

ISSN 2663-3051
(ISSN 0555-2656 до 2019 р.)

БІОНІКА ІНТЕЛЕКТУ

ІНФОРМАЦІЯ, МОВА, ІНТЕЛЕКТ

№ 1 (102)

2025

НАУКОВО-ТЕХНІЧНИЙ ЖУРНАЛ

Заснований у жовтні 1967 р.

Засновник та видавець
Харківський національний університет радіоелектроніки

Періодичність видання – 2 рази на рік



Науково-технічний журнал
«БІОНІКА ІНТЕЛЕКТУ»

ISSN 2663-3051

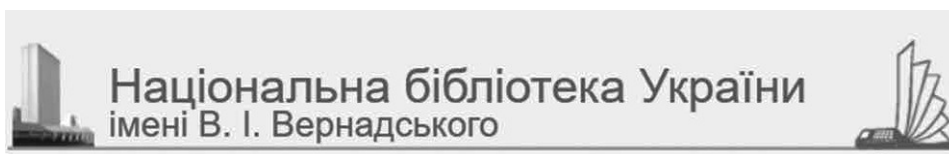
Заснований Харківським національним університетом
радіоелектроніки у 1967 році

Реферування та індексування:

Google Scholar



INDEX  COPERNICUS
I N T E R N A T I O N A L



Журнал включено до списку наукових спеціалізованих видань України
з технічних та фізико-математичних наук
згідно з наказом Міністерства освіти і науки України № 820 від 11.07.2016
(внесено зміни згідно з наказом МОНУ № 920 від 26.06.2024)

Alona Dorozhynska¹, Sergiy Dorozhynskyy²¹Ukrainian Lingua-Information Fund of NAS of Ukraine, Kyiv, Ukraine,
alonadrzh@gmail.com, ORCID iD: 0000-0001-6554-6731²Kyiv University of Intellectual Property and Law, Kyiv, Ukraine,
dorozhun1706@gmail.com, ORCID iD: 0000-0002-5395-6423**FROM DICTIONARY TO DATABASE: TRANSFORMATION AND PROTECTION**

The article outlines the importance of data digitization in modern society and in the context of countering the aggressor. It also shows the problems associated with paper-based media that lose their relevance and become unsuitable for work over time. The study proposes to use LiteDB as a lightweight NoSQL database for digitizing information due to its ease of use, embeddability, JSON/BSON data format, security assessment, and easy scalability. Particular emphasis is placed on the use of the HTMLAgilityPack library for working with HTML in the .NET environment. It is noted that it helps to parse HTML pages, extract structured data, and convert HTML to a structured format. An important priority of these tools is the ability to use built-in or connected information security systems, the best option of which is quantum cryptography and quantum key distribution. The advantage of such protection is its resistance to most modern attacks, even under conditions of uncertainty.

DATA DIGITIZATION, LITEDB, NOSQL DATABASE, HTMLAGILITYPACK, QUANTUM CRYPTOGRAPHY, QUANTUM KEY DISTRIBUTION PROTOCOLS, INFORMATION SECURITY, PAPER MEDIA, DATABASE (DB), DATABASE ARCHITECTURE, JSON/BSON FORMAT, PARSING, XML

А. Дорожнська, С. Дорожинський. Від словника до бази даних: трансформація та захист. У статті окреслено важливість оцифрування даних у сучасному суспільстві та в контексті протидії агресору. Також показано проблеми, пов'язані з паперовими носіями, які з часом втрачають свою актуальність і стають непридатними для роботи. У дослідженні пропонується використовувати `litedb` як легку `nosql`-базу даних для оцифрування інформації завдяки простоті використання, можливості вбудовування, формату даних `json/bson`, оцінці безпеки та легкій масштабованості. Особливий акцент зроблено на використанні бібліотеки `htmlagilitypack` для роботи з `html` в середовищі `.net`. Зазначається, що вона допомагає аналізувати `html`-сторінки, витягувати структуровані дані та конвертувати `html` у структурований формат. Важливим пріоритетом цих інструментів є можливість використання вбудованих або підключених систем захисту інформації, найкращим варіантом яких є квантова криптографія та квантовий розподіл ключів. Перевагою такого захисту є його стійкість до більшості сучасних атак, навіть в умовах невизначеності.

ОЦИФРУВАННЯ ДАНИХ, LITEDB, БАЗА ДАНИХ NOSQL, HTMLAGILITYPACK, КВАНТОВА КРИПТОГРАФІЯ, ПРОТОКОЛИ КВАНТОВОГО РОЗПОДІЛУ КЛЮЧІВ, ІНФОРМАЦІЙНА БЕЗПЕКА, ПАПЕРОВІ НОСІЇ, БАЗА ДАНИХ (DB), АРХІТЕКТУРА БАЗИ ДАНИХ, ФОРМАТ JSON/BSON, ПАРСИНГ, XML

Introduction

An important solution for society in the current conditions of existence and counteraction to the aggressor is the digitization of data in paper format. The problem is that paper media lose their relevance, are little used, and become unsuitable for work and processing over time. Their digital counterparts are more time-resistant and more accessible to the user. One of the most important stages of the digitization process is the formation of a database, its architecture and content. It should be remembered that paper-based media are usually unstructured or semi-structured. Therefore, databases that are easy to use and fill are suitable for such tasks. That's why we chose LiteDB, a lightweight NoSQL database designed for use in .NET applications, including C# and .NET Core. It can be an attractive option for some digitized data tasks for the following reasons:

- Ease of use: LiteDB has a simple API and integrates easily with .NET projects. Its syntax and data structure are well suited to working with C#.
- Embeddability: LiteDB can be easily embedded into an application because it does not require a separate

database server to be installed. You just need to add the library to the project and start using it.

- JSON/BSON data format: LiteDB uses the BSON (Binary JSON) data format, which makes it efficient for working with JSON objects. This is especially useful if you need to work with unstructured or semi-structured data.

- Security assessment: LiteDB has support for database encryption, which can be an important aspect for security, as it is possible to protect data with custom encryption keys, as well as apply the latest security methods, such as quantum cryptography and quantum key distribution.

- Easy scalability: In cases where the project requires a small database or does not require complex operations, LiteDB can be the best option due to its lightness and efficiency.

The use of this database makes it possible to use a large number of libraries to simplify the process of data digitization. One of them is HTMLAgilityPack, a library for working with HTML in the .NET environment. Its main purpose is to parse and process HTML documents. In the context of digitized data and databases, HTMLAgilityPack can be useful for extracting data before

saving it to a database. The main aspects of the benefits of using this library:

- Parsing HTML pages: HTMLAgilityPack provides a convenient way to access various elements of an HTML page, extracting text, attributes, and other data.

- Extraction of structured data: If you are searching for specific data on a website, you can use HTMLAgilityPack to sample and extract that data from the HTML code.

- Converting HTML to a structured format: If you receive an HTML document and need to convert it to a structured format for further processing or storage in a database, HTMLAgilityPack can be used to parse and create objects or data for further use.

HTMLAgilityPack helps to simplify the work with HTML and use the resulting data for further processing or storage in a database, which can be useful for obtaining information for further use in digitized data.

1. Basic technological stages

Conversion of electronic text of dictionaries into XML format

Converting electronic text of dictionaries into XML can be done by programming. However, this process is preceded by the extraction of all structural elements and the construction of the structure of a dictionary entry or text unit. Here is a general scheme of working with XML:

1. Reading the text. Read the electronic text of the dictionary according to the structure of its lexicography system.

2. Parsing the text. Parsing the text into various structural elements of the dictionary, such as words, definitions, examples of usage, etc. To do this, you can use regular expressions or text processing libraries such as NLTK (Natural Language Toolkit) in Python.

3. Create an XML structure. For each selected element, create a corresponding XML element, for example, <word>, <definition>, <example>, etc. Add these elements to the XML tree. Develop a general schema.

4. Save in XML format. Save the resulting XML structure in an XML file.

5. Additional options. Add the ability to process different dictionary formats (for example, JSON, CSV). Add validation of the entered data to make sure that it corresponds to the expected format.

This work can be implemented using programming languages such as Python, Java, or any other language that supports text and XML processing.

Converting electronic text from dictionaries to XML can be useful for several reasons:

1. Structured data: XML allows you to structure information in the form of a tree, which makes it easier to work with data and use it later.

2. Interoperability with other systems: XML is a common format and is used to exchange data between different programs and systems.

3. Ease of processing and analysis: The XML structure makes it easy to process and analyze data using a variety of programs, including programming languages and other tools.

4. Support for different formats: XML can be used to represent different kinds of information, including texts, numbers, dates, and more, which can be useful for dictionary data.

5. Extensibility: XML makes it easy to add new elements and extend the data structure without significant changes to the existing code.

Consequently, converting dictionary data to XML can make it easier to process, share, and use in different applications and systems.

For example, let's show two dictionary entries in the structure of a lexicographic system and in XML format. (Dictionary of Ukrainian Biological Terminology)

Example 1

адвенти#вний (рос. адвенти#вный, англ. adventive) 1. Який розвивається не з ембріональних тканин точки росту, а із старіших частин рослини; **адвенти#вна бру#нька** див. **бру#нька**: **бру#нька адвенти#вна**; **адвенти#вна ембріоні#я** див. **ембріоні#я**: **ембріоні#я адвенти#вна**; **адвенти#вна поліембріоні#я** див. **поліембріоні#я**: **поліембріоні#я нуцеля#рна [адвенти#вна]**; **адвенти#вний за#родок** див. **за#родок**: **за#родок адвенти#вний**; **адвенти#вний о#рган** див. **о#рган**: **о#рган адвенти#вний**; **адвенти#вний па#гін** див. **па#гін**: **па#гін адвенти#вний** 2. Занесена людиною рослина в ту місцевість, де вона раніше не росла; **адвенти#вна росли#на** див. **росли#на**: **росли#на адвенти#вна**; **адвенти#вна вид** див. **вид**: **вид адвенти#вний**.

According to the structure of the dictionary article, we will highlight all the structural elements

ТБ [термінологічний блок]: **адвенти#вний** (рос. адвенти#вный, англ. adventive)

ТК_У [термінологічний комплекс український]: **адвенти#вний**

ЗТ [заголовний термін]: **адвенти#вний**

ТК_Р [термінологічний комплекс російський]: *рос.* адвенти#вный

ММ[маркер мови]: *рос.*

Т_Р [термін російський]: *рос.* адвенти#вный

ТК_А [термінологічний комплекс англійський]: *англ.* adventive

ММ [маркер мови]: *англ.*

Т_А [термін англійський]: *англ.* adventive

СМБ [семантичний блок]: 1. Який розвивається не з ембріональних тканин точки росту, а із старіших частин рослини; **адвенти#вна бру#нька** див. **бру#нька**: **бру#нька адвенти#вна**; **адвенти#вна ембріоні#я** див.

ембріоні#я: **ембріоні#я адвенти#вна;** **адвенти#вна поліембріоні#я див. поліембріоні#я:** **поліембріоні#я нуцеля#рна [адвенти#вна]; адвенти#вний за#родок див. за#родок:** **за#родок адвенти#вний;** **адвенти#вний о#рган див. о#рган:** **о#рган адвенти#вний;** **адвенти#вний па#гін див. па#гін:** **па#гін адвенти#вний 2.** Занесена людиною рослина в ту місцевість, де вона раніше не росла; **адвенти#вна росли#на див. росли#на:** **росли#на адвенти#вна;** **адвенти#вна вид див. вид:** **вид адвенти#вний.**

БТ₁ [блок тлумачень]: 1. Який розвивається не з ембріональних тканин точки росту, а із старіших частин рослини;

НТ [номер тлумачення]: 1

ТЛ₁ [тлумачення]: Який розвивається не з ембріональних тканин точки росту, а із старіших частин рослини;

БТ₂ [блок тлумачень]: 2. Занесена людиною рослина в ту місцевість, де вона раніше не росла;

НТ [номер тлумачення]: 2

ТЛ₂ [тлумачення]: Занесена людиною рослина в ту місцевість, де вона раніше не росла;

БП₁ [блок посилань]: **адвенти#вна бру#нька див. бру#нька:** **бру#нька адвенти#вна;** **адвенти#вна ембріоні#я див. ембріоні#я:** **ембріоні#я адвенти#вна;** **адвенти#вна поліембріоні#я див. поліембріоні#я:** **поліембріоні#я нуцеля#рна [адвенти#вна]; адвенти#вний за#родок див. за#родок:** **за#родок адвенти#вний;** **адвенти#вний о#рган див. о#рган:** **о#рган адвенти#вний;** **адвенти#вний па#гін див. па#гін:** **па#гін адвенти#вний**

ПБП₁ [підблок посилань]: **адвенти#вна бру#нька див. бру#нька:** **бру#нька адвенти#вна;**

ПБП₂ [підблок посилань]: **адвенти#вна ембріоні#я див. ембріоні#я:** **ембріоні#я адвенти#вна;**

ПБП₃ [підблок посилань]: **адвенти#вна поліембріоні#я див. поліембріоні#я:** **поліембріоні#я нуцеля#рна [адвенти#вна];**

ПБП₄ [підблок посилань]: **адвенти#вний за#родок див. за#родок:** **за#родок адвенти#вний;**

ПБП₅ [підблок посилань]: **адвенти#вний о#рган див. о#рган:** **о#рган адвенти#вний;**

ПБП₆ [підблок посилань]: **адвенти#вний па#гін див. па#гін:** **па#гін адвенти#вний**

БП₂ [блок посилань]: **адвенти#вна росли#на див. росли#на:** **росли#на адвенти#вна;** **адвенти#вна вид див. вид:** **вид адвенти#вний.**

ПБП₁ [підблок посилань]: **адвенти#вна росли#на див. росли#на:** **росли#на адвенти#вна;**

ПБП₂ [підблок посилань]: **адвенти#вна вид див. вид:** **вид адвенти#вний.**

Each sub-block consists of an addressee1, a link marker and an addressee2.

ПБП₂ [підблок посилань]: **адвенти#вна вид див. вид:** **вид адвенти#вний.**

САНТ [адресант]: **адвенти#вна вид**

МП[маркер посилань]: **див.**

САТ [адресат]: **вид: вид адвенти#вний.**

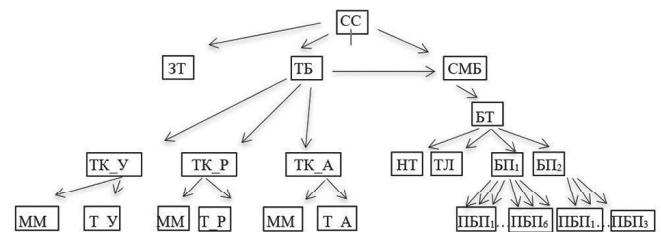


Figure 1: Outline of a dictionary article

Example 2

перетя#жка (*рос.* перетя#жка, *англ.* constriction) 1. Неспіралізована ділянка спіральної хромосоми, величина і локалізація якої варіюється; 2. (*англ.* strangulation) Дуже тонке, звужене місце; **перетя#жка втори#нна** (*рос.* перетя#жка втори#нная, *англ.* secondary chromosome strangulation) 1. Звужена частина, яка може локалізуватись у різних місцях хромосоми.

According to the structure of the dictionary article, we will highlight all the structural elements

ТБ₁ [термінологічний блок]: **перетя#жка** (*рос.* перетя#жка, *англ.* constriction)

ТК_У [термінологічний комплекс український]: **перетя#жка**

ЗТ [заголовний термін]: **перетя#жка**

ТК_Р [термінологічний комплекс російський]: *рос.* перетя#жка

ММ[маркер мови]: *рос.*

Т_Р [термін російський]: перетя#жка

ТК_А [термінологічний комплекс англійський]: *англ.* constriction

ММ [маркер мови]: *англ.*

Т_А [термін англійський]: constriction

ТБ₂ [термінологічний блок]: 2. (*англ.* strangulation)

ТК_У [термінологічний комплекс український]: **перетя#жка**

ЗТ [заголовний термін]: **перетя#жка**

ТК_Р [термінологічний комплекс російський]: *рос.* перетя#жка

ММ[маркер мови]: *рос.*

Т_Р [термін російський]: перетя#жка

ТК_А [термінологічний комплекс англійський]: *англ.* strangulation

ММ [маркер мови]: *англ.*

Т_А [термін англійський]: strangulation

СМБ₁ [семантичний блок]: 1. Неспіралізована ділянка спіральної хромосоми, величина і локалізація якої варіюється;

БТ [блок тлумачень]: 1. Неспіралізована ділянка спіральної хромосоми, величина і локалізація якої варіюється;

НТ [номер тлумачення]: 1

ТЛ [тлумачення]: Неспіралізована ділянка спіральної хромосоми, величина і локалізація якої варіюється;

СМБ₂ [семантичний блок]: Дуже тонке, звужене місце; **перетя#жка втори#нна** (рос. перетя#жка втори#нная, англ. secondary chromosome strangulation) 1. Звужена частина, яка може локалізуватись у різних місцях хромосоми.

БТ [блок тлумачень]: Дуже тонке, звужене місце

НТ [номер тлумачення]: 2

ТЛ [тлумачення]: Дуже тонке, звужене місце

ПБТС [підблок словосполучень]: **перетя#жка втори#нна** (рос. перетя#жка втори#нная, англ. secondary chromosome strangulation) 1. Звужена частина, яка може локалізуватись у різних місцях хромосоми.

БТС [блок словосполучень]: **перетя#жка втори#нна** (рос. перетя#жка втори#нная, англ. secondary chromosome strangulation) 1. Звужена частина, яка може локалізуватись у різних місцях хромосоми.

ТБСЛ [термінологічний блок словосполучень]: **перетя#жка втори#нна** (рос. перетя#жка втори#нная, англ. secondary chromosome strangulation)

ТКС_У [термінологічний комплекс словосполучення укр.]: **перетя#жка втори#нна**

ТС_У [термінологічне словосполучення укр.]: **перетя#жка втори#нна**

ТКС_Р [термінологічний комплекс словосполучення рос.]: *рос.* перетя#жка втори#нная

ММ [маркер мови]: *рос.*

ТС_Р [термінологічне словосполучення рос.]: перетя#жка втори#нная

ТКС_А [термінологічний комплекс словосполучення англ.]: *англ.* secondary chromosome strangulation

ММ [маркер мови]: *англ.*

ТС_А [термінологічне словосполучення англ.]: secondary chromosome strangulation

БТСЛ [семантичний блок словосполучень]: 1. Звужена частина, яка може локалізуватись у різних місцях хромосоми.

НТСЛ [номер тлумачення словосполучення]: 1

ТЛС [тлумачення словосполучення]: Звужена частина, яка може локалізуватись у різних місцях хромосоми.

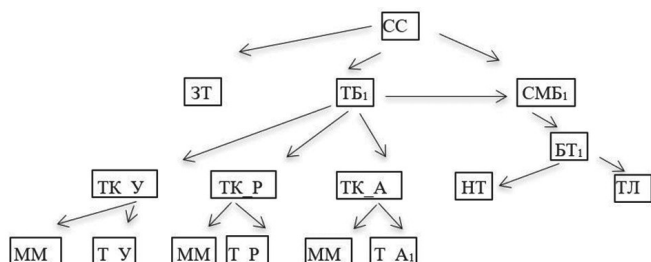


Figure 2.1: The scheme of articles is divided into two blocks, first

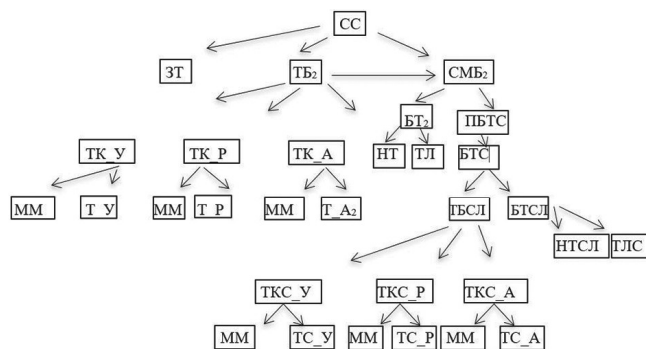


Figure 2.2: The scheme of articles is divided into two blocks, second

Example 1 to XML

```
<СС>
  <текст_СС><![CDATA[<В>адвенти#вний</В>
  (<I>рос.</I> адвенти#вный, <I>англ.</I> adventive) 1.
  Який розвивається не з ембріональних тканин точки
  росту, а із старіших частин рослини; <В>адвенти#вна
  бру#нька</В> <I>див.</I> <В>бру#нька: бру#нька
  адвенти#вна</В>; <В>адвенти#вна ембріоні#я</
  В> <I>див.</I> <В>ембріоні#я: ембріоні#я
  адвенти#вна</В>; <В>адвенти#вна поліембріоні#я</
  В> <I>див.</I> <В>поліембріоні#я: поліембріоні#я
  нуцеля#рна [адвенти#вна]</В>; <В>адвенти#вна
  за#родок</В> <I>див.</I> <В>за#родок: за#родок
  адвенти#вний</В>; <В>адвенти#вний о#рган</В>
  <I>див.</I> <В>о#рган: о#рган адвенти#вний</
  В>; <В>адвенти#вний па#гін</В> <I>див.</I>
  <В>па#гін: па#гін адвенти#вний</В> 2. Занесена
  людиною рослина в ту місцевість, де вона раніше не
  росла; <В>адвенти#вна росли#на</В> <I>див.</
  I> <В>росли#на: росли#на адвенти#вна</В>;
  <В>адвенти#вна вид</В> <I>див.</I> <В>вид: вид
  адвенти#вний</В>.]></текст_СС>
  <ЗТ>адвенти#вний</ЗТ>
  <ТБ номер="1">
    <ТК_У номер="1">
      <Т_У>адвенти#вний</Т_У>
      <ММ>укр.</ММ>
    </ТК_У>
    <ТК_Р номер="1">
      <Т_Р>адвенти#вный</Т_Р>
      <ММ>рос.</ММ>
    </ТК_Р>
    <ТК_А номер="1">
      <Т_А>adventive</Т_А>
      <ММ>англ.</ММ>
    </ТК_А>
  </ТБ>
  <СМБ номер="1">
    <БТ номер="1">
      <НТ>1</НТ>
      <ТЛ>Який розвивається не з ембріональних тка-
      нин точки росту, а із старіших частин рослини</ТЛ>
```

```

</БТ>
<БТ номер="2">
  <НТ>2</НТ>
  <ТЛ>Занесена людиною рослина в ту місцевість,
де вона раніше не росла</ТЛ>
</БТ>
<БП номер="1">
  <САНТ>адвенти#вна бру#нька</САНТ>
  <САТ>бру#нька: бру#нька адвенти#вна</САТ>
  <МП>див.</МП>
</БП>
<БП номер="2">
  <САНТ>адвенти#вна ембріоні#я</САНТ>
<САТ>ембріоні#я: ембріоні#я адвенти#вна</САТ>
  <МП>див.</МП>
</БП>
<БП номер="3">
  <САНТ>адвенти#вна поліембріоні#я</САНТ>
<САТ>поліембріоні#я: поліембріоні#я нуцеля#рна
[адвенти#вна]</САТ>
  <МП>див.</МП>
</БП>
<БП номер="4">
  <САНТ>адвенти#вна за#родок</САНТ>
<САТ>за#родок: за#родок адвенти#вний</САТ>
  <МП>див.</МП>
</БП>
<БП номер="5">
  <САНТ>адвенти#вний о#рган</САНТ>
<САТ>о#рган: о#рган адвенти#вний</САТ>
  <МП>див.</МП>
</БП>
<БП номер="6">
  <САНТ>адвенти#вний па#гін</САНТ>
<САТ>па#гін: па#гін адвенти#вний</САТ>
  <МП>див.</МП>
</БП>
<БП номер="7">
  <САНТ>адвенти#вна росли#на</САНТ>
<САТ>росли#на: росли#на адвенти#вна</САТ>
  <МП>див.</МП>
</БП>
<БП номер="8">
  <САНТ>адвенти#вна вид</САНТ>
  <САТ>вид: вид адвенти#вний</САТ>
  <МП>див.</МП>
</БП>
</СМБ>
</СС>

```

Example 2 to XML

```

<СС>
<текст_СС><![CDATA[<В>перетя#жка</В>
(<I>рос.</I> перетя#жка, <I>англ.</I> constriction)

```

1. Неспіралізована ділянка спіральної хромосоми, величина і локалізація якої варіюється; 2. (<I>англ.</I> strangulation) дуже тонке, звужене місце; <В>перетя#жка втори#нна</В> (<I>рос.</I> перетя#жка втори#чная, <I>англ.</I> secondary chromosome strangulation) звужена частина, яка може локалізуватись у різних місцях хромосоми.]]></текст_СС>

```

<ЗТ>перетя#жка</ЗТ>
<ТБ номер="1">
  <ТК_У номер="1">
    <Т_У>перетя#жка</Т_У>
    <ММ>укр.</ММ>
  </ТК_У>
  <ТК_Р номер="1">
    <Т_Р>перетя#жка</Т_Р>
    <ММ>рос.</ММ>
  </ТК_Р>
  <ТК_А номер="1">
    <Т_А>constriction</Т_А>
    <ММ>англ.</ММ>
  </ТК_А>
</ТБ>
<ТБ номер="2">
  <ТК_У номер="1">
    <Т_У>перетя#жка</Т_У>
    <ММ>укр.</ММ>
  </ТК_У>
  <ТК_Р номер="1">
    <Т_Р>перетя#жка</Т_Р>
    <ММ>рос.</ММ>
  </ТК_Р>
  <ТК_А номер="1">
    <Т_А>strangulation</Т_А>
    <ММ>англ.</ММ>
  </ТК_А>
</ТБ>
<СМБ номер="1">
<БТ номер="1">
  <НТ>1</НТ>
  <ТЛ>дуже тонке, звужене місце;</ТЛ>
</БТ>
<БТС номер="1">
  <ТБСЛ номер="1">
    <ТКС_У номер="1">
      <ТС_У>перетя#жка втори#нна</ТС_У>
      <ММ>укр.</ММ>
    </ТКС_У>
    <ТКС_Р номер="1">
      <ТС_Р>перетя#жка втори#чная</ТС_Р>
      <ММ>рос.</ММ>
    </ТКС_Р>
    <ТКС_А номер="1">
      <ТС_А>secondary chromosome strangulation</
ТС_А>

```

```

    <ММ>англ.</ММ>
  </ТКС_А>
</ТБСЛ>
<БТСЛ номер="1">
  <НТСЛ>1</НТСЛ>
  <ТЛС>звужена частина, яка може локалізуватись
у різних місцях хромосоми</ТЛС>
  </БТСЛ>
</БТС>
</СМБ>
<СМБ номер="2">
  <БТ номер="1">
    <НТ>1</НТ>
    <ТЛ>Неспіралізована ділянка спіральної хромо-
соми, величина і локалізація якої варіюється</ТЛ>
  </БТ>
</СМБ>
</СС>
<СС>

```

2. Stages of database formation

Creating a database (DB) from an XML file can be done in several steps:

Creating a database schema. Define the data structure that needs to be stored in the database. The structure of the XML file is taken into account and the tables and fields for displaying this data are determined.

Creating tables in the database. Using SQL or another query language, tables are created in the database that correspond to the data structure from the XML file.

Loading data from XML. Tools or programming languages that support working with XML are used (for example, Python with the ElementTree library or .NET-based programming languages using XmlDocument). The XML file is parsed and the data is inserted into the database schema.

Using the HTMLAgilityPack, the entire input text is divided into articles and structural elements are extracted from them (stems).

Examples (Dictionary Lexicon of Polish and Ukrainian Active Phraseology):

```

HtmlDocument HtmlDoc = new HtmlDocument();
1. Load the text from the file HtmlDoc.
LoadHtml(input);

```

```

Selecting articles XmlNodeCollection Nodes =
HtmlDoc.DocumentNode.SelectNodes("//article");

```

After each selection of structural elements, perform this check

```

if (Nodes != null)
{
  foreach (XmlNode S_Node in Nodes)// Each article
  XmlNode tmp_node = S_Node.
  SelectSingleNode("./p[@class = 'j']"); // selecting one
element from the article by type <p class="t">. The dot
here means that we are "looking" at only one article

```

```

String text = tmp_node.InnerHtml; // Get the text
that is in (between) the tags

```

```

Int ID_Ukr = tmp_node.Attributes["href"].Value; //
We got the value of the attribute of type <... href="***"
...>

```

2. Transferring the received data to an instance of the class. With a check whether there is already a "reverse" article (1 copy for both pol-UKR and ukr-pol. We look to see if there is, for example, the same case among the processed ones and, if so, just add the ID and letter in the language) ATTENTION the class should have a separate one (for the database), like this:

```

[BsonId]
public int ID { get; set; }
3. Loading data into the database. Examples:
private LiteDatabase DB { get; set; }
private LiteCollection<ReestrUnitMember> R_
Members { get; set; }
private string DB_Path { get; set; }

```

```

DB_Path = path; // Name and (optionally) path of
the database file

```

```

DB = new LiteDatabase(path);
R_Members = DB.GetCollection<ReestrUnitMemb
er>("r_members"); // ReestrUnitMember - class name,
r_members - collection name

```

```

R_Members.EnsureIndex("Inner_ID", "$.Inner_
ID"); // Create an index. The 1st is the name of the in-
dex, the 2nd is the name of the parameter in the class, $.
is required

```

```

R_Members.EnsureIndex("Orig_Inner_ID", "$.Orig_
Inner_ID");

```

```

R_Members.Insert(Member); // Add a new instance
of the class (article) to the database.

```

Automation of the process. Consider automating this process, especially if you have a large number of XML files or if you plan to update data on a regular basis.

Verification and optimization. You need to make sure that the data is loaded correctly into the database and perform optimization, if necessary, such as using indexes to improve query performance.

The general conclusion is that the formation of a database from an XML file includes several key stages, from creating schema and tables to loading data and its further processing in the database, using appropriate tools and libraries.

3. Database protection

In today's digitalized society, an important criterion is the use of modern methods and means of protecting information, as any digitized data can be compromised or stolen. One of the most reliable methods of protection is the use of quantum key distribution protocols, one of which is Twin field (Table 1). The use of the Twin Field protocol to protect LiteDB involves the use of technology

based on the properties of a dual field of quantum states to create a secure key exchange channel. This allows you to achieve a high level of protection in the exchange of keys between the sender and the recipient in a quantum environment.

Table 1.

Principles of the TFmod quantum key distribution protocol

Criteria	Uncertainty principle	Measurement independence principle	Information transfer fidelity principle
1. Using quantum properties	Takes advantage of quantum uncertainty	Applies independent measurements	Guarantees flawless photon transmission
2. Security of key transfer	Uses quantum properties to protect keys	Ensures unconditional security of key transfer	Applies interception detection mechanisms
3. Interception detection	Detects any attempted interception	Detects any unauthorized interference	Provides detection of any changes or distortions
4. Measurement independence	The measurement is carried out independently	Ensures independence between measurements	Uses independent channels for photon transmission
5. Ensuring transmission accuracy	Uses error correction mechanisms	Provides high transmission accuracy	Applies error correction methods
6. Efficiency and speed	Uses optimized algorithms	Provides high transmission speed	Highly productive and efficient
7. Scalability	Implemented for large networks	Scalable for extended networks	Applies to different network scales

The essence of Twin Field is to use two parameters (polarization and time) to encode and transmit quantum bits (qubits). This provides a double level of information that can be used to create a cryptographically secure key. One of the important advantages of the protocol is its ability to detect key interception attempts. According to the principles of quantum mechanics, any attempt to measure a quantum state leads to its change, which is automatically detected by the system, ensuring the security of key exchange. In this context, the use of Twin Field to protect LiteDB includes the creation of cryptographically secure keys that are used to encrypt and decrypt the database. This approach helps to ensure a high level of protection and prevents attacks on encryption keys in a quantum environment.

To ensure the process of applying the Twin Field protocol to protect LiteDB, the following steps should be followed:

1. Quantum key exchange: A quantum key exchange process is performed between the sender and the receiver using Twin Field. During this exchange, the double field must be considered to provide an additional layer of information and security.

2. Interception detection: Using a property of quantum mechanics, such as the non-determinism of measurements, to detect any attempts to intercept keys. This can be realized by using quantum properties that change with any observation.

3. Encrypting and decrypting data in LiteDB: using the obtained quantum keys to encode and decode data. This will ensure the confidentiality and integrity of information in the database.

4. Ongoing monitoring and support of the system: monitoring the state of the system and detecting any anomalies or attack attempts in time. Ensure that keys are regularly updated and the system is improved to reflect the latest advances in quantum security.

When considering the integration of a quantum key distribution device on the same server as the LiteDB database, the key aspect is to ensure the comprehensive protection and efficiency of this infrastructure. For this purpose, it is important to address physical, technical, and network aspects. In terms of physical protection, you should place the device in a secure room using restricted access mechanisms.

Efficient use of server resources should include optimizing the configuration for quantum signal processing. It is also important to consider network aspects, using modern encryption tools and firewalls to protect against external attacks. Maintaining the continuous operation of the system includes the implementation of a monitoring and maintenance system to detect possible malfunctions and anomalies. The use of logging and auditing methods allows you to maintain event logs and identify potential security threats.

The level of modern security bypass systems should also be taken into account. level of modern security bypass systems should also be taken into account. Therefore, the standard Twin Field protocol should be enhanced with additional methods to ensure resilience. One of the best ways is to improve the security system using quantum identification and quantum state multiplexing methods, which includes several promising aspects. First of all, there is a high resistance to cyber threats due to the application of quantum identification principles to generate cryptographic keys. This is important because this approach makes it difficult to crack encryption algorithms based on classical computing. The second aspect is the use of quantum state multiplexing, which allows more information to be transmitted simultaneously over communication channels. This can lead to improved performance and efficient processing of large amounts of data. The third aspect involves the use of quantum states to transmit

keys, reducing the likelihood of interception. Observing a quantum state leads to a change in the state itself, which makes detecting attacks and hacking even more difficult. In addition, the use of quantum methods can increase the level of confidentiality and data security, as they are based on the physical principles of quantum mechanics. This approach makes the keys unique and unpredictable, which makes the system more relevant to modern cybersecurity challenges.

Conclusion

In this article, we have considered important aspects of forming digitized data in the LiteDB database, using the HTMLAgilityPack library to efficiently process HTML in the .NET environment and ensure the reliability of this system through the use of quantum cryptography.

Digitization of data is becoming an integral part of modern society, and using LiteDB provides us with a simple and effective solution for storing this data. Its embeddability, JSON/BSON data format, ease of use, and scalability make it an excellent choice for projects of all sizes.

The HTMLAgilityPack library allows you to efficiently process HTML pages in the .NET environment, making it an ideal tool for extracting structured information from web resources.

Special attention is paid to the use of quantum cryptography to ensure system security. Its high resistance to various cyber threats, even under conditions of uncertainty, makes it an advanced tool for protecting confidential information. An important aspect is the use of the Twin field protocol, which, due to its resistance to attacks under conditions of uncertainty, forms a reliable foundation for data security.

References

- [1] *Grodzinsky, D. M., Simonenko, L. O. et al.* (2012). Ukrainian biological terminology Dictionary. Kyiv: KMM, 2012.
- [2] Linguistic and information studies: works of the Ukrainian language and information fund of the NAS of Ukraine: in 5 volumes. (2018). V. A. Shyrokov et al. Vol. 1: Kyiv.
- [3] *Ben Bongalon, Joel Ilao, Ethel Ong, Rochelle Irene Lucas, Melvin Jabar.* (2021). Using Open-Source Tools to Digitise Lexical Resources for Low-Resource Languages. Proceedings of the eLex 2021 conference. 5–7 July 2021, virtual. Brno: Lexical Computing CZ, s.r.o.
- [4] *Shyrokov, V. A. (Ed.)* (2011). Computer lexicography. Kyiv: Naukova dumka
- [5] *Karpova, O.* (2009). Lexicography and Terminology: A Worldwide Outlook / Olga Karpova, Faina Kartashkova. Cambridge: Cambridge Scholars Publishing.
- [6] *Shyrokov V. A. etc.* (2018). Linguistic and information studies: works of the Ukrainian Language and Information Fund NAS of Ukraine: in 5 vols. Vol. 5: Virtualization of linguistic technologies. Kyiv: Ukrainian Lingua-Information Fund of NAS of Ukraine. URL: https://movoznavstvo.org.ua/files/Ling_inf_studio_TOM_5_umif_B5.pdf doi: 10.33190/978-966-02-8683-2/8690-0
- [7] *Baiisa, V., Blahuš, M., Cukr, M., Herman, O., et al.* (2019). Automating Dictionary Production: a Tagalog-English-Korean Dictionary from Scratch. Proceedings of the eLex 2019 conference. 1-3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, s.r.o.
- [8] *Shyrokov V.A.* (2018). Grammatical systems: phenomenological approach. V. A. Shyrokov, T. P. Lyubchenko, I. V. Shevchenko, K. V. Shyrokov. Kyiv: Naukova dumka.
- [9] *Gnatyuk, S., Okhrimenko, T., Dorozhynskyy, S., Fesenko, A.* (2019). Review of modern quantum key distribution protocols, Scientific and Practical Cyber Security Journal (SPCSJ), 4(1), 56–60. ISSN 2587-4667.
- [10] *Buck, L. E., Bodenheimer, B.* (2021). Privacy and Personal Space: Addressing Interactions and Interaction Data as a Privacy Concern. 2021 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Lisbon, Portugal, 399-400, doi: 10.1109/VRW52623.2021.00086.
- [11] *Kernerman, I.* (2015). A multilingual trilogy: Developing three multi-language lexicographic datasets. Electronic Lexicography in the 21st Century: Linking lexical data in the digital age. Proceedings of eLex 2015, 11–13 August 2015, 372-383. URL: <https://elex.link/elex2015/>
- [12] *Liu, X., Xie, J., Kang, J., Zhang, M.* (2022). Twin-Field Quantum Key Distribution Protocol with Heralded Single-Photon Source, 2022 IEEE 12th International Conference on Electronics Information and Emergency Communication (ICEIEC), Beijing, China, 94-97, doi: 10.1109/ICEIEC54567.2022.9835076.

