на відновлювальні джерела енергії. Водночас відзначається швидкий розвиток технічних, ринкових та економічних змін у галузі електроенергетики. Впровадження нових підходів до виробництва електроенергії та її розподілу, які спрямовуються на зменшення використання викопних видів палива та задоволення зростаючого попиту на електроенергію, є необхідністю.

Суттєвий вплив на сферу електроенергетики має цифрова трансформація, яка полягає у впровадженні сучасних передових технологій. Цифрова трансформація енергетичного сектора являє собою інтеграцію найсучасніших технологій, спрямованих на автоматизацію та покращення ефективності процесів управління енергією. Поширення набувають мережі «smart grid», які забезпечують інтелектуальне управління енергосистемою режимі реального часу [1]. Змінюється також характер споживання та генерації електроенергії – відбувається перехід від класичного централізованого до більш автономного розподіленого, через впровадження мікромереж.

Збільшується частка компаній, які надають послуги із встановлення систем відновлюваної енергетики для малого бізнесу та приватних домогосподарств. Змінюється парадигми виробництва електроенергії з централізованого на децентралізоване, що потребує застосування новітніх програмних рішень для управління енергоспоживанням та функціональної оптимізації енергосистеми [2]. Необхідним є впровадження систем управління енергетичними мікромережами для забезпечення балансу між споживанням та виробництвом електричної енергії, а також систем підтримки прийняття рішень [3, 4].

Користувачі енергетичних мікромереж матимуть до них довіру за умови повного і своєчасного одержання інформації про їх стан, ефективність управління, а також прогнозування функціонального стану мікромереж на майбутні періоди часу, що можуть забезпечити зручні та зрозумілі інтерфейси подання даних [5]

Для моніторингу поточного стану роботи енергетичних мікромереж та відслідковування прогнозного стану необхідним є розроблення програмного додатку, який буде забезпечувати такий сервіс на мобільних пристроях.

Аналіз останніх досліджень і публікацій.

Розроблення мобільних додатків для моніторингу та енергетичного менеджементу «розумних» енергетичних мереж представлене в роботах [6-8]. Досліджуються переваги використання підключення по мережі Wifi для передачі даних функціонування енергетичних мереж та розроблення мобільного інтерфейсу систем енергетичного менеджменту. У статті [8] наведено опис архітектури системи енергетичного менеджементу на стороні споживача вигляді, яка складається у тому числі й з мобільного додатку.

Нижче наведено опис комерційних мобільних додатків, які можуть бути використані для підтримки надання послуг від енергетичних мікромереж. Програмний додаток Enphase Enlighten від компанії Enphase Energy [9], який має у тому числі й мобільний інтерфейс, пропонує широкий спектр функцій, які дозволяють споживачам електроенергії моніторити їхнє споживання електроенергії та вироблення електроенергії з фотоелектричних модулів компанії Enphase.

Програмний продукт μGrid Manager від компанії GridWhiz Thailand – мобільний додаток, який дозволяє операторам керувати їхньою мікромережею [10]. Цей додаток дозволяє користувачам отримувати інформацію про поточний стан мікромережі, включаючи споживання електроенергії, виробництво електроенергії та баланс потужностей, надає користувачам інформацію про стан їхньої мікромережі за певний період часу, наприклад, за день, тиждень, місяць або рік, надсилає повідомлення користувачам про можливі проблеми з їхньою мікромережею, наприклад, про перевантаження мережі або про відключення електроенергії.

Мобільний додаток mySunPower від компанії SunPower дозволяє власникам систем сонячної енергії SunPower моніторити їхню систему та керувати нею [11]. Мобільний додаток SMA Energy від компанії SMA Solar Technology AG дозволяє споживачам електроенергії, які використовують інвертори SMA, моніторити їхнє споживання електроенергії, а також вироблення електроенергії з фотоелектричних модулів [12].

Проведений огляд дозволив виділити функціональні вимоги до мобільних додатків підтримки надання енергетичних послуг. Розглянуті додатки здебільшого є платними та орієнтованими на використання для конкретних постачальників енергетичних послуг, тобто не є універсальними, що обмежує їх використання підприємствами малого бізнесу та приватними домогосподарствами. Окрім цього жодний з досліджених мобільних додатків не надає інформації щодо прогнозних показників стану мікромереж.

Постановка завдання. Метою роботи є розроблення мобільного додатку для надання користувачам енергетичних мікромереж з відновлювальними джерелами енергії доступу до інформації про

поточний та прогнозний стан їх функціонування. Мобільний додаток має відображати дані з бази даних системи підтримки прийняття рішень при управлінні енергетичними мікромережами. Доступ до даних має здійснюватися через API-сервіс.

Виклад основного матеріалу. Мобільний додаток розроблено мовою Kotlin під операційну систему Android за принципом Clean Architecture, за яким програмний код поділяється на шари, кожен з яких має свою власну відповідальність. Роботу з API-інтерфейсами реалізовано за допомогою Retrofit2 та OkHttp3. Взаємодія з локальною базою даних SQLite реалізована через RoomDB, що використовує ефективні алгоритми для доступу до бази даних, а також прості та зрозумілі анотації для опису моделей і взаємодії з базою даних.

Перед розробкою проведено моделювання сценарії використання мобільного додатку, а також послідовності дій основних акторів побудовано відповідні діаграми в нотації UML. Діаграму варіантів використання зображено на рис.1.

Акторами є зареєстрований користувач, в якого є доступ до перегляду даних енергетичної мікромережі, програмний API-інтерфейс сервісу мікромереж, який надає дані про функціональний стан мікромережі, а також програмний API-інтерфейс сервісу погоди, який надає дані прогнозу погоди та поточної погоди на місцевості, де встановлена мікромережа.

Варіанти використання:

– Авторизація – дозволяє користувачу авторизуватися в мобільному додатку;

– Отримання списку мікромереж – відображає користувачу мікромережі, до перегляду даних про які в нього є доступ;

– Отримання даних про структуру – надає користувачу дані про структуру мікромережі;

– Отримання даних про мікромережі – дозволяє користувачу переглядати поточний статус генерації електроенергії та її витрати, а також інформацію щодо накопиченої електроенергії в акумуляторах;

– Отримання даних про погоду – дозволяє користувачу переглядати поточну погоду та її прогноз.

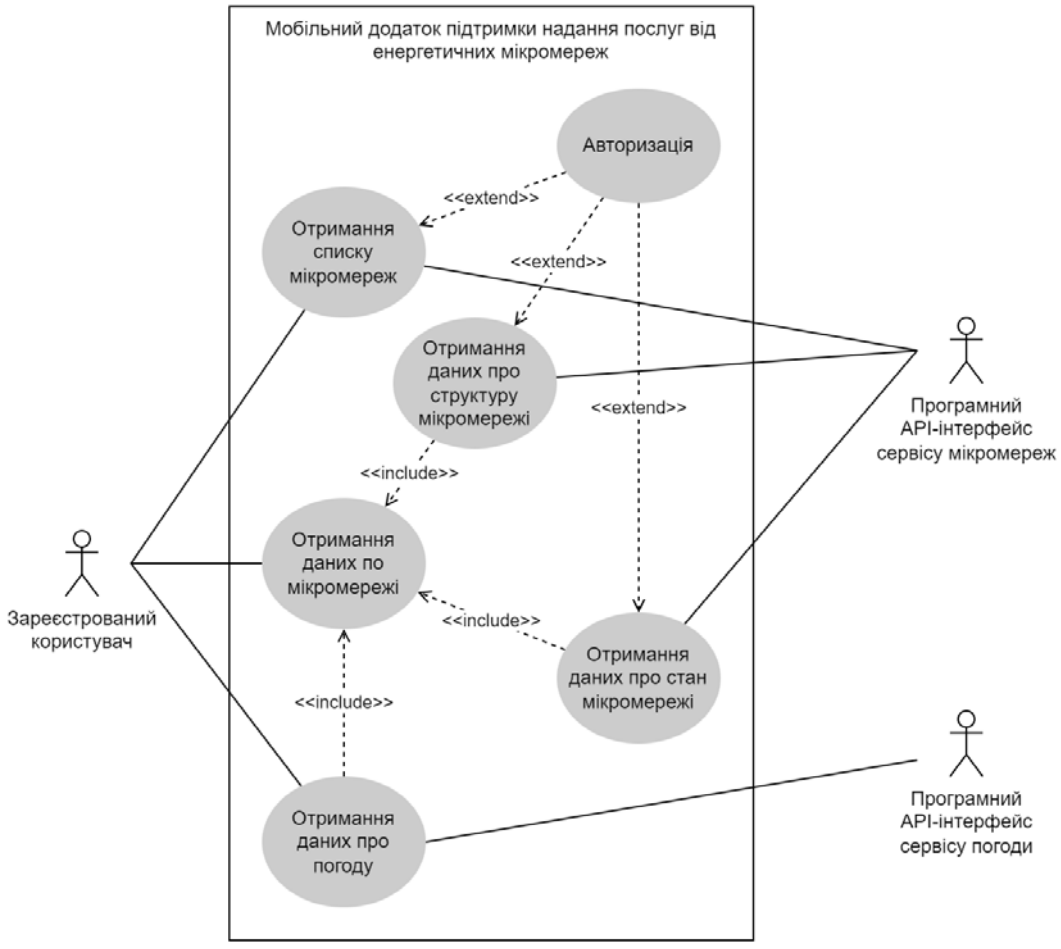


Рис. 1. Діаграма варіантів використання мобільного додатку

Діаграму послідовності дій користувача при роботі з мобільним додатком показано на рис.2.