The article was delivered to editorial staff on the 17.01.2025

О.В. Лозинська¹, О.О. Марків², В.А. Висоцька³¹НУЛП, м. Львів, Україна, olha.v.lozynska@lpnu.ua, ORCID iD: 0000-0002-5079-0544²НУЛП м. Львів, Україна, oksana.o.markiv@lpnu.ua, ORCID iD: 0000-0002-1691-1357³НУЛП, м. Львів, Україна, victoria.a.vysotska@lpnu.ua, ORCID iD: 0000-0001-6417-3689

МЕТОД ВИЯВЛЕННЯ ДЖЕРЕЛ ДЕЗІНФОРМАЦІЇ НА ОСНОВІ АНСАМБЛЕВИХ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ

У представленому дослідженні розроблено метод виявлення джерел дезінформації на основі ансамблевих моделей машинного навчання. Проаналізовано сучасні методи боротьби з дезінформацією та виявлення неправдивого контенту. В рамках роботи реалізовано систему ідентифікації фейків, побудовану на ансамблевому підході, а також описано її архітектурну структуру. Детально описано основні етапи очищення текстових даних, отриманих із соціальних мереж і новинних, зокрема нормалізацію категоріальних змінних. Проведено статистичний аналіз тексту та аналіз критеріїв виявлення джерел поширення дезінформації. Здійснено аналіз балансу цільових і допоміжних змінних, що дало змогу виявити залежності між мовою повідомлення та достовірністю. Для моделювання використано два різні типи текстових ембедингів та відповідні моделі класифікації: лінійну регресію та логістичну регресію. Підсумковим етапом стало застосування ансамблю моделей, що дало змогу поєднати прогностичну здатність обох моделей. Результати показали, що комбінація підходів покращує класифікаційну якість, особливо в умовах незбалансованих даних. Використання ансамблю моделей дало змогу збільшити точність з 73% (модель 1) та 71% (модель 2) до 78%.

ДЕЗІНФОРМАЦІЯ, ДАТАСЕТ, МАШИННЕ НАВЧАННЯ, АНСАМБЛЕВІ МОДЕЛІ, ЛІНІЙНА РЕГРЕСІЯ, ЛОГІСТИЧНА РЕГРЕСІЯ, ЕМБЕДИНГ

O.V. Lozynska, O.O. Markiv Oksana, V.A. Vysotska. Method for detecting sources of disinformation based on ensemble machine learning models. The presented study developed a method for detecting sources of disinformation based on ensemble machine learning models. Modern methods of combating disinformation and detecting false content were analyzed. A fake news identification system based on the ensemble approach was implemented as part of the work, and its architectural structure was described. The main stages of cleaning text data obtained from social networks and news are described in detail, in particular, the normalization of categorical variables. Statistical analysis of the text and analysis of the criteria for identifying sources of disinformation distribution are carried out. An analysis of the balance of target and auxiliary variables was performed, which made it possible to identify dependencies between the language of the message and reliability. Two types of text embeddings and corresponding classification models were used for modeling: linear regression and logistic regression. The final stage was the application of an ensemble of models, which made it possible to combine the predictive ability of both models. The results showed that the combination of approaches improves classification quality, especially in conditions of unbalanced data. Using an ensemble of models allowed us to increase the accuracy from 73% (model 1) and 71% (model 2) to 78%.

DISINFORMATION, DATASET, MACHINE LEARNING, ENSEMBLE MODELS, LINEAR REGRESSION, LOGISTIC REGRESSION, EMBEDDING

Вступ

Основним джерелом новин та даних для більшості людей є соціальні мережі та онлайн-новини, оскільки вони є легкодоступні. Однак одночасно вони також сприяють поширенню фейкових новин, тобто повідомлень, які містять дезінформацію та мають значний негативний вплив на суспільство. Такі повідомлення зазвичай мають спільні риси, такі як граматичні помилки, неправдива інформація, використання подібно обмеженого набору слів та містять емоційно забарвлену інформацію, яка впливає на думку читача [1]. Щоб вирішити цю проблему, дослідженням з ідентифікації фейкових новин останнім часом приділяють багато уваги. Незважаючи на численні обчислювальні рішення, доступні наразі для виявлення фейкових новин, відсутність комплексної бази даних фейкових новин стала однією з суттєвих перешкод. Масштабне поширення новин через соціальні мережі унеможливує ручну перевірку, що

сприяє розробці та впровадженню автоматизованих систем виявлення дезінформації [2].

Фейкові новини можуть функціонувати як пропаганда або дезінформація, але вони завжди апелюють до емоцій громадськості та мають намір приховати раціональні реакції, аналіз та порівняння інформації з кількох джерел.

Виявлення фейкових новин – це багаторівневий процес, який включає аналіз змісту новин для визначення їхньої правдивості. Новини можуть містити інформацію в різних форматах, таких як текст, відео, зображення тощо. Комбінації різних типів даних ускладнюють процес виявлення дезінформації. Оскільки фейкові новини утворюють великі, неструктуровані дані [3], попередня обробка таких даних надзвичайно важлива для очищення та структурування даних перед їх використанням у моделі виявлення.

Однією з проблем аналізу фейкових новин є правильна обробка відсутніх даних. Оскільки фейкові

новини не створюються для досліджень, більшість інформації в соціальних мережах та новинах не структурована заздалегідь. Відповідно, виникають відсутні значення, що призводить до суперечливих або упереджених статистичних результатів при застосуванні аналізу або класифікації [4].

У даній роботі запропоновано метод виявлення джерел дезінформації на основі ансамблевих моделей машинного навчання. Розглянуто відомі методи ідентифікації дезінформації та фейкових новин. Наведено основні етапи ідентифікації дезінформації, використовуючи датасет текстових новин з різних українських соцмереж та новинних сайтів. Зокрема, проведено препроцесинг датасету, який включає очищення певних непотрібних полів, нормалізацію категоріальних змінних, уніфікацію, аналіз балансу цільової змінної та моделювання. У межах моделювання застосовано два різні типи ембедингів для представлення текстових даних та відповідні моделі класифікації: лінійну регресію та логістичну регресію. Крім того, використано ансамбль цих двох моделей. Такий підхід може привести до покращення метрик класифікації інформації у різних соцмережах.

1. Методи ідентифікації дезінформації

Для ідентифікації дезінформації та фейкових новин можна ефективно використовувати такі підходи: машинне навчання, аналіз настроїв або їх комбінацію. Ці підходи можна використовувати для двох різних завдань — розпізнавання фейкових новин або їх авторів.

У роботі [5] машинне навчання та аналіз настроїв використовуються як паралельні підходи для ідентифікації не лише фейкових новин, але й їх авторів. Використання методу опорних векторів у цій роботі дало змогу досягнути значення 0,98 для F1-оцінки.

Наукова праця [6] зосереджена на виявленні ненадійних авторів онлайн-дописів за допомогою згорткової нейронної мережі (CNN), а також довгої короткочасної пам'яті (LSTM). Автори використовували дані з Twitter, а також класичні методи машинного навчання, такі як метод опорних векторів (SVM) та метод k-найближчих сусідів (KNN). Використання нейронної мережі дало змогу досягнути точності 0,93. Цей результат виявився кращим на кілька відсотків порівняно з SVM та на 10% кращим, ніж KNN. Іншим прикладом використання комбінації машинного навчання та аналізу настроїв є дослідження [7].

У науковій роботі [8] розглянуто такі методи машинного навчання для виявлення фейкових новин: метод на основі логістичної регресії, метод на основі нейронних мереж (багатосаровий перцептрон (MLP) та згорткові нейронні мережі), методи дерев рішень, баєсівські методи, класифікація C-опорних векторів. Згорткові нейронні мережі були оцінені як

найкращий варіант серед інших методів, незважаючи на їх значно довший час навчання та вимогу до більших наборів даних.

У статті [9] запропоновано такі підходи для розпізнавання фейкових новин в онлайн-просторі, як логістична регресія, метод опорних векторів (SVM), k-NN, дерева рішень, випадковий ліс, згорткові нейронні мережі, вентильні рекурентні мережі та довгу короткочасну пам'ять. В експериментах було використано два набори даних, а саме ISOT та KDnugget. Завдяки використанню підходу стекування, авторам вдалося досягти найкращої точності 0,99 для моделі випадковий ліс і набору даних ISOT та 0,96 — для логістичної регресії та набору даних KDnugget.

У інших роботах, таких як [10], додатково досліджувалася методологія виявлення фейкових новин, яка враховує не лише інформацію про зміст новин, але й додаткову інформацію щодо використання соціальних мереж. Для генерації представлення новин було використано тензорну факторизацію. Було представлено порівняльний аналіз трьох підходів: перший — на основі новинних текстів, другий — на основі способу використання соціальних мереж, і третій — на основі їх комбінації. Було виявлено, що комбінований підхід, який базується на контенті та контексті, забезпечує кращі результати. Використання глибокої нейронної мережі забезпечило покращення точності та повноти (recall) порівняно з класифікатором XGBoost.

У дослідженні [11] розроблено архітектуру моделі FakeDetector, яка являє собою графову нейронну мережу. У роботі автори поєднали явні та неявні атрибути, отримані з новинних текстів, для виявлення фейкових новин. FakeDetector — це глибока дифузійна нейронна мережа, яка одночасно представляє та оцінює новини, авторів та тему статті. Таким чином, модель прогнозування може визначати достовірність авторів. У статті представлено нову модель дифузійної одиниці, а саме GDU. Модель являє собою нейронну мережу, засновану на принципі поширення інформації шляхом дифузії. Модель GDU може обробляти кілька вхідних даних одночасно та ефективно комбінувати вхідні дані для генерації виходів. Проведені експерименти на наборах даних показали задовільний рівень ефективності запропонованого підходу у виявленні фейкових новин та їх авторів в онлайн-просторі.

Авторами дослідження [1] застосовано такі методи машинного навчання як метод опорних векторів, метод k-найближчих сусідів, згорткові нейронні мережі, а також ансамблеві методи. Використання ансамблевих навчальних методів дало змогу досягти кращих оцінок за всіма показниками продуктивності порівняно з окремими навчальними методами.

У роботі [12] запропоновано новий підхід DocEmb

для виявлення фальшивих повідомлень за допомогою вбудовування текстових даних. У статті представлені експерименти з різними текстовими представленнями, такими як TF-IDF або з використанням вбудовування слів та трансформаторів: Word2Vec SG та CBOW, FastText SG та CBOW, GloVe, BERT, BART та RoBERTa. Автори навчали моделі на цих текстових представленнях, використовуючи методи машинного навчання, такі як наївний байєсівський метод, градієнтно-підсилени дерева та моделі глибокого навчання – Perceptron, Multi-Layer Perceptron, [Bi]LSTM та [Bi]GRU. Результати дослідження показують, що представлення слів у документі відіграє важливу роль для досягнення більшої точності.

Більшість спроб виявлення фейкових новин зазвичай зосереджені лише на текстовій інформації. Мультиmodalні підходи менш поширені та зазвичай класифікують повідомлення як правдиві або хибні. У роботі [13] описано мультиmodalний підхід до виявлення фейкових новин. Виявлення проводилося з використанням як уніmodalного, так і мультиmodalного підходів. Результати роботи показали, що мультиmodalний підхід досяг досить задовільних результатів у випадку, коли він базувався на архітектурі згорткової нейронної мережі, де поєднувалися лише текстові та графічні дані. Таким чином, використання як текстових, так і графічних даних покращує виявлення фейкових новин.

У дослідженні [12] наведено огляд існуючих інструментів, придатних для виявлення фейкових новин, та огляд веб-сайтів, які можна використовувати для перевірки фактів. Вони вказують на важливість донесення до громадськості методів виявлення дезінформації.

Авторами [15] представлено методи глибокого навчання для створення ансамблю моделей. У статті [16] описано підхід з використанням різноманітних ансамблевих стратегій з метою підвищення ефективності виявлення дезінформації для певного набору моделей.

У науковій роботі [17] розглянуто метод ідентифікації фейкових новин шляхом використання архітектури BERT на даних із соціальних мереж. Поряд із текстовим вмістом публікацій, автори також включили дев'ять додаткових атрибутів, що стосуються користувача або контенту, створеного користувачем, таких як кількість друзів, перегляди або використання хештегів у публікаціях користувача. Отримані результати дали змогу досягти оцінки F1 приблизно 85%. У випадку [18] виявлення фейкових повідомлень проводилося з використанням архітектури на основі CNN та LSTM. Для векторного представлення автори використовували W2V. Набір даних містив заголовки та текстовий вміст новинних статей, що супроводжувалися пов'язаними статтями, позначеними

як джерела. Метою цього дослідження було оцінити, чи мають статті схожість, а також зв'язок з відповідними джерелами. Автори досягли результатів для F1-оцінки, що становить 0,97.

2. Пропонований виявлення джерел дезінформації з використання ансамблевих моделей

Для реалізації методу виявлення джерел дезінформації авторами запропоновано та розроблено систему ідентифікації дезінформації з використання ансамблю моделей, яка включає в себе попередню обробку (препроцесинг) даних датасету, вибір текстових ознак, векторизацію (ембединг), використання моделей для класифікації тексту та оцінку даних моделей. Архітектура даної системи ідентифікації зображена на рис. 1.



Рис. 1. Архітектура системи ідентифікації дезінформації

2.1. Нормалізація та препроцесинг датасету новин

Для забезпечення достовірності аналізу та точності побудови моделі ідентифікації фейкових новин використовуються нормалізація та препроцесинг датасету україномовних новин. Препроцесинг включає в себе очищення певних полів датасету. Для аналізу використано датасет, розроблений авторами [19] (рис. 2).

Очищення датасету є одним з ключових етапів підготовки даних до аналізу, що включає виявлення та усунення некоректних, відсутніх або зайвих значень. Цей процес необхідний для покращення якості даних, підвищення достовірності результатів та точності побудови моделей машинного навчання.

У межах попередньої обробки даних було виконано наступне:

Рис. 2. Датасет після завантаження

Видалення неінформативних колонок: після імпорту датасету, який складався з 14 колонок, залишено лише ті, що є релевантними для подальшого аналізу (текст повідомлень, мітка, post/repост, автор/group, джерело, мова).

Нормалізація категоріальних змінних: оскільки значення деяких категоріальних полів були неоднорідними, їх було приведено до уніфікованого формату — логічних значень типу True/False.

Нормалізацію застосовано до таких колонок як «мітка», «пост/репост» та «мова», які потрібно було привести до стандартизованого вигляду, з огляду на велику кількість варіантів формулювання. На рис. 3-5 зображено значення колонок «мітка», «пост/репост» та «мова» до уніфікації, на рис. 6 — значення колонки «мова» після уніфікації.

```
print(df['Мітка'].unique())
✓ [4] < 10 ms
['FALSE' 'TRUE' 'фейк' 'істина' 'Фейк' 'Істина' 'False' 'True' 'F' 'T'
 'T' 'F' 'фейк' 'правда' 'Брехня' 'брехня' nan 'TRUE' 'FAKE' 'TRUE'
 'Fake' 'True' 'False' 'Реальна' 'Фейкова']
```

Рис. 3. Значення колонки "мітка" до уніфікації

```
print(df['Post/Repост'].unique())
✓ [6] < 10 ms
['Post' 'Reel' 'post' 'reporter' 'repost' 'Publication' 'Post' 'Repost'
 'POST' 'post' 'video' 'Post' 'REPOST' 'News' 'Пост']
['Post' 'Reel' 'post' 'reporter'
 'repost' 'Publication' 'Post' 'Repost'
 'POST' 'post' 'video' 'Post' 'REPOST' 'News' 'Пост']
```

Рис. 4. Значення колонки «пост/репост» до уніфікації

```
print(df['Мова'].unique())
✓ [8] < 10 ms
['Russian' 'French' 'German' 'English' 'Ukrainian' 'Russian' 'Українська'
 'російська' 'польська' nan 'russian' 'Українська' 'Українська'
 'англійська' 'Український' 'Російська' 'УКРАЇНСЬКА' 'РОСІЙСЬКА' 'Чеська'
 'Англійська' 'RUSSIAN' 'UKRAINIAN' 'Polish' 'укр' 'Slovak' 'Poland'
 'Ukranian' 'ukranian' 'Українська' 'Польська' 'Російська'
 'Ukrainian (the original language \nis Russian)' 'ukrainian' 'Латвійська'
 'Французька' 'Українська переклад з російської']
```

Рис. 5. Значення колонки «мова» до уніфікації

```
print(df['Мова'].unique())
✓ [9] < 10 ms
['RUS' 'FRA' 'GER' 'ENG' 'UKR' 'POL' 'CZE' 'SVK' 'LVA']
```

Рис. 6. Значення колонки «мова» після уніфікації

У результаті виконаних трансформацій датасет було очищено від зайвих колонок та нормалізовано всі основні категоріальні змінні (рис. 7).

Рис. 7. Датасет після етапів очищення та нормалізації

На наступному кроці було здійснено очищення полів імен авторів та назв джерел, що є критично важливим під час обробки текстових даних, особливо якщо джерела — відкриті або неструктуровані. Часто імена чи назви одного об'єкта представлені у різних варіаціях, містять спеціальні символи, емоційні вирази чи граматичні відхилення. У нашому випадку це стосується поля «автор/group», що містить значну кількість варіацій, включно зі смайлами, знаками «@», переносами рядків, порожніми або нечіткими значеннями.

Для уніфікації:

- видалено зайві символи;
- текст переведено у нижній регістр;
- порожні значення замінено на стандартне значення «unknown»;

— об'єднано поля «джерело» та «веб-сторінка» у нове поле з витягнутим доменом як репрезентативним ідентифікатором джерела;

— застосовано правила перетворення, наприклад, заміна «t» на «telegram».

2.2. Аналіз категоріальних змінних

Цей етап спрямований на вивчення якісних характеристик даних, що представлені у вигляді категорій. Для комплексного аналізу було проведено:

- 1) аналіз дублікатів — за допомогою функції duplicated() виявлено 10 повторюваних записів;
- 2) аналіз пропущених значень — використано функцію isnull().sum(), результати якої показують відсутність пропусків у колонках (рис. 8);
- 3) оцінку балансу категорій — виконано візуалізацію для полів «пост/репост», «мова» та «мітка»

(target), що виявила суттєву диспропорцію (рис. 9);

4) розподіл значень відносно поля «мітка» — здійснено аналіз залежності між мовою публікації та ймовірністю її правдивості (рис. 10);

5) крос-табуляцію — побудовано таблиці для оцінки співвідношення категоріальних змінних до цільової ознаки (рис. 11 та рис. 12).

```
df.isnull().sum()
Текст повідомлень    0
Мітка                 0
Post/Repost          0
Автор/Group          0
Джерело              0
Мова                 0
dtype: int64
```

Рис. 8. Кількість пропущених значень у кожній з колонок

Побудова крос-таблиць дала змогу оцінити числове співвідношення між категоріальними змінними. Цей метод є ефективнішим у порівнянні з графічною візуалізацією у випадках, коли маркування на діаграмах не є чітким. Було сформовано крос-таблиці для зв'язків між полями «мова» та «мітка», а також «пост/репост» та «мітка».

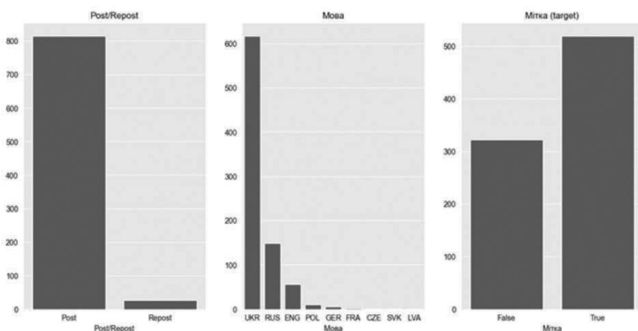


Рис. 9. Баланс категорій для трьох колонок: «пост/репост», «мова», «мітка»

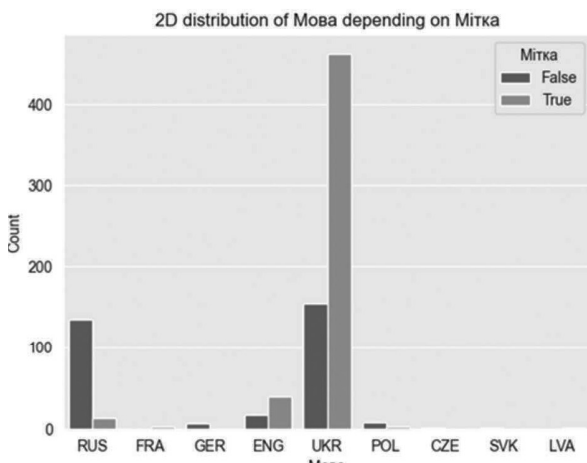


Рис. 10. Розподіл значень поля «мова» відносно значень поля «мітка»

Мова	CZE	ENG	FRA	GER	LVA	POL	RUS	SVK	UKR
Мітка									
False	1	17	0	6	0	8	135	1	154
True	0	39	2	0	1	2	13	0	462

Рис. 11. Співвідношення між кількістю значень полів «мова» та «мітка»

Post/Repost	Post	Repost
Мітка		
False	311	11
True	503	16

Рис. 12. Співвідношення між кількістю значень полів «мітка» та «пост/репост»

2.3. Оцінка балансу цільової змінної

Аналіз балансу цільової змінної («мітка») дає змогу візуально ідентифікувати можливий дисбаланс класів, що може призводити до упередженості у роботі моделей машинного навчання. Цей крок важливий для прийняття рішень щодо використання технік балансування перед навчанням моделей. На рис. 13 зображено кругову діаграму, яка показує переважання правдивих повідомлень у вибірці.

Баланс поля "мітка"

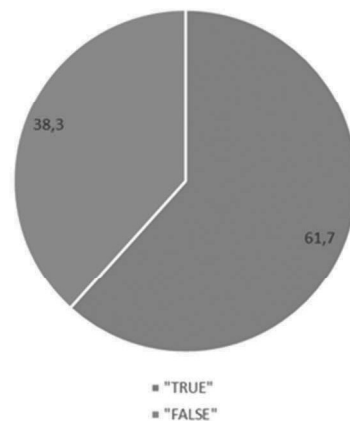


Рис. 13. Діаграма балансу поля «мітка»

Для коректної обробки тексту було проведено видалення всіх зайвих елементів, таких як емоджі, посилання, хештеги, тощо. Це забезпечує чистоту вхідних даних та підвищує якість подальшого аналізу.

2.4. Статистичний аналіз тексту

Використання статистики тексту дає змогу кількісно охарактеризувати зміст повідомлень. Було оцінено:

- довжину текстів,
- кількість слів,
- розподіл за реченнями,
- загальну складність тексту.

За допомогою гістограм із KDE-графіком візуалізовано щільність розподілу довжин текстів. Для

виявлення відмінностей між класами застосовано Boxplot, що дало змогу оцінити типову кількість слів, розкид та наявність аномалій (рис. 14–16).

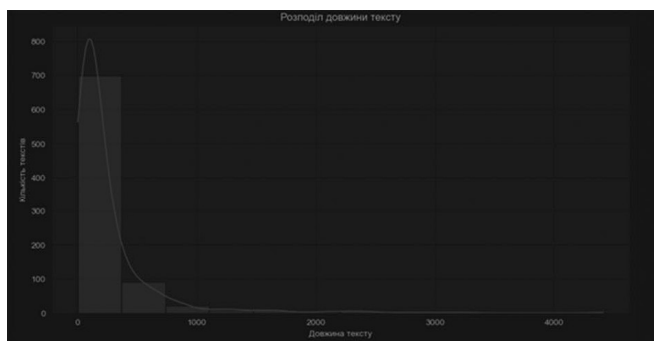


Рис. 14. Гістограма розподілу довжини тексту

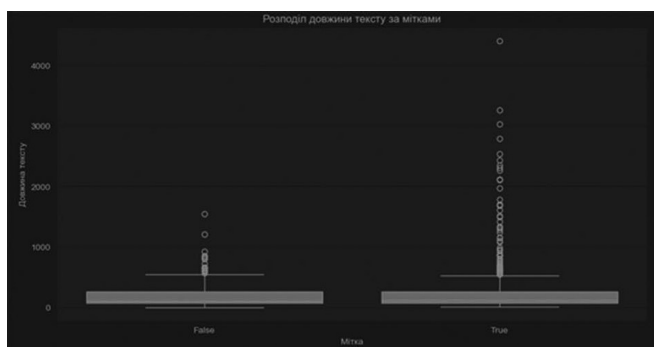


Рис. 15. Діаграма розподілу довжини тексту за мітками

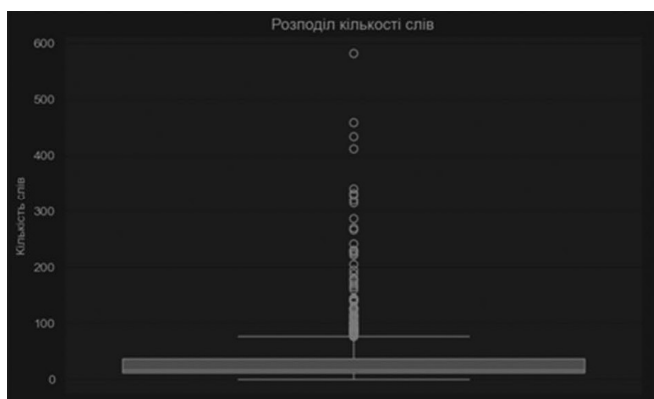


Рис. 16. Діаграма розподілу кількості слів

На рис. 17 зображено хмару слів для ілюстрації найчастіше вживаних слів у корпусі, що дає змогу інтуїтивно ідентифікувати основні теми та контекст повідомлень.



Рис. 17. Хмара слів

2.5. Критерії та параметри для виявлення джерел дезінформації

Крім статистичного аналізу тексту проведено аналіз критеріїв та параметрів, які можуть впливати на виявлення саме джерел дезінформації. Серед критеріїв, які дають змогу виявляти джерела поширення дезінформації обрано основні наративи пропаганди (у датасеті таких 40). Проведено аналіз датасету і обрано найпоширеніші наративи: «НАТО, ЄС», «БІЛЯ ДНІПРА», «АЕС», «ОХМАТДИТ», «КУРСЬК», «МОБІЛІЗАЦІЯ». Діаграма розподілу даних наративів у датасеті зображена на рис. 18. Проведений аналіз дає змогу окреслити наступні висновки:

- якщо у новині використано наратив «НАТО, ЄС» і мова даної новини російська – то дана новина потенційно є фейком;
- якщо більшість новин (постів) на сторінці джерела фейкові, швидше за все дане джерело чи аккаунт є також фейковим
- потрібно відслідковувати час і дату поширення фейкової новини від першоджерела.

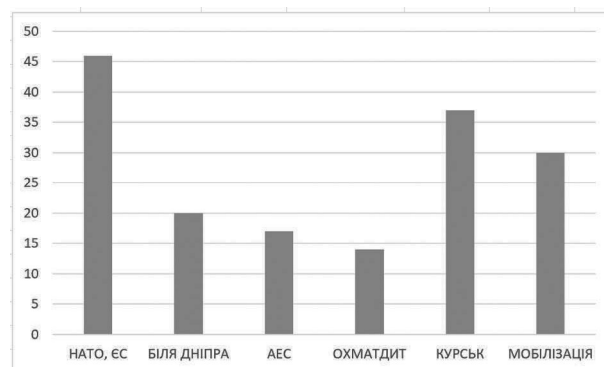


Рис. 18. Діаграма розподілу найпоширеніших наративів у датасеті

3. Експериментальні дослідження

У межах моделювання застосовано два різні типи ембедингів для представлення текстових даних:

1. intfloat/multilingual-e5-base
2. sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2

На основі кожного з ембедингів було побудовано окремі моделі класифікації із застосуванням лінійної та логістичної регресії, які прогноують ймовірність правдивості тексту. Для покращення якості класифікації результати обох моделей було інтегровано, а також проведено порівняння їх ефективності за допомогою відповідних метрик.

3.1. Модель 1 (лінійна регресія)

На першому етапі ембединги були згенеровані за допомогою інструменту intfloat/multilingual-e5-base. Тексти було попередньо підготовлено, де кожен рядок отримав префікс «query:», відповідно до специфікації моделі. В результаті обробки було отримано вектори з 768 ознаками.

Для зменшення розмірності векторів застосовано метод головних компонент (PCA). Розподіл значень у двох головних компонентах подано на рис. 19. Аналіз двовимірної проєкції показав значне перекриття між класами, що свідчить про неефективність PCA у даному випадку. Тому у подальших етапах навчання було збережено повну розмірність ембедингів.

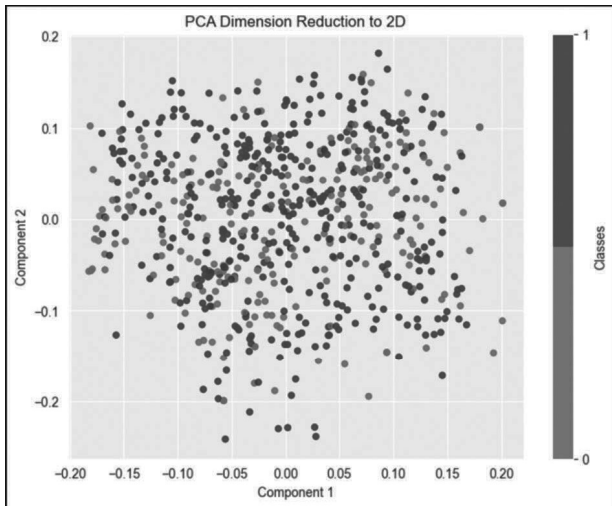


Рис. 19. Графічне зображення зменшення вимірності до 2D з використанням PCA

У процесі навчання моделі лінійної регресії було застосовано метод оптимізації Randomized Search, який виконує випадковий підбір гіперпараметрів із заздалегідь визначеного діапазону. Для забезпечення

стабільності та достовірності результатів модель оцінювалася за допомогою крос-валідації. Визначена конфігурація найкращих параметрів була використана для фінального навчання моделі, що позитивно вплинуло на її здатність до узагальнення та точність ідентифікації дезінформації.

Для оцінки роботи моделі побудовано три типи матриці неточностей (confusion matrix):

- 1) класична (абсолютна) — відображає кількість випадків кожного типу класифікації;
- 2) нормалізована за рядками — показує, наскільки точно модель класифікує правдиві та фейкові тексти;
- 3) нормалізована за стовпцями — дає змогу оцінити точність передбачень по кожному з прогнозованих класів.

Ці візуалізації (рис. 20) надають змогу глибше зрозуміти сильні та слабкі сторони моделі.

Оцінка результатів виявила високий рівень помилок першого роду (False Negatives) — велика кількість фейкових повідомлень була неправильно класифікована як правдиві. Це критично у контексті інформаційної безпеки, адже модель пропускає потенційно шкідливий контент. Водночас рівень помилок другого роду (False Positives) є низьким — менше 1% правдивих новин помилково віднесено до фейкових. Це вказує на відносну стриманість моделі у «звинуваченні» достовірних джерел.

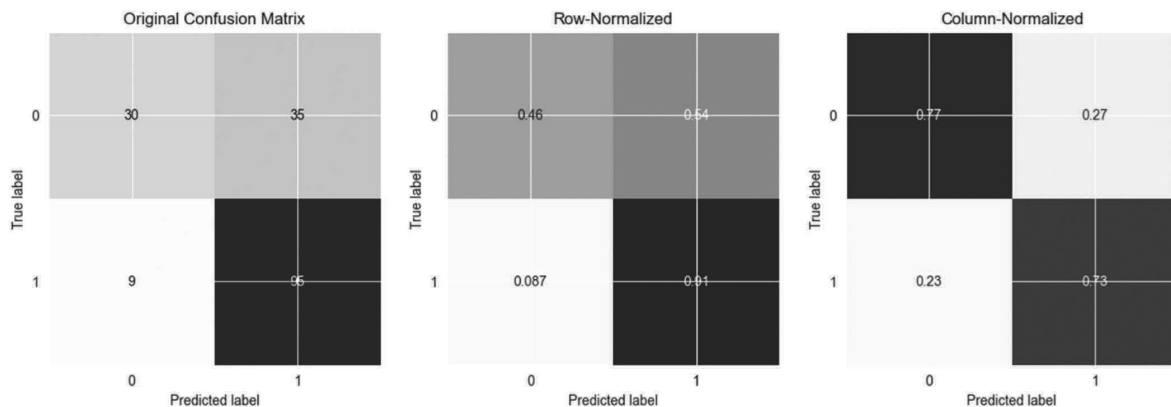


Рис. 20. Матриці невідповідностей для першої моделі

Оцінка метрик класифікації. З огляду на незбалансованість класів у вибірці, загальний показник точність (Accuracy) є малорелевантним. Натомість для оцінки ефективності було використано такі метрики:

- Precision, що відображає точність визначення фейкових новин;
- Recall показує здатність моделі виявити усі фейкові повідомлення;
- F1-метрику, яка показує узагальнений баланс precision та recall;
- Fβ-метрику, адаптовану для підвищеної чутливості до precision.

Отримане значення $F1 = 0.81$ свідчить про загалом

ефективну роботу моделі, попри наявність певних упущень. Додаткова Fβ-метрика допомогла акцентувати увагу на важливості мінімізації помилок другого типу. Результати класифікаційних метрик наведено у табл. 1.

Для оцінки здатності моделі розрізняти між класами False Negatives та False Positives було використано ROC AUC і побудовано ROC-криву (див. рис. 21). Хоча значення ROC AUC не досягло високих показників, це свідчить про потенціал до подальшого вдосконалення моделі, зокрема в аспектах зменшення частоти хибнопозитивних значень.

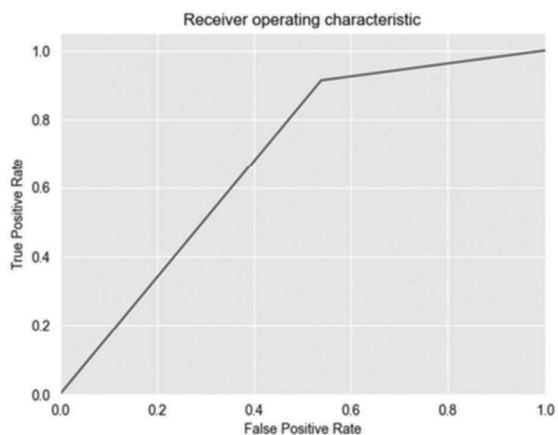


Рис. 21. ROC крива для моделі 1

3.2. Модель 2 (логістична регресія)

У другій моделі як інструмент для векторизації тексту було використано sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2.

Після генерації ембедингів дані поділено на тренувальну та тестову вибірки у співвідношенні 80:20.

Наступним кроком став підбір гіперпараметрів за допомогою Grid Search, після чого оптимальні параметри було застосовано до логістичної моделі. Отримана точність (Accurasy) моделі становить 0.716.

Також було побудовано ROC-криву (рис. 22), яка підтвердила, що загальна якість моделі потребує покращення (низький ROC AUC).

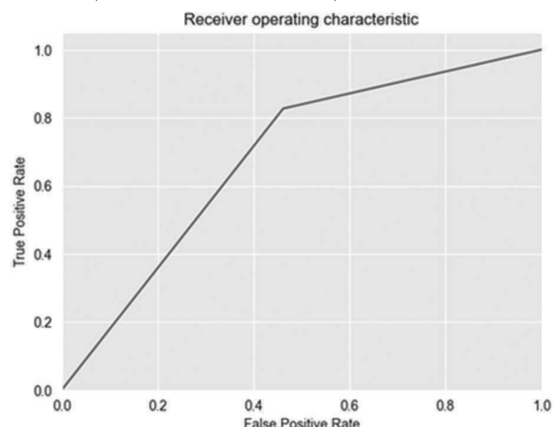


Рис. 22. ROC-крива для моделі 2

Побудована матриця невідповідностей показала (рис. 23):

- а) високий рівень помилок першого роду – значна частина фейкових новин була визначена як правдиві;
- б) низький рівень помилок другого роду – менше 2% достовірних новин класифіковано як фейки.

Додатково обчислено основні метрики класифікації: Recall становить 0.8269, Precision – 0.7414, метрика F1= 0.7818, метрика Fβ-score = 0.7570.

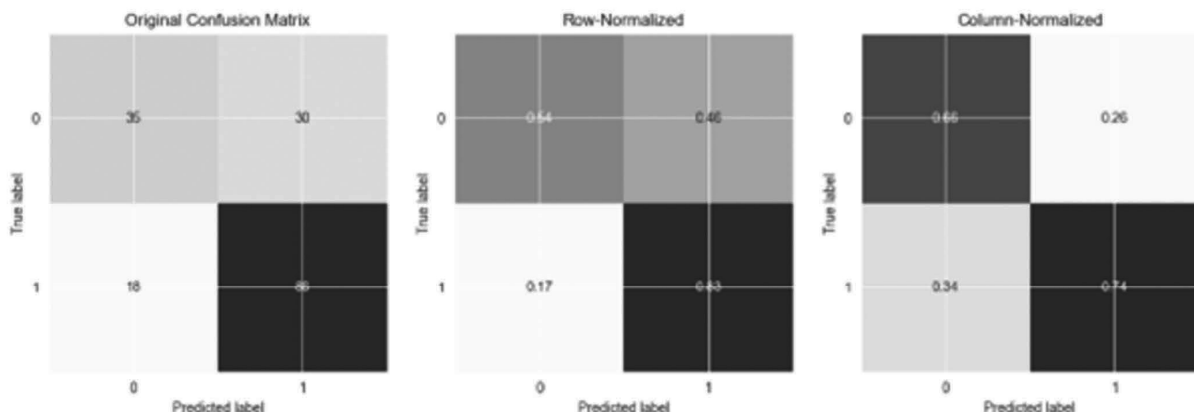


Рис. 23. Матриці невідповідностей для другої моделі

3.3. Використання ансамблю моделей

Для підвищення ефективності ідентифікації фейкових новин реалізовано метод стекування моделей. Для цього об'єднано результати двох попередньо навчених моделей – без повторного навчання, лише на основі тестової вибірки.