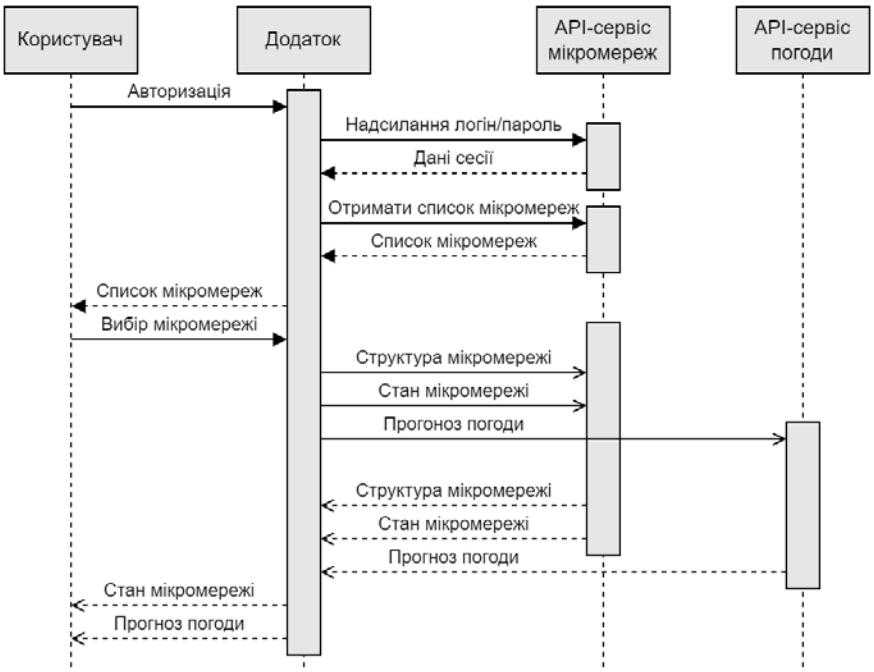


Рис. 2. Діаграма послідовності

Для перегляду інформації в інтерфейсі мобільного додатку користувачу спершу потрібно авторизуватися, після чого обрати зі списку мікромережу, дані про яку він бажає переглянути.

При розробленні мобільного додатку було застосовано архітектуру «Clean Architecture» для розмежування відповідальності між різними частинами додатку. Це досягається шляхом розбиття коду на кілька шарів, кожен з яких має свою власну відповідальність і не залежить від інших шарів. Таки підхід дозволяє зробити додаток більш гнучким, масштабованим та зручним для тестування.

Мобільний додаток розподілений на такі шари:

– презентація (Presentation) – відповідає за відображення даних так отримання команд від користувача.

– бізнес-логіка (Domain) – відповідає трансформацію даних і має бути незалежний від користувацького інтерфейсу.

– шар даних (Data) – відповідає за доступ до даних як з локальної бази даних, так і з API-інтерфейсів.

Дані для відображення у мобільному інтерфейсі беруться через API інтерфейс з бази даних SQLite системи підтримки прийняття рішень при управлінні енергетичними мікромережами. Для роботи з базою даних обрано бібліотеку RoomDB, яка автоматично створює таблиці на основі описів сутностей, виконує всі операції з базою даних у фоновому потоці та має детальну документацію.

Для взаємодії з REST API при розробці було використано бібліотеку Retrofit2, яка автоматично генерує код для виконання запитів до API на основі інтерфейсів з анотаціями, має механізми обробки помилок та може виконувати запроси в фонових потоках. Бібліотека OkHttp3 використовується бібліотекою Retrofit2 і потрібна для доступу до керування заголовками запитів (наприклад, для додавання даних про авторизацію) та для можливості мати доступ до детальної інформації по запитам і відповідям до web-API під час розробки додатку.

У мобільному додатку класи даних розділені на три групи: моделі даних для отримання даних з web-API (DTO-моделі), моделі даних для зберігання даних в локальній базі даних (Entity-моделі) та моделі даних для виводу інформації користувачу (Domain-моделі). Для інтеграції шаблону проектування впровадження залежностей (Dependency Injection, DI) використано бібліотеку мови програмування Kotlin – Koin. Для роботи з запитами та відповідями web-API використано формат даних JSON і для зручної конвертації даних між JSON та внутрішніми класами додатку використано бібліотеку Gson.

Інтерфейс мобільного додатку складається з трьох екранів, а саме:

– екран авторизації, на якому користувач вводить свій логін та пароль і надсилає запит на авторизацію для того, щоб перейти на наступний екран;

– екран вибору локацій, на якому користувач має можливість перемикатися між своїми мікромережами, для отримання детальної інформації по ним;

– головний екран, на якому буде виводиться детальна інформація по вибраній енергетичній мікромережі – її поточний та прогнозований стан роботи, а також інформація про прогноз погоди.

Перший екран авторизації, на якому користувач має можливість ввести свій логін та пароль для авторизації та переходу на наступний екран, показано на рис. 3. У випадку невірно введеного логіна чи пароля на екрані має відображатися помилка.

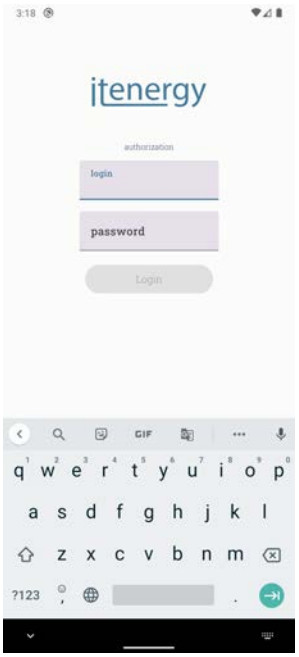


Рис. 3. Екран авторизації **Рис.4. Екран вибору локацій**

Користувач може мати доступ до більше ніж однієї мікромережі, тому необхідно мати екран для вибору локації з мікромережею для детальної інформації по ній. Цей екран реалізовано у вигляді спливаючого діалогового вікна та являє собою список з переліком локацій, кожен елемент списку якого містить в собі назву локації, назву типу локації та її адресу (рис. 4). На головному екрані (рис. 5, 6) відображається інформація про поточний та прогнозний стан роботи обраної енергетичної мікромережі. Основним елементом головного екрану є структурні компоненти енергетичної мікромережі.

Рис. 5. Головний екран **Рис. 6. Головний екран**

Дані з сонячних панелей Дані з вітроустановки

Для компонентів з сонячними панелями та вітроустановками відображаються такі дані, як максимальна генеруюча потужність, поточна та прогнозна потужність у вигляді фактичних значень та у вигляді графічного зображення частки генеруючою потужності, список пристроїв які використовуються для генерації електроенергії з їх назвою та їх кількістю.

Для компонента з акумуляторами на екран виводиться максимальний запас накопичення енергії, поточний запас накопиченої енергії у вигляді фактичних значень та у вигляді графічного зображення частки накопиченої енергії, список пристроїв які використовуються для накопичення енергії з їх назвою та їх кількістю.

Висновки. Розроблено мобільний додаток підтримки надання послуг від енергетичних мікромереж. З урахування сучасних трендів у сфері розробки мобільних додатків для операційної системи Android розглянуті підходи та вимоги до проектування мобільних додатків, обрана сучасна гнучка архітектура побудови додатку та визначені необхідні допоміжні бібліотеки. Під час програмної реалізації проекту були використані мова програмування Kotlin, бібліотеки Retrofit2 і OkHttp3 для взаємодії з web-API та бібліотеку Gson для перетворення JSON даних. Мобільний додаток протестовано і підготовлено до використання користувачами послуг від енергетичних мікромереж як засіб інформування про стан мікромережі. Подальші дослідження полягають у розширенні функцій мобільного додатку, доповнення відображення даних прогнозу генерації електроенергії на різні часові інтервали.

Список використаних джерел:

1. Zawada M., Pabian A., Kucęba R., Bylok F. Impact of Smart Grid Intelligent Networks on Energy Efficiency Improvement. Proceedings of the 2019 10th International Conference on E-business, Management and Economics. 2019.

2. Ameur A., Berrada A., Emrani A. Intelligent energy management system for smart home with grid-connected hybrid photovoltaic/ gravity energy storage system. Journal of Energy Storage. 2023.

3. Shendryk V., Boiko O., Parfenenko Yu., Shendryk S., Tymchuk S. Decision Making for Energy Management in Smart Grid. Research Anthology on Clean Energy Management and Solutions. 2021. P. 1742–1776. https://doi.org/10.4018/978-1-7998-9152-9.ch077

4. Shendryk S., Shendryk V., Parfenenko Yu., Drozdenko O., Tymchuk S. Decision Support System for Efficient Energy Management of MicroGrid with Renewable Energy Sources. Proceedings of the 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications, IDAACS 2021. 2021. P. 225–230. https://doi.org/10.1109/IDAACS53288.2021.9660966.

5. Boiko Olha, Shendryk Vira, Parfenenko Yuliia, Pavlenko Petro, Titariev Artem. Information support of stakeholders in the management of energy systems: development and implementation of interfaces. Eastern-European Journal of Enterprise Technologies, 2023, vol. 6, P. 15–24. https://doi.org/10.15587/1729-4061.2023.292186.

6. Sebasthirani D., Maruthupandi D. Devolpment of Novel Mobile Application for Smart Meter Monitoring in Smart Grid. Gedraz & Organisatie Review. 2020, vol 33, no 155,

7.Hossain I., Islam M.S., Sultana R., Rahman M.R. IoT Based Home Automation System Using Renewable Energy. American Journal of Agricultural Science, Engineering, and Technology. 2022.

8.L'hadi I., Sikal M.B., Lahtani S., Khallaayoun A., Lghoul R. Development of a mobile application for home energy management in smart grids. 2015 World Congress on Sustainable Technologies (WCST), 2015, 123-129.

9. Enlighten – the monitoring experience for the system owner. URL: https://play.google.com/store/apps/details?id=com.enphaseenergy.myenlighten (дата звернення 06.04.2024)

10. The μGrid Manager or The Microgrid Manager Play Google. URL: https://play.google.com/store/apps/details?id=th.co.gridwhiz.MicrogridManager (дата звернення 06.04.2024)

11. mySunPower. SunPower. URL: https://play.google.com/store/apps/details?id=com.mysunpower (дата звернення 06.04.2024)

12. SMA Energy. SMA Solar Technology AG. URL: https://play.google.com/store/apps/details?id=de.sma.energy (дата звернення 06.04.2024)

УДК 316.343.3
DOI https://doi.org/10.32689/maup.it.2024.1.12

Світлана ПЕТРЕНКО
науковий співробітник науково-організаційного центру
Національної академії Служби безпеки України
***ORCID:** 0000-0003-1219-2401*

Наталія НАЗАРЕНКО
старший викладач кафедри романо-германських мов
навчально-наукового гуманітарного інституту
Національної академії Служби безпеки України
***ORCID:** 0000-0001-6353-4761*

ПРАКТИЧНІ АСПЕКТИ ВЕДЕННЯ ІНФОРМАЦІЙНОЇ ВІЙНИ В ОН-ЛАЙН ПРОСТОРІ

***Анотація**. У статті розглядається інформаційна війна в он-лайн просторі як один з аспектів сучасного життя. Інформаційна війна, яка активно проводилася ворожою російською федерацією впродовж багатьох років в інформаційному просторі не лише України, а й багатьох країн світу, набула значних масштабів з повномасштабним вторгнення. Вивчення цієї проблеми є **актуальним**, і потребує нових підходів, що і визначає **новизну** представленої роботи.*

***Метою** роботи є аналіз способів, якими інформація може бути використана як зброя для досягнення різних цілей: політичних, економічних та соціальних. В статті акцентовано увагу на тому, що основна мета інформаційної війни – не фізичне знищення людей, а руйнування їх як соціальної групи.*

***Методологія.** Проаналізувавши численні медіа ресурси, в статті розглядаються ключові підходи до ведення інформаційної війни в он-лайн просторі, такі як створення та поширення дезінформації, використання соціальних медіа, кібератаки та пропаганда. Зокрема, підкреслено важливість критичного мислення та перевірки фактів, у боротьбі з дезінформацією. Стаття також описує способи використання інформації як зброї в он-лайн просторі, акцентуючи увагу на фабриках тролів, ботах, соціальних медіа та кібератаках. Обгунтовано, як ці елементи можуть бути використані для маніпулювання громадською думкою, формування соціальних рухів та впливу на політичні, соціальні та культурні відносини. Зокрема, розглядається вплив кібершпигунства та хакерських атак на різні об'єкти, включно з урядовими установами, корпоративними мережами та персональними комп'ютерами громадян. У контексті військового конфлікту між Україною та росією хакерські атаки та кібершпигунство стали ключовими інструментами ведення війни. Російські зловмисники активно використовують ці методи для атак на українські інформаційні системи, з метою викрадення або знищення конфіденційної інформації. Інформаційна війна переважно спрямована на порушення обміну достовірною інформацією та створення паніки серед українців. Російська пропаганда активно поширює маніпулятивні новини та використовує гіперболізацію, для негативного висвітлення діяльності українських військових та переселенців.*

***Висновки.** В статті наголошується на необхідності розробки, удосконалення ефективних стратегії захисту від кібератак, а саме: використання надійного програмного забезпечення, регулярне оновлення систем, резервне копіювання даних та підвищення обізнаності населення щодо безпеки і інформаційної гігієни. Автори статті наголошують на необхідності збалансованого підходу до захисту національної безпеки та дотримання основних прав людини, а також підкреслюють важливість створення ефективних засобів для боротьби з кіберзагрозами.*

***Ключові слова:** кібератака, онлайн простір, інформаційна війна, медіаманіпуляція, дезінформація, пропаганда.*

Svitlana PETRENKO, Natalia NAZARENKO. PRACTICAL ASPECTS OF INFORMATION WARFARE IN ONLINE DOMAIN

***Abstract.** The given article studies information warfare in the online domain as one of the aspects of contemporary life. Information warfare, which has been carried out by the hostile russian federation for a long time not only in the information space of Ukraine, but other countries all over the world, got a full swing with the full-scale invasion. Studying this problem from a new perspective is **crucial**, which defines the **novelty** of the given work*

***The aim** of the article is to analyse the ways of using information as a weapon to achieve diverse objectives, including political, economic and social. It emphases that the primary goal of information warfare is not the physical destruction of individuals, but the dismantling of their social cohesion.*

***Metodology.** Having analysed numerous media sources, there have been revealed the main approaches to conducting information warfare in the online domain, such as the creation and dissemination of disinformation, the use of social media, cyberattacks, and propaganda. The article emphasizes the significance of critical thinking and fact-checking in combating misinformation. It also describes the use of information as the information warfare weapon, focusing on troll factories, bots, social media, and cyberattacks, and how they can be used to manipulate public opinion, encourage social movements, and influence political, social and cultural relations in the society. The article considers the impact of cyber espionage and hacker attacks on various targets, particularly governmental institutions, corporate networks, and personal computers of ordinary users. In the context russian-Ukrainian war hacker attacks and cyber espionage have become pivotal tools of warfare. Russian malicious actors actively employ these methods to target Ukrainian information systems to steal or destroy confidential information. The main aim of information warfare predominantly is to disrupt the exchange of reliable information and induce*

panic among the Ukrainians. Russian propaganda actively disseminates manipulative news and employs hyperboles to show Ukrainian military personnel and displaced persons in a negative light.

Conclusions. *The article focuses on necessity to develop and improve efficient strategies for dealing with cyberattacks, such as the use of robust software, regular system updates, data backup and raise of public awareness about information security and hygiene. The authors of the article call for a balanced approach to ensuring national security and preserving fundamental human rights, as well as creating effective ways to resist various cyber threats.*

Key words: *cyberattack, online domain, information warfare, media manipulation, disinformation, propaganda.*

Постановка проблеми. У сучасному світі, де інформація є ключовим ресурсом, інформаційна війна стає все більш актуальною. Ця проблема набуває особливого значення в контексті повномасштабного вторгнення росії в Україну, де інформаційні війни в он-лайн-просторі стають важливим фронтом боротьби.

Інформаційна війна в он-лайн-просторі включає в себе різноманітні тактики та стратегії, які використовуються для маніпулювання громадською думкою, формування настрою та впливу на політичні процеси. Згадані тактики та стратегії можуть включати в себе все: від пропаганди та дезінформації до кібератак та використання соціальних медіа для поширення певних повідомлень або ідей.

Аналіз останніх досліджень і публікацій. Питання інформаційного впливу, інформаційних війн, їх стратегій та методів широко висвітлюються в наукових роботах як вітчизняних так і закордонних науковців, зокрема О. В. Курбана, В. І. Башманівського, Н. М. Шульської, Д. М. Солоденка та інших. Але, зважаючи на стрімкий розвиток інформаційних технологій та широке використання інформаційного простору, ця проблема не втрачає своєї актуальності, а набуває нових образів і характеристик, що і визначає **актуальність** цього дослідження.

Метою дослідження є аналіз практичних аспектів ведення інформаційної війни в он-лайн просторі.

Виклад основного матеріалу. Інформаційна війна в он-лайн просторі стала важливим аспектом сучасного життя. Це поле, де інформація використовується як зброя для досягнення політичних, економічних або соціальних цілей.

Он-лайн простір – це віртуальне середовище, у якому люди взаємодіють й обмінюються інформацією. Він включає в себе соціальні медіа, веб-сайти, електронну пошту та інші платформи.

Інформаційна війна – це сучасний спосіб беззбройного конфлікту, який набирає шалених обертів, і її наслідки є загрозливими та непередбачуваними . Її основна мета – не фізичне знищення людей, а руйнування їх як соціальної групи. Таким чином, людство почало використовувати переваги прогресу в зловмисних цілях – проведенням інформаційних воєн, тобто поширення інформації з метою формування потрібних думок, настроїв та системи поглядів щодо певних питань, подій або людей на користь організатора конкретної інформаційної або пропагандистської кампанії. Основне завдання такої діяльності – це маніпулювання масами, тобто внесення потрібних ідей та поглядів у свідомість ворога, дезорієнтація та дезінформація цільової аудиторії, залякування власного народу образом ворога та створення почуття страху у супротивника, пропагуючи своєю могутністю.

Швидкість проведення інформаційної кампанії – це запорука успіху інформаційної війни, оскільки вона впливає на здатність ворога швидко та оперативно приймати рішення, реагувати на ситуації та вести війну на реальному полі бою. Вдала інформаційна кампанія проти ворога призводить до прийняття ним помилкових рішень і забезпечує невиконання запланованих завдань [3, с. 70].

О. В. Курбан виділяє наступні аспекти ведення інформаційної війни в он-лайн просторі :

– створення та поширення дезінформації для введення в оману людей або для створення певного враження;

– використання соціальних медіа для маніпулювання громадською думкою або для створення соціального руху;

– кібератаки для збору інформації, пошкодження інфраструктури або зламу систем безпеки;

– пропаганда для формування певного враження або впливу на громадську думку або для підтримки певної політичної агенди або ідеології [7].

Генерування дезінформації відіграє центральну роль в контексті інформаційної війни. Дезінформація, як правило, використовується для маніпулювання публічною свідомістю та формування специфічного сприйняття. Після генерування дезінформації наступним етапом є її дисемінація (з англ. «*dissemination*» – розповсюдження), яка здійснюється через різні канали, такі як соціальні медіа, веб-сайти, блоги, форуми тощо. Її мета полягає в тому, щоб донести цю інформацію до якомога більшої аудиторії.

Дезінформація може мати значний вплив на громадську думку та поведінку людей. Вона може використовуватися для формування певного враження, маніпулювання громадською думкою або навіть впливу на різні соціально-політичні події, такі як результати соцопитувань, вибори тощо.

Важливо пам'ятати, що дезінформація – це потужний інструмент, який може бути використаний для маніпулювання громадською думкою та формування певного поглядів, настроїв в суспільстві. Тому важливо завжди бути насторожі та критично ставитися до отриманої інформації [5]. Боротьба з дезінформацією вимагає, в першу чергу, критичного мислення та ретельної перевірки фактів, достовірності джерел, порівняння інформації, що надається іншими джерелами, та пошук офіційних або незалежних джерел перевірки фактів.

Використання соціальних медіа в інформаційній війні відіграє вирішальну роль, оскільки ці платформи служать потужними каналами для дисемінації інформації. Вони можуть бути використані для маніпулювання громадською думкою, шляхом поширення дезінформації, пропаганди або інших форм маніпулятивного контенту за допомогою ботів, тролів або інших автоматизованих засобів для поширення інформації, які підтримують певну агенду або враження [9]. Фабрики тролів представляють собою організації, де працівники створюють коментарі в Інтернеті, відповідно до завдань замовника, використовуючи фальшиві профілі в соціальних мережах. Основними характеристиками таких тролінг-провокацій є конфіденційність, маскування, тіньова позиція та безкарність. По відношенню до цих провокацій практично немає ефективних методів протидії, крім блокування форумів та обмеження можливості коментувати і поширювати далі інформацію. Боти – це програми, які автоматично відправляють повідомлення, особливо відгуки на появу конкретного ключового слова. Однак проблема полягає в тому, що одна людина може управляти десятками ботів одночасно, і це ускладнює їх виявлення та заборону. Принцип роботи ботів можна описати так: спершу троль розміщує пост, велика кількість інших ботів починає виражати вподобання, коментувати та репостити його. Алгоритм соціальної мережі реєструє – це як зацікавленість реальних користувачів у даному пості. У результаті він потрапляє в стрічки вже не ботів, а справжніх людей. Якщо пост вдалий, його розповсюджують далі звичайні користувачі, а потім його можуть використати і журналісти [2, с. 116-117].