Попередньо для кожної моделі обчислено ймовірності приналежності об'єктів до позитивного класу. Після цього ці ймовірності було усереднено і утворено новий комбінований прогноз. На наступному кроці проведено оптимізацію порогу класифікації (threshold) у діапазоні 0.1–0.9 шляхом вибору значення, при якому оцінка F1 є максимальною. Після цього було обчислено загальну оцінку якості ансамлевої моделі, яка включає такі основні метрики, як Accurasy, Recall, Precision, оцінки F1 та Fβ, ROC AUC.

Усі значення були зібрані у порівняльну таблицю 1 для аналізу відносної ефективності кожної моделі та їх комбінації. Результати свідчать, що стекування дало змогу підвищити якість класифікації новин, уникаючи окремих слабких сторін кожної з моделей.

Таблиця 1

Показники метрик для різних моделей

Метрика	Модель 1	Модель 2	Модель 3
Accurasy	0,739645	0,715976	0,781065
Recall	0,913462	0,826923	0,913462
Precision	0,730769	0,741379	0,772358
F1	0,811966	0,781818	0,837004
Fβ	0,869963	0,808271	0,881262
ROC AUC	0,687500	0,682692	0,741346

Подяка

Дана стаття підготована завдяки грантової підтримки Національного Фонду Досліджень України, реєстраційний номер проекту 33/0012 від 3/03/2025 (2023.04/0012) «Розроблення інформаційної системи автоматичного виявлення джерел дезінформації та неавтентичної поведінки користувачів чатів» за конкурсом «Наука для зміцнення обороноздатності України».

Висновки

У цій роботі запропоновано метод виявлення джерел дезінформації на основі ансамблевих моделей машинного навчання. Проведено огляд актуальних методів ідентифікації дезінформації та неправдивого контенту. У межах дослідження реалізовано систему ідентифікації дезінформації з використанням ансамблевої моделі. Наведено архітектуру даної системи ідентифікації. Описано основні етапи аналізу текстових даних, отриманих із соціальних мереж і новинних ресурсів, зокрема: очищення датасету від нерелевантних полів, нормалізацію категоріальних змінних, уніфікацію даних, вивчення балансу цільової змінної та побудову моделей. Початкові етапи очистки дали змогу отримати структурований датасет із нормалізованими значеннями, що підвищило якість подальшого аналізу. Для моделювання було застосовано два підходи до генерації ембедингів тексту, а також реалізовано класифікацію за допомогою лінійної та логістичної регресії. Моделювання за допомогою двох типів ембедингів і різних класифікаторів виявило сильні сторони кожної моделі, зокрема високе значення повноти (recall), проте наявність false positives і false negatives вказує на необхідність додаткових підходів до балансування.

На завершальному етапі реалізовано комбінування моделей через метод ансамблювання, що дало змогу підвищити якість класифікації новинного контенту у соціальних медіа. Використання ансамблю моделей дало змогу збільшити точність з 73% до 78%. Результати дослідження демонструють ефективність поєднання класичних підходів до попередньої обробки тексту із сучасними методами машинного навчання для задач інформаційної безпеки.

Перспективним продовженням досліджень є розширення ансамблю моделей та використання на інших датасетах запропонованої моделі.

Список літератури:

- [1] Ahmad, I., Yousaf, M., Yousaf, S., & Ahmad, M. (2020). Fake news detection using machine learning ensemble methods. *Complexity* 2020, 1–11.
- [2] Harb, J.G., Ebeling, R., & Becker, K. (2020). A framework to analyse the emotional reactions to mass violent events on Twitter and influential factors. *Inform Process Manag*, 57(6).
- [3] Akinyemi, B. (2020). An improved classification model for fake news detection in social media. *International Journal of Information Technology and Computer Science*, 12(1), pp. 34–43. <https://doi.org/10.5815/ijitcs.2020.01.05>.
- [4] Maniruzzaman, M., Rahman, M.J., Al-Mehedi Hasan, M., Suri, H.S., Abedin, M.M., et al. (2018). Accurate diabetes risk stratification using machine learning: role of missing value and outliers. *J Med Syst*, 42(5), pp. 92. <https://doi.org/10.1007/s10916-018-0940-7>.
- [5] Machova, K., Mach, M., & Vasilko, M. (2022). Comparison of Machine Learning and Sentiment Analysis in Detection of Suspicious Online Reviewers on Different Type of Data. *Sensors*, 22, 155.
- [6] Sansonetti, G., Gasparetti, F., D'aniello, G., & Micarelli, A. (2020). Unreliable Users Detection in Social Media: Deep Learning Techniques for Automatic Detection. *IEEE Access*, 8, 213154–213167.
- [7] Kandasamy, V., Trojovský, P., Machot, F.A., Kyamakya, K., Bacanin, N., Askar, S., & Abouhawwash, M. (2021). Sentimental Analysis of COVID-19 Related Messages in Social Networks by Involving an N-Gram Stacked Autoencoder Integrated in an Ensemble Learning Scheme. *Sensors*, 21, 7582.
- [8] Papakostas, D., Stavropoulos, G., & Katsaros, D. (2022). Evaluation of Machine Learning Methods for Fake News Detection. In *Combating Fake News with Computational Intelligence Techniques*, Studies in Computational Intelligence; Lahby, M., Pathan, A.K., Maleh, J., Shafer-Yafooz, W.M., Eds.; Springer International Publishing: Berlin/Heidelberg, Germany, Volume 1001, pp. 163–183.
- [9] Jiang, T., Li, J.P., Haq, A.U., Saboor, A., & Ali, A. (2021). A Novel Stacking Approach for Accurate Detection of Fake News. *IEEE Access* 2021, 9, 22626–22639.
- [10] Kaliyar, R.K., Goswami, A., & Narang, P. (2021). DeepFakE: Improving fake news detection using tensor decomposition-based deep neural network. *J. Supercomput.* 77, 1015–1037.
- [11] Zhang, J., Dong, B., & Yu, P.S. (2020). FakeDetector: Effective fake news detection with deep diffusive neural network. In *Proceedings of the International Conference on Data Engineering*, Dallas, USA, pp. 1826–1829.
- [12] Truică, C.O., & Apostol, E.S. (2023). It's All in the Embedding! Fake News Detection Using Document Embeddings. *Mathematics*, 11, 508.
- [13] Deepak, P., Tanmoy, C., & Cheng, L. (2021). Santhosh Kumar, G. Multi-modal Fake News Detection. *Inf. Retr. Ser.*, 42, 41–70.
- [14] Sharma, D.K., Garg, S., & Shrivastava, P. (2021). Evaluation of tools and extension for fake news detection. In *Proceedings of the International Conference of Innovative Practices in Technology and Management (ICIPTM 21)*, India, pp. 227–232.
- [15] Hruz, M., Gruber, I., Kanis, J., Boháček, M., Hlaváč, M., & Kr̃noul, Z. (2022). One Model is not Enough: Ensembles for Isolated Sign Language Recognition. *Sensors*, 22, 5043.
- [16] Atitalah, S.B., Driss, M., & Almomani, I. (2022). A Novel Detection and Multi-Classification Approach for IoT-Malware Using Random Forest Voting of Fine-Tuning Convolutional Neural Networks. *Sensors*, 22, 4302.
- [17] Heidari, M., Zad, S., Hajibabae, P., Malekzadeh, M., Hekmati Athar, S., Uzuner, O., & Jones, J.H. (2021). BERT Model for Fake News Detection Based on Social Bot Activities in the COVID-19 Pandemic. In *Proceedings of the IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*, New York, USA, pp. 0103–0109.
- [18] Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A. Choi, G.S., & On, B.W. (2020). Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM). *IEEE Access* 2020, 8, 156695–156706.
- [19] Лозинська, О., Марків, О., Висоцька, В., Романчук, Р., & Назаркевич, М. (2024). Інформаційна технологія розроблення та наповнення датасету дезінформації з використанням інтелектуального пошуку дипфейків та клікбейтів. *Herald of Khmelnytskyi National University. Technical Sciences*, № 343, т. 6(1), с. 158–167. DOI: 10.31891/2307-5732-2024-343-6-24.



І.В. Кириченко¹, Г.Ю. Терещенко², В.В. Ніколенко³

¹ Харківський національний університет радіоелектроніки, м. Харків, Україна, iryna.kyrychenko@nure.ua, ORCID iD: 0000-0002-7686-6439

² Харківський національний університет радіоелектроніки, м. Харків, Україна, hlib.tereshchenko@nure.ua, ORCID iD: 0000-0001-8731-2135

³ Харківський національний університет радіоелектроніки, м. Харків, Україна, vitalii.nikolenko@nure.ua

ДОСЛІДЖЕННЯ МЕТОДІВ ЗБЕРІГАННЯ ВІЗУАЛЬНИХ ДАНИХ

Стаття присвячена дослідженню зберігання великих обсягів візуальних даних в програмних системах з використанням різних технологій та підходів до зберігання. В роботі проаналізовано актуальні методи зберігання даних, переваги та недоліки використання реляційних та нереляційних СКБД, наявних вбудованих технологій для роботи з даними великого розміру, MS SQL Server FILESTREAM, порівняно зі зберіганням цих даних в окремому сховищі. Для оцінки методів виконується бенчмаркінг (порівняння продуктивності) запису, читання та фільтрації даних різного розміру та обсягу та порівняння використання системних ресурсів. В теорії, використання FILESTREAM повинно було підвищити продуктивність роботи програмної системи, в той час, як за результатами дослідження при великих навантаженнях воно збільшило час запису на 110%, послідовного читання на 118% та час фільтрації на 32%, порівняно зі зберіганням зображень напряму в базі даних. Зберігання файлів напряму в базі даних підвищило фрагментацію даних та знизило ефективність індексів, що негативно вплинуло на масштабованість програмної системи та результати експериментів з великим обсягом даних. Гібридні підходи показали найкращий результат; серед них Microsoft SQL Server в середньому в 1.95 рази швидший за MongoDB, при цьому MongoDB показав меншу втрату продуктивності для запису та фільтрації з великим обсягом даних.

ASP.NET CORE, BIG DATA, ENTITY FRAMEWORK, FILESTREAM, MONGO DB, MS SQL SERVER, БЕНЧМАРКІНГ, ГІБРИДНЕ СХОВИЩЕ, ЗОБРАЖЕННЯ, НЕРЕЛЯЦІЙНІ БАЗИ ДАНИХ, РЕЛЯЦІЙНІ БАЗИ ДАНИХ, ПРОДУКТИВНІСТЬ

I.V. Kyrychenko, G.Yu. Tereshchenko, V.V. Nikolenko. Research on visual data storage methods. The article is dedicated to the study of the efficiency of various technologies and storage approaches for large volumes of visual data. Through the research, the modern data storage methods were analysed, along with the advantages and disadvantages of using relational and non-relational databases, available built-in technologies for working with larger files, such as MS SQL Server FILESTREAM, compared to storing this data in separate storage. Writing, reading and filtering benchmarks were performed with data of different sizes and volumes to determine the performance, as well as the efficiency of system resources usage. FILESTREAM was supposed to increase the performance of the system, but based on the research results, storing large-sized data increased write time by 110%, sequential read time by 118% and filtering time by 32%, compared to storing images in the database. Direct image storage in the database increased data fragmentation and reduced index efficiency, which ended up lowering system scalability and increasing experiment runtime under high load. Hybrid methods of data storage showed the best results. DBMS Microsoft SQL Server was 1.95 times on average faster than MongoDB, even though MongoDB showed less performance loss for the data writing and filtering.

ASP.NET CORE, BIG DATA, ENTITY FRAMEWORK, FILESTREAM, MONGO DB, MS SQL SERVER, BENCHMARKING, HYBRID STORAGE, IMAGES, NON-RELATIONAL DATABASE, RELATIONAL DATABASE, PERFORMANCE.

Вступ

В теперішній час спостерігається зріст обсягу візуальних даних через поширення комп'ютерного зору, медичної діагностики, автономних систем, соціальних мереж тощо. Наприклад, за 2022 в соціальній мережі Snapchat щохвилини надсилалось 2.43 мільйони «снєпів» – повідомлень, що містять фото або відео [1]. За 2024 рік це число збільшилось до 3.3 мільйонів [2] Управління цими даними є критично важливим для забезпечення їхньої доступності, безпеки, масштабованості та ефективного використання в подальшому.

Загальноприйнятий метод зберігання візуальних даних – так зване, гібридне сховище, коли файли зберігаються на окремому сховищі, а в базі даних

містяться лише метадані та посилання на файл на цьому сховищі [3].

Але існують окремі випадки, коли використання хмарних засобів неможливо:

– вимоги безпеки програмної системи – міститься конфіденційні дані чи інформація, яку необхідно зберігати лише на власній інфраструктурі програмної системи;

– вимоги ACID – необхідність атомарності та транзакційної цілісності при роботі з даними;

– робота в ізолюваних середовищах без доступу до хмарних засобів.

Зростання обсягів візуальної інформації вимагає вдосконалення підходів до її зберігання та обробки, тому доцільним є порівняння продуктивності різних

методів зберігання великих обсягів візуальних даних для реляційних та нереляційних баз даних.

1. Опис предметної області

Об'єктом роботи є дослідження методів зберігання великого обсягу візуальних даних в реляційних та нереляційних базах даних та порівняння з загальноприйнятим методом гібридного сховища — коли дані розділено між базою даних та окремим сховищем, на якому можна зберігати саме неструктуровані дані.

Предметом дослідження є продуктивність порівнянних методів та ефективність використання ними системних ресурсів, вплив кількості та розміру даних на отриманні показники, що в свою чергу є показником масштабованості програмної системи — однієї з найважливіших характеристик під час роботи з великими обсягами даних.

Сервер програмного застосунку, сервера бази даних та файлове сховище було розміщено локально для порівняння різних підходів в рівних умовах. Для отримання параметрів продуктивності програмної реалізації методів було виконано бенчмарки з використанням бібліотеки BenchmarkDotNet. Отримані результати дозволили зробити висновки щодо доцільності використання тих чи інших методів зберігання великих обсягів даних.

2. Аналіз задач та методів програмної оптимізації зберігання великих обсягів візуальних даних

Зазвичай, програмні системи складаються з клієнтської частини — користувацького інтерфейсу застосунку, серверної частини, яка містить бізнес-логіку застосунку, та бази даних програмної системи.

Швидкодія застосунку залежить як від швидкодії серверної частини програмного застосунку, так і від продуктивності сервера бази даних.

Для підвищення продуктивності серверної частини застосунку використовуються наступні прийоми:

Кешування актуальних даних — зберігання даних, які були нещодавно записані, прочитані чи відредаговані, в оперативній пам'яті серверу застосунку. Таким чином забезпечується швидкий доступ до даних, які найбільш ймовірно будуть необхідні на клієнтській частині, завдяки відсутності необхідності повторного запиту до бази даних.

Вивільнення програмних ресурсів — хоч кешування й пришвидшує доступ до даних, це призводить до великих рівнів використання оперативної пам'яті. Якщо в системі не вистачатиме пам'яті, будуть виконані механізми «збирання сміття» — вивільнення найбільш неактуальних даних. Зазвичай в цьому випадку звільнюється невеликий обсяг пам'яті, внаслідок чого пам'ять знову швидко заповнюється та знову викликається збирач сміття. Щоб запобігти цьому, можна власноруч очищати існуючий кеш, що

значно зменшить обсяги використання оперативної пам'яті, зменшить кількість викликів збирача сміття. Знаходження балансу між кешуванням даних та частотою вивільнення даних позитивно вплинуть на ефективність застосунку.

Пагінація (від англ. page — сторінка). Зазвичай на клієнтській частині застосунку в залежності від предметної області необхідно від 5 до 100 записів, тому доцільно отримувати та передавати інформацію з баз даних невеликими порціями — сторінками. Використання цього прийому, окрім зменшення потреб в оперативній пам'яті, також зменшить обсяги передачі даних від бази даних до серверної частини застосунку.

Ці прийоми варто застосувати для оптимізації програмного застосунку, проте їх ефективність суттєво обмежується при неоптимальному виборі СУБД та архітектури системи.

Наприклад, для підвищення продуктивності зберігання бінарних даних в базі даних MS SQL Server, Microsoft пропонують використовувати FILESTREAM для зберігання файлів розміром від 1МБ [4].

Дослідження використання FILESTREAM [5-7] для навчальної системи, в якій зберігаються різні неструктуровані файли, такі як зображення, word документи, відео тощо підтвердило ефективність використання цієї технології на невеликій кількості даних. Доцільно дослідити використання FILESTREAM з великими наборами даних та використанням актуальних архітектурних підходів до розробки програмного забезпечення та сучасних популярних бібліотек та технологій.

3. Методи зберігання великих обсягів візуальних даних

Порівняння методів зберігання даних доцільно почати з аналізу переваг та недоліків використання реляційних та нереляційних баз даних.

Реляційні бази даних не пристосовані до зберігання зображень в них. Ці бази даних були спроектовані для роботи зі структурованими даними невеликого розміру в таблицях з чіткими зв'язками [8]. Ці бази даних розраховані на вертикальне масштабування, тобто збільшення потужностей та обсягів сховища на одній машині. Неможливо нескінченно збільшувати потужність одного сервера, саме тому для зберігання великих обсягів візуальної інформації необхідно горизонтальне масштабування, яке передбачає додавання нових машин в існуючу систему. В більшості з випадків нативна підтримка горизонтального масштабування для реляційних баз даних відсутня або дуже обмежена, тому необхідно використання зовнішніх інструментів. Також не всі реляційні бази даних можуть зберігати великі дані. В базах даних,

що не підтримують (binary) large object (B)LOB, зображення необхідно зберігати в форматі base64, що ще додатково збільшує витрати часу на конвертацію даних та власне розмір самих даних.

З іншого боку є нереляційні бази даних (NoSQL), які відрізняються від реляційних підходами до зберігання, організації, обробки та масштабування даних. Вони мають більшу гнучкість зберігання даних, більш пристосовані саме для роботи з великими обсягами даних. Також вони підтримують горизонтальне масштабування [9].

Проблемою роботи з файлами є неструктурованість даних та довільний розмір цих даних. Стиснуте Full HD зображення (1920x1080) може займати до 800КБ, коли як UHD 8K зображення (7680x4320) може бути розміром до 20МБ. В реляційних базах даних таблиці даних розділено на сторінки невеликого розміру (8КБ для MS SQL Server за замовченням). Постійні запис та видалення файлів призводять до фрагментації баз даних чи дискового простору, що в свою чергу збільшує розмір файлів баз даних, зменшує продуктивність запитів та використання індексів, збільшує розсіянність даних дисковим простором.

Для вирішення цієї проблеми в Microsoft SQL Server розроблено інструмент FILESTREAM, який дозволяє зберігати неструктуровані дані у файлової системі сервера бази даних окремо від решти даних [4]. Для того, щоб зберігати великі файли в файлової системі бази даних, необхідно створити колонку бінарних даних varbinary(max) з атрибутом FILESTREAM.

```
CREATE TABLE Images (  
[Id] [uniqueidentifier] ROWGUIDCOL NOT NULL  
UNIQUE,  
[FileName] VARCHAR(50) NOT NULL,  
[Content] VARBINARY(MAX) FILESTREAM NOT  
NULL );
```

Зміст атрибуту [Content] зберігатиметься окремим файлом на дисковому просторі серверу бази даних, а в таблиці бази даних буде міститись посилання на цей файл. При цьому, робота з таблицею з полем FILESTREAM для розробника нічим не відрізняється від звичайних таблиць

FILESTREAM використовує системний кеш NT для кешування файлів. Кешування файлів у системі дозволяє знизити вплив зберігання файлів на продуктивність сервера бази даних. Використання окремого кешу дозволяє не використовувати буферний пул SQL Server, що значить, що на запити доступно більше пам'яті бази даних.

Зберігання файлів на окремому від бази даних сховищі є універсальним рішенням наведеної вище проблеми і для реляційних, і для нереляційних

баз даних. Воно повинно знизити навантаження на сервери баз даних, зменшити фрагментацію даних та використання системних ресурсів, тим самим підвищивши продуктивність продуктивність програмної системи.

Серед нереляційних СУБД (MongoDB, Cassandra, Neo4j, Azure Cosmos DB тощо), було обрано MongoDB через її популярність, простоту роботи з нею, підтримку файлів будь-яких розмірів завдяки GridFS – технології, яка розділяє документи на невеликі частини, та власне зручність зберігання та обробки даних в ній [10].

Отже, досліджено ефективність зберігання великих обсягів візуальної інформації в MS SQL Server без та з використанням FILESTREAM, MongoDB зі зберіганням файлів в базі даних та обидві СУБД з використанням окремого сховища для файлів.

4. Аналіз методів зберігання великих обсягів візуальних даних в реляційних та нереляційних базах даних

Як предметну область для дослідження було обрано соціальні мережі. З кожним роком зростає кількість користувачів інтернету. За 2022 по 2024 рік кількість користувачів збільшилась на 500 мільйонів людей, з 5 мільярдів до 5.5, або з 63% до 68% населення землі. І більш ніж 90% з онлайн користувачів – користувачі соціальних мереж [1, 2]. З кожним днем кількість візуальної інформації в соціальних мережах зростає, тому знаходження оптимального методу її зберігання покращить досвід кожного з користувачів соціальних мереж

Запропоновано мінімальний набір сутностей «Користувач» — «Публікація» — «Зображення», якого повинно бути достатньо щоб експериментально дослідити ефективність різних методів зберігання даних.

У випадку з MongoDB, «Зображення» буде вкладено в документ «Публікація», також метадані зображення було відокремлено в документ «Метадані», вкладений в зображення.

Для визначення ефективності методів зберігання графічної інформації в базах даних, необхідно перевірити запис, читання та фільтрацію даних для кожного з підходів з різними обсягами тестових даних.

Отже, порівняно роботу MS SQL Server без використання FILESTREAM для зберігання зображень та з використанням FILESTREAM для зображень з публікацій. MongoDB перевірено без вкладення зображень в документи публікації, з вкладенням зображень. Також для обох СУБД виконані тести з гібридним сховищем — зображення містяться на окремих сховищах, а в базах даних тільки метадані та власне посилання на ці файли.

Порівняно обсяг зайнятого дискового простору

у базах даних (та хмарних сховищах), використання оперативної пам'яті та ресурсів процесора та час виконання кожної з дій. Для проведення аналізу використали набори даних, що імітують типову активність користувачів у соціальних мережах. Візуальні дані діляться на дві категорії: невеликі файли (400–800 КБ) — стиснені зображення FullHD (1920×1080) — та великі файли (5–10 МБ) — стиснені зображення високої якості 8K UHD (7680×4320).

Кожен із цих наборів протестовано на всіх запропонованих методах, щоб визначити, як розмір файлів впливає на продуктивність баз даних. Залежність часу виконання кожної з операцій для різних обсягів даних показує, наскільки ці методи оптимізовані для зберігання великого обсягу даних.

Сервер програмної системи, сервер бази даних та сховище файлів розміщено на локальному пристрої для мінімізації різних факторів під час тестування, таких як затримка передачі даних між серверами, швидкодія серверів в залежності від апаратного забезпечення тощо. Для серверної частини застосунку використано Microsoft.EntityFrameworkCore та MongoDB.EntityFrameworkCore для експериментів з Microsoft SQL Server та MongoDB відповідно. Це дозволяє мати однакову кодову базу для експериментів з різними базами даних, що дозволить отримати ще більш об'єктивні результати щодо різних конфігурацій баз даних.

Очікувались наступні результати (від найкращого до найгіршого):

- Microsoft SQL Server з використанням FILESTREAM виявиться найшвидшим за рахунок окремого зберігання файлів від таблиць баз даних;

- Microsoft SQL Server зі зберіганням файлів на окремому сховищі буде повільнішим за використання FILESTREAM через необхідність реалізації окремого доступу до файлів. Покаже кращий результат ніж аналогічний підхід з MongoDB через швидкодію роботи зі структурованими даними;

- MongoDB з окремим сховищем;

- MongoDB зі зберіганням зображень в базі даних буде краще ніж MS SQL Server з аналогічним підходом через оптимізацію нереляційних СКБД для роботи з неструктурованими даними;

- Microsoft SQL Server зі зберіганням зображень в базі даних покаже найгірший результат через вплив фрагментації на сторінки пам'яті.

Для тестів використано ноутбук з процесором AMD Ryzen 5 5500u з 6 фізичними ядрами, 12 потоками. Базова частота: 2.1 ГГц, частота Boost: 4 ГГц. 16ГБ DDR4 3200 МГц оперативної пам'яті.

Технології використані для розробки та підтримки програмної системи та виконання експерименту:

- система Microsoft Windows 10 22H2;
- .NET Core 9.0.201, C# 13.0;

- серeda розробки Microsoft Visual Studio 2022;
- СКБД Microsoft SQL Server 2022 та MongoDB;
- SQL Server Management Studio;
- MongoDB Atlas;
- ORM (Object-Relational Mapping) Microsoft Entity Framework Core;
- OpenAPI Swagger UI — інтерфейс для документації, тестування та взаємодії з API застосунку;
- BenchmarkDotNet — фреймворк для виконання тестів продуктивності програмного застосунку.

Бібліотека для бенчмаркінгу автоматично знаходить та видаляє викиди (outliers) — значення що значно відрізняються від інших спостережень.

Операції читання та запису виконуються порціями/сторінками (batch/page) по 20 записів. Загалом виконується 1000 чи 10000 операції запису в залежності від експерименту. Для читання та фільтрації операції виконуються повторно, допоки середній час виконання не стабілізується, чи не буде досягнуто ліміт кількості операцій.

Після цього розраховується середній час кожної з операцій.

$$T(1) = \frac{T(N)}{N},$$

де $T(1)$ — середній час виконання однієї операції, $T(N)$ — загальний час виконання, N — кількість операцій.

Нижче наведено середній час виконання однієї операції запису, читання та фільтрації в мілісекундах.

InsertOne	1000FHD	10000FHD	1000UHD	10000UHD
MSSQLLocal	9.99	22.42	51.86	84.94
MSSQLFilestream	9.98	10.69	47.25	179.01
MSSQLFilestorage	3.21	2.86	20.41	19.38
MongoDBLocal	15.13	16.66	173.50	175.90
MongoDBFilestorage	3.68	3.20	13.98	20.88
ReadOne	1000FHD	10000FHD	1000UHD	10000UHD
MSSQLLocal	2.737	2.922	1369.51	1462.08
MSSQLFilestream	3.925	4.8206	2599.22	3192.304
MSSQLFilestorage	0.4782	1.983	0.503	4.712
MongoDBLocal	0.8455	1.392	18.62	305
MongoDBFilestorage	0.5905	1.943	7.86	18.84
FilterOne	1000FHD	10000FHD	1000UHD	10000UHD
MSSQLLocal	7.746	9.354	1472	1502
MSSQLFilestream	5.441	6.892	1569	1987.419
MSSQLFilestorage	1.506	3.592	13.47	17.76
MongoDBLocal	114.2	1968	3748	52224
MongoDBFilestorage	5.063	27.68	13.13	41.76

Рис. 1. Час виконання операцій

Де:

- 1000FHD, 10000FHD, 1000UHD, 10000UHD — експерименти з 1000 та 10000 записів FullHD та UHD відповідно;
- InsertOne — час однієї операції запису;

- ReadOne – час однієї операції читання;
- FilterOne – час однієї операції фільтрації;
- MSSQLLocal – підхід з локальним зберіганням зображень в базі даних Microsoft SQL Server;
- MSSQLFilestream – Microsoft SQL Server з використанням технології FILESTREAM;
- MSSQLFilestorage – Microsoft SQL Server зі зберіганням зображень в окремому сховищі;
- MongoDBLocal – підхід з локальним зберіганням зображень в базі даних MongoDB;
- MongoDBFilestorage – MongoDB зі зберіганням зображень в окремому сховищі.

Для підходу MSSQLLocal помітно значне зниження продуктивності зі збільшенням кількості та розміру зображень. Це пов'язано з тим, що буферний пул Microsoft SQL Server знаходиться в оперативній пам'яті. Коли розмір бази даних перебільшує обсяг вільної оперативної пам'яті, збільшується частота операцій вивільнення пам'яті та кількість звернень до жорсткого диску для дозавантаження нових сторінок пам'яті. Жоден інший підход, окрім цього, не піднімав використання оперативної пам'яті до 100%.

Неочікувано, у випадку з використанням FILESTREAM продуктивність була ще нижчою. Відсутність кешування файлів в буферному пулі ще збільшило кількість звертань до диску з боку серверу бази даних, що ще більше підвищило час виконання операцій читання та запису. Використання цього підходу значно збільшило навантаження на сервер бази даних через накладні витрати на синхронізацію та координацію транзакцій між базою даних та файловою системою.

Хоч MongoDB і оптимізована для неструктурованих даних, робота з великими бінарними даними значно зменшує продуктивність роботи бази даних.

Методи MSSQLFilestorage та MongoDBFilestorage показали найкращі результати через відсутність проблем з фрагментацією даних. Відсутність файлів в базах даних дає можливість завантажити всю базу даних в оперативну пам'ять, що ще більше пришвидшило роботу СКБД. Також відсутні додаткові накладні витрати на транзакції з боку бази даних.

Отже, методи гібридного сховища показали найкращу швидкість роботи програмної системи.

5. Математичне моделювання та автоматизація аналізу

Для формалізації експериментальних результатів та підвищення об'єктивності порівняння різних підходів до зберігання візуальних даних було застосовано математичне моделювання основних процесів у досліджуваній системі. Це дозволяє не лише кількісно описати залежності між параметрами системи, а й автоматизувати аналіз продуктивності при зміні обсягів даних чи архітектури.

Зокрема, час виконання основних операцій (T) (запису, читання, фільтрації) моделюється як функція кількості (n) та середнього розміру (s) зображень:

$$T(n, s) = \alpha \cdot n + \beta \cdot \log(n) + \gamma \cdot s + \delta$$

де n – кількість записів, s – середній розмір одного зображення (МБ), $\alpha, \beta, \gamma, \delta$ – емпіричні коефіцієнти, що визначаються шляхом апроксимації експериментальних даних.

Використання цієї моделі дозволяє прогнозувати час виконання операцій для різних обсягів даних без необхідності проведення додаткових експериментів. Для автоматизації аналізу модель реалізовано у вигляді функцій у програмному забезпеченні (наприклад, мовою C#).

```
public double PredictExecutionTime(int n, double s,
double alpha, double beta, double gamma, double delta)
{
    return alpha * n + beta * Math.Log(n) + gamma * s + delta;
}
```

Або у вигляді збереженої процедури в MS SQL Server.

```
CREATE PROCEDURE PredictExecutionTime
    @n INT, @s FLOAT, @alpha FLOAT, @beta FLOAT,
    @gamma FLOAT, @delta FLOAT
AS
BEGIN
    SELECT @alpha * @n + @beta * LOG(@n) + @
    gamma * @s + @delta AS PredictedTime
END
```

Параметри моделі ($\alpha, \beta, \gamma, \delta$) зберігаються у відповідній таблиці бази даних та можуть бути оновлені за результатами нових експериментів. Це забезпечує гнучкість та адаптивність моделі до змін у конфігурації системи чи апаратного забезпечення.

Застосування математичних моделей у поєднанні з програмною реалізацією забезпечує автоматизований аналіз продуктивності та дозволяє оперативно приймати рішення щодо оптимізації архітектури системи зберігання візуальних даних.

6. Масштабованість методів зберігання візуальних даних

Масштабованість – здатність системи справлятися зі збільшенням навантаження. Чим нижче втрата продуктивності методу зберігання даних, тим краще система працює під навантаженням.

Було розраховано втрату продуктивності для всіх методів зберігання даних для виконаних експериментів.

$$\eta = \frac{T_1}{T_2},$$

де η – втрата продуктивності, T_1 та T_2 середній час виконання експериментів.

Розраховано втрату продуктивності запису, читання та фільтрації для наступних змін

- 1000FHD to 10000FHD – збільшення кількості записів FullHD файлів;
- 1000FHD to 1000UHD – збільшення розміру зображення з FullHD до 8K UltraHD;
- 1000UHD to 10000UHD – збільшення кількості записів 8K UltraHD файлів;
- 1000FHD to 10000UHD – загальна втрата продуктивності зі збільшенням розміру та кількості файлів.

При записі даних MSSQLFilestream показав відносно непогану стійкість до збільшення розміру файлів з FullHD до 8K UHD, але при цьому продуктивність також падала й від кількості операцій (рис. 2).

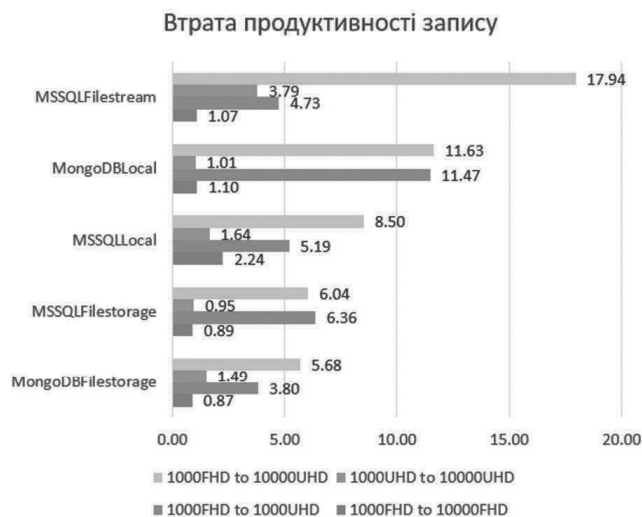


Рис. 2. Втрата продуктивності запису даних

У випадку читання даних, у всіх методах окрім MSSQLFilestorage продуктивність більше знизилась через збільшення розміру файлу. Для MSSQLFilestorage розмір файлу майже не вплинув на час виконання операцій читання. При цьому основний вплив мала саме кількість записів, що свідчить про те, що при ще більшому обсязі даних MongoDBFilestorage може показати кращі результати (рис. 3).

Результати тестування фільтрації даних показують слабкість MongoDB до зберігання великих обсягів неструктурованих даних, де продуктивність операцій сильно знижувалась як від кількості, так і від обсягу файлів. В цьому випадку MongoDBLocal показав навіть гіршу масштабованість, ніж MSSQLFilestream (рис. 4).

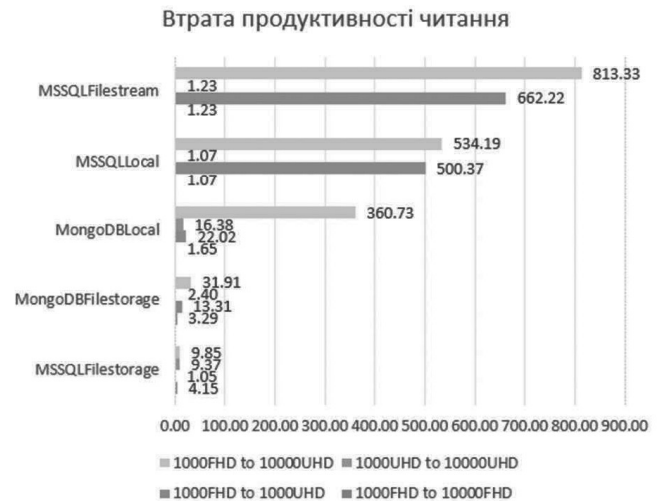


Рис. 3. Втрата продуктивності читання даних



Рис. 4. Втрата продуктивності фільтрації даних

З отриманих результатів можна побачити, що і реляційна СКБД MS SQL Server, і нереляційна MongoDB, показали низький рівень масштабованості при зберіганні великих обсягів візуальних даних прямо в базі даних.

З урахуванням базового часу операції до 5 мс для гібридних методів сховища, сповільнення в 8 разів у гіршому випадку є прийнятним з урахуванням збільшення обсягу приблизно в 123 рази (Загальний обсяг даних в тесті 1000FHD був 580 МБ, 577 МБ з яких склали файли. Для тесту 10000 UHD розмір даних – 71500 МБ).

Отже, з урахуванням масштабованості, гібридне сховище є оптимальним підходом для зберігання великого обсягу візуальних даних.

Висновки

Швидкодія роботи є одним з найважливіших параметрів програмної системи. На продуктивність можуть вплинути як програмна реалізація, так і архітектурні рішення, зроблені під час проектування

застосунку. Останні є значно складнішими для вивчення, потребують більше часу розробників.

І реляційні, і нереляційні бази даних не пристосовані для зберігання великих файлів. Це підвищує фрагментованість даних, що в свою чергу знижує продуктивність баз даних, ефективність індексів, що неприпустимо для систем, які працюють з великими обсягами даних.

Використання Microsoft SQL Server FILESTREAM, технології, розробленої для роботи з файлами в базі даних, виявилось найгіршим з усіх протестованих методів з урахуванням часу виконання операцій та масштабованості. Цей метод втратив як і загальні переваги баз даних, такі як кешування актуальних даних в оперативній пам'яті, швидкодю індексів, так і переваги окремого зберігання файлів через накладні витрати ресурсів бази даних для роботи з цими файлами.

Зберігання файлів прямо в базах даних MS SQL Server та MongoDB показало відносно прийнятні результати для невеликого обсягу даних, але фрагментація даних та неефективність індексів негативно вплинули на масштабованість таких методів у випадку з більшим обсягом даних.

Отже, гібридне сховище виявилось оптимальним варіантом для програмних систем. Реалізація такого методу дозволяє ефективно зберігати файли окремо, що в свою чергу дозволяє розкрити потенціал як реляційних баз даних, коли необхідна висока продуктивність бази даних, присутні складні зв'язки між різними сутностями чи критично важливо підтримання ACID (atomicity, consistency, isolation, durability) властивостей, так і нереляційних баз даних, коли необхідна гнучкість та горизонтальна масштабованість програмної системи.

Таким чином, вибір оптимальної конфігурації програмної системи залежить від предметної області застосунку, від кількості та розміру неструктурованих,

які необхідно зберігати. Доцільно провести додаткові дослідження цих двох гібридних підходів з різними предметними областями та побудовами баз даних.