Соціальні медіа також можуть бути використані для створення або підтримки соціальних рухів. З допомогою хештегів, мемів або інших форм вірального контенту відбувається мобілізації аудиторії навколо певної проблеми або агенди. Використання певних алгоритмів для визначення інтересів, переглядів або поведінки користувачів дозволяє соціальним медіа точно визначити конкретні цільові групи, що забезпечує ефективне поширення інформації. Соціальні медіа також надають можливість моніторити та аналізувати реакцію аудиторії на поширену інформацію, шляхом відслідковування таких метрик, як кількість переглядів, вподобань, коментарів або репостів. Важливо пам'ятати, що використання соціальних медіа в інформаційній війні вимагає глибокого розуміння цих платформ та їх динаміки, що передбачає знання того, як інформація поширюється в цих мережах, як вона сприймається користувачами, та як вона може впливати на громадську думку та поведінку [9].

Одним з найдієвіших методів інформаційної війни є кібератаки, оскільки вони спрямовані на збір інформації, ураження інфраструктури або зламу систем безпеки. Кіберзлочинці використовують цілий арсенал засобів для досягнення своїх цілей, а саме фішинг, DDoS-атаки, віруси, трояни та інше, для несанкціонованого доступу до систем, крадіжки даних, знищення або зміни інформації, а також перешкоджання нормальному функціонуванню мереж. Ці атаки можуть мати значні наслідки, такі як порушення конфіденційності, втрату даних, фінансові втрати та підрив репутації [8]. У контексті війни України з росією, хакерські атаки та кібершпигунство стають ключовими інструментами ведення війни. Російські зловмисники активно використовують ці методи для атак на українські інформаційні системи, метою яких є викрадення або знищення конфіденційної інформації. Хакерські атаки переважно спрямовані на урядові установи, корпоративні мережі, а також персональні комп'ютери та мобільні пристрої громадян України. Кібершпигунство, з іншого боку, зосереджується на отриманні доступу до конфіденційних даних або слідкуванні за діями користувачів за допомогою шпигунського ПЗ, фішингу, соціальної інженерії та інших тактик для отримання доступу до приватної інформації або для слідкування за діями користувачів. Сучасна війна набула нових рис, оскільки агресор прагне не лише завдати конкретних втрат противнику на полі бою, але й, вдаючись до засобів інформаційної війни, суттєво вплинути на поширення достовірної інформації, на механізми прийняття важливих державних рішень, а також викликати паніку серед українців, створити у населення відчуття страху та дезорієнтації, спонукати його до швидкої капітуляції. Завдання полягає в тому, щоб громадяни думали про втечу, витрачаючи свій час на поширення фейкових повідомлень, замість надання реальної допомоги війську та своїй державі [1, с. 274].

Пропаганда в інформаційній війні є стратегічним інструментом, що використовується для формування специфічного сприйняття реальності та впливу на громадську думку через поширення інформації, яка підтримує певну політичну агенду або ідеологію. Пропаганда широко використовується для маніпулювання громадською свідомістю, формування соціальних настроїв і рухів, а також для зміни

або підтримки певних політичних, соціальних або культурних відносин. Це вимагає глибокого розуміння психології мас, медіа-ландшафту та динаміки соціальних медіа [7].

Російські засоби пропаганди дуже активно поширюють маніпулятивні новини на різні теми, але здебільшого вони стосуються українських військових та наших переселенців із територій, де ведуться активні бойові дії. Кремлівські ЗМІ у своїх текстах намагаються виставити Збройні сили України в негативному світлі та дискредитувати їх в очах громадськості. Якщо ж сприймати інформацію критично й перевіряти дані в офіційних джерелах, то абсолютно нескладно зрозуміти, що ці текстові матеріали неправдиві [3, с. 71].

Показовим засобом російської пропаганди слугує виразна гіперболізація, як-от у фейкових новинах про те, що на Херсонщині був дуже великий ажіотаж серед місцевих мешканців, які буцімто масово хотіли отримати документи про набуття російського громадянства. Такого типу новини характеризуються додатковими маркерами маніпуляції, а саме лексемами «*великий ажіотаж*» (підсилення контексту надає ненормативний плеоназм), «*дуже*», «*масово*» і т. і. Після окупації Маріуполя російська влада почала поширювати неправдиву інформацію, що України вже нібито «не існує», а російська армія просувається далі. Російські медіа поширювали хибні «вкиди» про те, що начебто в Запоріжжі формують списки добровольців, які прагнуть захищати місто від українських збройних сил. Зрозуміло, що подібного роду інформація видається цілком абсурдною [1, с. 277].

Ще один спосіб, який широко застосовується окупантами на тимчасово окупованих територіях, це блокування або обмеження доступу до Інтернету та інших джерел інформації, що має великий вплив на суспільство. Це дозволяє контролювати або маніпулювати інформацією, яка надходить до громадян, і є дієвим засобом безпеки для запобігання поширенню правдивої інформації. Беззаперечно, такі дії порушують право на свободу вираження думки та доступ до інформації, що є основними правами людини. Крім того, вони перешкоджають нормальному функціонуванню суспільства, оскільки багато сфер життя, включно з освітою, охороною здоров'я, бізнесом та урядовими службами, залежать від доступу до Інтернету та інших засобів комунікації.

З іншого боку, обмеження доступу до певних джерел інформації або неоприлюднення певної інформації є необхідністю для забезпечення певних аспектів національної безпеки під час війни. Однак, це питання створює певні суперечки і несприйняття в суспільстві. Тому важливо знайти баланс між потребою в захисті національної безпеки та захисті основних прав людини. Необхідно розробити і запровадити прозорі та обґрунтовані процедури для блокування або обмеження доступу до комунікаційних засобів, а також забезпечити можливість оскарження таких дій та відновлення доступу, коли це безпечно та доцільно. Крім того, важливо контролювати, щоб такі дії були тимчасовими та пропорційними, а не стали постійними або надмірними обмеженнями на свободу вираження думки та доступ до інформації [6].

Отже, зважаючи на те, що інформаційна війна може мати серйозні, а й подекуди катастрофічні наслідки для національної безпеки країни, особливо під час війни необхідно вживати невідкладні заходи і використовувати ефективні методи протидії, а саме:

– активно розвивати власні інформаційні технологій та систем, щоб забезпечити свою здатність до протидії інформаційним атакам;

– підвищувати обізнаність громадян щодо інформаційних війн та їх наслідків;

– застосовувати норми міжнародного та національного законодавства в боротьбі проти інформаційних війн, а також для притягнення до відповідальності тих, хто проводить інформаційні атаки. Створювати відповідні закони та норми, які регулюють інформаційний простір;

– співпрацювати з міжнародними партнерами для обміну інформацією, координації дій та спільної розробки стратегій протидії інформаційній агресії;

– розроблювати ефективні стратегій протидії інформаційним атакам, з урахуванням специфіки інформаційного простору та особливостей сучасних інформаційних війн;

– здійснювати постійний моніторинг інформаційного простору;

– використовувати асиметричні моделі інформаційної боротьби – непередбачувані, нестандартні методи протидії інформаційній агресії [10].

Особливої уваги потребує проблема кіберзахисту, яка вимагає системного підходу. Захист від кібератак та забезпечення цілісності та конфіденційності інформації – це безперервний процес, що вимагає постійного моніторингу, оновлення та адаптації до нових загроз [6]. В Україні розроблено Стратегію кібербезпеки, яка включає в себе ряд заходів для забезпечення кібербезпеки в країні. Ця стратегія передбачає розробку надійних систем безпеки, освіту користувачів та постійний моніторинг і оновлення заходів безпеки для відповіді на нові загрози. Крім того, Україна співпрацює з міжнародними партнерами, такими як USAID, для підвищення рівня кібербезпеки в країні.

Розвиток власних інформаційних технологій є важливим елементом стратегії України проти інформаційної війни, яку веде росія. Створення та впровадження власних програмних рішень та систем сприятимуть ефективному виявленню, блокуванню та попередженню можливих кіберзагроз. Ці технології допомагають Україні виявляти та відстежувати інформаційні атаки, а також розробляти стратегії для їх протидії. Вони також дозволяють забезпечити доступ до точної та актуальної інформації, що є важливим елементом протидії інформаційній війні. Однак, розвиток власних інформаційних технологій вимагає постійного аналізу потенційних загроз, оновлення та адаптації до нових. Тому Україна повинна продовжувати інвестувати в розвиток своїх інформаційних технологій та підвищувати свою здатність до протидії інформаційним атакам [4].

Висновки. Отже, інформаційна війна в он-лайн просторі визначається низкою складних та динамічних аспектів, що вимагають комплексного підходу до вивчення та контролю. Здійснення таких операцій стає неодмінною складовою стратегій впливу в сучасному геополітичному середовищі. В ході аналізу було виявлено, що практичні аспекти ведення інформаційної війни в он-лайн просторі визначаються широким спектром технічних, соціальних та політичних факторів.

У сучасному інформаційному полі кіберпростір відіграє головну роль у формуванні громадської думки, впливі на глобальну політику та формуванні образу країни в світі. Використання сучасних технологій, алгоритмів штучного інтелекту, а також вміння ефективно маніпулювати інформаційними потоками дозволяє надзвичайно швидко й ефективно досягати визначених стратегічних цілей: формувати погляди суспільства та його реакції на події, що відбуваються, зловживати психологічними та емоційними аспектами для впливу і маніпулювання масовим суспільством, поширювати неправдиву інформацію та багато іншого.

Загальним висновком є те, що інформаційна війна в он-лайн просторі визначається високою складністю та необхідністю вивчення та аналізу всіх її аспектів. Ефективна протидія цьому явищу вимагає поєднання технічних, правових, соціальних та політичних заходів для забезпечення цілісності та безпеки сучасного інформаційного простору.

Список використаних джерел:

1. Башманівський В. І., Шульська Н. М., Зінчук Р. С. Фейкоінструментарій ведення інформаційної війни в Україні: на матеріалі мови сучасних медіа. *Вчені записки Таврійського національного університету імені В. І. Вернадського.* 2023. №34(73). С. 274–280.

2. Солоденко Д. М. Методи ведення війни у віртуальному просторі. *Актуальні проблеми соціальних комунікацій :* матер. восьмої всеукр. студ. наук. конф. 2022. С. 116–120.

3. Шульська Н. М. Медіаманіпуляції в умовах російсько-української війни (на прикладі локальних ЗМІ). *Південний архів* (*філологічні науки*). Херсон, 2022. Вип. 90. С. 68–76.

4. Як українці створили мапу, яка розповідає про війни та протести в усьому світі. *Liveuamap*. URL: https://www.bbc.com/ukrainian/-news-55543745 (дата звернення: 27.01.2024).

5. У віртуальної війни – нове обличчя». *Детектор Медіа*. URL: https://detector.media/withoutsection/article/135987/2018-03-26-u- virtualnoi-viyny-nove-oblychchya/ (дата звернення: 27.01.2024).

6. Інформаційна війна проти України та методи її ведення. *PolUkr.* URL: https://www.polukr.net/uk/blog/2021/04/informacijna-vijna-proti-ukraini/ (дата звернення: 27.01.2024).

7. Курбан О. В. Сучасні інформаційні війни в мережевому он-лайн просторі. UKR: http://www.interinf.chnu.edu.ua/res//interinf/Inf%20vijny.pdf.

8. Невельська-Гордєєва О. Нечитайло В. Феномен «fake news» в контексті забезпечення інформаційної безпеки держави. *Вісник НЮУ імені Ярослава Мудрого. Сер. : Філософія, філософія права, політологія, соціологія*. 2022. №1(52). URL: https://doi.org/10.21564/2663-5704.52.250655 (дата звернення: 25.01.2024).

9. Федотенко К. «Інформаційна війна» та «інформаційний фронт»: наукове осмислення понять. *Вісник НЮУ імені Ярослава Мудрого. Сер. : Філософія, філософія права, політологія, соціологія*. 2023. №3(58): веб-сайт. URL: https://doi.org/10.21564/2663-5704.58.285787 (дата звернення: 24.01.2024).

10. Information Warfare and the Changing Face of War. URL: https://www.rand.org/pubs/monograph_reports/MR661.html (дата звернення: 26.01.2024).

УДК 004.77
DOI https://doi.org/10.32689/maup.it.2024.1.13

Сергій ТИМЧУК
доктор технічних наук, доцент, професор кафедри інформаційних технологій,
Сумський державний університет, s.tymchuk@itp.sumdu.edu.ua
***ORCID:** 0000-0002-8600-4234*

Ірина БАРАНОВА
кандидат технічних наук, доцент, доцент кафедри інформаційних технологій,
Сумський державний університет, i.baranova@cs.sumdu.edu.ua
***ORCID:** 0000-0002-3767-8099*

Олексій ПІСКАРЬОВ
кандидат технічних наук, доцент,
доцент кафедри електронних обчислювальних машин,
Харківський національний технічний університет радіоелектроніки, oleksii.piskarov@nure.ua
***ORCID:** 0000-0002-6980-984X*

Станіслав РАДЧЕНКО
асистент кафедри комп'ютерних інформаційних систем і технологій,
Харківський національний технічний університет радіоелектроніки, stanislav.radchenko@nure.ua
***ORCID:** 0000-0003-2520-6120*

Тарас ЮРЧЕНКО
здобувач РВО магістр-науковець,
Харківський національний технічний університет радіоелектроніки, taras.iurchenko@nure.ua
***ORCID:** 0009-0002-5535-6998*

ПОЛІПШЕННЯ ЗАВАДОСТІЙКОСТІ ТА ЗБІЛЬШЕННЯ ШВИДКОСТІ ПЕРЕДАЧІ ДАНИХ У WI-FI-МЕРЕЖАХ

***Анотація**. На теперішній час відбувається інтенсивний розвиток систем бездротового зв'язку, у тому числі комп'ютерних Wi-Fi мереж. У каналах радіозв'язку таких систем діє комплекс перешкод і спотворень. Для покращення таких параметрів як швидкодія та перешкодостійкість, особливо в умовах щільного використання достатньо обмеженого частотного каналу, існує необхідність вдосконалення існуючих методів, та створення принципово нових. **Метою статті** є огляд методів передачі інформації в сучасних системах бездротового доступу та дослідження алгоритмів підвищення пропускної здатності мережі за рахунок застосування методів адаптивної просторової обробки сигналів і пошук балансу між підвищенням пропускної здатності технології MIMO та зменшення імовірності помилки на прийомі. **Методи дослідження**: під час дослідження використовуються методи передачі інформації в сучасних системах бездротового доступу та алгоритми підвищення пропускної здатності мережі. **Наукова новина** дослідження полягає в тому, що проведений аналіз сучасних методів бездротової передачі інформації, виявив, що просторово-часове кодування вдало поєднує переваги методів просторового рознесення з можливостями виправлення помилок коригувальним кодом при використанні оптимальних алгоритмів декодування, при цьому ефективність досліджень та розробок нових методів просторово-часового кодування значною мірою залежить від того, наскільки моделі каналів відповідають реальним умовам. Одним з перспективних методів підвищення параметрів якості мережі – є метод синтезу згортково-блокових сигнально-кодових конструкцій з використанням внутрішніх сигналів з класу просторово-часового блокового кодування й зовнішніх сигнальних конструкцій, який є ефективною технікою для зменшення впливу завмирання на сигнали, поліпшення якості та пропускної спроможності системи Wi-Fi зв'язку. **Висновки**. Розробка цих алгоритмів та методів відкриває широкі перспективи для майбутнього розвитку бездротових комунікаційних систем. Однією з ключових перспектив є подальше вдосконалення методів адаптивної просторової обробки сигналів та оптимізація балансу між підвищенням пропускної здатності та зменшенням ймовірності помилок на прийомі. Додатково, можливості згортково-блокових сигнально-кодових конструкцій можуть бути розширені шляхом дослідження та впровадження нових технологій, наприклад, використання машинного навчання для оптимізації параметрів кодування та декодування сигналів. Також існують можливості застосування цих методів у високошвидкісних мережах мобільного зв'язку, таких як мережі п'ятого покоління (5G) і майбутні покоління, де висока пропускна здатність та ефективність передачі даних стають ключовими вимогами.*

***Ключові слова:** Wi-Fi мережа, просторово-часове кодування, технологія MIMO, декодування Аламоуті, бінарна фазова маніпуляція*

Sergiy TYMCHUK, Iryna BARANOVA, Oleksiy PISKAROV, Stanislav RADCHENKO, Taras YURCHENKO.
IMPROVING NOISE IMMUNITY AND INCREASING DATA TRANSMISSION SPEED IN WI-FI NETWORKS

Abstract*. Wireless communication systems, including computer Wi-Fi networks, are currently undergoing intensive development. The radio communication channels of such systems are subject to a complex of interference and distortion. To improve such parameters as performance and interference resistance, especially in conditions of dense use of a rather limited frequency channel, there is a need to improve existing methods and create fundamentally new ones.* ***The purpose of the article*** *is to review the methods of information transmission in modern wireless access systems and to study algorithms for increasing network capacity by applying adaptive spatial signal processing methods and finding a balance between increasing the throughput of MIMO technology and reducing the probability of reception errors.* ***Research methods****: the study uses methods of information transmission in modern wireless access systems and algorithms for increasing network capacity.* ***The scientific novelty*** *of the study is that the analysis of modern methods of wireless information transmission revealed that space-time coding successfully combines the advantages of spatial diversity methods with the ability to correct errors with a corrective code when using optimal decoding algorithms, while the effectiveness of research and development of new methods of space-time coding largely depends on how well the channel models match real-world conditions. One of the promising methods for improving network quality parameters is the method of synthesizing convolutional-block signal-code structures using internal signals from the class of space-time block coding and external signal structures, which is an effective technique for reducing the effect of fading on signals, improving the quality and throughput of the Wi-Fi communication system.* ***Conclusions****. The development of these algorithms and methods opens up broad prospects for the future development of wireless communication systems. One of the key prospects is to further improve the methods of adaptive spatial signal processing and optimize the balance between increasing throughput and reducing the probability of reception errors. Additionally, the capabilities of convolutional-block signal-code designs can be expanded by researching and implementing new technologies, such as using machine learning to optimize signal coding and decoding parameters. There are also opportunities to apply these techniques to high-speed mobile networks, such as fifth generation (5G) and future generations, where high bandwidth and data efficiency are becoming key requirements.*

Key words: *Wi-Fi network, space-time coding, MIMO technology, Alamouti decoding, binary phase manipulation.*

Вступ і постановка проблеми. На сьогодні спостерігається активний прогрес у розвитку бездротових комунікаційних систем, що охоплюють мобільний радіозв'язок, системи безпровідного доступу до Інтернету, бездротові комп'ютерні мережі, що функціонують всередині будівель та інші подібні технології. Канали радіозв'язку в таких системах піддаються впливу різноманітних перешкод і спотворень. Особливу увагу слід звернути на багатопроменевість, що виникає через відбиття радіохвиль на шляху їх поширення. У діапазоні коротких хвиль виникають повторні відбиття від неоднорідностей іоносфери. У метрових (дециметрових) діапазонах хвиль спостерігаються відбиття від будівель, нерівностей рельєфу (особливо при організації зв'язку на відкритих місцевостях) та відбиття від стін і конструкцій (зокрема під час зв'язку всередині будівель). Сильні завмирання сигналу у каналі створюють труднощі у визначенні переданих повідомлень та можуть спричинити спотворення переданої інформації [11].

Аналіз досліджень і публікацій. Вперше методи статистичної теорії зв'язку для рознесеного прийому були використані наприкінці 1930-х років, однак для розробки основних теоретичних концепцій статистичної теорії розподіленого прийому знадобилося ще 15-20 років. Ідея використання рознесеного прийому для подолання завмирань полягає в спільному прийомі кількох сигналів на приймачі, які несуть однакову інформацію, але були прийняті різними шляхами. Рознесення має застосовуватися таким чином, щоб імовірність одночасного завмирання сигналів у всіх каналах, що використовуються, була значно меншою, ніж у будь-якого окремого каналу.

Були розроблені методи комбінування сигналів із різних рознесених каналів з втратами [12]. Радикальним методом подолання завмирання радіосигналів є розподілення приймальних антен, а найефективнішим методом об'єднання є оптимальне "вагове" комбінування, яке базується на критерії максимального відношення сигналу до шуму. Ця комбінація (рознесення та оптимальна обробка на прийомі) стала настільки поширеною, що до приходу "нової ери" бездротового зв'язку вона була включена до арсеналу засобів, описаних у спеціалізованому посібнику з бездротового зв'язку [14].

Були спроби вирішити проблему завадостійкого передавання інформації каналами з втратами, "обходячи" традиційні методи рознесення. Одним із таких підходів є використання структурних властивостей переданих сигналів, наприклад, передача широкосмугових сигналів, які дозволяють розділяти промені на прийомі – це система RAKE [11].