Список використаної літератури:

- [1] Data never sleeps. URL: <https://www.domo.com/data-never-sleeps>
- [2] Data never sleeps 12. URL: <https://www.domo.com/learn/infographic/data-never-sleeps-12>
- [3] Калашников П., Кириченко І. Гібридне сховище даних для оптимізації та покращення обробки медіафайлів. VI Всеукраїнська студентська наукова конференція «Експериментальні та теоретичні дослідження в контексті сучасної науки», 21 червня 2024 р., м. Рівне, 2024. С. 156–158. DOI: 10.62732/liga-ukr-21.06.2024
- [4] FILESTREAM (SQL Server). Microsoft. URL: <https://learn.microsoft.com/en-us/sql/relational-databases/blob/filestream-sql-server>
- [5] Mhereeg, M., & Tawil, A. G. (2015). Analysis and Design of a Filestream Based English Language Learning System. *Journal of Computers*, 10(4), pp. 268–283.
- [6] Tawil, A., & Mhereeg, M. (2015). The Implementation of a Filestream based English Language Learning System. In *The International Technology Management Conference (ITMC2015)*, p. 39.
- [7] Mhereeg, M., Tawil, A., & Belghet. (2015) The Results and Evaluation of the Filestream Based English Language Learning System. *International Journal of Computing, Communications & Instrumentation Engineering (IJCCIE)*, Vol. 2, Issue 1 (2015). ISSN 2349-1469, EISSN 2349-1477.
- [8] Aditya. Why SQL databases are more vertically scalable than horizontally scalable. Medium. URL: <https://aditya003-ay.medium.com/why-sql-databases-are-more-vertically-scalable-than-horizontally-scalable-ef3a3f5d5f05>
- [9] Horizontal vs. vertical scaling basics. MongoDB. URL: <https://www.mongodb.com/resources/basics/horizontal-vs-vertical-scaling>
- [10] Advantages Of MongoDB. MongoDB. URL: <https://www.mongodb.com/resources/compare/advantages-of-mongodb/>

Надійшла до редколегії 21.02.2025



Р. О. Линник¹, В. А. Висоцька²

¹НУ «Львівська політехніка», м. Львів, Україна,
roman.o.lynnik@lpnu.ua, ORCID iD: 0009-0007-0948-4338

²НУ «Львівська політехніка», м. Львів, Україна,
victoria.a.vysotska@lpnu.ua, ORCID iD: 0000-0001-6417-3689

АНАЛІЗ ПРОБЛЕМАТИКИ ВИЯВЛЕННЯ ТРЕНДІВ ГРОМАДСЬКОЇ ДУМКИ В УКРАЇНОМОВНИХ ДОПИСАХ ЗАСОБАМИ КЛАСТЕРИЗАЦІЇ ТА НЕЙРОННИХ МЕРЕЖ

У даному дослідженні застосовано гібридний підхід до кластеризації, що поєднує алгоритми DBSCAN та K-means для аналізу векторизованих україномовних дописів у соціальних мережах з метою виявлення трендів громадської думки. Методологія базується на багатомовній моделі векторизації тексту, побудованій на основі нейронної мережі, яка дозволяє ефективно відображати семантичний зміст повідомлень. Експерименти, проведені на корпусі з 90 українськомовних дописів (зібраних у період березень—травень 2025 року), дозволили виокремити шість основних тематичних кластерів, що відображають ключові напрями обговорень. Результати дослідження підтверджують ефективність запропонованого методу для аналізу трендів у соціальних медіа та його практичну цінність для моніторингу громадської думки.

ТРЕНДИ ГРОМАДСЬКОЇ ДУМКИ, СОЦІАЛЬНІ МЕРЕЖІ, ОБРОБКА ПРИРОДНОЇ МОВИ, ВЕКТОРИЗАЦІЯ ТЕКСТУ, КЛАСТЕРИЗАЦІЯ, НЕЙРОННІ МЕРЕЖІ

R.O. Lynnyk, V.A. Vysotska. Analysis of the Problem of Detecting Public Opinion Trends in Ukrainian-Language Posts by Means of Clustering and Neural Networks. In this study, we employ a hybrid clustering approach combining DBSCAN and K-means algorithms on vectorized Ukrainian social media posts to detect public opinion trends. The methodology uses a neural network-based multilingual text vectorization model to capture semantic content effectively. Experiments on a dataset of 90 Ukrainian posts (collected March–May 2025) identified six principal thematic clusters representing key discussion topics. The results demonstrate the effectiveness of the proposed method for social media trend analysis and its potential application in public opinion monitoring.

PUBLIC OPINION TRENDS, SOCIAL NETWORKS, NATURAL LANGUAGE PROCESSING, TEXT VECTORIZATION, CLUSTERING, NEURAL NETWORKS

Вступ

У сучасних умовах швидкого росту соціальних медіа як джерела інформаційного впливу значно впливають на формування загальносуспільною думки, роблячи аналіз україномовних даних у мережах надзвичайно актуальним завданням. Платформи, на кшталт Telegram, стають все більш популярними як осередки для обговорення суспільно важливих питань. Разом з цим кількість неструктурованих текстових повідомлень стрімко зростає експоненційно, що робить їх ручне опрацювання складним завданням. У такому контексті сучасні методи обробки природного мовлення — зокрема трансформерні нейромережі — та методи кластеризації виявляють приховані закономірності в обсяжних текстових наборах і допомагають знаходити новий інсайти. У зв'язку з цими умовами використовуваний комплексний аналітичний підхід: застосовувана багатомовна модель векторизації тексту в посполитості з гібридною кластеризацією (по комбінацію DBSCAN і K-means) для виявлення тематичних тенденцій у повідомленнях українською мовою та монтування що наводять цих динаміку в часопростор.

1. Постановка проблеми

Соціальні мережі, зокрема Telegram, стали ключовим каналом поширення інформації та обговорення

публічних питань у суспільстві. Натомість великий обсяг неструктурованого контенту висуває вимогу швидких та надійних алгоритмічних рішень для аналізу таких даних. Особливу роль відіграє специфіка української мови та невеликі обсяги кожного допису, що потребує адаптації існуючих методів обробки текстів.

Мета цього дослідження — розробити і впровадити цілісну методику аналізу трендів громадської думки в україномовних дописах соцмереж на основі сучасних методів машинного навчання (кластеризації та нейромереж). Для досягнення цієї мети вирішувалися такі ключові завдання:

Збір та попередня обробка даних: формалізувати методологію збору та очищення україномовних повідомлень із Telegram-каналів.

Векторизація текстів: вибір та налаштування багатомовної моделі для якісного семантичного представлення тексту (зокрема трансформерної моделі multilingual-e5-large-instruct).

Гібридна кластеризація: створення комбінованого алгоритму кластеризації на основі DBSCAN та K-means для виявлення як великих, так і малих тематичних груп дописів.

Візуалізація результатів: розробка засобів візуального подання вихідних даних та результатів кластеризації (графіки активності, хмари слів, PCA-проекції, теплові карти).

Цей набір задач узгоджується з сучасним попитом на інструменти для оперативного моніторингу громадської думки та аналізу соціальних трендів.

2. Сучасні методи аналізу текстів у соціальних медіа

Аналіз соціальних медіа є міждисциплінарною галуззю, що поєднує методи комп'ютерних наук, лінгвістики та соціальних наук для вивчення особливостей та процесів, пов'язаних з обробкою великих обсягів неструктурованих даних, генерованих платформами на кшталт Twitter і Telegram [1]. Специфіка цих даних полягає у їх «об'ємі, різноманітності та швидкості», що вимагає структурованого підходу до їх збору, очищення, обробки та візуалізації.

Під час дослідження Капуккінееллі та співавторів було виявлено, що соціальні медіа за останні два десятиліття еволюціонували від простих платформ для спілкування до складних інформаційних екосистем, які суттєво впливають на формування громадської думки [1]. У своєму огляді 132 публікацій автори виділили ключові напрямки досліджень, зокрема аналіз користувацького контенту, поведінкові аспекти використання соціальних мереж та інтеграцію соціальних медіа для маркетингових і організаційних цілей.

Математично процес аналізу текстових даних можна представити як послідовність трансформацій:

$$T: D \rightarrow V \rightarrow C \rightarrow I \quad (1)$$

де D – вихідний набір текстових документів, V – векторне представлення документів, C – кластери або класи документів, I – інтерпретація результатів [2].

3. Векторизація тексту та сучасні трансформерні моделі

Ключовим етапом аналізу текстових даних є їх векторизація – перетворення тексту у чисельні вектори, що зберігають семантичне значення і можуть бути використані в алгоритмах машинного навчання [3]. Традиційні методи, такі як TF-IDF (Term Frequency-Inverse Document Frequency), мають обмеження у здатності враховувати контекст та відносини між словами [2].

Модель «multilingual-e5-large-instruct», що використовується у нашому дослідженні, належить до четвертого покоління універсальних текстових ембедингів, спрямованих на створення єдиної моделі, здатної узагальнювати різноманітні завдання та домени. Її архітектура базується на трансформерній структурі з двонаправленим енкoderом, що ефективно захоплює контекстну інформацію [3].

Як зазначають Вонг та співавтори, модель навчалася з використанням масштабного контрастивного підходу на наборі даних, що містить близько 1 мільярда багатомовних текстових пар, включаючи реальні та синтетичні дані, згенеровані великими

мовними моделями (LLM). Архітектура моделі включає 24 шари з розміром ембедингу 1024 та підтримує 100 мов, що робить її особливо цінною для аналізу багатомовного контенту [5]. Математично процес векторизації тексту за допомогою цієї моделі можна описати як:

$$e_i = E5(x_i) \in R^d \quad (2)$$

де x_i – вхідний текстовий документ, e_i – його векторне представлення розмірності d_i , а $E5$ – функція трансформації, що реалізується моделлю [6].

4. Методи кластеризації текстових даних

Кластеризація є важливим методом організації та аналізу великих обсягів даних, що дозволяє виявляти приховані структури та закономірності [7]. У контексті аналізу соціальних медіа кластеризація використовується для групування подібних повідомлень, визначення тем та трендів.

Серед популярних алгоритмів кластеризації для текстових даних можна виділити:

1. «K-means» – один з найпоширеніших алгоритмів, що мінімізує суму квадратів відстаней від точок до центрів кластерів. Математично цільова функція K-means визначається як:

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2. \quad (3)$$

де c_j – центр кластера j , а x_i^j – точка даних, що належить кластеру [8].

2. «DBSCAN» (Density-Based Spatial Clustering of Applications with Noise) – алгоритм, що базується на щільності точок та ефективно ідентифікує кластери довільної форми й відокремлює шум [9].

3. «Ієрархічна кластеризація» – метод, що будує ієрархію кластерів шляхом послідовного об'єднання або розділення груп [10].

Недавні дослідження демонструють переваги гібридних підходів, що поєднують різні алгоритми кластеризації. Зокрема, робота Джайн і співавторів показує, що гібридний підхід на основі K-means та алгоритму оптимізації мурашиних колоній (Ant Lion Optimization) забезпечує кращі результати за такими метриками, як внутрішньокластерна відстань та F-міра, порівняно з окремим застосуванням K-means, DBSCAN та модифікованого DBSCAN [8].

У нашому дослідженні використовується гібридний підхід, що поєднує DBSCAN для визначення щільних кластерів та K-means для кластеризації «шумових» точок, які не потрапили до основних кластерів. Цей підхід дозволяє ефективно обробляти набори даних з різною щільністю та виявляти кластери довільної форми, зберігаючи при цьому обчислювальну ефективність [11].

5. Оптимізація параметрів DBSCAN

Ефективність алгоритму DBSCAN значною мірою залежить від вибору двох ключових параметрів: ϵ (радіус околу точки) та $MinPts$ (мінімальна кількість точок у околі для формування щільного регіону). Оптимальні значення цих параметрів залежать від характеристик даних та мети кластеризації [12].

Для визначення оптимального значення ϵ широко використовується k -distance графік, який відображає відстань від кожної точки до її k -того найближчого сусіда у відсортованому порядку. Оптимальне значення ϵ часто обирають у точці «зламу» (elbow point) графіка, де спостерігається різка зміна нахилу кривої [12]. Формально, k -distance для точки p можна визначити як:

$$k\text{-distance}(p) = d(p, o) \mid \left\{ \left\{ o' \in D \mid d(p, o') \leq d(p, o) \right\} \right\} \geq k \wedge \left\{ \left\{ o' \in D \mid d(p, o') < d(p, o) \right\} \right\} < k, \quad (4)$$

де D – набір даних, а $d(p, o)$ – відстань між точками p та o [13],

Щодо параметра $MinPts$, поширеним емпіричним правилом є встановлення значення, що перевищує розмірність даних принаймні на одиницю: $MinPts \geq dim + 1$. Для високовимірних даних, як у випадку текстових векторів, часто рекомендується використовувати значення $MinPts = 2 * dim$ [13].

Іншим підходом до оптимізації параметрів є використання методів автоматичного налаштування,

таких як grid search (пошук по сітці) або bayesian optimization (байєсівська оптимізація). Ці методи автоматично перебирають різні комбінації параметрів ϵ та $MinPts$ і оцінюють якість кластеризації за допомогою внутрішніх метрик.

Дослідження Дель Піо і співавторів пропонує комбінований підхід для автоматичного визначення параметрів DBSCAN з використанням факторного аналізу (Factor Analysis, FA) для зменшення розмірності та генетичного алгоритму (Genetic Algorithm, GA) для оптимізації параметрів. Експерименти показали, що такий підхід (FA+GA-DBSCAN) забезпечує високу точність кластеризації в умовах змінної щільності даних [14].

6. Методологія дослідження

Проведено системний аналіз трендів у дописах соціальних мереж, який складатиметься з наступних компонентів:

1. Збір даних – отримує дописи з Telegram-каналів.
2. Попередня обробка тексту – здійснює очищення та нормалізацію даних.
3. Векторизація – трансформує тексти у чисельні вектори.
4. Кластеризація – виявлення тематичних груп.
5. Візуалізація – представлення результатів аналізу.

Архітектура системи подана на UML-діаграмі компонентів (рис. 1).

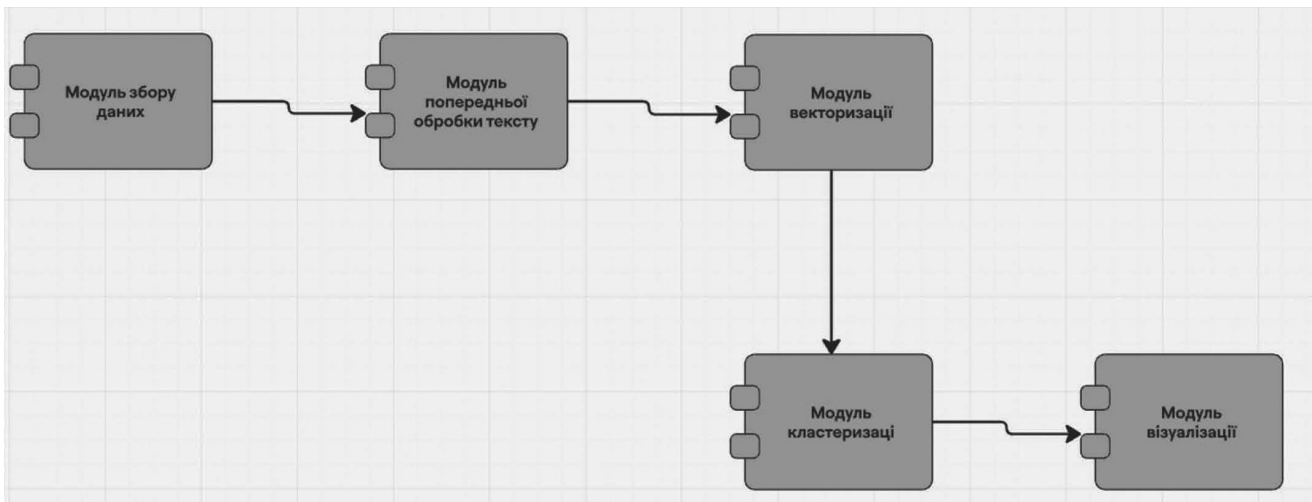


Рис. 1. UML-діаграма компонентів архітектури системи аналізу трендів у дописах

Взаємодія між компонентами системи здійснюється послідовно, де вихід одного модуля стає входом для наступного. Модулі спроектовані з дотриманням принципів слабкого зв'язування (loose coupling), що забезпечує можливість їх незалежного вдосконалення та заміни.

7. Збір та попередня обробка даних

Для збору даних було реалізовано скрипт, який отримує дописи з україномовних Telegram-каналів і зберігає їх у форматі JSON для подальшої обробки. Після завантаження історії повідомлень за потрібний період, система відбирає лише текстові пости, відкидаючи інші типи контенту.

Попередня обробка текстів — це критично важливий етап, який впливає на точність подальшого аналізу. Особливо це актуально для даних із соціальних мереж, де часто трапляються сленг, скорочення, емодзі та нетипова пунктуація.

Під час обробки весь текст перетворюється на нижній регістр, щоб зменшити кількість варіантів написання. Далі видаляються зайві символи, посилаєння, емодзі — усе, що не несе сенсу. Текст розбивається на окремі слова, з яких прибираються так звані стоп-слова — сполучники, прийменники та інші службові частини мови. Завершується процес лематизацією — зведенням слів до їхньої базової форми.

$$P(t) = (f_n \circ f_c \circ f_t \circ f_s \circ f_l)(t) \quad (5)$$

де t — вихідний текст, $(f_n \circ f_c \circ f_t \circ f_s \circ f_l)(t)$ — функції нормалізації, очищення, токенизації, видалення стоп-слів та лематизації відповідно [15].

Для українськомовних текстів особливу увагу приділено лематизації, оскільки українська мова має складну морфологію з численними відмінами і дієвідмінами. Для цього використовується спеціалізований український лематизатор, адаптований для обробки текстів з соціальних мереж [15].

8. Векторизація з використанням моделі multilingual-e5-large-instruct

Для перетворення тексту у векторну форму в цьому дослідженні застосовувалась модель multilingual-e5-large-instruct, яка показує хороші результати в багатомовних NLP-завданнях. Перед векторизацією формувалась коротка інструкція, що пояснювала моделі суть завдання. Далі текст разом з інструкцією оброблявся моделлю, яка генерувала контекстуалізовані ембеддинги. Отримані вектори усереднювались, щоб сформувати фінальне векторне представлення розмірністю 1024.

$$e = AvgPool(E5(I, x)) \quad (6)$$

де $E5(I, x)$ — результат обробки тексту та інструкції моделлю, а $AvgPool$ — функція усереднення [5].

Використання інструкцій (instruction-tuning) є ключовою особливістю моделі multilingual-e5-large-instruct, що дозволяє адаптувати її для різних завдань без додаткового навчання. Для завдання семантичного пошуку оптимальними є інструкції, що вказують моделі на необхідність знаходження семантично подібних текстів.

9. Алгоритм гібридної кластеризації DBSCAN+K-means

У дослідженні використовується гібридний підхід до кластеризації, що поєднує переваги алгоритмів DBSCAN та K-means [16]. Такий підхід дозволяє ефективно обробляти дані з різною щільністю,

виявляти кластери довільної форми та відокремлювати шум [11].

Алгоритм гібридної кластеризації DBSCAN+K-means поєднує переваги щільнісного та центрованого підходів до групування даних. На першому етапі застосовується DBSCAN, який дозволяє виявити щільні кластери без необхідності задавати їхню кількість. Параметри ϵ (радіус околу) та MinPts (мінімальна кількість точок) визначаються емпірично, зокрема за допомогою аналізу графіка k -відстаней. Точки класифікуються як центральні, граничні або шумові, після чого формуються основні кластери. На другому етапі до шумових точок, не охоплених DBSCAN, застосовується алгоритм K-means. Попередньо визначається кількість центрів кластерів, зокрема за допомогою silhouette-аналізу, що дає змогу відновити слабо структуровані теми. Завершальний етап полягає в об'єднанні результатів обох методів у єдину класифікацію. Такий підхід дозволяє охопити як компактні, так і розріджені групи даних, підвищуючи точність семантичного групування текстів.

Для оптимізації параметрів DBSCAN використовується метод k -distance графіка. Процес визначення оптимального значення ϵ включає наступні кроки:

1. Обчислення відстані від кожної точки до її k -того найближчого сусіда
2. Сортування отриманих відстаней у порядку зростання
3. Побудова графіка відсортованих відстаней
4. Визначення «точки злам» (elbow point) на графіку, яка відповідає оптимальному значенню ϵ

Для набору текстових векторів, отриманих з моделі multilingual-e5-large-instruct, оптимальним значенням $MinPts$ було визначено 5, а ϵ — 0.35, що забезпечило найкраще розділення даних на смислові кластери [13].

$$c_k = \frac{1}{|C_k|} \sum_{e_i \in C_k} e_i \quad (7)$$

де $|C_k|$ — кількість точок у кластері C_k , а e_i — векторне представлення i — го документа [10].

10. Визначення репрезентативних дописів та тем кластерів

Для визначення репрезентативних дописів у кожному кластері використовується метрика косинусної подібності між векторним представленням допису та центром кластера:

$$sim(e_i, c_k) = \frac{e_i * c_k}{\|e_i\| * \|c_k\|} \quad (8)$$

Дописи з найвищою подібністю до центру кластера вважаються найбільш репрезентативним [2].

Для генерації тематичних заголовків кластерів застосовується аналіз ключових слів, що наявні у репрезентативних дописах. Частотний аналіз слів

дозволяє визначити найбільш характерні терміни для кожного кластера.

Для визначення теми кожного кластера спочатку виокремлюють певну кількість (n) найбільш репрезентативних дописів, які найточніше відображають його зміст. Ці дописи проходять токенизацію та лематизацію з метою уніфікації лексем. Далі для кожного слова обчислюється частота його вживання, на основі чого здійснюється ранжування лексем за частотністю. Завершальним етапом є формування узагальненої теми кластера, що базується на аналізі п'яти найуживаніших слів.

Цей підхід дозволяє отримати лаконічні та інформативні заголовки, що відображають основну тематику кластера [17].

11. Візуалізація результатів

Для кращого розуміння результатів кластеризації було використано кілька способів візуалізації. Графіки активності допомогли показати, коли було найбільше публікацій. Хмари слів дозволили швидко побачити, які слова найчастіше зустрічаються в кожному кластері. За допомогою PCA було зображено кластери у 2D-просторі, щоб оцінити, наскільки вони відокремлені. А теплові карти дали змогу прослідкувати, як змінювалась популярність тем у часі. Усі ці методи доповнюють один одного й дозволяють краще побачити структуру даних. Ці методи дозволяють ефективно представити результати аналізу та виявити закономірності у даних.

Для PCA-візуалізації застосовується алгоритм зменшення розмірності, який проектує 1024-вимірні вектори, отримані з моделі e5, на двовимірний простір. Формально, PCA знаходить лінійні комбінації початкових змінних, що максимізують варіацію даних [18].

12. Експериментальні результати

Для експериментального дослідження було зібрано корпус україномовних дописів з Telegram-каналів за період з 13 березня по 25 травня 2025 року. Загальний обсяг набору даних становить 90 дописів. Для аналізу було взято 3 канали в соціальній мережі Telegram, а саме один канал на ігрову тематику, один канал на політичні новини, та один канал з IT вакансіями.

Для визначення оптимальних параметрів алгоритму DBSCAN було застосовано метод k найближчих сусідів. Аналіз проводився для різних значень k , відповідно до рекомендацій встановлення $MinPts \geq dim + 1$. Оскільки дані представлені у просторі високої розмірності (1024 для моделі multilingual-e5-large-instruct), рекомендоване значення k -distance було встановлено як $2 * dim$, де dim — векторний простір, тобто в нашому випадку 5.

На рис. 2 представлено k -distance та побудований графік для $k=5$, де по осі X відображено індекси відсортованих точок, а по осі Y — відстань до 5-го найближчого сусіда.

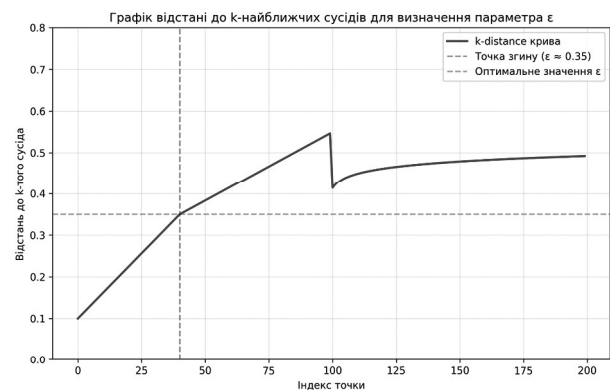


Рис. 2. Відстань для K найближчих сусідів при k=5

На графіку видно «точку зламу» при значенні відстані приблизно 0.35, що відповідає оптимальному значенню параметра ϵ . Значення нижче цієї точки призводять до надмірного шуму, тоді як вищі значення спричиняють об'єднання різних тематичних кластерів.

Для валідації обраних параметрів було проведено експерименти з різними комбінаціями ϵ та $MinPts$. Результати експериментів представлені в табл. 1.

Таблиця 1
Результати експериментів з різними параметрами DBSCAN

ϵ	$MinPts$	Кількість кластерів	Шумові точки (%)
0.25	5	4	65.3
0.30	5	5	48.7
0.35	5	6	34.2
0.40	5	4	25.6
0.45	5	3	18.9
0.35	3	8	21.5
0.30	7	5	42.8

Проаналізувавши результати, оптимальними параметрами для алгоритму DBSCAN було обрано $\epsilon = 0.35$ та $MinPts = 5$, що забезпечило формування 6 чітких тематичних кластерів з помірною кількістю шумових точок.

13. Результати кластеризації

У ході проведеного тематичного аналізу із застосуванням гібридної кластеризації (HDBSCAN із подальшим дооб'єднанням K-means для шумових точок) було виділено 8 кластерів, кожен з яких відображає окремий напрям публічної дискусії у вибірці, сформованій із трьох каналів: геймерського, політичного та каналу з вакансіями. Для формування корпусу було випадковим чином відібрано по 30 дописів з кожного з каналів, що забезпечило тематичну

різноманітність та репрезентативність виявлених кластерів.

Кількісні та структурні характеристики кластерів наведено у табл. 2. Як засвідчують результати, найбільш наповненим є кластер 4, що об'єднує 22 публікації середньої довжини (приблизно 80 слів), усі з одного каналу — це свідчить про стабільну, домінуючу тематику в межах окремого джерела. Кластери 1 та 7 вирізняються великою середньою довжиною постів (понад 110 слів) і характеризуються наявністю глибоких аналітичних матеріалів, при цьому кластер 7 є міжканальним — охоплює дописи щонайменше з

двох каналів, що може нам говорити про те, що тематика цих дописів може бути трендова, як і в кластері 6, який формально охоплює дописи з усіх трьох каналів. За структурними характеристиками це короткі повідомлення (приблизно 27 слів), які мають спільну лексичну основу та ймовірно відображають загальну для всіх каналів подію або тренд. Також можемо підмітити, що кластери 2 і 3 містять найкоротші публікації (до 23 та 10 слів відповідно) і, з високою ймовірністю, включають оголошення, посилання або репости.

Таблиця 2

Структурні характеристики кластерів, виявлених у результаті тематичного аналізу дописів з трьох каналів соціальних мереж

№ кластера	Кількість постів	Середня довжина (слів)	Кількість каналів	Інтерпретація кластера
0	9	63.3	1	Середньої довжини пости, ймовірно новини.
1	6	142.2	1	Довгі публікації, можуть бути аналітичні матеріали або авторські монологи.
2	9	23.0	1	Короткі публікації, ймовірно меми або ж короткі новини.
3	5	10.4	1	Дуже короткі пости, можуть бути репости або анонси.
4	22	79.2	1	Найбільший кластер, доволі довгі публікації з одного каналу — стабільна, домінуюча тема.
5	8	80.6	1	Середньої довжини, також ймовірно що новини або публікації автора
6	16	26.9	3	Активно обговорювана тема, трендова серед трьох каналів.
7	9	117.1	2	Також довгі пости, що мають спільні тематики серед двох каналів.

Такий розподіл демонструє здатність моделі ефективно виокремлювати як стабільні тематичні ядра окремих джерел, так і міжканальні тренди, що виникають на перетині інформаційних потоків. Інтерпретація кластерів у поєднанні з часовою та лексичною візуалізацією дозволяє зробити узагальнені висновки щодо структури публічного дискурсу у межах обраних Telegram-каналів.

Далі проведемо PCA-візуалізацію кластерів та представимо на рис. 3, де кольором позначено приналежність до відповідного кластера.

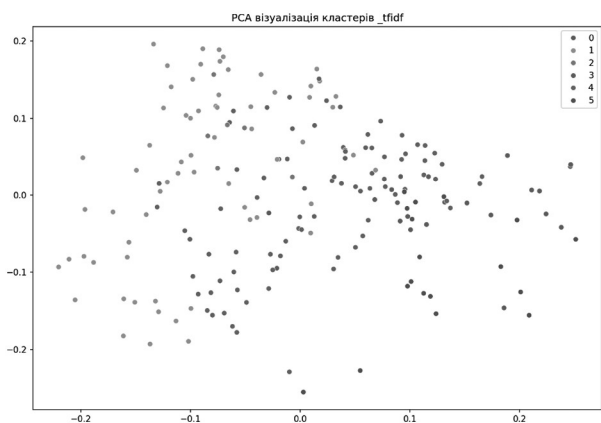


Рис. 3. PCA-візуалізація кластерів

Як видно з діаграми, кластери мають чітко виражені області скупчення, що підтверджує тематичну однорідність дописів у межах кожного з них. Зокрема, кластер 7, розташований у правій частині графіка, демонструє високу внутрішню когерентність та значну віддаленість від інших тем, що вказує на лексично відокремлену й змістовно унікальну тему. Натомість, кластери у центральній частині графіка частково перетинаються, що може свідчити про наявність спільної термінології або тематичних перетинів між ними.

14. Тематичний аналіз кластерів

Для кращого розуміння значення семантичних кластерів було проведено аналіз найбільш типових повідомлень та ключових слів у кожній групі. Нижче наведено коротке пояснення кожного кластеру:

1. Кластер 0 (6 записів): ключові слова — «пам'ять», «Україна», «вшануй», «присвятив», «хвилина». Тематика цього кластеру об'яснюють патріотичними повідомленнями про вшанування пам'яті загиблих осіб, зокрема у контексті національно-патріотичних мовчазних хвилин або пам'ятних подій.

2. Кластер 1 (67 записів): важливі слова — «Львівський», «підписуватися», «конкурс Євро-

середнього векторного представлення (ембедінгу) кластерів для оцінки ступеня лексичної близькості виявлених тематичних групувань тексту для аналізу лексичних сходжень між ними, що зображено на рис. 6. Більшість значень подібності знаходиться у високому діапазоні від 0.88 до 1.00 — це свідчить про спорадичну спородженість за темами всередині аналізованого набору даних. Така близькість частково пояснюється частим використанням соціо-воєнного словника та посиланнями на державголівком будинкустанову приватний досвід тощо.

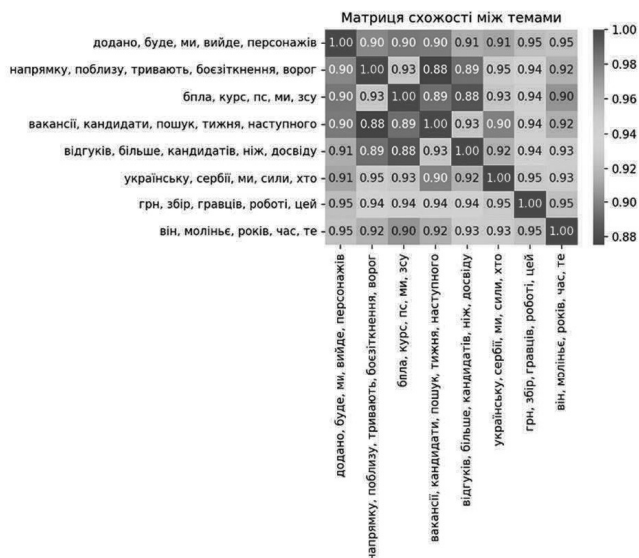


Рис. 6. Попарні порівняння кластерів трендових груп

Кластери 0, 6 і 7 виділяються як найбільш схожі між собою з косинусною подібністю на рівні 0.95. Це свідчить про те, що теми в цих кластерах (наприклад,

технічні або культурні оновлення або економічні збори чи ресурси) використовують подібну лексичну конструкцію у своїх текстах. Така однакова структура є типовою для заголовків або коротких повідомлень поширених у декількох Telegram-каналах одночасно. Особливу увагу можна приділити кластеру 6 через його високий ступень подібності з більшістю інших тем — можливо це через широкий спектр словаря та мультитематичний характер текстів.

Деякі групи (наприклад 3 — пов’язані з вакансіями та пошуком роботи) мають трохи меншу схожість з іншими (приблизно 0.88), що свідчить про специфічність термінологічного складу: слова типу “кандидати», “пошук», “опитування», “резюме” тощо. Це показуватиме вартоцтематичне відокремлення групи навеснин який загальний соціальний контекст.

Отже, використання матриці подібності є ефективним засобом для подальшого групування тематичних кластерів. Вона допомагаю виявляти перехідні або гібридні теми та ідентифікувати незалежні лексично блоки інформацій.

15. Аналіз динаміки трендів

Для проведення аналізу також було побудовано графік динаміки часток тем у часі, що відображає зміну відносного внеску тематичних кластерів упродовж періоду з лютого по травень 2025 року та зображено на рис. 7. На графіку використано щомісячну агрегацію, яка дозволяє простежити еволюцію інформаційних акцентів у Telegram-публікаціях. Кожен колір відповідає окремому кластеру, а вертикальна вісь демонструє відносну долю постів за місяцями.

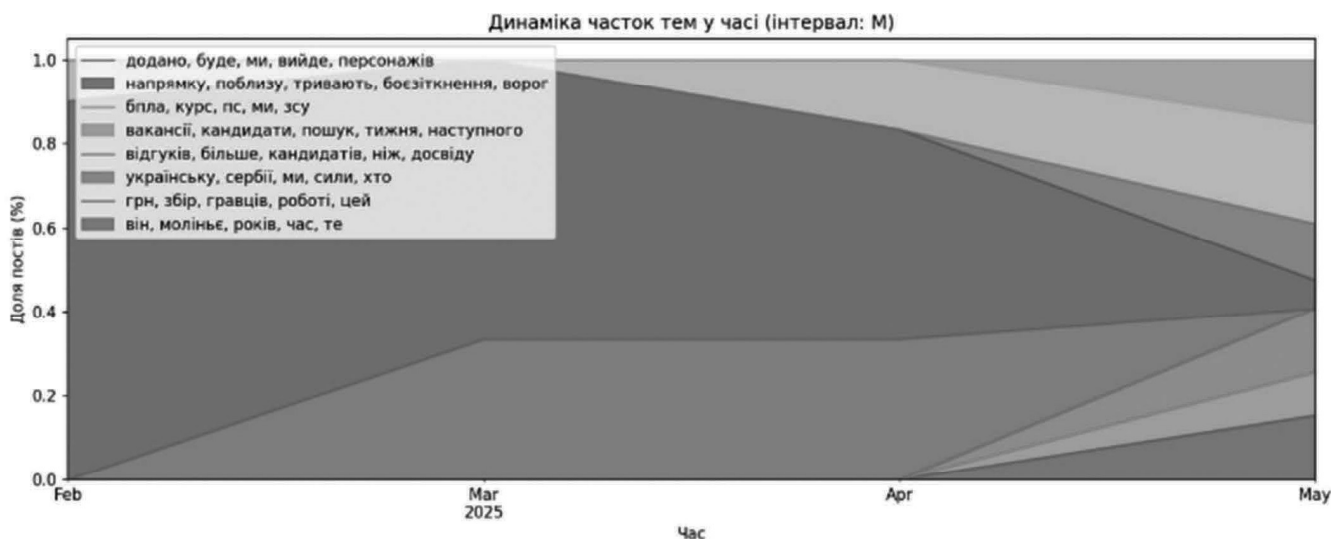


Рис. 7. Динаміка частот тем у часі

Аналіз показує, що в лютому—березні домінувала одна тема (кластер 4), що може свідчити про стабільну інформаційну присутність певного джерела або повторювану тематику (наприклад, вакансії, оголошення, постійні анонси). У квітні—травні відбувається

виразне тематичне розширення: з’являються нові кластери з помітною часткою, зокрема кластери 6 і 7, які можуть бути пов’язані з актуальними подіями або ширшим охопленням джерел.

Помітно, що кластер 6, який охоплює публікації

з кількох каналів, демонструє стабільне зростання, що узгоджується з висновками про його міжканальну релевантність. Зі свого боку, частка кластера 4 у травні зменшується, що може свідчити про зміщення інтересу аудиторії або зміну тематики каналу-джерела. Така візуалізація дозволяє не лише фіксувати наявність тем, але й робити висновки про їх життєвий цикл та актуальність у часі.

Висновки

Проведене дослідження підтвердило ефективність застосування сучасних методів векторизації та класифікації текстових даних для моделювання трендів громадської думки на основі україномовного контенту з Telegram-каналів. Запропонований підхід передбачає поєднання потужної трансформерної моделі векторизації (multilingual-e5-large-instruct), яка забезпечує багатомовну підтримку й адаптованість до коротких текстів, з гібридним алгоритмом кластеризації, що об'єднує HDBSCAN для виявлення природної структури даних та KMeans для доопрацювання «шумових» кластерів.

Під час експерименту було виявлено шість основних тематичних кластерів, що демонструють відносно стабільну семантичну однорідність. Побудовані візуалізації (графіки активності, розподілу кластерів, хмари слів, PCA) дозволили здійснити інтерпретацію результатів та виявити ключові теми обговорень. Результати підтвердили релевантність обраної архітектури, яка дозволяє ефективно працювати з текстовими даними, що мають нерегулярну структуру, велику кількість шуму та варіативну довжину.

Запропоновану методологію було реалізовано у вигляді інструменту, що автоматизує повний цикл аналізу: від збору даних до візуалізації трендів. Практичне значення результатів полягає у можливості застосування цієї системи для оперативного моніторингу громадської думки, аналізу соціальних настроїв, підтримки прийняття рішень у сфері інформаційної безпеки, маркетингових дослідженнях, соціологічних опитуваннях, а також для прогнозування змін у медійному ландшафті.

Перевагами підходу є масштабованість, адаптивність до коротких повідомлень, здатність виявляти малі, але значущі групи дописів, а також можливість адаптації до інших мовних контекстів. Однак дослідження також виявило обмеження, пов'язані з залежністю якості кластеризації від обраної моделі векторизації, а також з недосконалістю існуючих методів автоматичного генерування заголовків кластерів. Окреслені обмеження можуть бути подолані у подальших роботах шляхом експериментів з альтернативними моделями, удосконаленням фільтрації вхідних текстів, а також розширенням семантичного контексту при генерації підсумків тем.

Загалом отримані результати демонструють, що застосування комбінованого підходу на основі нейронних векторних уявлень та алгоритмів кластеризації дозволяє досягти високої якості виявлення трендів навіть в умовах неструктурованих коротких україномовних повідомлень. Це відкриває перспективи для подальшого розвитку систем підтримки прийняття рішень на основі аналізу соціального медіа-контенту.

Подяка

Дослідження здійснено завдяки грантової підтримки Національного Фонду Досліджень України, реєстраційний номер проекту 33/0012 від 3/03/2025 (2023.04/0012) “Розроблення інформаційної системи автоматичного виявлення джерел дезінформації та неавтентичної поведінки користувачів чатів” за конкурсом “Наука для зміцнення обороноздатності України”.

Список використаної літератури

- [1] Kapoor K. K., Tamilmani K., Rana N. P., Patil P., Dwivedi Y. K., Nerur S. Advances in social media research: past, present and future // Information Systems Frontiers. — 2018. — Vol. 20, No. 3. — P. 531–558. — DOI: <https://doi.org/10.1007/s10796-017-9810-y> (дата звернення: 13.04.2025)..
- [2] Petukhova A., Matos-Carvalho J. P., Fachada N. Text clustering with large language model embeddings // International Journal of Cognitive Computing in Engineering. — 2024. — DOI: <https://doi.org/10.1016/j.ijcce.2024.11.004> (дата звернення: 13.04.2025).
- [3] Snowflake Inc. Vector Embeddings // Snowflake Documentation. — Режим доступу: <https://docs.snowflake.com/en/guides/ai-ml/llm/vector-embeddings>. — Дата звернення: 19.05.2025.
- [4] Cao H. Recent advances in universal text embeddings: A comprehensive review of top-performing methods on the MTEB Benchmark [Електронний ресурс] // arXiv preprint, 2024. — arXiv:2406.01067. — Режим доступу: <https://arxiv.org/abs/2406.01067> (дата звернення: 16.04.2025).
- [5] Wang L., Yang N., Huang X., Yang L., Majumder R., Wei F. Multilingual-E5 Text Embeddings: A Technical Report [Електронний ресурс] // Hugging Face. — 2024. — Режим доступу: <https://huggingface.co/intfloat/multilingual-e5-large-instruct> (дата звернення: 17.04.2025).
- [6] Devins J. Multilingual vector search with the E5 embedding model [Електронний ресурс] // Elastic Search Labs Blog. — 12.09.2023. — Режим доступу: <https://elastic.co/search-labs/blog/multilingual-vector-search-e5-embedding-model> (дата звернення: 11.05.2025).
- [7] Nazeri S. Comparing the state-of-the-art clustering algorithms [Електронний ресурс] // Medium. — 19.07.2023. — Режим доступу: <https://medium.com/@sina.nazeri/comparing-the-state-of-the-art-clustering-algorithms-1e65a08157a1> (дата звернення: 12.04.2025).
- [8] Majhi S. K., Biswal S. Optimal cluster analysis using hybrid K-Means and Ant Lion Optimizer // Karbala International Journal of Modern Science. — 2018. — Vol. 4, No. 3. — P. 347–360. — DOI: <https://doi.org/10.1016/j.kijoms.2018.09.001>

- [9] Satpati S. Clustering by DBSCAN (Density-Based Spatial Clustering of Applications with Noise) Clearly Explained with Coding in Python [Електронний ресурс] // Medium. — 07.12.2023. — Режим доступу: <https://medium.com/@satpatishrimanta/clustering-by-dbscan-density-based-spatial-clustering-of-applications-with-noise-clearly-f93c5cf27f06> (дата звернення: 13.05.2025).
- [10] Bansal A. Optimizing customer segmentation for enhanced recommendation systems through comparative analysis of K-Means, Hierarchical Clustering, and DBSCAN algorithms [Електронний ресурс] // ResearchGate. — Травень 2023. — Режим доступу: <https://www.researchgate.net/publication/384604526> (дата звернення: 12.04.2025).
- [11] Pishro A. A., Zhang S., L'Hostis A., Liu Y., Hu Q., Hejazi F., Shahpasand M., Rahman A., Oueslati A., Zhang Z. Machine learning-aided hybrid technique for dynamics of rail transit stations classification: a case study [Електронний ресурс] // *Scientific Reports*. — 2024. — Vol. 14. — Article number: 23929. — Режим доступу: <https://doi.org/10.1038/s41598-024-23929-2> (дата звернення: 11.05.2025).
- [12] Mullin T. DBSCAN Parameter Estimation Using Python [Електронний ресурс] // Medium. — 10.07.2020. — Режим доступу: <https://medium.com/@tarammullin/dbscan-parameter-estimation-ff8330e3a3bd> (дата звернення: 19.05.2025).
- [13] Sefidian A. M. How to determine epsilon and MinPts parameters of DBSCAN clustering [Електронний ресурс] // *sefidian.com*. — 18.12.2022. — Режим доступу: <https://sefidian.com/2022/12/18/how-to-determine-epsilon-and-minpts-parameters-of-dbscan-clustering> (дата звернення: 22.04.2025).
- [14] Perafán-López J. C., Ferrer-Gregory V. L., Nieto-Londoño C., Sierra-Pérez J. Performance analysis and architecture of a clustering hybrid algorithm called FA+GA-DBSCAN using artificial datasets [Електронний ресурс] // *Entropy*. — 2022. — Vol. 24, No. 6. — Article number: 875. — Режим доступу: <https://doi.org/10.3390/e24070875> (дата звернення: 23.04.2025).
- [15] Jamin R. J., Talukder M. A. R., Malakar P., Kabir M. M., Nur K., Mridha M. F. Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review [Електронний ресурс] // *Natural Language Processing Journal*. — 2024. — Vol. 6. — Article number: 100059. — Режим доступу: <https://doi.org/10.1016/j.nlpj.2024.100059> (дата звернення: 19.05.2025).
- [16] Stroud R. S., Al-Saffar A., Carter M., Moody M. P., Pedrazzini S., Wenman M. R. Testing outlier detection algorithms for identifying early stage solute clusters in atom probe tomography [Електронний ресурс] // *Microscopy and Microanalysis*. — 2024. — Vol. 30. — P. 853–865. — DOI: <https://doi.org/10.1093/mam/ozae076> (дата звернення: 19.05.2025).
- [17] Raman R., Nair V. K., Nedungadi P., Sahu A. K., Kowalski R., Ramanathan S., Achuthan K. Fake news research trends, linkages to generative artificial intelligence and sustainable development goals [Електронний ресурс] // *Heliyon*. — 2024. — Vol. 10. — Article number: e24727. — Режим доступу: <https://doi.org/10.1016/j.heliyon.2024.e24727> (дата звернення: 25.04.2025).
- [18] Han M., Zhou Y. Exploring trends and emerging topics in oceanography (1992–2021) using deep learning-based topic modeling and cluster analysis [Електронний ресурс] // *nprj Ocean Sustainability*. — 2024. — Article number: 97. — Режим доступу: <https://doi.org/10.1038/s44183-024-00097-z> (дата звернення: 28.05.2025).

Надійшла до редколегії 10.03.2025

УДК 004.8

DOI 10.30837/bi.2025.1(102).05

В. В. Кириї¹, О. В. Калиниченко²¹Харківський національний університет радіоелектроніки, м. Харків, Україна, valentyna.kyryi@nure.ua²Харківський національний університет радіоелектроніки, м. Харків, Україна, olga.kalynychenko@nure.ua

ПРОГНОЗУВАННЯ ФІНАНСОВИХ ПОКАЗНИКІВ РИНКУ КІНОВИРОБНИЦТВА ЗАСОБАМИ ШТУЧНОГО ІНТЕЛЕКТУ

Предметом дослідження є проблема прогнозування фінансових показників на послугоорієнтованих ринках за надзвичайних умов.

Метою роботи є розробка ефективного підходу для прогнозування фінансових показників ринку кіновиробництва, що базується на рекурентних та згорткових нейронних мережах і використовує засоби обробки природної мови для конвертації описових даних в числовий вигляд.

У статті вирішуються такі завдання: визначення набору індикаторів, які здатні описати стану ринку кіновиробництва з точки зору компанії, зовнішнього оточення та цільової аудиторії; формування алгоритмів для передобробки числової та текстової інформації; визначення переліку цільових нейромереж та розкриття особливостей їх імплементації; визначення найбільш ефективного прогностичного підходу за допомогою вирішення задачі лінійної оптимізації.

Використовуються такі методи: аналітичний метод для визначення набору нейромереж; експертне оцінювання для формування найбільш важливих незалежних індикаторів та визначення факторів ефективності; експериментальний, багатокритеріального оцінювання для визначення найбільш ефективної моделі.

Отримано наступні результати: сформовано набір алгоритмів передобробки даних для подальшого їх використання в рекурентних та згорткових нейронних мереж. Імплементовано декілька поширених архітектур із залученням технології MapReduce. Визначено, що найефективнішою моделлю є двонаправлена рекурентна нейромережі з підтримкою довго- та короткострокової пам'яті. Показана доцільність використання технології паралелізації та визначено набір відкритих питань для подальшого дослідження.

Висновки: проведення аналізу щодо алгоритмів прогнозування фінансових показників заснованих на штучному інтелекті з подальшою експериментальною перевіркою дозволило сформуванню відносно ефективний спосіб передбачення стану індикаторів ринку кіноіндустрії за надзвичайних умов. Отримані результати дозволяють стверджувати доцільність впровадження запропонованого підходу, що може вплинути на формування політики компанії кіновиробника чи фонду, який оперує фінансовими інструментами компанії. Водночас передбачено шляхи подальшого покращення результатів із залученням альтернативних підходів як до паралелізації, так і до прогнозування загалом.

АНАЛІЗ НАСТРОЇВ, НЕЙРОННІ МЕРЕЖІ, ОБРОБКА ПРИРОДНИХ МОВ, СОЦІАЛЬНІ ЗСУВИ, ЧАСТОТНІ ПАТЕРНИ.

V. V. Kyrii, O. V. Kalynychenko. Forecasting financial indicators of the film production market using artificial intelligence. The subject of the research is the problem of forecasting financial indicators for service-oriented markets under extraordinary conditions.

The goal of the work is to develop an effective approach for forecasting the financial indicators of the film production market, which is based on recurrent and convolutional neural networks and uses natural language processing tools to convert descriptive data into numerical form.

The article addresses the following tasks: defining a set of indicators capable of describing the state of the film production market from the point of view of the company, the external environment, and the target audience; development of algorithms for processing numerical and textual information; determining the list of target neural networks and revealing the specifics of their implementation; determination of the most effective prognostic approach by solving the problem of linear optimization.

The following methods used are — an analytical method for determining a set of neural networks; expert assessment for the formation of the most important independent indicators and determination of efficiency factors; experimental, multi-criteria evaluation to determine the most effective model.

The following results were obtained: a set of data reprocessing algorithms was formed for their longer use in recurrent and convolutional neural networks. Several common architectures involving MapReduce technology have been implemented. It was determined that the most effective model is a bidirectional recursive neural network with support for long- and short-term memory. The expediency of using parallelization technology is shown and a set of open questions for further research is defined.

Conclusions: conducting an analysis of algorithms for forecasting financial indicators based on artificial intelligence with subsequent experimental verification made it possible to form a relatively effective way of predicting the state of indicators of the film industry market under extraordinary conditions. The obtained results allow us to affirm the feasibility of implementing the proposed approach, which can influence the policy formation of the film producer's company or the fund that operates the company's financial instruments. At the same time, there are ways to further improve the results with the involvement of alternative approaches both to parallelization and to forecasting in general.

SENTIMENT ANALYSIS, NEURAL NETWORKS, NATURAL LANGUAGE PROCESSING, SOCIAL SHIFTS, FREQUENCY PATTERNS

Вступ

Протягом останніх років світова економіка поступово цифровізується. У дослідженні [1] автори всебічно аналізують це питання, зосереджуючись на тому, що, крім появи нових благ і послуг, розвиток технологій істотно вплинув на вже розвинені ринки. Зокрема, вони зазначають, що "комп'ютерна революція" сприяла появі нових бізнес-моделей, полегшила доступ до глобальних ринків і підвищила ефективність виробничих процесів. Одним з таких є субринок розважальних послуг – кіноіндустрія. З початку минулого століття цей вид дозвілля швидко здобув популярність і поступово сформувався як один із найбільш прибуткових [2]. Завдяки інноваціям, таким як звукове кіно, кольорове зображення і спеціальні ефекти, кіноіндустрія постійно оновлювалася, залучаючи все більше глядачів. З появою стабільного і швидкого інтернет-з'єднання по всьому світу почали з'являтися нелегальні ресурси для перегляду фільмів, що створило нові виклики для цього ринку. Втрата доходів через піратство, розвиток соціальних мереж та загальне здешевлення технологій і кіновиробництва стимулювали появу легальних стрімінгових платформ, таких як Netflix, Amazon Prime, Disney+ чи HBO Max [3].

1. Аналіз останніх досліджень і публікацій

Основою подібних сервісів є підписна модель ціноутворення за якої люди платять за певний час користування. Ця модель має багато переваг як для компаній, так і для споживачів. Дослідники з університету міста Грац зазначають, що поширення цього підходу пояснюється можливістю компаній краще планувати свої трансферти, а споживачам – отримувати доступ до різноманітних контентів без значних одноразових витрат. Водночас, ця модель стимулює компанії постійно оновлювати і покращувати свої сервіси, щоб утримати підписників. Не зважаючи на подібні позитивні риси, вагомим недоліком для бізнесу є чутливість рівня доходів від поведінки споживачів. У випадку якихось надзвичайних ситуацій, як-от соціальних зсувів чи катастроф, відтік клієнтів може стимулювати процес банкрутства. У той же час при суттєвому збільшенні користувачів цифрові платформи можуть не витримати навантаження.

В якості підтвердження можна навести роботу групи азієцьких вчених, яка дослідила швидке зростання популярності стрімінгових сервісів у 2020 році під час активної фази пандемії COVID-19. Науковці довели, що подібна ситуація ускладнює аналіз інформації для поліпшення роботи зі споживачами і потребує обробки значних обсягів даних [4], оскільки соціальний зсув може впливати на поведінку цільової аудиторії і після її завершення. У той же час, пандемія спричинила зупинку виробництва нових фільмів,

що створило дефіцит контенту для стрімінгових платформ, змушуючи їх переглядати свої стратегії та інвестувати в оригінальний контент.

Крім того, сплески користувацької активності можуть виникати без соціальних катастроф, наприклад, у відповідь на певні новини або появу популярного фільму на платформі. Непрогнозовані ситуації можуть викликати перенавантаження, негативно впливаючи на користувацький досвід і зменшуючи ймовірність продовження підписки. Тому компанії змушені інвестувати в інфраструктуру та технології для забезпечення надійної роботи своїх сервісів. Прогнозування обсягів цільового ринку є складним завданням і потребує застосування сучасних підходів до інтелектуального аналізу даних

Загалом при прогнозуванні показників прийнято виділити наступні три великі групи алгоритмів [5, 6]:

- машинне навчання;
- ймовірнісний підхід;
- авторегресійний підхід.

Ймовірнісний підхід дозволяє отримати високу точність за обмежених обсягів даних, але вимагає значної кількості інформації про зовнішні фактори. Представники Корнелівського університету зазначають, що за обмеженої зовнішньої інформації точність цього підходу суттєво знижується [7]. Крім того, ймовірнісний підхід може бути менш надійним у разі раптових змін або кризових ситуацій, коли минулі дані можуть не відображати поточних тенденцій.

Авторегресійні моделі за своєю сутністю не мають вищезазначеної проблеми з обмеженістю вхідних даних. Це було доведено в ході досліджень проведених на базі університетів Шанхаю та Пекіна [8], а також ученими з Японії та Бангладешу [9]. Вони показали, що авторегресійні моделі можуть ефективно враховувати часові залежності і дозволяють отримувати точні прогнози навіть за умов нестабільності. Однак дослідження присвячені цьому класу моделей вже існують у вітчизняних працях. Зокрема, українські дослідники показали, що авторегресія рухомого середнього забезпечує точність понад 95% за нестабільних економічних умов [10]. Ці моделі також дозволяють швидко адаптуватися до нових даних і враховувати зміни в поведінці споживачів.

Розглядаючи інші можливі варіанти вирішення споріднені з авторегресією, слід зазначити, що періоди соціальної нестабільності ускладнюють використання альтернативних методів. Група китайських учених описала подібну проблему для багатовимірної лінійної регресії [11]. Вони відзначили, що такі методи можуть бути ефективними в стабільних умовах, але їхня точність суттєво знижується за наявності непередбачуваних змін. З огляду на це, авторегресійні моделі виглядають найбільш перспективними для прогнозування в умовах соціальної нестабільності.

Стосовно машинного навчання, то подібний підхід може потребувати значних обсягів даних для досягнення точності прогнозування понад 95%, і є чутливим до викидів. Це показали економісти з Пакистану та Південної Кореї [12]. Однак подібні проблеми можна вирішити шляхом додаткової передобробки даних та їх часткової аугментації.

2. Визначення не вирішених раніше частин загальної проблеми. Мета роботи, завдання

Враховуючи вказане вище, можна зазначити, що питання застосування засобів машинного навчання, залишається відкритим для задачі прогнозування стану специфічних ринків послуг. Прикладом такого ринку можна вважати ринок стримінгових сервісів. Особливо потребує опрацювання проблема врахування поведінки цільової аудиторії за надзвичайних ситуацій, які здатні викликати ажіотажний рух попиту. Базуючись на описаному науковому міжнародному досвіді, можна визначити, що найбільш ефективними засобами штучного інтелекту для вирішення поставлених питань є нейронні мережі. Особлива увагу варто приділити згортковим та рекурентним архітектурам, пристосованим для обробки великих обсягів числової та текстової інформації за рахунок можливості врахування контексту. У чинній роботі вирішено зробити акцент на чотирьох різних моделях:

- класична згорткова нейромережа (CNN);
- класична рекурентна нейромережа (RNN);
- рекурентна мережа з підтримкою довго- та короткострокової пам'яті (LSTM);
- двонаправлена рекурентна мережа з підтримкою довго- та короткострокової пам'яті (BiLSTM);

Однак треба зауважити, що вказані моделі потребують значних обсягів вихідної інформації а також часу для навчання моделі. Задля вирішення зазначених проблем можна використовувати принципи аугментації даних та паралелізму відповідно. Для останнього було вирішено використати технологію MapReduce, яка дозволяє розподілити навантаження між вузлами. Ефективність цього підходу була доведена групою українських вчених [13].

Метою роботи є дослідження ефективності використання засобів штучного інтелекту, зокрема рекурентних та згорткових нейронних мереж, для прогнозування фінансових показників ринку кіновиробництва. Задля досягнення цієї мети, визначено наступний перелік *завдань*:

- визначити набір базових характеристик ринку кіновиробництва та зовнішнього оточення і сформулювати алгоритми передобробки цих даних;
- розглянути існуючі архітектури рекурентних та згорткових нейромереж;
- визначити задачу багатокритеріального вибору задля порівняння запропонованих підходів;
- сформулювати план експерименту та провести

його для виокремлення найбільш ефективної моделі для досягнення поставленої мети.

3. Матеріали та методи

3.1. Обмеження та припущення

Перед наведенням опису нейронних мереж та алгоритмів передобробки, необхідно зазначити ряд обмежень, які накладаються на цільові показники та компанії в рамках чинного дослідження:

- тип компанії: ті, що надають послуги стримінгу кіно, наприклад: Netflix, Hulu, Amazon Prime;
- тип ціноутворення: сервіси, які притримуються моделі підписки з необмеженим використанням за фіксованою ставкою;
- поширеність: враховуючи описаний глобальний контекст, було вирішено розглядати лише ті платформи, які функціонують на декількох національних ринках; тому вітчизняні сервіси такі як Sweet.tv не враховуються;
- спеціалізація: лише платформи розраховані на аудиторії різного віку та гендеру, без неясності певної специфіки в контенті.

3.2. Вибір цільових показників

Задля вибору набору показників зовнішнього становища та алгоритмів передобробки було вирішено скористатися методом експертного оцінювання. Загалом опитування було проведено серед 200 IT-фахівців, менеджерів з поширення кінопродукції та менеджерів із впровадження інновацій з Харкова, Києва, Львова, Відня, Лісабона, Праги та Кракова. Анкета складалася з 3 блоків: цільові фактори фінансового стану компанії, фактори зовнішнього впливу та фактори цільової аудиторії. Для респондентів пропонувалося обрати з наданого переліку від 3 до 5 найбільш важливих індикаторів кожної групи. Якщо індикатор не набрав 10 балів (тобто хоча б 10 людей його не обрали), він відсіювався. Самі ж переліки формувалися на основі аналізу останніх міжнародних наукових публікацій [14, 15], що стосувалися прогнозування економічних показників діяльності бізнесу (без обмеження на кіноіндустрію чи ринок розважальних послуг).

Для цільових факторів фінансового стану компанії було обрано:

- індикатор фінансової стабільності компанії;
- ступінь впровадження інновацій;
- індекс інфляції на вартість підписки в період за 5 років;
- індекс збільшення наданого контенту;
- наявність лістингу на американській чи європейській біржі.

Зазначені показники є суто числовими, однак задля пом'якшення проблеми чутливості даних, було вирішено додати в загальний пайплайн з обробки операцію обрізки даних – занадто великі чи малі

показники відносно середнього будуть замінені на певні константи. Після додаткової консультації з експертами поза проведенням опитуванням, було вирішено здійснювати заміну всіх значень, які більші за 97 перцентиль чи менші за 3 перцентиль. Окрім цього вирішено застосувати операцію скейлінгу відносно стандартного відхилення та арифметичного середнього.

Серед індикаторів зовнішнього впливу було обрано наступні 5 показників:

- ступінь гостроти соціального зсуву визначений за алгоритмом, який було запропоновано групою українських вчених [10];

- ступінь монополізації ринку визначений за індексом Герфіндаля-Гіршмана;

- агрегований зважений показник стабільності національних економік в яких працює платформа, в якості вагів використовується обсяг доходу компанії в країні відносно загального значення;

- показник фінансової стабільності світової економічної системи.

Три останні показники розраховуються з використанням класичного математичного апарату. Перший же, використовує засоби обробки природних мов, аби сформувати профіль цільової надзвичайної ситуації і ступінь готовності компанії до можливого ажіотажного росту цін на комплементарні товари чи збільшення навантаження на цифрові ресурси. Треба зауважити, що навіть умовно незначні події для світу чи національних економік можуть ставати соціальним зсувом в рамках кіноіндустрії. Як приклад, слугує ситуація зі страйком спілки акторів США, що суттєво вплинуло на виробництво нових фільмів і серіалів, а відповідно, і на обсяги нового контенту на платформах.

Заключною групою індикаторів є ті, що стосуються цільової аудиторії:

- середній індекс платоспроможності населення по відношенню до ціни підписки на стрімінгову платформу;

- показник настрою населення сформований на основі алгоритму пошуку частотних патернів серед 200 найбільш популярних новин у соціальних мережах, які стосуються відповідного сервісу;

- показник дезінформації, щодо роботи кіноплатформи.

Задля визначення показника настрою населення було вирішено використати алгоритм Априорі для пошуку частотних патернів. Вибір цього способу ґрунтувався на простоті алгоритму, легкості його імплементації за допомогою мови програмування Python та можливостям паралелізації. Він має всього чотири кроки:

- перший: визначити рівень підтримки для кожного елемента в тексті. Підтримкою називають

частоту появи елемента в певному наборі даних;

- другий: використавши обмеження на рівень підтримки відфільтрувати набір слів;

- третій: побудувати асоціативні правила на основі найбільш частих закономірностей;

- четвертий: відсортувати отримані значення у порядку спадання ліфту.

Нижче наведено фрагмент псевдокоду для реалізації алгоритму (рис. 1).

```

while Lk-1 is not empty
  Ck ← Apriori_gen(Lk-1, k)
  for transactions t in T
    Dt ← {c in Ck : c ⊆ t}
    for candidates c in Dt
      count[c] ← count[c] + 1
  Lk ← {c in Ck : count[c] ≥ ε}
    
```

Рис. 1. Псевдокод для реалізації алгоритму Априорі

Стосовно показника дезінформації, як і у випадку з настроєм, було вирішено здійснювати аналіз тексту 200 найбільш поширених новин. Визначення "фейковості" по відношенню до роботи певного сервісу здійснювалась на основі наступних особливостей:

- використання риторичних запитань у суспільно значущих темах: Лінгвістичні дослідження показують, що у офіційно-діловому та публіцистичному стилях, які використовуються засобами масової інформації, риторичні запитання застосовуються рідко;

- відсутність заперечних конструкцій для зниження когнітивного навантаження: В поєднанні з песимістично забарвленими словами, наприклад, заміна слова "негаразд" на "катастрофа", відсутність заперечних конструкцій знижує когнітивне навантаження на читача. Слід зазначити, що у подальших текстах ненормативна лексика буде виключена, оскільки вона ускладнює процес аналізу емоційного забарвлення;

- використання закликів і заохочень у неналежному контексті та надмірна кількість займенників: Імітація публіцистичного стилю досягається через заклики і заохочення, вжиті у неналежному контексті, а також через надмірну кількість займенників;

- висока частота використання коротких речень та граматичних помилок: Часте використання коротких речень і слів з граматичними помилками.

Хоча подібні характеристики не є вичерпними, як показують міжнародні дослідження [13], вони можуть слугувати апроксимуючою функцією до функції дезінформації.

Розглянувши особливості визначення ключових індикаторів кожної групи перейдемо до розгляду нейронних мереж, які були використані в ході дослідження.

3.3. Огляд архітектур нейронних мереж

Почнемо з рекурентних нейромереж (рис. 2). Їхня сутність полягає у використанні декількох послідовних прихованих шарів. Це дозволяє враховувати

результат попереднього шару в наступного. Подібну особливість прийнято називати короткостроковою пам'яттю.

В середині кожного прихованого шару відбувається оптимізація градієнту. Через це у випадку не передоброблених даних може виникнути проблема вибохового (того, що прямує до нескінченності) та зникаючого (того, що прямує до 0) градієнту. Задля уникнення подібної проблеми було вирішено використати вже вказані функції скейлінгу та обрізки.

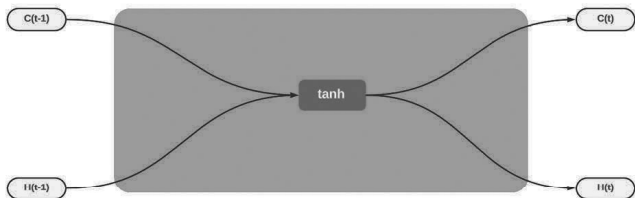


Рис. 2. Схематичне зображення прихованого шару архітектури RNN

Однак, з точки зору архітектурної складової можна використовувати більш просунуту модель з підтримкою довго- та короткострокової пам'яті (рис. 3) — її було вирішено обрати в якості другого алгоритму для дослідження. Сутність його математичного апарату полягає у поступовому використанні декількох сигмоїд та гіперболічних тангенсів, які дозволяють корегувати значення аргументу градієнтної функції. Послідовність складається з трьох ключових етапів — Forget Gate, Input Gate і Output Gate. На першому етапі відбувається множення сігма-функцію активації, що дозволяє отримати результат в межах від 0 до 1 і помножити його на значення з довгострокової пам'яті. Другий етап подібний до першого, однак включає в себе корекцію із застосуванням гіперболічного тангенсу. Подібна операція дозволяє нівелювати проблему спрямування до нескінченності при здійсненні додавання до значення з каналу довгострокової пам'яті. У результаті роботи двох вказаних етапів формується стан пам'яті, що разом із вхідними та попередніми вихідними даними, слугує базисом для визначення нового показника короткотривалої пам'яті на третьому етапі.

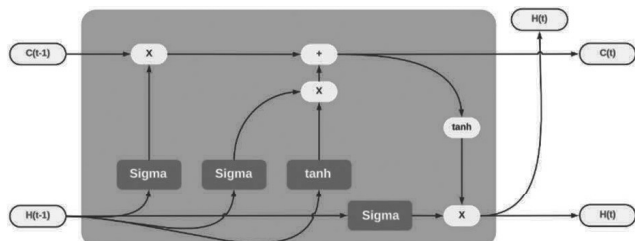


Рис. 3. Схематичне зображення прихованого шару архітектури LSTM

Хоча вказана архітектура дозволяє уникнути проблеми градієнтів, вона все ж має один істотний недолік — неможливість врахувати майбутній контекст. Зазначені вище цільові індикатори для кожної групи

є передобробленими числовими значеннями, проте для кращого розуміння вказаного недоліку розглянемо речення "Apple is something that...". Для обраної архітектури нейромережі невідомо, що саме мається на увазі під "Apple" — фрукт чи компанія, бо вона не має інформація про закінчення цього речення. Для неї варіанти "Apple is something that competitors simply cannot reproduce" та "Apple is something that I like to eat" є ідемпотентними.

Вирішення описаної проблеми з точки зору архітектури відбувається додавання другої мережі, яка спрямована в інший бік. Подібна модель називається Двонаправлена рекурентна нейромережа з підтримкою довго- та короткострокової пам'яті. Після відпрацювання двох підмереж результат обох рівнів поєднується, спочатку шляхом простої конкатенації, а після цього за допомогою лінійних трансформацій, визначених за допомогою крос-валідації. У нашому випадку — використання усереднених значень.

Перейдемо до заключної архітектури — згорткової нейромережі (рис. 4). У порівнянні з попереднім випадком, ця модель не запам'ятовує результати попередніх шарів. Натомість вона використовує шар згортки, що дозволяє суттєво зменшити розмірність вихідних даних. Редукція відбувається за допомогою проходження фільтра між даних і заміною цих даних на скалярний добуток значень між фільтром і даними.

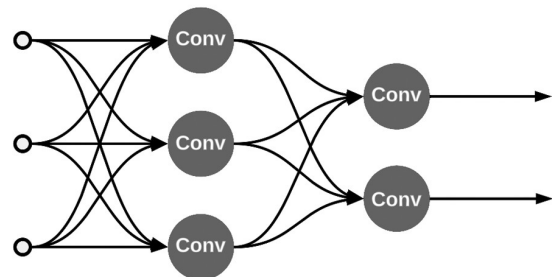


Рис. 4. Схематичне зображення архітектури CNN

Для задавання цієї моделі потрібні значення її гіперпараметрів встановлених в ході крос валідації, зокрема:

- було обрано фільтр розмірності $5 \times 5 \times 12$, де 12 — це кількість дескрипторів, тобто цільових змінних, які описані вище;
- глибина, як і випадку останнього показника розмірності фільтра має бути рівна кількості факторів — 12;
- розмір ядра (при крос-валідації було використано ядро в межах від 2 до 5 і встановлено оптимальне значення рівне 4);
- розмір кроку при розгляді. Виходячи із рекомендацій, які вказують на небажаність використання кроку більше 3 для тексту, було визначено оптимальне значення рівне 1;
- виходячи із встановленого кроку, параметр додавання неістотних нулів не застосовуватиметься;

– виходячи із специфіки предметної області було вирішено не застосовувати параметр зміщення.

Варто зауважити, що кількість шарів згорткової мережі для аналізу текстової інформації має бути рівна 1.

Проблемою зазначеної архітектури при її використанні для обробки природних мов є обмеженість у врахуванні контексту. Звичайно, проходження фільтра дозволяє взяти до уваги окіл проте часові межі певного соціального зсуву можуть бути істотними. Таким чином існує ймовірність, що визначний контекст знаходитиметься поза фільтром CNN моделі.

3.4. Розпаралелювання алгоритмів

Як зазначалося вище, для реалізації паралелізації було вирішено застосувати технологію MapReduce, яка базується на розподілі вихідного набору даних на окремі вузли. Виходячи з цієї концепції, ключовими компонентами моделі є функції мапінгу та редукції. Найбільш поширеними варіантами реалізації є платформи Spark та Hadoop. У цій роботі було обрано Hadoop, оскільки він включає додаткові функції всередині кожного вузла, що прискорює роботу з базами даних. Це є значною перевагою обраного підходу з огляду на велику кількість різноманітної інформації, що потребує обробки. Графічне зображення запропонованого рішення представлено нижче (рис. 5).

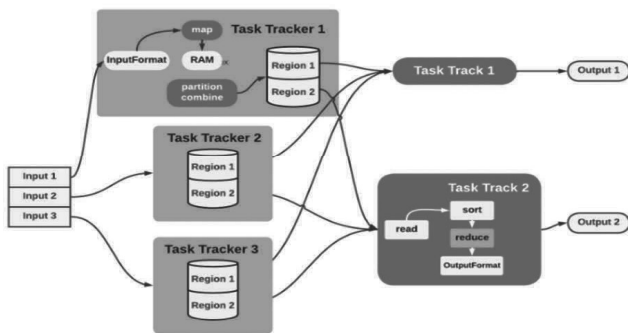


Рис. 5. Схематичне зображення MapReduce

У випадку передобробки даних усі показники трансформуватимуться паралельно один до одного на окремих вузлах. Особливості реалізації роботи з регіонами пам'яті пришвидшать розрахунок показника дезінформації та настрою населення. Зокрема для алгоритму Априорі ми розподіляємо зчитану з бази даних інформацію по різних блоках. Кожен із цих блоків виконує послідовну форму алгоритму. Далі об'єднувач агрегує значення в кожному блоці та передає їх редуктору, який, у свою чергу, фільтрує дані. Варто зазначити, що процес упорядкування, необхідний для коректної роботи алгоритму, виконується автоматично.

У випадку нейронних мереж використання паралелізації повноцінно доступне лише для BiLSTM, аби надати можливість працювати двом різнонаправленим мережам одночасно

Загалом запропонований підхід має наступний набір переваг:

- масштабованість, яка поліпшується завдяки розподілу навантаження по вузлах;
- відмовостійкість: MapReduce автоматично керує відмовами вузлів, що забезпечує надійність в обробці даних, оскільки при виникненні проблем навантаження буде перенесено на допоміжний вузол;
- простота програмування: у випадку Hadoop реалізації достатньо створити функції редукції та мепінгу а також додати конфігуруючий файл;
- суттєве покращення швидкодії: як вже було зазначено, технологія MapReduce дозволяє досягти пришвидшення в декілька разів.

Розглянувши ключові показники та алгоритми обробки й прогнозування перейдемо до формування плану експерименту.

3.5. План експерименту

Задля релаізації описаних алгоритмів було вирішено використати мову програмування Python3 і відповідні пакети:

- tensorflow – задля реалізації пайплайнів нейронних мереж;
- polars – задля структуризації даних та пришвидшення векторних операцій (обрізання, скейлінг);
- scikit-learn – задля реалізації кастомних трансформерів;
- nltk – задля конвертації текстових даних, зокрема знаходження частоти слів та інших операцій;
- numpy – задля допоміжної конвертації даних, оскільки на момент написання чинної статті polars підтримується не всіма зазначеними вище бібліотеками;
- datetime – для фіксації часу роботи.

Кількість вузлів MapReduce була встановлена на рівні 10, при цьому кожен мав наступну конфігурацію:

- частота процесора – 2.8 ГГц;
- кількість ядер – 1;
- оперативна пам'ять – 8 Гб;
- оточення – Linux Ubuntu 22.04.

Задля визначення найбільш ефективної моделі було вирішено обрати принцип лінійної адитивної згортки з ваговими коефіцієнтами. Хоча подібний підхід має певні недоліки пов'язані з можливістю покращення "малоефективних" альтернатив шляхом корекції лише одного із факторів, він є апікабельним за умови, що ці альтернативи є близькими за всіма критеріями.

Наступний кроком у формування плану експерименту є визначення критеріїв ефективності, як і у випадку з цільовими індикаторами було проведено експертне оцінювання (серед тої самої групи експертів). В наслідок чого обрані наступні фактори (сума показників важливості рівна 100):

– точність прогнозу – оскільки прогнозування фінансових показників прямо впливає на корпоративні рішення щодо політики компанії (або у випадку трейдингу на рішення сторонніх осіб щодо купівлі/продажу цінних паперів) цей показник було обрано найголовнішим у питанні ефективності запропонованого підходу із присвоєнням показника важливості рівним 70. Під точністю у цьому випадку мається на увазі обернений нормалізований показник кореня середньоквадратичної похибки (RMSE);

– економія часу роботи моделі – беручи до уваги специфіку показників, що розглядається та надзвичайність умов, наступний за важливістю є час роботи моделі. Варто зауважити, що поняття "час" не враховує тривалість тренування моделей, оскільки це прямо не впливає на швидкість прийняття рішень. Однак, передобробка даних перед передачею їх на вхід – враховується. Показник важливості цього фактора рівний 20;

– економія обсягу даних – оскільки кожна компанія на ринку кіновиробництва може мати свої особливі риси обсяг даних необхідний для досягнення точності у 80%. Показник важливості цього фактора рівний 10;

Враховуючи зазначене отримуємо наступні вагові коефіцієнти:

- точність прогнозу – 0.7;
- економія часу роботи моделі – 0.2;
- економія обсягу даних – 0.1.

Для подальшого дослідження було сформовано власний набір даних щодо Netflix та Amazon Prime. Усі описані вище характеристики були взяті або з соціальних мереж та новин з BBC (у випадку текстових описів) й з відкритих джерел, зокрема Statista та Yahoo Finance. Для кожного набору даних в якості таргету було вирішено обрати ціну закриття акцій компанії.