Поява робіт, що реалізували ідею просторово-часового кодування (ПЧК), відзначила нову еру в теорії методів передавання інформації каналами із завмиранням. Зокрема в роботі [5] пропонується оточити джерело сигналу багатопроменевості (середовище поширення радіосигналів) безліччю передавальних та приймальних антен і відповідно організувати передачу та обробку сигналів на прийомі.

У будь-яких теоретичних дослідженнях методів передавання інформації через специфічні канали ключовим є питання математичної моделі цього каналу [2]. Для теорії інформації, як науки про

телекомунікації, сьогодні важливо є те, що основні принципи К. Шеннона про пропускну здатність каналу можна застосовувати до нових моделей каналів. У [15] було розглянуто питання широкосмугових бездротових мереж передавання інформації з описом інженерних рішень. Передові ефективні методи передавання інформації каналами з втратами можуть бути представлені та проаналізовані як сигнально-кодові конструкції (СКК), де завадостійке кодування забезпечує наближення до пропускної здатності каналу, а сигнали використовуються для узгодження кодування з каналом.

Метою цієї роботи є огляд методів передачі інформації в сучасних системах бездротового доступу, а також дослідження алгоритмів підвищення пропускної здатності мережі шляхом застосування методів адаптивної просторової обробки сигналів і знаходження балансу між підвищенням пропускної здатності технології MIMO та зменшенням ймовірності помилки на прийомі.

Виклад основного матеріалу. Сучасні системи радіозв'язку відзначаються високими показниками пропускної здатності. Різке зростання класичної межі Шеннона для пропускної здатності систем зв'язку стало можливим завдяки впровадженню просторової обробки сигналів, або технології MIMO (Multiple Input-Multiple Output — множинний вхід-множинний вихід), яка використовує ефект множинного поширення, де передані сигнали відбиваються від різних об'єктів та перешкод, а приймальна антена приймає сигнали під різними кутами і в різний час. Застосування цієї технології дозволяє через рознесення сигналу при передачі та прийомі зменшити відсоток біт, що отримані з помилками, тим самим підвищуючи завадостійкість каналів зв'язку.

Методи передачі інформації в сучасних системах широкосмугового бездротового доступу ґрунтуються на використанні технології MIMO. Переваги та особливості цього методу були продемонстровані на лабораторному прототипі ще у 1998 році, і згодом він був включений до стандартів широкосмугового доступу IEEE 802.11n і IEEE 802.16e [9]. Основна ідея технології MIMO показана на рисунку 1.

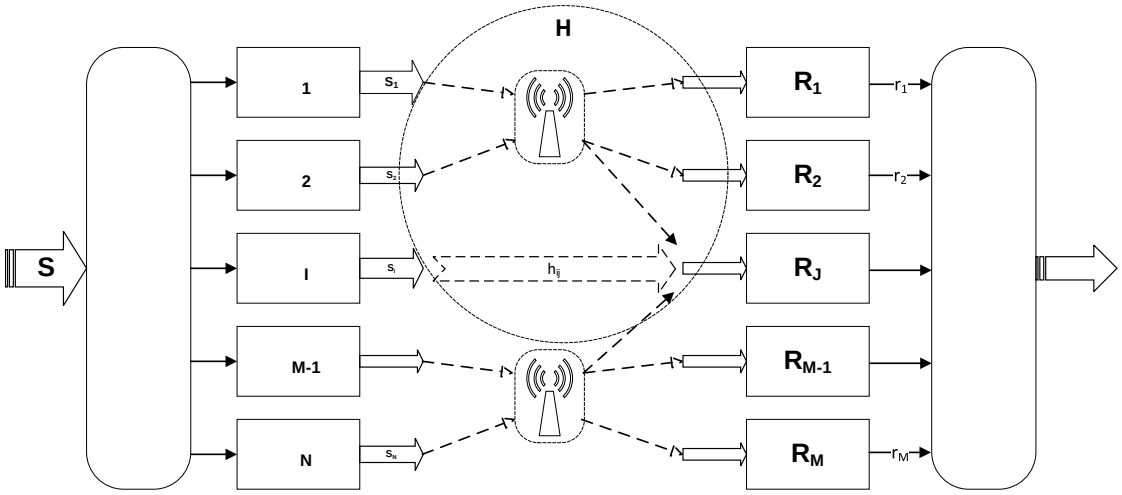


Рис. 1. Загальна структура роботи системи MIMO

Передавальна частина системи включає M передавачів ($T_1 \ldots T_M$) із передавальними антенами, тоді як приймальна частина включає N приймачів і приймальних антен ($R_1 \ldots R_N$). Вважається, що завмирання породжується розсіювальним середовищем H поширення радіосигналу. На рисунку 1 стрілками показано, що сигнал будь-якого з передавачів T_i може досягати входу будь-якого з приймачів ($R_1 \ldots R_N$), зазнаючи завмирання. Функціонування системи забезпечується мультиплексором на передачі, демультиплексором на прийомі та приймачем максимальної правдоподібності на приймальній стороні.

Отже вноситься просторова надмірність, завдяки якій вдається «пронизувати» турбулентне середовище поширення радіосигналу та при цьому уникнути впливу завмирань при відповідній обробці прийнятої сукупності сигналів. Дана структура забезпечує передачу «від об'єму до об'єму» (from volume to volume). Це визначення підкреслює суттєву різницю між об'ємно-багатовимірною просторовою моделлю каналу та звичайною одновимірною моделлю, яка розглядається як проста лінія між передавачем та приймачем Такі структури зазвичай позначають як MIMO (NxM) (N – кількість передавальних антен, M – кількість приймальних антен).

При цьому можлива велика різноманітність варіантів систем. Можна визначити такі варіанти структури MIMO (2x2):

1. Структура MIMO (1x2), іменована як SIMO (Single Input-Multiple Output – один вхід-багато виходів). Традиційна система радіозв'язку з одним передавачем і двома рознесеними в просторі антенами і приймачами.

2. Структура MIMO (2x1), іменована як MISO (Multiple Input-Single Output – багато входів – один вихід).

Термін «просторово-часове кодування» (Space-Time Coding – STC) – в теорії інформації під завадостійким кодуванням прийнято розуміти процедуру, відповідно якої в передані повідомлення вноситься надмірність, яка дає можливість виправляти канальні помилки при адекватному декодуванні. Зазвичай у традиційних методах кодування для введення надмірності використовується часовий ресурс (введення додаткових символів при блоковому або згортковому кодуванні). При цьому платою за підвищення завадостійкості є зниження швидкості передачі інформації. У розглянутих на рис.1 багатоантенних системах MIMO крім часового ресурсу (традиційне завадостійке кодування можливе також у будь-якому каналі "T_I-R_J") виникає можливість використовувати просторовий ресурс та вирішувати завдання оптимального введення надмірності, тобто використовувати оптимальні методи ПЧК, для забезпечення якнайкращого обміну надмірності на завадостійкість.

При практичному застосуванні систем MIMO потрібно вирішувати питання організації мультиплексної передачі сигналів від передавальних антен до приймальних. В основному застосовують часовий поділ сигналів. Тобто організовується «кадр» передачі з усіма необхідними в таких випадках атрибутами кадрової синхронізації («синхрослово» тощо). Обираючи метод модуляції сигналів-переносників можна вирішити питання швидкості передачі інформації в системі загалом [23].

У переважній більшості методів ПЧК у каналах MIMO неодмінною умовою теоретичного аналізу є квазістаціонарність каналу в такій формі [14]:

– в структурі MIMO передачу інформації можливо організувати кадрами (frame), які періодично передаються і мають спеціальну структуру;

– при зміні місця розташування передавальних і приймальних антен коефіцієнти передачі змінюються;

– коефіцієнти передачі на інтервалах декількох (зазвичай двох) поряд розташованих символів залишаються незмінними. Пропонована при цьому структура кадру (рис. 2) може бути подібна до широко використовуваної форми кадру в стандарті США системи стільникового зв'язку IS-136 [4].

Навчальна послідовність	**Дані**	**Пілот-сигнали**	**Дані**	**Пілот-сигнали**	**Дані**	**Пілот-сигнали**	**Дані**

Рис. 2. Типова структура кадру просторово-часового кодування

Кадр включає в себе початкову навчальну (traning) послідовність (НП) і періодично повторювані блоки переданих даних, які розділені пакетами пілот-сигналів (ПС). Структура НП може забезпечувати синхронізацію за кадрами. Використання пілот-сигналів ПС перед даними зумовлено необхідністю організації в демодуляторі когерентного прийому. Отже, увесь міжрядний простір передавальних і приймальних антен охоплено системою тимчасового мультиплексування, яка давно добре освоєна в системах стільникового мобільного зв'язку з часовим розподілом каналів TDMA.

Є дві групи методів ПЧК у каналах MIMO: просторово-часове решітчасте кодування (ПЧРК) та просторово-часове блокове кодування (ПЧБК).

Метод ПЧРК може поєднувати в собі переваги методів просторового різноманіття з можливістю виправлення помилок за допомогою виправлення коду та використовуючи оптимальні алгоритми декодування, які одночасно реалізують оптимальний алгоритм об'єднання сигналів різноманітності. Використовуючи традиційне коригувальне кодування додаючи при цьому надмірність у часовій області. У системах із ПЧРК вводиться надмірність також і в просторовій області, яка утворена кількома передавальними антенами й однією приймальною антеною (рис. 3). За рахунок ускладнення способів передачі і обробки сигналів на прийомі можливе отримання переваги у перешкодостійкості. Можна обрати для реалізації згортковий код зі швидкістю $R=k/n$. Кодер такого коду генерує послідовності, що утворюють кодову сітку, за допомогою якої алгоритм Вітербі в процесі декодування знаходить найбільш правдоподібний шлях.

На рисунку 3 показано модель системи з просторово-часовим решітчастим кодуванням, що містить кодер ПЧРК з n виходами, які під'єднані до n передавальних антен. Приймання виконується на одну

антену, приймач містить декодер просторово-часового коду. Зазначимо, що представлено систему типу MIMO (*nx1*), у якої обсяг рознесення *m=n*.

Приклади кодеру ПЧРК, конфігурації сигнального сузір'я восьмипозиційної фазової модуляції ФМ-8, і решітчастої діаграми коду показано на рисунку 4. На рис.4,а показано кодер згорткового коду з одиничною пам'яттю [9], двома виходами (*n=2*) і погоджувальними многочленами $g_1=5$ та $g_2=1$ (D – символ затримки). Кодування виконується в алфавіті алгебраїчного кільця Z8 (кільце цілих чисел з операціями додавання і множення за модулем 8). Під час синтезу такої сигнально-кодової конструкції широко застосовується ізоморфізм між символами алгебраїчного кільця Z_8 та сигналами фазової модуляції ФМ-8 [9]. Виходи кодера c_1 та c_2 під'єднано до відповідних входів передавачів рознесених передавальних антен. Для підвищення питомої швидкості передавання інформації в наведеному прикладі використовується фазова модуляція ФМ-8. Один крок решітчастої діаграми ПЧРК (рис. 4,в) включає набори попередніх і наступних станів кодеру (1...7) та гілок, які їх з'єднують.

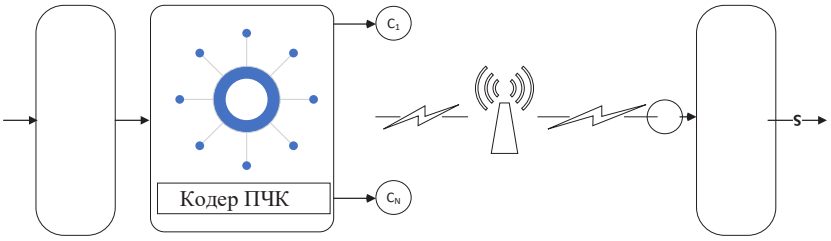


Рис. 3. Загальна модель роботи системи з просторово-часовим решітчастим кодуванням

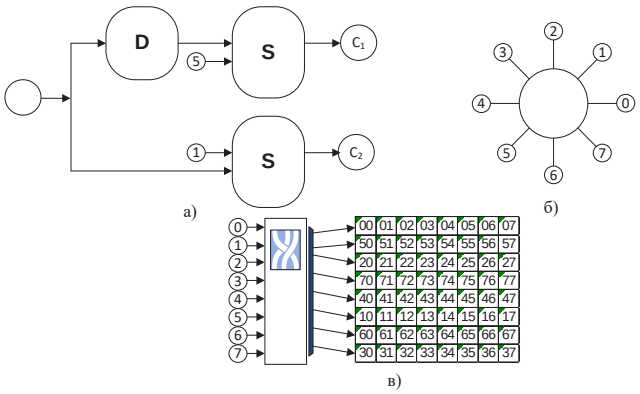


Рис. 4. Приклади кодеру ПЧРК: а – кодер, б – сигнальне сузір'я, в – решітчаста діаграма ПВРК

Принцип ПЧРК наведено на рисунку 5 – представляє систему кодування в каналі MIMO (2x1), що містить дві передавальні антени та одну приймальну. Відповідно до [6] вхідний потік символів, що передаються, розбивається на пари (c_1-c_2): на першому напівтактовому інтервалі символ c_1 відправляється через антену T_1, а символ c_2 – через антену T_2. На другому напівтактовому інтервалі порядок передачі змінюється: через антену T_1 передається інверсія символу c_2, а символ c_1 передається через антену T_2. Блоковий код Alamouti – це розташування символів *c* у вигляді матриці (рис.5) в структурі кодера [6].

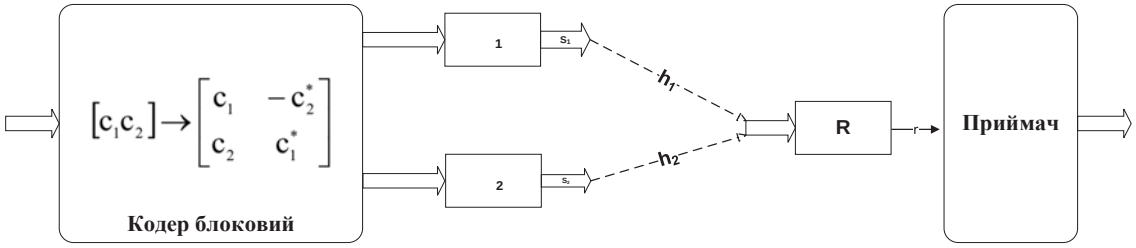


Рис. 5. Загальна модель роботи блокового кодування у каналі MIMO

Коефіцієнти передачі каналу на сусідніх інтервалах залишаються сталими. Теоретично завдання пошуку оптимальних блокових кодів просторово-часової обробки редукується до використання ортогональних матриць відповідних форматів.

З перших публікацій, які розкривають сутність та особливості просторово-часового кодування [4; 6; 17], розробники методів передавання інформації каналами з втратами швидко взяли методи ПЧК на озброєння, зважаючи на такі переваги [7]:

1. У просторово-часовому кодуванні для сигналів-переносників, на відміну від широкосмугових сигналів, не потрібно значного розширення смуги частот за умови збереження однакового рівня стійкості до завад. Ця перевага є критичною для операторів систем бездротового зв'язку в умовах зростаючого попиту на послуги бездротового зв'язку, де дефіцит спектра стає гострішим, а вартість виділення частотних смуг для бездротових систем постійно зростає.

2. Універсальність та гнучкість методів ПЧК [19] надають найкращі можливості для обміну енергоефективності на частотну ефективність у багатопроменевих каналах.

3. Можливість комбінування методів ПЧК разом з високошвидкісними сигналами цифрової модуляції гарантує високі показники частотної ефективності.

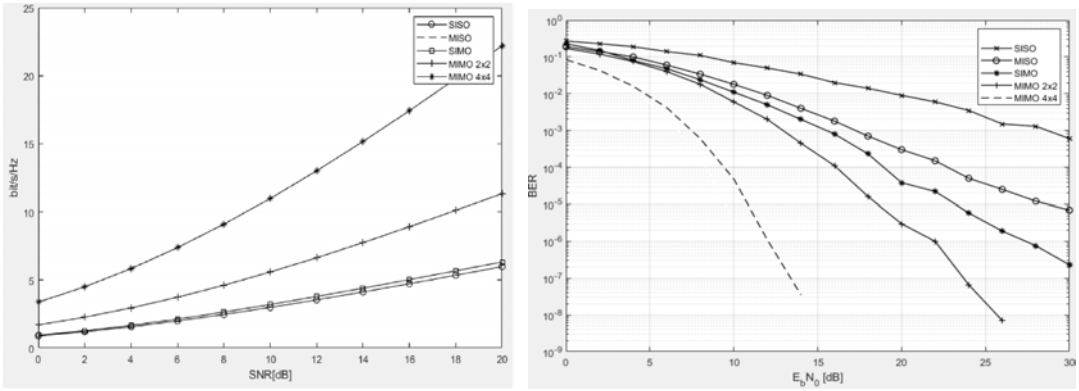
4. Можливість подальшого підвищення стійкості до перешкод у системах із ПЧК передбачається за умови впровадження адаптивного регулювання рівнів переданих сигналів.

5. Можливість вбудовування просторово-часових конструкцій у структуру сигналів у багатокористувацьких мережах.

Розглянемо алгоритми підвищення пропускної здатності мережі шляхом використання методів адаптивної просторової обробки сигналів та пошуку балансу між підвищенням пропускної здатності технології MIMO. Збільшення кількості незалежних радіоканалів призводить до зниження енергії на біт переданого повідомлення, що в свою чергу може підвищити ймовірність помилок на біт при прийнятті повідомлення.

Передбачається, що сигнали на передавальній стороні випромінюються одночасно та в одній смузі частот через M передавальних антен, а приймач має повну інформацію про характеристики каналу. Дані про параметри каналу передаються безпосередньо до пристрою декодування Аламоуті [4; 6], який витягує та інтерполює їх для отримання оцінки каналу для кожного переданого корисного символу. Під час моделювання, бінарний генератор Бернуллі створює випадковий двійковий сигнал, який подальше модулюється за допомогою різних методів: бінарна фазова маніпуляція (BPSK), квадратурна фазова маніпуляція (QPSK), 16-квадратурна амплітудна модуляція (16-QAM). Далі кодується кодером Аламоуті для передавання каналом MIMO з релеївськими завмираннями та адитивним білим шумом. Декодер Аламоуті [4; 6] об'єднує сигнали, отримані від приймальних антен, у єдиний потік для проведення демодуляції, а блок обчислення помилок порівнює демодульовані дані з вхідними.

Під час моделювання систем зв'язку (SISO – одна передавальна антена, одна приймальна антена; SIMO – одна передавальна антена, дві приймальні антени; MISO – дві передавальні антени, одна приймальна антена; MIMO 2×2 – дві передавальні антени, одна приймальна антена [22]; MIMO 2×2 – дві передавальні антени, дві приймальні антени; MIMO 4×4 – чотири передавальні антени, дві приймальні антени; MIMO 4×4 – чотири передавальні антени, чотири приймальні антени) всі вихідні параметри, такі як частота дискретизації, методи модуляції, швидкість передавання та ін., обрані в повній відповідності зі стандартом системи радіозв'язку IEEE 802.16e. Результати дослідження [18] відображено на рисунках 6, 7.

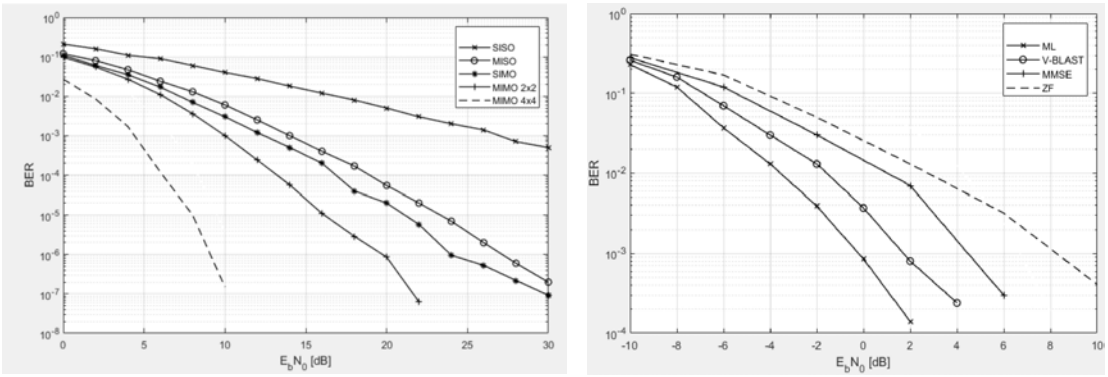


а) б)

Рис. 6. Залежність пропускної здатності систем від відношення сигнал/шум (а) та імовірності помилки на біт від відношення сигнал/шум для різних варіантів рознесення на прийомі та модуляція QPSK (б)

Видно, що в бездротових системах зв'язку швидкість передавання інформації помітно зростає лише зі збільшенням кількості передавальних і приймальних антен [13].