Заключним елементом плану експерименту є визначення можливих похибок. Виходячи із описаного плану можна виокремити наступні фактори, що здатні вплинути на результат: під час перевірки економії часу: людський фактор та інструментальна похибка; під час перевірки точності: проблема даних.

Аби пом'якшити зазначені невизначеності вимірювання показників здійснюватиметься декілька разів.

4. Результати досліджень

Почнемо з показника економії часу роботи алгоритмів. Результати наведені нижче (табл. 1). Усі значення економії порашовані відносно найповільнішого алгоритму – двонаправленої нейромережі з підтримкою довго- та короткострокової пам'яті. Варто зауважити, що різниця у часі з LSTM присутня не зважаючи на застосування паралелізації в наслідок накладних витрат з редукції даних.

Таблиця 1

Економія часу роботи алгоритма (у мілісекундах)

Спроба	CNN	RNN	LSTM	BiLSTM
I	58	37	11	0
II	55	55	12	0
III	57	48	10	0
IV	61	42	13	0
V	59	38	9	0
Середнє	58	44	11	0

Як бачимо алгоритм згорткової нейромережі працює швидше, оскільки має незначну глибину і містить в собі простіші алгоритмічні операції. Подібна логіка прослідковується і при порівнянні RNN з LSTM. Стосовно ефективності MapReduce підходу варто зауважити, що після декількох вимірювань було встановлено, що пришвидшення складає близько 1.89. Перейдемо до результатів точності прогнозування. У таблиці нижче (табл. 2) наведені обернене нормалізовані значення кореня середньоквадратичного відхилення нормалізованого в межах від 0 до 1. Подібна нормалізація можна застосовувати оскільки було прийнято рішення здійснювати скейлінг усіх незалежних та залежних змінних.

Таблиця 2

Точність прогнозування

Дані	CNN	RNN	LSTM	BiLSTM
Netflix	0.74	0.81	0.83	0.89
Amazon	0.72	0.79	0.84	0.91
Середнє	0.73	0.8	0.835	0.9

Треба зауважити, що використання MapReduce не вплинуло на точність BiLSTM. З огляду на отриманий результат саме BiLSTM є найефективнішою моделлю для обраного фактора.

Заключною метрикою є економія обсягу даних, необхідних для досягнення точності щонайменше 80%. Для цього було проведено кілька ітерацій з поступовим збільшенням кількості записів від 5000 до 10000 (оскільки дані були згенеровані на щоденній основі проводилась додаткова аугментація за допомогою векторної авторегресії). Результати наведено у таблиці нижче (табл. 3).

Таблиця 3

Економія обсягу даних для досягнення точності у 80%

Дані	CNN	RNN	LSTM	BiLSTM
Netflix	0	100	500	700
Amazon	0	100	600	600
Середнє	0	100	550	650

Тепер можемо систематизувати отримані значення метрик та визначимо альтернативи оптимальні за Парето (табл. 4). Усі значення були нормовані та округлені до сотих.

Таблиця 4
Значення критеріїв при аналізі аудіо оптимальних за Парето

Модель	Економія часу	Точність	Економія обсягу даних	Згортка
CNN	1.00	0.73	0.00	0.711
RNN	0.76	0.80	0.15	0.727
LSTM	0.19	0.84	0.85	0.711
BiLSTM	0.00	0.90	1.00	0.730

Враховуючи наведене вище, можна зазначити, що найбільш ефективною моделлю є двонаправлена нейромережа з підтримкою довго- та короткострокової пам'яті. Вона ж гарантує отримання найбільшою точністю щодо прогнозування даних. Варто зауважити, що питання швидкодії цієї моделі можна вирішити за допомогою паралелізації внутрішніх розрахунків, зокрема визначення стану обох типів пам'яті. Важливим є також те, що модель LSTM виявилось гіршою за RNN. Однак це пов'язано з особливостями побудови чинної задачі з оптимізацією функції ефективності. Загалом отриманий результат відповідає міжнародним дослідженням проведеним для інших, не кіно, ринків.

Висновки

Метою чинної роботи було дослідження ефективності використання засобів штучного інтелекту, зокрема рекурентних та згорткових нейронних мереж, для прогнозування фінансових показників ринку кіновиробництва. Задля цього здійснено аналіз сучасних наукових публікацій, присвячених обраній тематиці, та проведено ряд експертних опитувань, що дозволило сформулювати набір цільових прогностичних алгоритмів і послідовність незалежних змінних. До останніх відноситься:

- індикатор фінансової стабільності компанії;
- ступінь впровадження інновацій;
- індекс інфляції на вартість підписки в період за 5 років;
- індекс збільшення наданого контенту;
- наявність лістингу на американській чи європейській біржі;
- ступінь гостроти соціального зсуву;
- ступінь монополізації ринку визначений за індексом Герфіндала-Гіршмана;
- агрегований зважений показник стабільності національних економік в яких працює платформа;
- показник фінансової стабільності світової економічної системи;
- середній індекс платоспроможності населення по відношенню до ціни підписки на стрімінгову платформу;
- показник настрою населення сформований на основі алгоритму пошуку частотних патернів;
- показник дезінформації, щодо роботи кіноплатформи.

У якості цільових моделей було розглянуто класичну згорткову та рекурентну нейромережу, а також більш просунуті версії останньої, зокрема, рекурентну нейромережу з підтримкою довго- та короткострокової пам'яті та її двонаправлену варіацію. Для передобробки описаних даних було застосовано алгоритм пошуку частотних патернів Априорі та принципи контент-аналізу.

Задля подолання проблеми пов'язної з часом роботи та, частково, навчання моделі, було вирішено використати технологію MapReduce. Аби визначити найбільш ефективну нейронну мережу було сформовано набір критеріїв, який дозволив використати принцип лінійної адитивної згортки з ваговими коефіцієнтами.

У ході експериментальної перевірки із залученням даних щодо роботи Netflix та Amazon Prime виявлено, що двонаправлена рекурентна нейромережа з довго та короткостроковою пам'яттю є найбільш ефективною, не зважаючи на повільність по відношенню до інших моделей. При цьому застосування MapReduce було виправдним, наблизивши час роботи вказаного підходу до простої версії.

Таким чином можемо стверджувати, що використання BiLSTM та описаного набору незалежних змінних у поєднанні з алгоритмами їх передобробки і технологією MapReduce дозволяє ефективно прогнозувати фінансові показники діяльності компаній на кіноринку. Це в свою чергу відкриває можливості для корекції політики компанії та передбачення стану фінансового ринку. Відкритим залишається питання подальшого прискорення запропонованого підходу і можливість застосування більш просунутих гібридних моделей.

Список літератури

- [1] Bagnoli C., Biazzo S., Biotto G. Digital business models for Industry 4.0. How innovation and technology shape the future of companies. Springer, Cham. 2022. 268 с. DOI: 10.1007/978-3-030-97284-4.
- [2] Oyewola D. O., Dada E. G. Machine Learning Methods for Predicting the Popularity of Movies. *Journal of Artificial Intelligence and Systems*. 2020. Т. 4. С. 65–82. DOI: 10.33969/AIS.2022040105.
- [3] Wang W., Guo Q. Subscription strategy choices of network video platforms in the presence of social influence. *Electronic Commerce Research*. 2021. Т. 23. С. 577–604. DOI: 10.1007/s10660-021-09504-w.
- [4] Shin Z., Moon J., Rho S. A Comparative Analysis of Ensemble Learning-Based Classification Models for Explainable Term Deposit Subscription Forecasting. *Journal of Society for e-Business Studies*. Т. 3 (26). С. 1–21. URL: <http://www.jsebs.org/jsebs/index.php/jsebs/article/view/457> (дата звернення: 08.07.2024).
- [5] E-commerce Sales Forecast Based on Ensemble Learning / C. Zhan. *IEEE International Symposium on Product Compliance Engineering-Asia*. 2020. С. 1–5. DOI: 10.1109/ISPCE-CN51288.2020.9321858.

- [6] Masini R. P., Medeiros M. C., Mendes E. F. Machine learning advances for time series forecasting. *Journal of Economic Surveys*. 2021. Т. 37 (1). С. 76–111. DOI: 10.1111/joes.12429.
- [7] Ning C., You F. Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming. *Computers & Chemical Engineering*. 2019. Т. 125. С. 434–448. DOI: 10.1016/j.compchemeng.2019.03.034.
- [8] Qi X.-Z., Ning Z., Qin M. Economic policy uncertainty, investor sentiment and financial stability—an empirical study based on the time varying parameter-vector autoregression model. *Journal of Economic Interaction and Coordination*. 2022. Т. 17. С. 779–799. DOI: 10.1007/s11403-021-00342-5.
- [9] Shibasaki M., Witayangkurn A., Rahman M. M. Comparison of life patterns from mobile data in Bangladesh. *Smart Technology & Urban Development*, м. Chiang Mai, 13 груд. — 14 груд. 2019 р. 2019. С. 1–7. DOI: 10.1109/STUD49732.2019.9018795.
- [10] Parallelization of the VAR Algorithm Family to Increase the Efficiency of Forecasting Market Indicators During Social Disaster / A. Khovrat та ін. *Information Technology and Implementation*, м. Київ, 30 лист. — 2 груд. 2022 р. 2022. С. 222–233. URL: https://ceur-ws.org/Vol-3347/Paper_19.pdf (дата звернення: 08.07.2024).
- [11] Wang W., Guo Q. Subscription strategy choices of network video platforms in the presence of social influence. *Electronic Commerce Research*. 2021. Т. 23. С. 577–604. DOI:10.54691/bcpbm.v34i.3018.
- [12] A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector / I. Ullah та ін. *IEEE Access*. 2019. Т. 7. С. 60134–60149. DOI: 10.1109/access.2019.2914999.
- [13] Yakovlev S., Khovrat A., Kobziev V. Using Parallelized Neural Networks to Detect Falsified Audio Information in Socially Oriented Systems. *Information Technology and Implementation*, м. Київ, 20 лист. — 21 лист. 2023 р. 2023. С. 220–238. URL: https://ceur-ws.org/Vol-3624/Paper_19.pdf (дата звернення: 08.07.2024).
- [14] Chen Y. C., Huang W. C. Constructing a stock-price forecast CNN model with gold and crude oil indicators. *Applied Soft Computing*. 2021. Т. 112. № 107760. DOI: 10.1016/j.asoc.2021.107760.
- [15] Sousa, A. M. J., Braga, A. C., Cunha J. Impact of macroeconomic indicators on bankruptcy prediction models: case of the Portuguese construction sector. *Quantitative Finance and Economics*. 2022. Т. 6 (3). С. 405–432. DOI: 10.3934/QFE.2022018.
- plainable Term Deposit Subscription Forecasting", *Journal of Society for e-Business Studies*, Vol. 26(3), P. 1–21, available at: <http://www.jsebs.org/jsebs/index.php/jsebs/article/view/457> (last accessed 08.07.2024).
- [5] Li, J., Zhan, C., Sha, W., Jiang, W., Guo, Y. (2020), "E-commerce Sales Forecast Based on Ensemble Learning", *IEEE International Symposium on Product Compliance Engineering-Asia*, P. 1–5. DOI: 10.1109/ISPCE-CN51288.2020.9321858.
- [6] Masini, R., Medeiros, M., Mendes, E. (2021), "Machine learning advances for time series forecasting", *Journal of Economic Surveys*, Vol. 37 (1), P. 76–111. DOI: 10.1111/joes.12429.
- [7] Ning, C., You, F. (2019), "Optimization under uncertainty in the era of big data and deep learning: When machine learning meets mathematical programming", *Computers & Chemical Engineering*, No. 125, P. 434–448. DOI: 10.1016/j.compchemeng.2019.03.034.
- [8] Qi, X.-Z., Ning, Z., Qin, M. (2022), "Economic policy uncertainty, investor sentiment and financial stability — an empirical study based on the time varying parameter-vector autoregression model", *Journal of Economic Interaction and Coordination*, Vol. 17, P. 779–799. DOI: 10.1007/s11403-021-00342-5.
- [9] Shibasaki, M., Witayangkurn, A., Rahman, M. M. (2019), "Comparison of life patterns from mobile data in Bangladesh". *Smart Technology & Urban Development (STUD-2019): 1st International Conference, Chiang Mai, 13 December — 14 December 2019: IEEE*. P. 1–7. DOI: 10.1109/STUD49732.2019.9018795.
- [10] Khovrat, A., Kobziev, V., Nazarov, A., Yakovlev, S. (2022), "Parallelization of the VAR Algorithm Family to Increase the Efficiency of Forecasting Market Indicators During Social Disaster". *Information Technology and Implementation (IT&I 2022): 9th International Conference, Kyiv, 30 November — 2 December 2022: CEUR Workshop Proceedings*. No. 3347, P. 222–233. available at: https://ceur-ws.org/Vol-3347/Paper_19.pdf (last accessed: 08.07.2024).
- [11] Wang, G., Wang, Zh., Xie, Y. (2022), "Subscribers forecasting of netflix based on multiple linear models", *BCP Business & Management*, Vol. 34, P. 229–236. DOI:10.54691/bcpbm.v34i.3018.
- [12] Ullah, I., Raza, B., Malik, A. K., Imran, M., Islam, S. U., Kim, S. W. (2019), "A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector", *IEEE Access*, Vol. 7, P. 60134–60149. DOI: 10.1109/ACCESS.2019.2914999.
- [13] Yakovlev, S., Khovrat, A., Kobziev, V. (2023), "Using Parallelized Neural Networks to Detect Falsified Audio Information in Socially Oriented Systems". *Information Technology and Implementation (IT&I 2023): 10th International Conference, Kyiv, 20 November — 21 November 2023: CEUR Workshop Proceedings*. No. 3624, P. 220–238. available at: https://ceur-ws.org/Vol-3624/Paper_19.pdf (last accessed: 08.07.2024).
- [14] Chen, Y. C., Huang W. C. (2021), "Constructing a stock-price forecast CNN model with gold and crude oil indicators", *Applied Soft Computing*, Vol. 112, Article 107760. DOI: 10.1016/j.asoc.2021.107760.
- [15] Sousa, A. M. J., Braga, A. C., Cunha J. (2022), "Impact of macroeconomic indicators on bankruptcy prediction models: case of the Portuguese construction sector", *Quantitative Finance and Economics*, Vol. 6 (3), P. 405–432. DOI: 10.3934/QFE.2022018.

References

- [1] Bagnoli, C., Biazzo, S., Biotto, G. (2022), *Digital business models for Industry 4.0. How innovation and technology shape the future of companies*, Springer, Cham. 268 p. DOI: 10.1007/978-3-030-97284-4
- [2] Oyewola, D. O., Dada, E. G. (2022), "Machine Learning Methods for Predicting the Popularity of Movies", *Journal of Artificial Intelligence and Systems*, Vol. 4, P. 65–82. DOI: 10.33969/AIS.2022040105.
- [3] Wang, W., Guo, Q. (2021), "Subscription strategy choices of network video platforms in the presence of social influence", *Electronic Commerce Research*, Vol. 23, P. 577–604. DOI: 10.1007/s10660-021-09504-w.
- [4] Shin, Z., Moon, J., Rho, S. (2021), "A Comparative Analysis of Ensemble Learning-Based Classification Models for Ex-

УДК 004.42

DOI 10.30837/bi.2025.1(102).06

О.С. Назаров¹, С.А. Биковська², Н.В. Назарова³¹ Харківський національний університет радіоелектроніки, м. Харків, Україна, oleksii.nazarov1@nure.ua, ORCID iD: 0000-0001-8682-5000² Харківський національний університет радіоелектроніки, м. Харків, Україна, sofia.bykovska@nure.ua³ Харківський національний університет радіоелектроніки, м. Харків, Україна, nataliia.nazarova@nure.ua, ORCID iD: 0009-0007-7816-7088

ДОСЛІДЖЕННЯ МЕТОДІВ АНІМАЦІЙ ВИКОРИСТОВУЮЧИ ФРЕЙМВОРК FLUTTER

Стаття присвячена дослідженню методів реалізації анімацій у мобільних застосунках на основі фреймворку Flutter з метою покращення користувацького досвіду. У межах роботи було проаналізовано сучасні підходи до створення анімованих інтерфейсів, зокрема імпліцитні (implicit) та експліцитні (explicit) анімації, а також використання анімованих віджетів і кастомних трансформацій. Особлива увага приділялася не лише технічним аспектам реалізації, а й впливу анімацій на сприйняття інтерфейсу користувачем, емоційне залучення та інклюзивність цифрового продукту. Оцінювання ефективності впроваджених рішень здійснювалося шляхом аналізу продуктивності, плавності відтворення анімацій, відповідності принципам Material Design, а також доступності для користувачів з особливими потребами. У дослідженні застосовувалися як інструменти профілювання Flutter, так і експертне оцінювання взаємодії користувача із застосунком. Результати дослідження показали, що вдумливе використання анімацій не лише покращує естетичне сприйняття інтерфейсу, але й сприяє інтуїтивному розумінню навігації, підвищенню залученості користувача та загальній ефективності мобільного рішення.

DEVTOOLS, FLUTTER, MATERIAL DESIGN, UX, АДАПТИВНІСТЬ, ЕКСПЛІЦИТНІ АНІМАЦІЇ, ІМПЛІЦИТНІ АНІМАЦІЇ, ІНКЛЮЗИВНІСТЬ, ПРОДУКТИВНІСТЬ ІНТЕРФЕЙСУ, ПРОФІЛЮВАННЯ ПРОДУКТИВНОСТІ

O. Nazarov, S. Bykovska, N. Nazarova. Research on animation methods using the Flutter framework. The article focuses on the study of methods for implementing animations in mobile applications using the Flutter framework, aiming to enhance the user experience. The paper analyzes modern approaches to creating animated interfaces, including implicit and explicit animations, as well as the use of animated widgets and custom transformations. Particular attention was paid not only to the technical aspects of implementation but also to the impact of animations on user experience, emotional engagement, and inclusiveness of the digital product. The effectiveness of the implemented solutions was evaluated by analyzing the performance, smoothness of animations, compliance with Material Design principles, and accessibility for users with special needs. The study used both Flutter profiling tools and expert evaluation of user interaction with the application. The results of the study showed that the thoughtful use of animations not only improves the aesthetic perception of the interface but also contributes to an intuitive understanding of navigation, increased user engagement, and overall effectiveness of the mobile solution.

DEVTOOLS, FLUTTER, MATERIAL DESIGN, UX, RESPONSIVENESS, EXPLICIT ANIMATIONS, IMPLICIT ANIMATIONS, INCLUSIVITY, INTERFACE PERFORMANCE, PERFORMANCE PROFILING

Вступ

У сучасну цифрову епоху користувачі очікують не лише функціонального додатку, але й інтуїтивно зрозумілого, приємного та простого у використанні інтерфейсу. Анімація відіграє важливу роль у цьому випадку: вона дозволяє краще зрозуміти логіку навігації, формує ілюзію плавності та знайомства, і навіть допомагає встановити емоційний зв'язок з продуктом. При правильному використанні анімація може зробити інтерфейс «живим» — інтерактивним, інформативним та емоційним. Водночас, надмірна або неякісна анімація може мати негативні наслідки, такі як інформаційне перевантаження, роздратування або погіршення продуктивності додатку. Flutter, відомий крос-платформний набір для розробки від Google, пропонує розробникам потужні інструменти для створення анімованих користувацьких інтерфейсів.

Він підтримує легке включення базової неявної анімації, яка автоматично запускається при зміні стану віджету, а також розширювані явні механізми, де розробник зберігає контроль над кожним кроком у процесі анімації. Це відкриває величезні можливості, але водночас і лякає: як досягти анімації, яка буде плавною, але не млявою? Очевидною, але не нав'язливою? Доступною для всіх користувачів, включаючи тих, хто має вади зору або чутливість до руху? У цьому дослідженні розглядаються різні техніки реалізації анімації в середовищі Flutter, їхній вплив на користувацький досвід, а також способи оцінки ефективності анімації з точки зору продуктивності, зручності та доступності. Ми порівнюємо приклади як простих, так і складних технік анімації, використовуємо інструменти профілювання (наприклад, Flutter DevTools) і робимо акцент на дотриманні

принципів інклюзивного дизайну, щоб інтерфейс був не тільки естетично привабливим, але й зручним для якомога ширшої аудиторії. Мета цієї книги – не просто проілюструвати технічні особливості Flutter, а досягти рівноваги між зовнішнім виглядом і функціональністю, привабливістю і продуктивністю, «вау-фактором» і реальною зручністю для користувача.

1. Опис предметної області

У фокусі цього дослідження – методи інтеграції анімації в мобільні додатки з конкретним застосуванням механізмів розробки неявної та явної анімації за допомогою фреймворку Flutter. Матеріал охоплює не тільки вже існуючі анімовані віджети (наприклад, `AnimatedContainer`, `AnimatedOpacity`, `Hero`), але й елементи, які використовуються для ручного управління циклом анімації, тобто `AnimationController`, `Tween`, `AnimationBuilder` тощо [1]. Основна увага в цьому дослідженні зосереджена на вивченні впливу різних форм анімації на досвід користувача при взаємодії з мобільним інтерфейсом, насамперед з точки зору продуктивності, простоти навігації та доступності цифрового середовища. Пріоритетним завданням є пристосування анімації до потреб користувачів з різним сприйняттям, наприклад, чутливістю до руху, при дотриманні стандартів доступності інтерфейсу відповідно до сучасних керівних принципів. Основна мета дослідження – вивчити та проаналізувати ефективність анімаційних рішень, розроблених за допомогою Flutter, з акцентом на поєднанні естетики, покращеної продуктивності та зручності інтерфейсу для широкого кола користувачів. У дослідженні використовувався Flutter DevTools для аналізу продуктивності анімації, а також якісні показники залучення користувачів на основі принципів UI/UX-дизайну [2]. Це дозволило нам поглиблено дослідити вплив анімації на загальний користувацький досвід мобільного додатку.

2. Аналіз задач та методів програмної оптимізації

Метою цього дослідження є вивчення ключових методів реалізації анімації у фреймворку Flutter та їхнього впливу на користувацький досвід. Воно спрямоване на виявлення найкращих практик у досягненні плавного, зручного та доступного інтерфейсу, не впливаючи на рівень продуктивності програми. У сучасному дизайні мобільних додатків все більше уваги приділяється не лише поверхневому графічному дизайну програми, але й емоційному досвіду, який користувачі отримують під час її використання. Анімація в цьому контексті є не просто декоративним елементом, а фактично життєво важливим компонентом, який сприяє покращенню навігації та полегшує розуміння дій користувача. Flutter пропонує два основні методи генерації анімації: неявні

анімації, які автоматично відбуваються у відповідь на зміну властивостей віджету, та явні анімації, в яких стан, час та поведінка анімації керуються розробником. Неявна анімація – це простий метод створення швидких візуальних ефектів, включаючи колір, розмір або прозорість, і використовується за допомогою таких віджетів, як `AnimatedContainer`, `AnimatedOpacity` або `AnimatedAlign`. Явна анімація забезпечує повний контроль над анімаційними операціями за допомогою класів `AnimationController`, `Tween`, `CurvedAnimation` та інших. Одним з найважливіших аспектів роботи з анімацією є її вплив на продуктивність інтерфейсу. Некоректне використання анімованих об'єктів може призвести до перевантаження потоків інтерфейсу, фризів і зниження FPS, що особливо критично для пристроїв з обмеженими ресурсами. Тому Flutter надає інструменти профілювання, включаючи `Performance Overlay`, `Timeline` і `DevTools`, які дозволяють виявляти «важкі» віджети, аналізувати використання кадрів і оптимізувати відтворення анімації. Поряд з технічною реалізацією систем важливо звертати увагу на психологічну обробку, пов'язану зі сприйняттям руху. Дослідження підтвердили, що анімація може зменшити когнітивне навантаження, полегшити навігацію між елементами інтерфейсу та підвищити емоційну залученість у додаток. З іншого боку, занадто довгі або кричущі анімації, або ті, що не мають очевидного функціонального виправдання, мають зворотний ефект, викликаючи роздратування або навіть розгубленість. Це особливо важливо з огляду на принципи інклюзивного дизайну, які передбачають створення інтерфейсів, доступних для людей з різною чутливістю до руху або когнітивними порушеннями.

У фреймворку Flutter є конфігурації для підтримки обмежень середовища користувача, такі як `MediaQuery.of(context).disableAnimations` та `AccessibilityFeatures.disableAnimations` [1]. Завдяки їх використанню додаток може адаптуватися в реальному часі до потреб користувача, відключаючи складні анімаційні ефекти, що робить його більш інклюзивним і поважає індивідуальні уподобання.

Отже, ефективне використання анімації у Flutter – це делікатний баланс трьох основних факторів: вибір правильного набору технічних прийомів (неявних чи явних), постійна перевірка ефективності за допомогою інструментів профілювання та дотримання рекомендацій щодо доступності. Врахування всіх цих факторів разом дозволяє створювати візуально інтуїтивно зрозумілі мобільні інтерфейси, які водночас є ефективними, простими і доступними для різних груп користувачів.

3. Методи реалізації анімацій у Flutter з урахуванням продуктивності та інклюзивності

Використання хорошої анімації в мобільних додатках є важливим аспектом UI/UX дизайну сьогодні. Вони можуть допомогти користувачеві краще відчувати інтерфейс, полегшити навігацію та встановити емоційний зв'язок між користувачем і продуктом. Але якщо ці анімації погано оптимізовані, вони можуть спричинити низьку продуктивність, сповільнити взаємодію та погіршити доступність для деяких груп користувачів. Таким чином, у сучасній практиці мобільної розробки мова йде не лише про створення красивих анімацій, але й про дотримання принципів інклюзивного дизайну та юзабіліті.

Flutter має величезний інструментарій для розробки анімації, який включає деякі з найбільш важливих категорій: неявна анімація, явна анімація, створення кастомної анімації та сумісність з апаратними засобами виконання. У цій доповіді ми розглянемо стандартні підходи до розробки анімації у Flutter, їхній вплив на продуктивність користувацьких інтерфейсів, а також те, як сприяти інклюзивності.

Flutter надає ряд практичних віджетів для неявної анімації, таких як `AnimatedContainer`, `AnimatedOpacity` та `AnimatedAlign`. Ці віджети дозволяють легко реалізувати прості переходи без необхідності створювати спеціалізований контролер анімації. Вони реагують на зміну стану, автоматично анімуючи відповідні властивості.

Такий підхід менш ресурсоємний і зручний для початкової розробки, але має обмеження в гнучкості. Неявна анімація добре підходить для забезпечення плавних переходів у простих елементах інтерфейсу, таких як кнопки, блоки та контейнери. Крім того, завдяки інтеграції з можливостями платформи Flutter, ці анімації забезпечують високий рівень продуктивності навіть на пристроях середнього класу.

Явні анімації з ручним керуванням використовуються для більш складних сценаріїв: `AnimationController`, `Tween`, `CurvedAnimation`, `AnimatedBuilder`. Такий підхід дозволяє задавати тривалість, криву згладжування, цикл повторення, запускати або зупинити анімацію тощо [4].

Явні анімації вимагають делікатного поводження з життєвими циклами віджетів і профілюванням, оскільки їх неправильне використання може призвести до витоку пам'яті або низької частоти кадрів в секунду. Однак вони надають більше свободи для створення взаємодій, необхідних для кастомних елементів, анімації переходів між екранами або складних інтерактивних сцен. Вбудований рушій Hero забезпечує плавні переходи між екранами для стандартних елементів інтерфейсу. Ця функція допомагає підтримувати контекст з точки зору користувача та створювати відчуття навігації. Однак важливо регулювати як

розмір, так і кількість одночасно анімованих об'єктів, щоб уникнути візуального перевантаження інтерфейсу.

Flutter дотримується системних обмежень щодо анімації, які встановлюються користувачами за допомогою налаштувань доступності. Наприклад, властивість `MediaQuery.of(context).disableAnimations` посилається на вимогу вимкнути анімацію для людей, чутливих до руху [9]. Ігнорування цього параметра може порушити комфорт користувачів з нейро-сенсорними особливостями або захворюваннями, пов'язаними з вестибулярним апаратом. Інтеграція цієї практики є ознакою відповідального та інклюзивного дизайну.

Профілювання продуктивності є необхідною умовою для створення складних анімацій. Flutter має ряд інструментів, включаючи `DevTools`, `Performance Overlay` та `Timeline`, які можна використовувати для визначення ділянок коду, які перевантажені. Типовими причинами втрати продуктивності є надто складні збірки в методах `build()`, відсутність адекватних механізмів кешування або перебудова всього дерева віджетів, а не конкретних цільових компонентів.

Оптимальними практиками є:

- використання `RepaintBoundary` для ізоляції анімованих елементів;
- анімація лише того, що дійсно змінюється;
- мінімізація кількості викликів `setState()` під час анімації.

Інклюзивний підхід до анімації охоплює розробку інтерфейсів, які враховують вимоги різних груп користувачів. Це включає в себе:

- контрастність і чіткість візуальних переходів;
- здатність адаптуватися до обмеженого руху;
- анімовані підказки для навігації по складному інтерфейсу;
- врахування принципів універсального дизайну та рекомендацій WAI-ARIA щодо рухомого контенту [9].

Отже, використання анімації в середовищі Flutter передбачає об'єднання креативного дизайну та технічної зручності використання. Балансуючи між продуктивністю, доступністю та гнучкістю, можна розробити інтерфейси, які будуть не лише естетично привабливими, але й зручними, інтуїтивно зрозумілими та стійкими на широкому спектрі пристроїв.

4. Аналіз застосування анімацій у Flutter – застосунках з урахуванням продуктивності та доступності

Використання анімації в мобільних додатках стало поширеною практикою для покращення візуальної взаємодії, мінімізації когнітивного навантаження та забезпечення інтуїтивної взаємодії з користувачем. У фреймворку Flutter анімація може бути реалізована як неявно (наприклад, `AnimatedContainer`,

AnimatedOpacity та AnimatedAlign), так і явно (наприклад, AnimationController, Tween та AnimatedBuilder). Незважаючи на те, що анімовані інтерфейси красиві, необхідно вивчити їхній вплив на продуктивність системи, швидкість взаємодії та зручність використання для людей з особливими потребами. Щоб оцінити цей вплив, ми розглянули зразок веб-додатку Flutter (на основі PWA), що містить прості неявні та адаптовані анімації. Вимірювання проводилося за допомогою Google Lighthouse – інструменту, що дозволяє оцінити як продуктивність додатку, так і дотримання правил доступності та найкращих практик [7].

На рис. 1 представлено зведену інформацію про показники, зібрані під час процедури оцінювання.

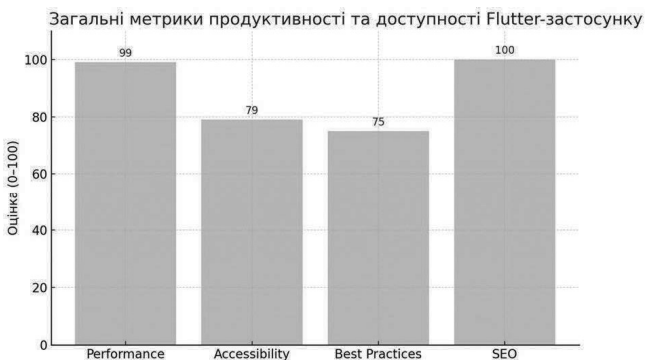


Рис. 1. Загальні метрики продуктивності та доступності

Результати тесту підтверджують, що продуктивність отримала рекордно високий бал 99, що свідчить про відсутність зависань, затримок рендерингу та коректну роботу графічного процесора навіть за умови одночасного використання анімації. Оцінка «Доступність» у 79 балів демонструє певну часткову невідповідність стандартам WCAG 2.1, що потенційно може спричинити перешкоди для людей з порушеннями зору, моторики або нейросенсорних функцій. Оцінка «Найкращі практики» в 75 балів свідчить про наявність невеликих прогалин у конфігурації JavaScript API, але оцінка SEO в 100 балів підкреслює відмінну семантичну структуру веб-сторінок. Щоб краще зрозуміти, як анімація пов'язана з продуктивністю користувацького інтерфейсу, на рис. 2 розглянуто ключові показники Web Vitals – показники, які відображають безпосередній досвід користувача [3].

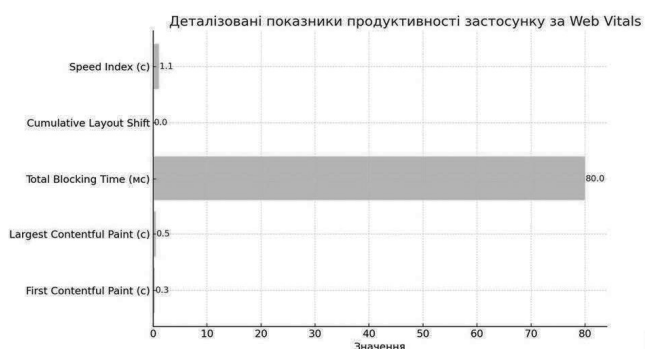


Рис. 2. Деталізовані показники продуктивності застосунку

Результати:

- FCP (First Contentful Paint): 0,3 секунди – початковий рендеринг контенту відбувається миттєво, що дозволяє уникнути відчуття «порожнього екрану»;

- LCP (Largest Contentful Paint): 0,5 секунди – основний контент завантажується швидко, що особливо важливо для веб-сторінок, які містять велику кількість медіа-об'єктів. Загальний час блокування (TBT): 80 мс – цей показник вказує на те, що взаємодія користувача з інтерфейсом не відкладається фоновими процесами або скриптами.

- CLS (Cumulative Layout Shift): 0 – елементи інтерфейсу стабільні під час завантаження, що свідчить про якісну верстку.

- Індекс швидкості: 1.1 секунди – швидкість візуального рендерингу стабільно висока.

Представлені факти (див. рис. 3) демонструють, що використання анімації, коли вона виконується відповідно до визначених найкращих практик, може призвести до дійсно високих результатів продуктивності (див. рис. 4) [8, 10]. Зокрема, використаний додаток використовував `RepaintBoundary` для ізоляції анімованих областей, мінімізував виклики `setState()` під час рендерингу та використовував лише легкі криві (`Curves.easeInOut`, `Curves.linear`).

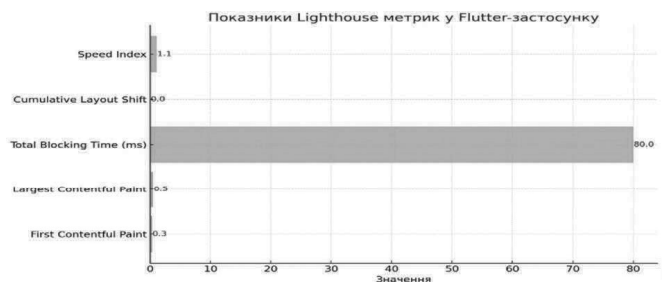


Рис. 3. Показники Lighthouse метрик

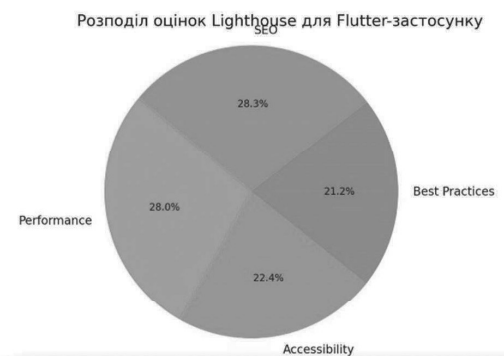


Рис. 4. Розподіл оцінок Lighthouse

Окрім вимірювання ефективності, важливо вимірювати доступність. Найчастіше зустрічаються такі проблеми: Відсутність належних альтернативних описів (alt) для анімованих кнопок і банерів.

Обмеження масштабування, встановлене директивою `user-scalable=«no»`, робить додаток недоступним для людей з порушеннями зору.

Відсутність достатнього контрасту в деяких анімованих елементах (наприклад, текст на кольоровій кнопці), що не відповідає вимогам WCAG.

Ці недоліки характерні для додатків, де анімація досягається за допомогою кастомних віджетів, які не мають належної семантики та достатньої підтримки атрибутів ARIA.

Для сприяння інклюзивності розробникам рекомендується використовувати віджети семантики для визначення анімованого контенту, повністю протестувати додаток з увімкненими функціями системної доступності (TalkBack, VoiceOver), додати логіку для вимкнення анімації, коли `MediaQuery.of(context).disableAnimations` оцінюється як `true`.

Отримані результати свідчать про те, що фреймворк Flutter підтримує надійний та універсальний інструментарій для створення анімації з прийнятними технічними показниками продуктивності. Водночас, щоб зробити користувацький досвід справді якісним, необхідно не лише стежити за такими показниками, як FPS або LCP, але й впроваджувати інклюзивні рішення, які задовольняють різноманітні сценарії використання, зокрема, вимоги людей з вадами зору або опорно-рухового апарату.

Тільки завдяки комплексному підходу можна гарантувати ефективність, стабільність і зручність використання цифрового продукту на всіх рівнях взаємодії.

5. Роль інклюзивних анімацій у Flutter-застосунках

Інклюзивність в анімаційному дизайні відіграє надзвичайно важливу роль у розробці сучасних мобільних додатків. Flutter, як кросплатформенний фреймворк, дозволяє створювати візуально привабливі, плавні та інтерактивні інтерфейси. У той же час, надмірне або неправильне використання анімації може стати бар'єром для деяких користувачів, особливо тих, хто має порушення зору, вестибулярні розлади або чутливість до руху.

Люди з такими порушеннями можуть відчувати дискомфорт або втрату контролю при взаємодії з анімованим інтерфейсом. Наприклад, анімація, яка не має логічного контексту або занадто швидка/повільна, може ускладнити навігацію.

Саме тому інклюзивний підхід до проектування анімації вимагає врахування користувацьких налаштувань, таких як `MediaQuery.of(context).disableAnimations`, які вказують на необхідність мінімізувати або відключити рухомі елементи.

Flutter також підтримує використання віджету «Семантика», який дозволяє озвучувати контент і проговорювати дії, викликані анімацією. Ця функція гарантує, що такі програми для читання з екрану, як TalkBack і VoiceOver, зможуть правильно інтерпретувати динамічні зміни в інтерфейсі.