Для ймовірності помилки 10^{-4} виграш у завадостійкості на 3 дБ мають система SIMO над MISO, система MIMO 2×2 над SIMO. За тієї самої ймовірності помилки виграш у завадостійкості системи MIMO 4×4 над системою MIMO 2×2 збільшується в 2 рази і становить 6,5 дБ. Порівняння характеристик для варіантів з однією передавальною антеною і двома приймальними антенами (рознесення на прийомі) та з двома передавальними антенами та однією приймальною антеною (рознесення на передачі) свідчить, що рознесення на прийомі забезпечує додатковий виграш у 3 дБ. Також рознесення на прийомі та передачі дозволяє отримати додатковий виграш у 3 дБ порівняно з рознесенням на прийомі. Отже, зі збільшенням кількості антен на передачі та прийомі поліпшується захищеність від завад [8].



а) б)

Рис. 7. Залежність імовірності помилки на біт від відношення сигнал/шум для різних варіантів рознесення на прийомі та передачі (а) та імовірності помилки на біт від відношення сигнал/шум за різними алгоритмами демодуляції (б)

Використовуючи модуляцію BPSK можна поліпшити завадостійкість. Для ймовірності помилки 10^{-4} SIMO виграє за завадостійкістю в системи MISO 3,2 дБ; система MIMO 2×2 виграє у SIMO 3,3 дБ; система MIMO 4×4 виграє у MIMO 2×2 7 дБ. Також додатковий виграш 3,2 дБ забезпечує рознесення на прийомі, рознесення на прийомі та передачі – 3,3 дБ, а збільшення кількості антен на передачі та на прийомі може поліпшити завадостійкість [20].

Алгоритми обробки сигналів у MIMO-системах без зворотного зв'язку відрізняються способом поділу переданих символів у приймальних антенах [14].

Основні з них [10; 24]: метод зведення до нуля (Zero Forcing (ZF)); оцінка за мінімумом середньоквадратичної помилки (МСКО-приймач); максимально правдоподібна (МП) оцінка прийнятих символів (МП-приймач); алгоритм BLAST (Bell Laboratories Layered Space-Time) просторового декодування, зокрема, вертикальний BLAST (V-BLAST).

Для ймовірності помилки 10^{-3} відношення сигнал/шум при демодуляції алгоритмом МП становить 0 дБ; алгоритмом V-BLAST – 1,8 дБ; алгоритмом МСКО – 4,5 дБ; алгоритмом ZF – 8,2 дБ. Алгоритм МП забезпечує найменше значення ймовірності помилки [1] порівняно з іншими, а відповідно, має найкращі властивості завадостійкості. Алгоритм МСКО не поступається в завадостійкості іншим і має меншу обчислювальну складність [21; 3].

Висновки. Після проведення досліджень можна зробити наступні висновки. Просторово-часове кодування успішно поєднує переваги методів просторового рознесення з можливостями виправлення помилок коригувальним кодом при застосуванні оптимальних алгоритмів декодування. Це реалізує одночасно оптимальний алгоритм об'єднання рознесених сигналів. Ефективність проведених досліджень та розроблених нових методів ПЧК для перспективних систем передачі даних значною мірою залежить від того, наскільки моделі каналів відповідають реальним умовам передачі. У статистичній теорії зв'язку широко та ефективно використовуються методи теорії інформації. Просторово-часові блокові коди надають значний виграш від рознесення за рахунок можливості забезпечення великого обсягу рознесення. Просторово-часові ґратчасті коди забезпечують значний енергетичний виграш у каналах із завмираннями завдяки кодуванню, проте мають обмежений виграш від рознесення через обмежений обсяг цього процесу.

Можливість подальшого підвищення завадостійкості решітчастих ПЧК стає можливою за умови використання як формувачів решітки кодерів сигнально-кодових конструкцій із нового перспективного

класу рекурсивних згорткових кодів [7]. Один з потенційно перспективних методів включає в себе синтез згортково-блокових сигнально-кодових конструкцій, де внутрішні сигнали використовуються з класу просторово-часових блокових кодів, а зовнішні – з кодів складної архітектури.

Застосування просторово-часового кодування в системах Wi-Fi зв'язку, підкріплене теоретичним аналізом та результатами моделювання, підтверджує його ефективність у зменшенні впливу завмирання на сигнали та покращенні якості і пропускної здатності. При використанні методів рознесення спостерігається покращення відношення сигнал/шум до десяти децибелів, що свідчить про перевагу цієї технології для систем зв'язку, які потребують підвищення пропускної здатності та ергодичної ємності, а також зменшення розмірів кінцевих пристроїв. Ефективність поліпшення відношення сигнал/шум виявляється при збільшенні кількості антен з однієї до двох як на передавальній, так і на приймальній стороні.

Список використаних джерел:

1. Губарев В.Ф., Романенко В.Д. Етапи та основні завдання столітньої теорії контролю і розробка системи ідентифікації. Частина 3. Проблема ідентифікації складних систем за неточними даними. *International Scientific Technical Journal "Problems of Control and Informatics"*. 69, 1 (Груд 2023), 5–23. DOI: 10.34229/1028-0979-2024-1-1.
2. Іщенко М.О. Сигнально-кодові конструкції для систем безпроводового зв′язку з просторово-часовим кодуванням: Автореф. дис. к.т.н. /. ОНАЗ. – Одеса, 2009. 150 с.
3. Могилевич Д., Погребняк Л. Аналітична модель OFDM-MIMO сигналу у нестаціонарних каналах зв’язку із завмираннями. *Collection "Information Technology and Security"*. Vol. 11. Iss. 1 (20) (Jun. 2023), 39–46. DOI: 10.20535/2411-1031.2023.11.1.283538.
4. Alamouti S. M. A Simple Transmit Diversity Technique for Wireless Communications. *IEEE J. Select. Areas Communication*. –Vol. 16. No. 8.– 1998. – P. 1451–1458.
5. Banket V.L. Downlink Processing Algorithms for Multi-Antenna Wireless Communications. *IEEE Communications Magazine*. – 2005. No.1 P. 45–48.
6. Calderbank A.R. Space-Time Block Coding from Orthogonal Designs. *IEEE Trans. on Inform. Theory*. – 1999. – Vol. 45. – No. 5. P. 1456–1467.
7. Calderbank A.R. Space-time coding and signal processing for high data rate wireless communications. *Wireless, Communications and Mobile Computing*. – 2001. – No.1. P. 13–34.
8. Duman T.M., Ghrayeb A. Coding for MIMO Communication Systems. – Chichester, UK: John Wiley & Sons, 2007. 338 p.
9. Erceg V. Channel models for fixed wireless applications. *IEEE Tech.Report*, IEEE 802.16 Work Group, 2001.
10. Erceg V. Channel models for fixed wireless applications. *Revised Version of the document IEEE 802.16.3c-01/29r4. IEEE Tech.Report,* IEEE 802.16 Work Group, 2003.
11. Feher K. Wireless digital communications. New Jersey: Prentice-Hall PTR. 1999. 520 p.
12. Foschini G. Layered space – time architecture for wireless communication in a fading environment when using multielement antennas. *Bell Laboratories Technical Journal*. – 1996. – Vol. 4, Autum. P. 41–59.
13. George T. MIMO System Technology for Wireless Communications. – CRC Press, USA, 2006.
14. Gesbert D. From Theory to Practice: An Overview of MIMO Space – Time Coded Wireless Systems. *IEEE Journal on selected areas in communications*. – 2003.– Vol. SAC – 21, No.3. P. 281–302.
15. History of MIMO in radiocommunications. URL: http://en.wikipedia.org/wiki/MIMO.
17. Hohwald B., Marzetta T. Systematic Design of Unitary Space – Time Constellations. *IEEE Trans. on Inform. Theory*. – 2000. – Vol. 46. – No. 6. P. 1962–1973.
18. Jankiraman M. Space-time codes and MIMO systems. – Artech House, 2004. P. 344.
19. Lau B. K., Ow S. M. S., Kristensson G., Molisch A. F. Capacity Analysis for Compact MIMO Systems. *IEEE Vehicular Technology Conference (VTC)* (ISSN; 1550-2251). – IEEE Xplore, 2005. – Vol. 1. P. 165–170.
20. Lee S.J. et al. A Space-Time Code with full Diversity and Rate 2 for 2 Transmit Antenna Transmission. IEEE C802.16e-04/434r2, 2004.
21. Mohammed M. A., Vodichev V. Modeling of MIMO systems with universal controller. *Electrotechnic and Computer Systems*. 37 (113) (Sep. 2023), 26–32. DOI: 10.15276/eltecs.37.113.2023.03.
22. Shcherbina A.A. Effect of antenna mutual coupling on MIMO channel capacity. *IX International Conference on Antenna Theory and Tech-niques (ICATT-2013)*, Odesa, Ukraine 16-20 September 2013, p. 178–180.
23. Wang D. Super-Ortogonal Differerntial Trellis Coding and Decoding. *IEEE Journal on Selected Areas in Communications* – 2003 – Vol. 23, No.9. P. 1768–1798.
24. WLAN Tests: According to Standard IEEE 802.11a/b/g. Rohde & Schwarz GmbH & Co. K URL: https://cdn.rohde–schwarz.com/pws/dl_downloads/dl_application/application_notes/1ma69/1MA69_2e.pdf.

НАУКОВЕ ВИДАННЯ

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ ТА СУСПІЛЬСТВО

INFORMATION TECHNOLOGY AND SOCIETY

ВИПУСК 1 (12)
ISSUE 1 (12)

2024

Коректура
Ірина Чудеснова

Комп'ютерна верстка
Наталія Кузнєцова

Формат 60x84/8. Гарнітура Cambria.
Папір офсет. Цифровий друк.
Підписано до друку 25.04.2024.
Ум. друк. арк. 11,16. Замов. № 0624/451. Наклад 300 прим.

Видавництво і друкарня – Видавничий дім «Гельветика»
65101, Україна, м. Одеса, вул. Інглезі, 6/1
Телефон +38 (095) 934 48 28, +38 (097) 723 06 08
E-mail: mailbox@helvetica.ua
Свідоцтво суб'єкта видавничої справи
ДК No 7623 від 22.06.2022 р.