На особливу увагу заслуговує специфікація ARIA Authoring Practices Guide (APG), адаптована для дизайну Flutter. Хоча Flutter не використовує HTML безпосередньо, принципи APG — такі як передбачуваність, логічна навігація та доступність для альтернативних пристроїв введення — залишаються актуальними.

Розробники можуть слідувати цим принципам, щоб забезпечити комфорт для всіх користувачів незалежно від їхніх фізичних можливостей.

Таким чином, метою анімації доступності є не лише мінімізація когнітивного навантаження, але й забезпечення того, щоб анімаційний шар надавав доступ до функціональності додатку, а не блокував його, а отже, сприяв кращій взаємодії з користувачем.

6. Інструменти для оцінки інклюзивності

Щоб оцінити якість доступності анімації Flutter, розробники можуть використовувати кілька інструментів, які прямо чи опосередковано підтримують тестування доступності:

Flutter DevTools дозволяє профілювати продуктивність анімації, що дає змогу виявити фризи, затримки та перевантаження графічного потоку.

Це важливо для забезпечення плавної анімації, особливо на пристроях з низькою обчислювальною потужністю.

Lighthouse (у веб-версіях додатків Flutter) — дозволяє аналізувати доступність, зокрема виявляти проблеми з контрастністю, нестандартні елементи керування, а також вказує на проблеми зі збільшенням масштабу або відключенням звуку в динамічному контенті.

TalkBack (Android), VoiceOver (iOS) — основні системні інструменти, що використовуються для тестування озвучення елементів, зокрема, після анімаційних переходів або появи нових віджетів [6].

Screen Reader Testing — ручне тестування за участю користувачів з особливими потребами або автоматизоване тестування за допомогою емуляторів, що дозволяє перевірити, наскільки логічною є структура анімованого інтерфейсу для сторонніх програм. Хоча у Flutter ще немає повністю нативного масштабованого інструменту тестування доступності анімації, який можна порівняти з ахе або WAVE у веб-середовищі, його вбудована інтеграція з Lighthouse, DevTools та відповідність керівним принципам Material Accessibility дає розробникам інструменти, необхідні для виявлення та усунення бар'єрів.

Висновки

Анімація у Flutter — чудовий спосіб підвищити візуальну привабливість, зробити навігацію ефективною та зменшити когнітивне навантаження на користувача.

Однак дуже важливо, щоб їх застосування не порушувало принципів доступності, оскільки надмірна або неконтрольована анімація може спричинити проблеми у взаємодії з інтерфейсом для людей з обмеженими можливостями.

Дослідження підтвердило, що за умови дотримання рекомендацій щодо продуктивності (RepaintBoundary, оптимізація setState, профілювання через DevTools) та доступності (Semantics, MediaQuery.disableAnimations, системні зчитувачі екрану) Flutter дозволяє створювати інтерактивні та інклюзивні інтерфейси.

Такі інструменти, як Google Lighthouse, Flutter DevTools, а також рекомендації Material Design Accessibility та ARIA APG надають розробникам засоби для всебічної оцінки та покращення анімованих інтерфейсів.

Таким чином, поєднання продуктивності, естетики та інклюзивності в реалізації анімації Flutter є основою для високоякісного мобільного досвіду, який враховує потреби найширшої аудиторії – незалежно від їхніх фізичних або когнітивних здібностей.

Конфлікт інтересів

Автори заявляють, що не мають конфлікту інтересів.

Список джерел:

- [1] Flutter documentation: Animations overview. URL: <https://docs.flutter.dev/ui/animations/overview>
- [2] Flutter Performance profiling with DevTools. URL: <https://docs.flutter.dev/tools/devtools>
- [3] Web Vitals. Google Developers. URL: <https://web.dev/articles/vitals>
- [4] Material Design – Motion principles. Google. URL: <https://m3.material.io/styles/motion/overview/how-it-works>
- [5] Accessibility in Flutter. Flutter documentation. URL: <https://docs.flutter.dev/ui/accessibility-and-internationalization/accessibility>
- [6] TalkBack screen reader for Android. Google Support. URL: <https://support.google.com/accessibility/android/answer/6007100?hl=en>
- [7] Lighthouse. Google Developers. URL: <https://developer.chrome.com/docs/lighthouse/overview/>
- [8] Semantics Widget. Flutter API Reference. URL: <https://api.flutter.dev/flutter/widgets/Semantics-class.html>
- [9] MediaQuery class – Flutter. URL: <https://api.flutter.dev/flutter/widgets/MediaQuery-class.html>
- [10] Animations in Flutter. URL: <https://codelabs.developers.google.com/advanced-flutter-animations#0>

Надійшла до редколегії 03.04.2025



С.Г. Удовенко¹, В.А. Затхей², О.В. Тесленко³, Л.Е. Чала⁴

¹ХНЕУ ім. С. Кузнеця, м. Харків, Україна, serhiy.udovenko@hneu.net,
ORCID iD: 0000-0001-5945-8647

²ХНЕУ ім. С. Кузнеця, м. Харків, Україна, zathey_va@ukr.net,
ORCID iD: 0000-0003-4426-7789

³ХНЕУ ім. С. Кузнеця, м. Харків, Україна, oleh.teslenko@hneu.net,
ORCID iD: 0000-0003-3105-9323

⁴ХНУРЕ, м. Харків, Україна, larysa.chala@nure.ua,
ORCID iD: 0000-0002-9890-4790

АВТОМАТИЗОВАНЕ ТЕСТУВАННЯ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ МОДУЛІВ ІНФОРМАЦІЙНИХ СИСТЕМ

Досліджено проблему автоматизованого тестування програмного забезпечення (ПЗ) модулів інформаційних систем (ІС) на прикладі тестування веб-застосунків. Розглянуто схему генерації тестових сценаріїв для тестування ПЗ модулів ІС (на прикладі веб-сайтів визначеної предметної області). Запропоновано технологію оптимізації процесів тестування веб-сайтів з використанням спеціалізованого тестового фреймворку та інструментів системи керування контентом Sitecore CMS. Розглянуто можливість застосування в запропонованій системі тестування концепції безперервної інтеграції та доставки (CI/CD), що дозволяє запускати різні типи тестів на кожному етапі тестування і завершувати їх розгортанням тестованого коду в кінцевий продукт. Працездатність запропонованої технології автоматизованого тестування модулів ІС досліджено на прикладі тестування веб-застосунків. Визначено, що використання цієї технології дозволяє скоротити витрати часу на тестування веб-додатків.

АВТОМАТИЗОВАНЕ ТЕСТУВАННЯ, ВЕБ-ЗАСТОСУНОК, ФРЕЙМВОРК, ТЕСТОВИЙ СЦЕНАРІЙ, ІНФОРМАЦІЙНА СИСТЕМА, ТЕСТОВА МОДЕЛЬ

S.G. Udovenko, V.A. Zathey, O.V. Teslenko, L.E. Chala. Automated software testing of information systems modules. The paper explores the problem of automated software testing (SW) of information system (IS) modules, using web application testing as a case study. A scheme for generating test scenarios for IS module testing is presented, specifically applied to websites within a defined subject area. The paper proposes a technology for optimizing website testing processes using a specialized test framework and Sitecore CMS tools. The integration of continuous integration and delivery (CI/CD) concepts into the testing system is examined, enabling the execution of various types of tests at different testing stages, culminating in the deployment of the tested code as a final product. The performance of the proposed automated testing technology is evaluated through web application testing, demonstrating a significant reduction in the time required for testing processes.

AUTOMATED TESTING, WEB APPLICATION, FRAMEWORK, TEST SCENARIO, INFORMATION SYSTEM, TEST MODEL

Вступ

Функціональні модулі інформаційної системи (ІС) — це взаємопов'язані частини системи, кожна з яких виконує певні функції, необхідні для досягнення загальної мети системи. Вони складаються з елементів, які визначають функціональні можливості системи та забезпечують її роботу [1].

Існує широка градація модулів ІС (МІС), що, насамперед, залежить від функціонального призначення інформаційних систем та їх структурних особливостей. Втім, такі модулі, як функціональні підсистеми та спеціалізовані веб-додатки (зокрема веб-сайти структурних одиниць ІС) є притаманними для переважної більшості розроблюваних ІС.

Це обумовлює важливість уважного ставлення до створення та удосконалення ефективних технологій побудови веб-додатків, як важливих модулів ІС.

Запорукою ефективності створюваних модулів ІС є розроблення працездатного програмного

забезпечення (ПЗ МІС). Важливим етапом розробки програмного забезпечення є етап тестування. Складність і різноманіття функцій сучасних програм вимагає застосування спеціальних методів аналізу стану і працездатності програмного забезпечення протягом усього циклу розробки, впровадження та супроводу. У широкому сенсі, тестування — це одна з технік контролю якості, яка включає планування, складання тестів, безпосереднє виконання тестування та аналіз отриманих результатів. На сьогодні забезпечення якості програмних продуктів, що контролюється на етапі тестування, є одним з провідних напрямів в ІТ-індустрії.

До найбільш важливих критеріїв вибору технології створення та тестування веб-сайтів слід віднести: розмір та тип проекту; складність проекту; швидкість розробки; вартість послуг спеціалістів; наявність доступних інструментів для розробки; гнучкість рішення; тенденції розвитку проекту; наявність детальної

документації; вартість підтримки; кросплатформеність; можливість інтеграції з іншими рішеннями тощо [2, 3].

Автоматизація процесів тестування представляє особливий інтерес, тому що дозволяє значно скоротити час проходження системою повного комплексу тестових сценаріїв та мінімізувати при цьому кількість трудових витрат на цей процес. Автоматизовані системи тестування вже стали невід'ємною частиною процесів розробки модулів ІС. Особливу увагу слід приділити сфері веб-розробок, що стрімко розвивається та стає все більш поширеною областю в інженерії програмного забезпечення.

В області веб-розробок на сьогоднішній день існує велика кількість різних систем автоматизації тестування, що відрізняються функціональними можливостями, ступенем стандартизації, опрацюванням призначеного для користувача інтерфейсу, правилами ліцензування [4, 5].

Незважаючи на те, що в області тестування веб-застосунків (зокрема, веб-сайтів МІС) здійснюється велика кількість теоретичних та практичних досліджень, в ній досі залишається чимало важливих питань, що потребують вирішення [6]. Одним з них є побудова систем автоматичного тестування веб-сайтів та генерація тестових сценаріїв для них. Таким чином, розроблення ефективної технології генерації тестових сценаріїв та універсальних шаблонів для автоматичного тестування веб-сайтів є актуальним завданням.

Слід також зазначити, що на сьогодні потужними засобами автоматизованого тестування створюваних веб-сайтів є застосування фреймворків (середовищ розробки для програмістів з готовими шаблонами та інструментами) та систем керування контентом CMS (Content Management System), призначених для організації веб-сайтів чи інших інформаційних ресурсів в Інтернеті чи окремих комп'ютерних мережах. В загальному випадку CMS використовується як готове рішення, в якому потрібно лише налаштувати контент, проте на основі фреймворку можна створити власну CMS під специфіку та функціональність проекту. Одною з таких систем є Sitecore CMS. Sitecore CMS використовує платформу ASP.NET і працює під управлінням ІІС (Internet Information Services), що відкриває чимало можливостей для розробки веб-застосунків.

Метою цієї статті є розроблення та дослідження технології оптимізації процесів автоматизованого тестування веб-сайтів з використанням запропонованого тестового фреймворку, засобів Sitecore CMS та генерації тестових сценаріїв.

Реалізація такої технології дозволить автоматизувати процедуру функціонального тестування розроблених веб-застосунків ІС та скоротити сумарні

витрати на реалізацію життєвого циклу їх проєктування.

Відповідно до поставленої мети вирішуються такі завдання:

- аналіз існуючих технологій автоматизації модульного тестування веб-застосунків інформаційної системи;
- розроблення схеми генерації тестових сценаріїв для автоматичного тестування ПЗ модулів ІС (на прикладі веб-сайтів визначеної предметної області);
- розроблення технології оптимізації процесів тестування веб-сайтів з використанням пропонованого тестового фреймворку Sitecore та засобів Sitecore;
- експериментальне дослідження запропонованого підходу.

1. Технології автоматизації тестування

Функціональне тестування модуля інформаційної системи – це процес перевірки того, як програмне забезпечення модуля виконує певні функції, порівнюючи його поведінку з вимогами. Мета такого тестування – переконатися, що модуль системи працює відповідно до специфікацій.

До основних типів функціонального тестування слід, зокрема, віднести модульне тестування (перевірка окремих модулів або компонентів ІС) та інтеграційне тестування (перевірка взаємодії між різними модулями ІС).

Модульне та інтеграційне тестування зазвичай автоматизуються, що дозволяє значно прискорити загальний процес тестування і знизити ризик помилок [7, 8].

У контексті тестування ПЗ модулів ІС функціональне автоматизоване тестування може включати: перевірку введення та обробки даних; перевірку правильності відображення інформації; тестування роботи з інтерфейсом користувача; перевірку обробки різних сценаріїв користувачів; тестування безпеки та авторизації.

На сьогодні ієрархію технологій автоматизованого тестування (від вищого рівня до нижнього) модулів ІС можна умовно відобразити наступним чином [9]:

- тестування під управлінням поведінкою;
- тестування під управлінням ключовими словами;
- тестування під управлінням даними;
- запис та відтворення;
- використання фреймворків.

Кожна з цих технологій автоматизації тестування, в свою чергу, базується на відповідному наборі технічних рішень (інструментальних засобів, мов програмування, способів взаємодії з тестованим застосунком тощо) та має свої переваги та недоліки.

Більшість сучасних технологій автоматизації тестування максимально сфокусовані на технічних

аспектах поведінки тестованих модулів та мають загальний недолік: з їх допомогою складно перевіряти високорівневі сценарії, що призначені для користувача (а саме в них і зацікавлені замовники і користувачі). Цей недолік частково виправляє тестування під керуванням поведінкою (Behavior-driven development, BDD), де акцент робиться не на окремих технічних деталях, а на загальній працездатності ПЗ модуля при вирішенні типових користувальницьких завдань.

Такий підхід не тільки спрощує виконання цілого класу перевірок, а й полегшує взаємодію між розробниками, тестувальниками, бізнес-аналітиками і замовником, тому що в основі підходу лежить достатньо проста формула «given-when-then»:

- given («за умови»): описує початкову ситуацію, в якій знаходиться користувач в контексті роботи з тестованим застосунком;

- when («якщо»): описує набір дій користувача в даній ситуації;

- then («тоді»): описує очікувану поведінку програми модуля ІС.

Такий принцип опису перевірок дозволяє навіть учасникам проекту, що не мають спеціальної підготовки, брати участь в розробці і аналізі тест-кейсів, а для фахівців з автоматизації спрощується створення коду автоматизованих тест-кейсів, тому що така форма є стандартною, єдиною і при цьому надає достатньо інформації для написання високорівневих тест-кейсів. Існують спеціальні технічні рішення (наприклад, Behat, JBehave, NBehave, Cucumber), що спрощують реалізацію тестування під керуванням поведінкою.

Недоліки тестування під управлінням поведінкою: високорівневі поведінкові тест-кейси пропускають багато деталей, а тому можуть не виявити частину проблем в застосунку або не надати необхідної для розуміння виявленої проблеми інформації.

Тестування під управлінням ключовими словами є логічним розвитком ідеї про винесення зовні тест-кейса даних та команд (опису дій). Прикладом інструментального засобу автоматизації тестування під керуванням ключовими словами є фреймворк Selenium IDE, де кожна операція тест-кейса описується в вигляді: дія (ключове слово) – необов'язковий параметр 1 – необов'язковий параметр 2.

Переваги та недоліки тестування під управлінням ключовими словами:

- переваги: максимальне усунення надмірності коду тест-кейсів; можливість побудови міні-фреймворків, які вирішують широку спектр завдань; підвищення рівня абстракції тест-кейсів і можливість їх адаптації для роботи з різними технічними рішеннями; зручне зберігання і зрозумілий для користувачів формат даних і команд тест-кейса; можливість повторного використання коду тест-кейсів для

вирішення нових завдань; можливість розширення (додавання нової поведінки тест-кейса на основі вже реалізованого варіанту);

- недоліки: загальна складність розробки; висока ймовірність наявності помилок в коді тест-кейса; висока складність виконання низькорівневих операцій, якщо код тест-кейса не підтримує відповідні команди; ефект від використання даного підходу настає далеко не відразу (спочатку йде тривалий період розробки та налагодження тест-кейсів і допоміжної функціональності); необхідність застосування мови ключових слів, що використовуються в тест-кейсах.

Тестування під управлінням даними (Data Driven Testing, DDT) є підходом до створення автоматизованих тестів, при якому тест має приймати набір вхідних параметрів, а еталонний стан, з яким він повинен порівняти результат, отримується під час прогонки вхідних параметрів.

До типових прикладів використання тестування під керуванням даними відносяться: перевірка авторизації і прав доступу на великому наборі імен користувачів і паролів; багаторазове заповнення полів форм різними даними і перевірка реакції модуля; виконання тест-кейса на основі даних, отриманих за допомогою комбінаторних технік.

Технологія запису і відтворення (Record & Playback) стала актуальною з появою достатньо складних засобів автоматизації, що забезпечують глибоку взаємодію з тестованим застосунком і операційною системою. Використання цієї технології, зазвичай, зводиться до наступних основних кроків: тестувальник самостійно виконує тест-кейс, а засіб автоматизації записує його дії; результати записів формуються у вигляді коду на високорівневій мові програмування; тестувальник редагує отриманий код; готовий код автоматизованого тест-кейса виконується для проведення тестування в автоматизованому режимі.

Ця технологія при високої складності внутрішньої реалізації є простою у використанні, тому часто застосовується для попереднього навчання фахівців з автоматизації тестування. Її основні переваги та недоліки:

- переваги: простота освоєння тестувальниками; швидке створення «скелета» тест-кейса за рахунок запису ключових дій з тестованим застосунком; автоматичний збір технічних даних про тестований застосунок (ідентифікаторів і локаторів елементів, написів, назв тощо); автоматизація рутинних дій (заповнення полів, натискань на посилання, кнопки);

- недоліки: лінійність тест-кейсів (в записі не буде циклів, умов, викликів функцій і інших характерних для програмування і автоматизації функцій); запис зайвих дій (як помилкових випадкових дій тестувальника з тестованим застосунком, так і перемикачів на інші застосунки та роботу з ними); незручні імена

змінних, незручне оформлення коду тесткейса, відсутність коментарів, що ускладнюють підтримку і супровід тест-кейса в подальшому; низька надійність самих тест-кейсів через відсутність перевірки умов.

Фреймворки автоматизації тестування зазвичай є успішно розвиненими рішеннями, що поєднують в собі переваги тестування під керуванням даними та ключовими словами, а також можливість реалізації додаткових рішень на високому рівні абстракції.

Фреймворків автоматизації тестування можуть суттєво різнитися, але їх об'єднує кілька спільних рис, а саме: висока абстракція коду (немає необхідності опису кожної елементарної дії) зі збереженням можливості виконання низькорівневих дій; універсальність використовуваних підходів; досить висока якість реалізації (для популярних фреймворків).

Зазвичай кожен фреймворк спеціалізується на конкретних видах, рівнях та технологіях тестування. Існують фреймворки для модульного тестування (наприклад, сімейство xUnit), тестування веб-застосунків (наприклад, сімейство Selenium), тестування мобільних додатків, тестування продуктивності тощо. Існують вузько і широко спеціалізовані, а також безкоштовні і платні фреймворки, що оформлені як бібліотеки на деякій мові програмування або як додатки з графічним інтерфейсом.

Основні переваги та недоліки тестування модулів ІС з використанням фреймворків:

- переваги: широке розповсюдження; універсальність в рамках обраного набору технологій; високий рівень абстракції; наявність набору готових рішень і описів відповідних кращих практик застосування того чи іншого фреймворка для вирішення конкретних завдань;

- недоліки: необхідність витрат часу на вивчення фреймворка тестувальниками; висока ймовірність необхідності модифікації існуючого або створення власного фреймворка для вирішення конкретних завдань; виникнення додаткових труднощів при переході на інший фреймворк; високий ризик вибору невідповідного фреймворка.

Існують певні напрями оптимізації тестування, метою яких є: прискорення виконання тестування; вивільнення людських ресурсів; збільшення тестового покриття; виключення людського фактору; надання можливості переглядати звіт з проходження тест-кейсів.

Їх можна, зокрема, реалізувати за допомогою автоматизації тестування на проекті. При цьому мають бути протестованими: поведінка програмного продукту з точки зору контентного менеджера (додавання, редагування та видалення сторінок, компонентів тощо) та з точки зору кінцевого користувача (правильність відображення компонентів стосовно співпадіння даних, які заповнив контент-менеджер,

з даними, що відображаються), а також візуальна відповідність компонент результатів вимогам. Інколи для оптимізації цих сфер тестування необхідно розробити фреймворк для UI тестування (перевірка правильності роботи компонентів з точки зору користувача). Для перевірки роботи програмного продукту з точки зору контент-менеджера достатнім є його тестування за допомогою API.

Розглянемо особливості вибору інструментів автоматизованого тестування модулів ІС. Згідно з метою статті необхідно проаналізувати існуючі тестові фреймворки, що найбільш поширено використовуються для оптимізації процесів тестування веб-сайтів ІС.

До таких фреймворків слід, насперед, віднести інструмент Selenium з відкритим вихідним кодом, призначений для автоматизації тестування веб-додатків.

Фреймворк Selenium підтримується операційними системами Windows, Mac, Linux, а також браузерами Chrome, Firefox, IE і Headless. Скрипти для даного фрейма можна написати на більшості популярних сьогодні мов програмування: Java, Groovy, Python, C #, PHP, Ruby і Perl [10, 11].

Фреймворк Katalon Studio також є достатньо ефективним інструментом для автоматизації процесу тестування веб-додатків, мобільних додатків і веб-сервісів. Він перейняв у Selenium переваги, пов'язані з інтегрованою автоматизацією тестування ПЗ.

Фреймворк Unified Functional Testing (UFT) є популярним комерційним інструментом для функціонального тестування. Він надає повний набір функцій для тестування API, веб-сервісів, а також для тестування графічного інтерфейсу десктопних, мобільних і веб-додатків на всіх існуючих платформах. Для даного інструменту передбачена розширена функція розпізнавання об'єктів на основі зображень, багаторазові тестові компоненти і документація по автоматичному тестуванню.

Фреймворк Watir — це інструмент з відкритим вихідним кодом для автоматизації тестування веб-застосунків, що використовує бібліотеки Ruby. Для Watir передбачена можливість крос-браузерного тестування для браузерів Firefox, Opera, headless-браузерів і IE. Він також підтримує кероване даними тестування і інтегрований з такими інструментами BDD, якRSpec, Cucumber і Test / Unit.

Фреймворк TestComplete є інструментом для тестування десктопних, мобільних і веб-додатків. TestComplete підтримує такі мови сценаріїв, як JavaScript, VBScript, Python і C ++ Script. За допомогою TestComplete тестувальники можуть виконувати тестування з використанням ключових слів і кероване даними тестування. В інструменті також передбачена зручна функція запису і відтворення процесу тестування.

На основі розглянутих фреймворків може бути створена система керування контентом CMS, призначена для автоматизованого тестування веб-сайтів. Серед існуючих систем цього типу в останні роки набирає значної популярності система Sitecore CMS. Система Sitecore CMS володіє надійним набором можливостей для автоматизованого тестування, персоналізації та оптимізації, що робить її ідеальною платформою для тестування та уточнення створюваних веб-застосунків [12]. Так як Sitecore використовує платформу ASP.NET, то для кращої інтеграції з проектом написання модифікованого тестового фреймворку доцільно використовувати мову програмування C# і відповідно платформу .NET і середовище розробки Microsoft Visual Studio 2017.

Розглянемо переваги платформи ASP.NET, що використовуються в системі Sitecore CMS: платформа .NET ґрунтується на єдиній об'єктно-орієнтованій моделі; до складу платформи .NET входить інструмент захисту програми тестування від втрат пам'яті і від необхідності звільняти ресурси; будь-яка програма, розроблена з допомогою .NET є автономною, в тому сенсі, що не залежить від інших програм та від ОС; встановлення програми може бути проведене звичайним копіюванням файлів; в платформі .NET використовуються безпечні типи даних, що підвищує надійність програм та їх сумісність; програма взаємодіє з єдиною моделлю обробки помилок; всі помилки обробляються механізмом виключних ситуацій, що дозволяє запобігти неоднозначностям [13].

Для підвищення можливостей автоматичного тестування ПЗ модулів ІС останнім часом застосовуються методи генерації тестових сценаріїв (на прикладі веб-сайтів визначеної предметної області) та створення відповідних універсальних шаблонів.

Таким чином, можна зробити висновок про доцільність побудови фреймворку автоматизованого тестування на основі оптимізації процесів тестування веб-сайтів з використанням засобів Sitecore CMS та схеми генерації тестових сценаріїв.

2. Схема генерації тестових сценаріїв для автоматичного тестування веб-застосунків визначеної предметної області

До важливих завдань автоматизованого тестування ПЗ модулів ІС слід віднести побудову тестових сценаріїв, що брали б до уваги специфічні вимоги до тестованих веб-застосунків та озволяли б ефективно виявляти помилки в розроблених кодах. Такі тестові сценарії можуть бути створені на основі моделей веб-застосунків [14]. Послідовність їх створення наведено на рис. 1.



Рис. 1. Генерація сценаріїв тестування на основі моделі веб-застосунку

Модель веб-застосунків дає можливість генерувати набори варіантів сценаріїв тестування, де визначаються умови, вхідні дані, очікувані результати роботи системи та оцінювання результатів тестування з використанням компаратора. Порівняння в компараторі очікуваних результатів та поточних спостережень дозволяє зробити висновок щодо наявності помилок в тестованому модулі та його відповідності до вимог. В разі необхідності можна модифікувати існуючу модель або ініціювати додаткову генерацію варіантів тестування.

Особливості автоматизованого тестування у веб-проектах обумовлюється наявністю коротких циклів розробки, розподіленою архітектурою, кросплатформністю та кросбраузерністю.

Для моделювання та подальшого тестування складних веб-застосунків доцільно використовувати ієрархічний підхід.

Відзначимо, що процес формування тестових сценаріїв є індивідуальним для кожного конкретного проекту та може відбуватися паралельно з розробленням ПЗ тестованого модуля. Достатньо специфічною особливістю проектів створення веб-застосунків є неодноразове повторення типових процедур, які мають лише незначні функціональні відмінності. До таких проектів можна віднести розробку інтернет-магазинів, блогів, порталів новин тощо. Звичайно, що за таких умов розробка тестових сценаріїв для кожного окремого проекту є недоцільною та раціональному використанню ресурсів.

Пропонований підхід передбачає побудову ієрархічної структури універсальних тестових сценаріїв. Верхній рівень цієї структури містить універсальні тестові сценарії, що характерні для досліджуваної предметної області. Найнижчий рівень ієрархічної структури відповідає області взаємодії з елементами візуального інтерфейсу конкретного проекту. Адаптація розроблених шаблонних сценаріїв до конкретних проектів здійснюється саме на цьому

ієрархічному рівні. Ієрархічна структура тестових сценаріїв дозволяє здійснювати тестування на основі моделей, де враховуються функціональні вимоги до веб-додатків обраної типової предметної області та налаштування для конкретної реалізації.

Після налаштування моделі під конкретний веб-застосунок остаточно генеруються тестові набори.

Узагальнена ієрархічна структура успадкованих тестових сценаріїв містить ядро та змінну частину. До ядра вносять опис тестових випадків, що залишається незмінним від проекту до проекту в рамках певної предметної області. Змінна частина структури (словник проекту) містить відсилання до існуючих точок входу проекту (наприклад, адреси сторінок, посилання на активні елементи інтерфейсу користувача, посилання на блоки певного призначення).

Загальну структуру системи автоматичного тестування з використанням набору шаблонів наведено на рис. 2.

Ядро системи тестування в свою чергу поділяється на набір шаблонів тестових сценаріїв та бібліотеку операторів. Такий поділ є корисний для виключення дублювання коду в разі повторення однакових кроків для різних тестових сценаріїв. Під тестовим сценарієм будемо розуміти сукупність тестових випадків, які ініціюють перевірку працездатності окремого фрагменту модуля, що тестується. Тестові сценарії зберігаються в окремих файлах і є автономними функціональними одиницями. Вміст та характеристики тестових сценаріїв відповідають конкретній предметної області.



Рис. 2. Структура системи автоматичного тестування з використанням шаблонів

Шаблони тестових сценаріїв об'єднують сукупності тестових випадків, які визначають послідовність дій потенційного користувача, що впливає певним чином на стан тестованого модуля системи та конкретний очікуваний результат. Перевірка працездатності

тестованого модуля відбувається за результатами порівняння очікуваного результату з реальним.

Система автоматичного тестування з використанням шаблонів має забезпечувати:

- універсальність шаблонів, що визначають повноту покриття тестовими випадками предметної області (тобто тестових сценаріїв ядра);

- мінімізацію кількості елементів словника проекту, що визначає складність формування словників для нових проектів.

Бібліотека операторів містить сукупність повторюваних дій, що можуть використовуватися в різних тестових сценаріях. Такі оператори взаємодіють з даними зі словника проекту для отримання доступу до необхідних елементів інтерфейсу користувача веб-застосунку. Всі оператори поділяються на оператори переходу, оператори дії та оператори перевірки.

Оператори переходу використовуються для переміщення між сторінками веб-застосунку і мають забезпечувати імітацію дій потенційних користувачів. Система автоматичного тестування має використовувати лише ті елементи управління, що доступні реальним користувачам. Кожен тестовий випадок починається з виклику оператора переходу до розділу системи, розглядається цей випадок.

Оператори дії визначають варіанти взаємодії з елементами управління інтерфейсу користувачів (наприклад, з кліками по активним елементам, введенням даних з клавіатури тощо);

Оператори перевірки реалізують різні методи перевірки стану тестованого модуля. Такі методи перевіряють, зокрема, наявність на сторінках певних елементів та вміст текстових фрагментів, а також порівнюють поточні значення параметрів системи з раніше збереженими значеннями.

Словник проекту використовується для модифікації системи автоматичного тестування під потреби конкретних проектів. Він містить специфічні для тестованого веб-застосунку дані, а саме: найменування елементів графічного інтерфейсу та посилання на них; відклики на результати виконання операцій; опис структур даних аналізованих об'єктів тощо. В цілому складові частини словнику проекту можуть бути поділені на такі групи:

- ідентифікатори елементів інтерфейсу користувача, що використовують CSS-селектори;

- текстові дані та повідомлення, характерні для різних тестових випадків (за їх наявністю тестова система тестування визначає успішність виконання тестового випадку);

- тригери (параметри, що визначають алгоритми проведення тестів та характеризуються булевими значеннями для тестових випадків проекту).

Метод побудови та використання ієрархічної структури шаблонів автоматизованого тестування

(для конкретної предметної області) передбачає реалізацію наступних кроків:

Крок 1. Аналіз предметної області та визначення основних функціональних вимог до для тестованого веб-застосунку.

Крок 2. Формування ядра системи тестування, де міститься опис універсальних тестових сценаріїв з урахуванням визначених функціональних вимоги, а також бібліотека операторів, що визначають сукупність повторюваних дій.

Крок 3. Формування словника (змінної частини для конкретного веб-застосунку, де містяться адреси веб-сторінок, ідентифікатори елементів інтерфейсу користувачів, параметри тестованого веб-застосунку тощо.

Крок 4. Інтеграція розроблених шаблонів до системи автоматичного тестування.

Крок 5. Модифікація словника для конкретного тестованого веб-застосунку.

Крок 6. Тестування веб-застосунку з використанням шаблонів кінцевих тестових сценаріїв.

Основними вимогами до програмної реалізації запропонованого методу є: наявність гнучкої системи команд і операторів; можливість автоматичної адаптації до зміни умов тестування з використанням циклів і розгалужень; підтримка ієрархічних структур і успадкування; підтримка оголошень функцій і базової можливості повторного використання коду.

Таким чином, шаблони тестування мають бути зручними для користувачів та тестувальників.

В роботі [14] був досліджений варіант програмної реалізації запропонованого методу з використанням фреймворку Selenium. Втім такий варіант характеризується наявністю певних недоліків, пов'язаних з недосконалістю інтерфейсу записи сценаріїв в порівнянні з комерційними платними аналогами, а також з відсутністю вбудованої системи формування звітів за результатами автоматизованого тестування.

Є доцільним розглянути можливість реалізації методу автоматичного тестування на основі використання шаблонів, що базується на застосуванні запропонованого тестового фреймворку Sitetest та засобів Sitecore CMS.

3. Архітектура тестового фреймворку Sitetest для автоматизації тестування веб-сайтів

Розглянуто можливість оптимізації процесів тестування веб-сайтів з використанням спеціалізованого тестового фреймворку та засобів Sitecore CMS.

Для створення тестового фреймворку було обрано наступний набір інструментів та мови програмування: мова програмування C#, середовище розробки Visual Studio 2017, інструмент для UI тестування та взаємодії з веб елементами Selenium, BDD інструмент SpecFlow, Sitecore API бібліотека для

можливості тестування функціональності з участю контент-менеджера.

Побудова фреймворків для тестування модулів ІС (зокрема, тестування веб-застосунків) має враховувати загальні принципи розроблення інструментів автоматизованого тестування. Зокрема, згідно з принципом інверсії залежностей (dependency inversion principle) класи верхніх рівнів фреймворку тестування не повинні залежати від класів нижніх рівнів. Класи нижнього рівня реалізують базові операції (передачі даних по мережі, підключення до бази даних тощо).. Класи високого рівня містять більш складну бізнес-логіку програми, яка спирається на класи низького рівня для здійснення більш простих операцій. відбувається проектування. Відповідно до принципу інверсії залежностей слід розширювати низькорівневі класи і використовувати їх разом з класами бізнес-логіки, не змінюючи код останніх. В процесі реалізації тестового фреймворку цей принцип має використовуватися для створення шарів SpecFlow. В архітектурі тестового застосунку ці шари визначають зв'язок між .feature файлами та C# класами. Крім того для побудови остаточного варіанту фреймворку тестування треба спочатку створювати тестові сценарії без прив'язки до C# коду. Потім необхідно створити класи, де визначаються кроки тестування, а інші класи, (наприклад класи Page Object та Web Component) мають містити лише логіку роботи методів, що використовуються для роботи верхнього шару фреймворку [15].

На рис. 3 наведено архітектуру запропонованого фреймворку Sitetest, що використовується для тестування веб-сайтів.

Архітектуру тестового фреймворку Sitetest містить такі модулі та шари:

- ядро (Framework Core). Цей модуль задає конфігурацію веб-драйвера, розширення для веб-драйвера та веб-елементу, різні конфігураційні класи, класи з власними обгортками над очікуваннями тощо;

- шар SpecFlow Tests. Цей шар містить feature-файли та імплементацію кроків для них;

- тестова модель (TestModel). Цей модуль фреймворка містить декілька шарів. Перший шар — це шар сторінок (Page Objects), що допомагає інкапсулювати роботу з окремими елементами сторінки, дозволяє зменшити розмір коду та полегшити його підтримку (якщо, наприклад, дизайн однієї зі сторінок буде змінений, то потрібно буде переписати тільки відповідний клас, що описує цю сторінку). Другий шар є сервісний компонентом (Page Service), що містить класи, які відповідають за взаємодію з елементами сторінок, а також класи, які відповідають за перевірки;

- шар Sitecore, що відповідає за взаємодію з Sitecore API та використовується для тестування правильності функціонування додатку зі сторони контент-менеджера.

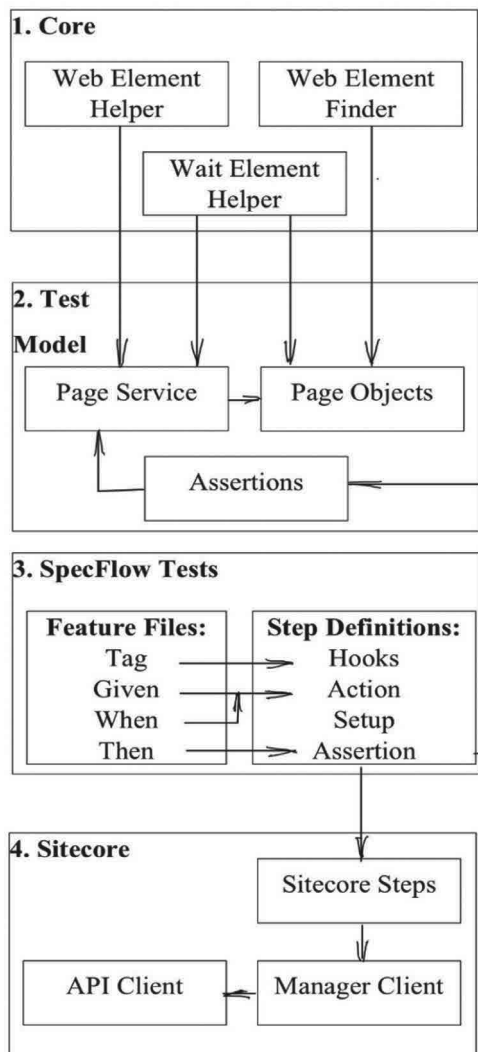


Рис. 3. Архітектура тестового фреймворку Sitetest для автоматизації тестування веб-сайтів

Функції та їх опис для шарів SpecFlow Tests, TestModel та Sitecore наведені в таблицях 1, 2, 3 відповідно.

Таблиця 1

Функції шару SpecFlow Tests

Функція	Опис функції
Extensions	Трансформація вхідних параметрів BDD сценаріїв до C# об'єктів та примітивних типів даних. Логіка порівняння вхідних параметрів сценаріїв.
Features	Тестові сценарії у вигляді послідовності кроків.
Step Definitions	Приведення BDD кроків до C# методів. Логіка порівняння очікуваного та актуального результатів. Логіка дій кінцевого користувача на веб-сайті.
App.config	Конфігурація тестового фреймворку. Змінні оточення. Глобальні тестові дані. Адреси веб-сторінок.

Тестові сценарії у вигляді послідовності кроків генеруються в шарі SpecFlow Tests на серверному

тестовому стенді в Sitecore CMS (в цьому випадку запуск і збір результатів тестування здійснюється планувальником тестового стенда) або на локальній машині тестувальника (у такому випадку локальна машина є тестовим стендом і запуск тестових сценаріїв здійснюється з командного рядка в консолі ОС, а результат роботи також буде відображено в консолі).

Таблиця 2

Функції шару TestModel

Функція	Опис функції
Web	C# шаблони веб-сторінок та веб-компонентів. Веб-сторінки – композиція веб-компонентів.
Utils	Допоміжні класи для спільної логіки роботи із різними типами даних (string, DateTime, int). Логіка генерації випадкових даних.
Entity	Представлення веб-компонентів у вигляді C# об'єктів. Логіка порівняння C# об'єктів (Equals та GetHashCode методи).
TestContext	Кеш тестових даних, згенерованих протягом запуску тестів. Стратегія ініціалізації тестових даних та компонентів в залежності від обраного режиму запуску тестів (Mobile та Desktop).
UserFlow	Логіка реєстрації нового користувача в системі. Логіка активації нового акаунту.
Constants	Глобальні константи. Перерахування для параметризації BDD кроків.

Таблиця 3

Функції шару Sitecore

Функція	Опис функції
Core	Логіка перетворення API відповідей із Sitecore до C# об'єктів.
Client	API клієнт для взаємодії із Sitecore REST сервісом. Методи, що відповідні REST CRUD операціям.
Entity	Конкретні класи-шаблони Sitecore компонентів. Представлення узагальнених полів (кнопки, текстові блоки, багатовибірні поля).
Models	Спільні C# моделі для роботи із Sitecore API (логін, створення, оновлення, видалення).
Utils	Допоміжні класи для роботи із різними типами полів на Sitecore (текстові блоки, кнопки, випадаючі меню).
Data Holder	Кеш даних, зібраних із Sitecore протягом запуску тестів. Дані, що чекають на публікацію у master базі.
Default Data	Тестові дані для конфігурації базових Sitecore компонентів, спільні для кожного запуску тестів.

Використання шарів в архітектурі фреймворку надає процесу тестування такі переваги:

- спрощує розробку тестів за рахунок перевикористання бізнес-функцій та шаблонів, які повторюються практично в кожному тесті;

- спрощує підтримку тестів (якщо змінюється верстка сторінки або навігація, то зміни достатньо внести лише в кілька класів сервісного і сторінкового шару, а інші тести залишаються без змін і продовжують успішно працювати);

- сприяє поділу відповідальності між елементами фреймворку Sitetest та забезпечує максимальне перевикористання коду.

4. Технологія оптимізації процесів тестування веб-сайтів з використанням фреймворку Sitetest та засобів Sitecore

Технологія оптимізації процесів тестування з використанням фреймворку Sitetest та базових Sitecore компонентів передбачає можливість запуску тестів на різних тестових оточеннях: локально та на сервері. Ініціалізація веб-драйвера здійснюється для цих двох випадків за стандартними алгоритмами, а на виході клієнти класа будуть працювати із `IWebDriver` об'єктом. Абстрактна стратегія при цьому задається класом `WebDriverProvider`, що використовує лише один абстрактний метод `InitWebDriver()`. Конкретними стратегіями будуть `RemoteWebDriverProvider` та `LocalWebDriverProvider`, що розширяють клас `WebDriverProvider`. Статичний клас `WebDriverResolver` формує та зберігає словник `IDictionary<string, WebDriverProvider>`, ключом до якого є значення змінної оточення, що буде надходити до додатку із конфігураційного файлу для двох різних реалізацій базової стратегії. Таким чином, якщо буде потреба додати ще одну логіку ініціалізації драйвера, це можна буде зробити за два простих кроки: створити ще одну імплементацію класу `WebDriverProvider`, а потім додати створений клас до словника, що зберігається у класі `WebDriverResolver`.

Тестування за допомогою Sitetest передбачає необхідність створення елементів, їх публікації в базі контент-менеджменту системи і перевірки відображення на веб-сайті. Для створення, оновлення, зміни та видалення елементів система тестування використовує API Sitecore. Для взаємодії з даними через API реалізуються CRUD операції. CRUD – акронім, що позначає чотири базові функції для роботи з базами даних: створення (create), читання (read), модифікація (update) та видалення (delete).

Всі CRUD операції з компонентами Sitecore виконуються з застосуванням нестандартного класу `SitecoreClient` за методами класу `RestSharp`.

`SitecoreClient` – це нестандартний клас, який може виконувати всі CRUD операції з компонентами

Sitecore та використовує методи класу `RestSharp`. У класі `SitecoreClient` є набір полів (`Language`, `AuthorizationCookie` і `Database`), які визначають параметри запиту на Sitecore. Параметр `Database` параметр передається динамічно під час прогону тестів, а параметр `Language` прописаний в файлі конфігурації як константа і не налаштовується при запуску автотестів (наприклад: `public const string AuthorizationCookieID = "SessionId"; public const string DatabaseParameterName = "database"; public const string LanguageParameterName = "language"`).

Логін і пароль адміна для Sitecore зберігаються як поле `SitecoreClient.AuthorizationCookie` поле задається після виконання логіну на Sitecore. Якщо у cookie файлу закінчується термін дії, то формуються 401 або 403 статус-коди для нашого запиту. Ці статус-коди обробляються всередині кожного методу CRUD операцій. Якщо повертається один з цих статус-кодів, то згідно з методом `Login()` оновлюється значення cookie файлу в поле `AuthorizationCookie`, а потім вже виконується запит.

Методи `GetItemById` і `GetItemByPath` – дозволяють повернення даних про компоненти Sitecore через `Id` або `Path` здійснюється за методами `GetItemById` і `GetItemByPath` відповідно.

Оскільки для роботи з Sitecore використовується клас `RestSharp`, то вхідні і вихідні дані при запиті на Sitecore представлені в форматі `.json`. Застосування різних шаблонів при створенні компонентів Sitecore викликає необхідність для кожного з макетів створення шаблону (C# клас), де вказуються імена полів.

Для серіалізації і десеріалізації відповідей і запитів на Sitecore сервері, використовується бібліотека `Newtonsoft.Json`, що надає гнучкі можливості роботи з `.json` рядками. Ці рядки можна перетворити в `JsonObject`, а потім в екземпляр об'єкта, представленого відповідним шаблоном.

Як приклад, розглянемо форму реєстрації, яка налаштована на веб-сайті із наступним набором полів: `EmailLabel`, `EmailPlaceholder`, `PasswordLabel`, `PasswordPlaceholder`.

Створюємо клас `RegistrationFormSitecoreItem`, який буде розширювати клас `SitecoreItem` (для цього додається атрибут `[ItemId ("1234567890")]`, в якому зазначено `id` веб-сайту). Для перевірки відображення створених компонентів на сайті будемо використовувати модель `WebComponentModel`, що задає поле моделей компонентів класу.

Для того, щоб порівняти між собою очікуваний результат (контент, налаштований на Sitecore) та актуальний результат (контент на веб-сайті), потрібно створити загальний тип для обох класів: `RegistrationFormWebDto`. Прошарок автоматизованих тестів `DTO (Data Transfer Object)`, де буде зберігатися модель, є спільний для елементів, набраних

із контент-менеджмент системи та із веб-фрагменту. Клієнт отримує дані, які безпосередньо відображаються у таблицях бази даних. Прошарок Data Transfer Object дозволяє оптимізувати тести таким чином: видалити кругові посилання; приховати певні властивості, які клієнти не повинні переглядати; опустити деякі властивості, щоб зменшити розмір корисного навантаження; вирівняти графіки об'єктів, що містять вкладені об'єкти, щоб зробити їх більш зручними для клієнтів; уникнути вразливостей, пов'язаних із надмірним розміщенням повідомлень; від'єднати рівень обслуговування від рівня бази даних.

Для того щоб уникнути дублювання коду в автотестах використовуються шаблони проектування Page Object та Page Elements.

Page Object — це шаблон проектування, який дозволяє розділити логіку виконання тестів від їх реалізації. Page Object моделює сторінки тестованого фрагменту програми як об'єкти в коді, що дозволяє визначати окремі класи, які відповідають за роботу з HTML, для кожної конкретної веб-сторінки. Такий підхід значно зменшує обсяг повторюваного коду, тому що одні й ті ж об'єкти сторінок можна використовувати в різних тестах. Основна перевага Page Object полягає в тому, що в разі зміни призначеного для користувача інтерфейсу можна виконати виправлення тільки в одному місці, а не виправляти кожен тест, в якому цей інтерфейс використовується.

Шаблон Page Elements дозволяє ділити сторінки на більш дрібні складові (блоки, віджети тощо), які надалі можна перевикористати в декількох сторінках.

Для візуалізації результату роботи тестів в пропонуваній системі тестування використовується фреймворк Allure. Allure Framework — це гнучкий багатомовний інструмент звіту про тести, який не лише формує стисле представлення результатів тестування у формі веб-звіту, але дозволяє тестувальникам отримати максимум корисної інформації з виконання тестів.

Життєвий цикл Allure налаштовується за допомогою файлу json (з ім'ям за замовчуванням allureConfig.json). Пакет NuGet встановлює allureConfig.Template.json, що можна використовувати як шаблон.

Процес тестування можна здійснювати різними способами. Передумовою ефективною автоматизації процесу тестування є тестування в рамках безперервного постачання ПЗ веб-застосунків.

В пропонуваній системі тестування може бути використана концепція безперервної інтеграції та доставки (CI/CD), що реалізується як конвеєр, полегшуючи злиття тільки що закоміченого коду в основну кодову базу. Концепція CI/CD дозволяє запускати різні типи тестів на кожному етапі тестування і завершувати його запуском з розгортанням тестованого коду в кінцевий продукт, що отримують кінцеві користувачі (виконання доставки) [16].

На виділеному сервері організується служба, до завдань якої входять: отримання коду з репозиторію; складання проекту; виконання тестів; розгортання готового проекту у тестовому середовищі; генерація та відправка звітів.

Для налаштування CI/CD обрано інструмент Jenkins. Jenkins надає можливість тестування коду в режимі реального часу та можливість отримувати звіти про окремі зміни в розгорнутій кодовій базі. Цей інструмент, головним чином, дозволяє Тестувальникам швидко позначати і виправляти помилки і баги в коді, а потім автоматично тестувати збірку коду.

На рис. 4 наведено узагальнену схему реалізації CI/CD процесу.

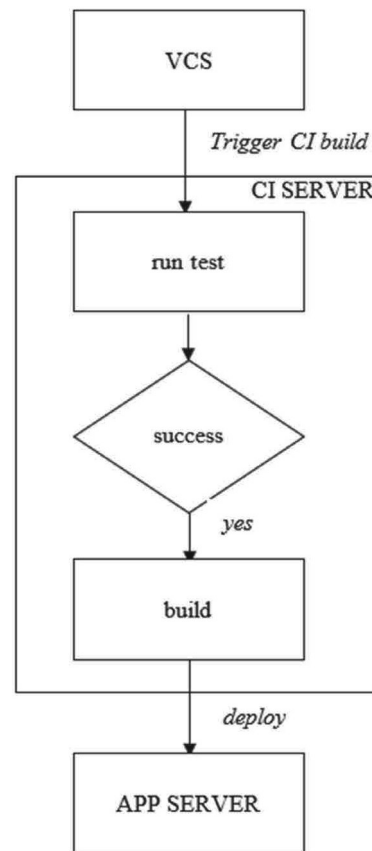


Рис. 4. Узагальнена схема реалізації CI/CD процесу

На рис. 4 використано такі позначення: VCS (англ. Version Control Systems) — система контролю версій; CI server — сервер безперервної інтеграції; APP server (application server) — сервер застосунків системи тестування.

Системи контролю версій VCS значно спрощують роботу з файлами системи тестування та забезпечують відстеження усіх змін. За допомогою VCS можна уникнути втрати даних або виправити помилки, що були допущені в коді або документації.

5. Тестування запропонованої технології

Тестування програмного забезпечення є критичним етапом життєвого циклу розробки ІС та її модулів, що дозволяє виявити недоліки в логіці,

перевірити відповідність функціоналу очікуваним вимогам, а також оцінити стабільність роботи системи за різних умов.

Працездатність запропонованої технології автоматизованого тестування модулів ІС була досліджена на прикладі тестування веб-застосунків (зокрема, інтернет-магазинів). Було розроблено серію тестових сценаріїв, які покривали ключові компоненти клієнтської та серверної частин системи естування. Тестування охоплювало як окремі елементи інтерфейсу, так і повноцінні ланцюги дій, починаючи з авторизації користувача та завершуючи генерацією звітів і аналізом виявлених помилок.

Якість результатів роботи запропонованої технології оцінювалась на основі стандартних метрик в галузі розробки веб-застосунків, з урахуванням особливостей функціонування тестованих модулів. Основна увага приділялася оцінці функціональної коректності, стабільності роботи, швидкодії, безпеки та зручності використання.

Було здійснено порівняння результатів ручного тестування та тестування веб-додатків типу «інтернет-магазин» (<https://tochkazp.com.ua>), наведених в роботі [14], з результатами тестування веб-сайтів з використанням запропонованого тестового фреймворку Sitetest, засобів Sitecore CMS та генерації тестових сценаріїв. Визначено, що використання запропонованої технології дозволяє скоротити витрати часу на тестування веб-додатку в середньому в 10 разів в порівнянні з ручним тестуванням та приблизно на 20% в порівнянні з комбінованим використанням генерації тестових сценаріїв і засобів фреймворку Selenium.

Були також змодельовані випадкові спроби завантажити дані для тестування без авторизації (у цих випадках система повертала статус 403 із повідомленням про заборону доступу).

Застосована стратегія тестування дозволила виявити та усунути ряд недоліків на ранніх етапах розробки, а також сформувані уявлення про реальну стабільність і надійність функціонування тестованих модулів.

Висновки

В статті наведено результати розроблення та дослідження технології автоматизованого тестування програмного забезпечення модулів інформаційних систем.

Досліджено різні типи сучасних технології автоматизації тестування, зокрема, тестові фреймворки, що найбільш поширено використовуються для оптимізації процесів тестування веб-застосунків ІС.

До найбільш важливих завдань реалізації запропонованої технології слід віднести: розроблення схеми генерації тестових сценаріїв для автоматичного

тестування ПЗ модулів ІС (на прикладі веб-сайтів визначеної предметної області); розроблення технології оптимізації процесів тестування веб-сайтів з використанням запропонованого тестового фреймворку Sitetest та засобів Sitecore.

Наведено схему побудови та використання шаблонів автоматизованого тестування веб-застосунків ІС (для конкретної предметної області).

Розглянуто можливість оптимізації процесів тестування веб-сайтів з використанням спеціалізованого тестового фреймворку та інструментів системи керування контентом Sitecore CMS.

Архітектура запропонованого тестового фреймворку Sitetest містить такі основні шари:

- ядро (Framework Core), що задає конфігурацію веб-драйвера та сукупність конфігураційні класів;
- шар SpecFlow Tests, що шар містить feature-файли та імплементацію кроків для них;
- тестову модель, що містить модуль сторінок, що допомагає інкапсулювати роботу з окремими елементами сторінки, дозволяє зменшити розмір коду та полегшити його підтримку, та сервісний компонент (Page Service), що містить класи, які відповідають за перевірки і взаємодію з елементами сторінок;
- шар Sitecore, що відповідає за взаємодію з Sitecore API та використовується для тестування правильності функціонування додатку зі сторони контент-менеджера.

Наведено докладний опис технології оптимізації процесів тестування веб-сайтів з використанням запропонованого фреймворку.

Працездатність запропонованої технології автоматизованого тестування модулів ІС була досліджена на прикладі тестування веб-застосунків (зокрема, інтернет-магазинів). Визначено, що використання цієї технології дозволяє скоротити витрати часу на тестування веб-додатку в середньому в 10 разів в порівнянні з ручним тестуванням та приблизно на 20% в порівнянні з комбінованим використанням генерації тестових сценаріїв і засобів фреймворку Selenium.

Список літератури

- [1] Проектування комп'ютерно-інтегрованих систем: монографія / І. Б. Албанський, Н. Я. Возна, П. В. Гуменний, А. Я. Давлетова, О. М. Заставний — Тернопіль: Університетська думка, 2023. — 494 с.
- [2] Поняття, структура та різновиди веб-сайтів. Автоматизоване розроблення веб-сайтів [Електронний ресурс]. — Режим доступу: <http://www.ndu.edu.ua/liceum/web.pdf>
- [3] Monier, M., & El-mahdy, M. M. (2015). Evaluation of automated web testing tools. *International Journal of Computer Applications Technology and Research*, 4.5, 405–408.
- [4] Sampath, S., & Sprenkle, S. (2016). Advances in web application testing, 2010–2014. *Advances in Computers*, 101, 155–191.

- [5] Garousi, V., Mesbah, A., Betin-Can, A., & Mirshokraie, S. (2013). A systematic mapping study of web application testing. *Information and Software Technology*, 55.8, 1374–1396.
- [6] Top 10 Automation Testing Tools in 2021 [Електронний ресурс] – Режим доступу до ресурсу: <https://www.netsolutions.com/insights/top-10automation-testing-tools>
- [7] Модульне тестування [Електронний ресурс]. – 2022– Режим доступу: 208290/mod_resource/content/Е2%84%964.pdf.
- [8] Єгорова О., Бичок В. Програмні засоби для тестування програмного забезпечення. *Молодий вчений*. 2019. № 11(75). С. 680–684.
- [9] Automation of software testing [Електронний ресурс] – Режим доступу: <https://www.sciencedirect.com/science/article/pii/S18770509>.
- [10] Satheesh, A., & Singh, M. (2017) Comparative study of open source automated web testing tools: Selenium and Sahi. *Indian Journal of Science and Technology*, 10(13), 1-9.
- [11] Kumar, Y. (2015). Comparative study of automated testing tools: Selenium, SoapUI, HP Unified Functional Testing and Test Complete. *Journal of Emerging Technologies and Innovative Research*, 2.9, 42–48
- [12] Sitecore content authoring. [Електронний ресурс]. – 2019 – Режим доступу: <https://doc.sitecore.com/users/92/sitecore-experience-platform/en/content-authoring.html/>
- [13] Test Deliverables in Software Testing – Detailed Explanation [Electronic resource] - Resource access mode: <https://www.softwaretestingmaterial.com/test-deliverables/>
- [14] Удовенко С.Г. Використання шаблонів автоматичного тестування в проектах з розробки веб-додатків / С.Г. Удовенко, Н.О. Міронова, Т.Д. Федорончак, К.К. Верещак // Системи управління, навігації та зв'язку. – 2017. – Вип. 5(45) – С. 111 – 118.
- [15] Тер-Карапетянц Т.С. Автоматизація процесу тестування веб-сайтів на платформі Sitecore / Т.С. Тер-Карапетянц, С.Г. Удовенко // Матеріали Міжнародної науково-практичної конференції "Інформаційні технології та системи": тези доповідей, 10-11 квітня 2020 Р. – Х.:ХНЕУ ім. С. Кузнеця, 2020. - С. 29
- [16] Rafique, N., Rashid, N., Awan, S., & Nayyar, Z. (2014). Model based testing in web applications. *International Journal of Scientific Engineering and Research*, 2.1, 56–60.

Надійшла до редколегії 15.04.2025

Л.М. Нуралієва¹, О.С. Назаров², Н.В. Назарова³¹ ХНУРЕ, м. Харків, Україна, liliia.nuraliieva@nure.ua² ХНУРЕ, м. Харків, Україна, oleksii.nazarov1@nure.ua, ORCID iD: 0000-0001-8682-5000³ ХНУРЕ, м. Харків, Україна, nataliia.nazarova@nure.ua, ORCID iD: 0009-0007-7816-7088

ДОСЛІДЖЕННЯ МЕТОДІВ ПРОГРАМНОЇ ОПТИМІЗАЦІЇ ЗАВАНТАЖЕННЯ САЙТІВ НА ФОРМУВАННЯ ДОСВІДУ КОРИСТУВАЧІВ

Стаття присвячена аналізу методів програмної оптимізації завантаження вебсайтів з метою покращення досвіду користувачів. У рамках дослідження було проаналізовано сучасні технології, спрямовані на зменшення часу завантаження сторінки, серед яких — відкладене завантаження ресурсів (lazy loading), кешування, оптимізація запитів до бази даних, а також мініфікація та стиснення статичних файлів. Увага приділялася не лише технічним аспектам підвищення продуктивності, а й принципам інклюзивності вебсайтів та формуванню доступності вебконтенту для людей з особливими потребами. Оцінювання ефективності впроваджених рішень здійснювалося за допомогою показників системи Web Vitals та інструменту Lighthouse, що дозволило комплексно оцінити як технічну продуктивність, так і користувацький досвід. Результати дослідження демонструють, що оптимізація швидкого завантаження без врахування критеріїв інклюзивного вебдизайну може призвести до втрати частини аудиторії, тоді як інтеграція практик доступності сприяє покращенню взаємодії всіх користувачів із сайтом, підвищенню SEO-показників та загальній ефективності цифрового продукту.

WEB VITALS, WEB ACCESSIBILITY, ОПТИМІЗАЦІЯ ЗАВАНТАЖЕННЯ, КЕШУВАННЯ, LAZY LOADING, LIGHTHOUSE, UX, ВЕБРОЗРОБКА, ІНКЛЮЗИВНІСТЬ ВЕБСАЙТІВ

Liliia Nuraliieva, Oleksii Nazarov, Nataliia Nazarova. Research on software optimization methods for website loading and their impact on user experience. The article is devoted to the analysis of software optimization methods for website loading with the aim of improving user experience. The study examined modern technologies aimed at reducing page load time, including lazy loading of resources, caching, database query optimization, as well as minification and compression of static files. Attention was given not only to the technical aspects of performance improvement but also to the principles of website accessibility for people with special needs. The effectiveness of the implemented solutions was evaluated using Web Vitals metrics and the Lighthouse tool, which allowed for a comprehensive assessment of both technical performance and user experience. The results of the research demonstrate that optimizing performance without considering inclusive web design criteria can lead to the loss of part of the audience, whereas integrating accessibility practices contributes to enhancing the interaction of all users with the website, improving SEO indicators, and increasing the overall effectiveness of the digital product.

WEB VITALS, WEB ACCESSIBILITY, LOAD OPTIMIZATION, CACHING, LAZY LOADING, LIGHTHOUSE, UX, WEB DEVELOPMENT, WEBSITE INCLUSIVITY.

Вступ

На сьогоднішній день швидкість завантаження виступає одним з найголовніших показників, який визначає ступінь задоволення користування вебсайтом та детермінує популярність веб-ресурсу. Зі зростанням кількості онлайн-сервісів та загально-го обсягу вебконтенту, користувачі стають усе менш лояльними до затримок і очікують майже миттєвого реагування з боку системи. Рекомендація утримувати затримки на рівні менше ніж 2 секунди була широко цитована і залишалася «золотим стандартом» вебдизайну аж до 1990-х років [1]. З того часу технології значно еволюціонували, однак загальна ідея збереглася — сайт повинен реагувати швидко, стабільно та без збоїв. Питання оптимізації продуктивності вебресурсів постає особливо гостро з огляду на сучасні вимоги до UX, SEO та зручності використання мобільних пристроїв.

У відповідь на ці виклики компанія Google представила концепцію Web Vitals — набір ключових метрик, що дозволяють розробникам вимірювати ефективність взаємодії користувачів із вебресурсами.

Web Vitals фокусуються на трьох основних аспектах: швидкості завантаження головного контенту, часу до першої взаємодії та візуальної стабільності сторінки. Ці показники дають змогу визначити, наскільки швидко та плавно користувач отримує доступ до інформації, і наскільки комфортно є навігація в момент завантаження сторінки. Таким чином, Web Vitals стали універсальним інструментом оцінки реального користувацького досвіду.

Метрики Web Vitals широко інтегровані в сучасні інструменти розробки та аналізу, зокрема Lighthouse, PageSpeed Insights і Chrome DevTools. Вони дозволяють не лише виявити непомітні на перший погляд проблеми, що стосуються завантаження сторінки, а й отримати персоналізовані рекомендації щодо покращення її продуктивності. Завдяки цим інструментам оптимізація стала більш прозорою, вимірюваною та орієнтованою саме на користувача.

Важливим є і те, що Web Vitals напряму впливають на пошукову оптимізацію: Google офіційно включив ці показники до факторів ранжування в результатах пошуку. Це означає, що технічна продуктивність

сайту стала не лише питанням зручності для користувача, а й елементом цифрової конкурентоспроможності на рівні індексації та видимості ресурсу.

Метою даної роботи є аналіз програмних методів оптимізації завантаження вебсайтів з урахуванням метрик Web Vitals як ключових показників ефективності. У центрі уваги — визначення найрезультативніших підходів до підвищення швидкості завантаження, зниження затримок, оптимізації обробки контенту та забезпечення стабільного інтерфейсу користувача. Особлива увага також приділяється врахуванню принципів інклюзивного дизайну, що дозволяє створювати вебсайти, які не лише швидко завантажуються, а й є доступними для користувачів з різними фізичними можливостями.

1. Опис предметної області

Об'єктом даного дослідження є методи програмної оптимізації завантаження вебсайтів, зокрема сучасні технології, що сприяють зменшенню часу завантаження сторінок, таєі як відкладене завантаження ресурсів (lazy loading), кешування, оптимізація запитів до бази даних, а також мініфікація та стиснення статичних файлів.

Предметом дослідження є вплив цих методів на формування досвіду користувачів під час взаємодії з вебсайтами, зокрема з огляду на технічну продуктивність та інклюзивність вебресурсів. Це включає формування доступності вебконтенту для людей з особливими потребами, а також забезпечення стабільної та швидкої взаємодії з користувачем.

Основним завданням роботи є дослідження та оцінка програмних рішень, спрямованих на покращення швидкості завантаження вебсторінок і підвищення якості користувацького досвіду з урахуванням не тільки продуктивності, а й інклюзивного вебдизайну. Оцінка ефективності впроваджених рішень проводилась за допомогою показників Web Vitals і інструменту Lighthouse, що дозволило комплексно оцінити як технічну продуктивність, так і загальний вплив на користувацький досвід.

2. Аналіз задач та методів програмної оптимізації

Проаналізуємо основні методи оптимізації вебресурсів та розглянемо їхній вплив на користувацький досвід. Метою такого аналізу є визначення найбільш ефективних підходів до забезпечення стабільності інтерфейсу та зменшення навантаження на серверну частину додатку. На сучасному етапі розвитку вебтехнологій особливу увагу приділяють не лише технічним характеристикам роботи вебдодатків, а й досвіду, який користувач отримує під час взаємодії з веб-сайтом. У цьому контексті Google запропонував комплексну систему метрик під назвою Web Vitals, яка дозволяє об'єктивно вимірювати продуктивність сайту з погляду користувача [2].

Основними індикаторами цієї системи є Largest Contentful Paint (LCP), який відповідає за час відмальовування найбільшого елемента на сторінці. First Input Delay (FID) — метрика, яка показує, наскільки швидко сайт реагує на першу взаємодію користувача; а також Cumulative Layout Shift (CLS) — показник, що характеризує стабільність візуального макету сторінки. Усі ці показники є фактичними індикаторами зручності та ефективності вебресурсу [3, 4, 5].

Низькі показники Web Vitals свідчать про поганий користувацький досвід, навіть якщо сам сайт функціонує коректно. Наприклад, якщо контент «стрибає» при завантаженні або кнопки реагують із затримкою, це створює відчуття нестабільності й викликає роздратування в користувачів. Тому оптимізація саме під ці метрики стала пріоритетом для сучасної frontend-розробки, а їх інтеграція в систему оцінки SEO підтверджує важливість врахування показників UX при створенні будь-якого вебпродукту.

У цьому контексті важливу роль відіграють такі методи, як відкладене (lazy) завантаження ресурсів, асинхронна обробка JavaScript, попереднє кешування даних, а також мініфікація CSS, JS та зображень. Наприклад, lazy loading зображень дозволяє значно зменшити початкове навантаження на сторінку, завантажуючи лише ті елементи, які дійсно необхідні в межах поточного перегляду. Це не тільки зменшує обсяг переданих даних, але й покращує LCP-метрику, оскільки найважливіший контент завантажується швидше.

Окрім клієнтської частини, значна увага приділяється серверним процесам. Сюди входить використання серверного кешування, зменшення затримок відповіді, оптимізація запитів до бази даних та балансування навантаження між серверами. Наприклад, застосування індексів у базах даних та кешування часто використовуваних результатів дозволяє скоротити час відповіді на запити, тим самим покращуючи загальну продуктивність системи.

Кешування на стороні клієнта також є вагомим чинником оптимізації. Збереження інформації у браузерних сховищах, таких як localStorage або sessionStorage, дозволяє уникати повторних запитів до сервера та пришвидшити повторне завантаження контенту. Це покращує відчуття миттєвої реакції інтерфейсу, особливо в односторінкових застосунках або при повторному використанні одних і тих самих даних у межах кількох сторінок.

Нарешті, важливим є інтеграційний підхід, коли усі методи оптимізації розглядаються в комплексі. Висока продуктивність досягається завдяки поєднанню ефективного керування ресурсами, коректного розподілу навантаження та постійного моніторингу основних метрик. Такий підхід дозволяє підтримувати високу якість роботи сайту навіть за умов активного росту аудиторії та зростаючих обсягів контенту.

Також треба пам'ятати і про забезпечення інклюзивного інтерфейсу, який відповідає за створенні рівних умов для взаємодії всіх категорій користувачів. Розробка веб-сайтів з урахуванням потреб людей з вадами сприяє розширенню цільової аудиторії та покращенню користувацького досвіду, що впливає на загальне задоволення використання ресурсу та неабияк підвищує його популярність.

3. Методи оптимізації завантаження сайту з урахуванням інклюзивності та доступності

Оптимізація швидкості завантаження вебсайтів є критично важливим аспектом сучасної веб-розробки, оскільки користувачі часто очікують майже миттєвого завантаження сторінок. Низька продуктивність сайту, що негативно впливає на користувацький досвід, призводить до неефективного ранжування у пошукових системах. Тому застосування ефективних методів оптимізації є ключовим для забезпечення високої продуктивності.

Проте не менш важливим є врахування інклюзивності вебсайтів, що забезпечує доступність для всіх користувачів, незалежно від їхніх фізичних можливостей. Вебсайти, що забезпечують швидке завантаження та враховують потреби користувачів з фізичними вадами, здатні суттєво підвищити якість взаємодії з аудиторією та охопити ширше коло відвідувачів. Інклюзивний дизайн передбачає правильне використання контрасту кольорів, текстових читачів, а також можливість навігації за допомогою клавіатури або альтернативних пристроїв вводу. Врахування цих аспектів у процесі оптимізації дозволяє забезпечити рівний доступ до інформації і функціоналу сайту для всіх користувачів, що, в свою чергу, підвищує загальну ефективність вебресурсу та сприяє покращенню репутації бренду.

1. Lazy Loading (відкладене завантаження)

Один з найефективніших методів оптимізації – це lazy loading. Цей метод полягає у завантаженні ресурсів (зображень, відео, скриптів) тільки тоді, коли вони стають видимими для користувача на екрані. Це дозволяє значно скоротити час завантаження сторінки, особливо коли на сторінці міститься велика кількість медіафайлів або важких елементів.

Застосування технології відкладеного завантаження (lazy loading) дає змогу зменшити кількість ресурсів, що підтягуються з серверу на етапі початкового рендерингу сторінки, що в свою чергу позитивно позначається на метриці Largest Contentful Paint (LCP) – часі, необхідному для відмальовування найбільшого візуального елемента на сторінці. Як правило, відкладене завантаження застосовується до зображень, відео та інших медіафайлів, що зменшує початкове навантаження і дозволяє ресурсу швидше відобразитися на екрані.

2. Асинхронне завантаження ресурсів (Async/Defer)

Ще одним ефективним методом оптимізації є асинхронне завантаження ресурсів, таких як JavaScript та CSS. За допомогою атрибутів async та defer можна змінити поведінку завантаження скриптів на вебсторінці, що дозволяє значно скоротити час відгуку сторінки.

Async – атрибут, який дозволяє завантажувати та виконувати скрипт без блокування рендерингу сторінки. Скрипти з атрибутом async завантажуються паралельно з іншими ресурсами та виконуються, як тільки вони будуть завантажені, не чекаючи на завершення завантаження інших елементів.

Defer – атрибут дозволяє відкласти виконання скрипта до моменту, поки весь HTML не буде завантажено і розпарсено. Це дозволяє зменшити час блокування рендерингу сторінки [6].

Це безпосередньо позначається на метриці First Input Delay (FID), яка відображає час між першою взаємодією користувача та моментом, коли браузер починає обробляти цю дію. Чим менше часу браузер витрачає на блокування скриптів, тим швидше відбувається перша взаємодія.

3. Мініфікація та компресія ресурсів

Мініфікація та компресія – ще одні важливі методи, що дозволяють значно зменшити час завантаження сторінки. Мініфікація передбачає видалення зайвих пробілів, коментарів та інші оптимізації коду, що дозволяє зменшити розмір файлів CSS, JavaScript та HTML.

Компресія файлів, зокрема, зображень та інших медіафайлів, є важливим кроком у зменшенні обсягу трафіку та прискоренні завантаження сайту. Використання алгоритмів стиснення дозволяє зменшити розмір файлів без втрати їх якості. У поєднанні з оптимізацією зображень можна досягти значного покращення часу завантаження.

4. Кешування на сервері та клієнті

Кешування – це одна з основних технік для підвищення швидкості завантаження сторінки. Воно дозволяє зберігати дані або ресурси, які не змінюються часто, в пам'яті або на диску, що дозволяє повторно використовувати їх при наступних запитах без необхідності завантажувати їх знову.

Кешування на сервері дозволяє зберігати частини даних, що використовуються на сторінці (наприклад, результати запитів до бази даних), в пам'яті серверу. Це забезпечує швидкий доступ до цих даних при повторних запитах користувачів, зменшуючи навантаження на сервер і час відповіді.

Кешування на клієнтській частині зберігає файли, такі як CSS, JS та зображення, у браузері користувача. Це дозволяє уникнути повторних запитів до сервера при навігації по сайту або при повторному візиті сторінки, що прискорює завантаження.

5. Оптимізація запитів до бази даних

Навантаження на сервер та базу даних часто стають проблемою при роботі з великими вебсайтами. Для покращення часу відповіді на запити необхідно правильно оптимізувати запити. Основні стратегії включають:

- Використання індексів для полів, які часто запитуються. Це дозволяє значно прискорити пошук даних у таблицях.

- Кешування результатів запитів на сервері або клієнті, що дозволяє уникнути повторних запитів до бази даних.

- Нормалізація та денормалізація даних, що дозволяє зменшити складність запитів та підвищити швидкість доступу до необхідних даних.

Ці методи допомагають скоротити час, необхідний для обробки запитів та покращують час завантаження сайту.

6. Використання сучасних форматів зображень

Існують нові формати зображень, які дозволяють досягти кращого балансу між якістю та розміром файлів. Формат WebP є однією з таких інновацій, який підтримує високу якість при значно меншому розмірі файлу порівняно з традиційними форматами (JPEG, PNG) [7]. Це дозволяє значно зменшити обсяг трафіку і прискорити завантаження сайту.

7. Забезпечення доступності для всіх користувачів

Інклюзивність в контексті вебдизайну означає створення сайтів, доступних для всіх користувачів, незалежно від їхніх можливостей чи обмежень. Це важлива складова сучасної веброзробки, оскільки вона включає в себе не лише покращення функціональності та дизайну для людей з особливими потребами, але й забезпечує рівний доступ до інформації для кожного користувача.

Інклюзивність передбачає адаптацію сайту до різних пристроїв і допоміжних технологій, зокрема програм для зчитування інформації з екрана, що використовуються людьми з порушеннями зору, альтернативні текстові описи для зображень, а також правильне використання кольорів та контрастів для користувачів з порушеннями кольоросприйняття. Інтеграція таких практик допомагає підвищити загальний досвід взаємодії з вебсайтом.

Починаючи з 1997 року, Ініціатива з вебінклюзивності (WAI) Консорціуму (W3C) відіграє ключову роль у підвищенні обізнаності щодо важливості доступності вебресурсів, а також у розробці ефективної моделі, що сприяє впровадженню доступних вебтехнологій в діяльність організацій [8].

4. Аналіз застосованих методів програмної оптимізації на вебсайтах

Розглянемо декілька популярних вебсайтів на предмет використання практик методів програмної оптимізації та врахування інклюзивності.

Rozetka.ua – один з найбільших онлайн-магазинів в Україні, на якому можна знайти безліч різноманітних товарів. Як можна побачити на рис. 1, Rozetka.ua демонструє відмінні результати за продуктивністю, з дуже швидким відображенням контенту, що забезпечує плавний і швидкий досвід для користувачів.

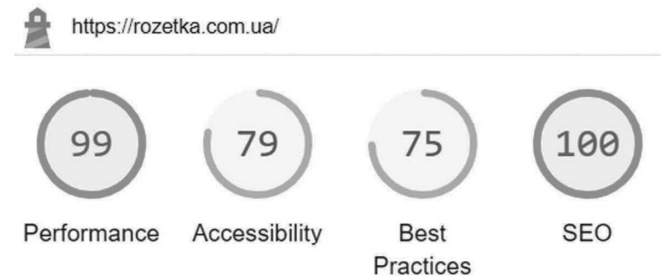


Рис. 1. Загальні метрики на сайті Rozetka.ua

За метриками на рис. 2 бачимо, що сайт Rozetka.ua демонструє відмінну продуктивність, зокрема завдяки дуже швидким показникам First Contentful Paint (0.3 с) та Largest Contentful Paint (0.5 с), що свідчить про те, що основний контент завантажується майже миттєво. Загальний час блокування (Total Blocking Time) складає всього 80 мс, що забезпечує плавну взаємодію з користувачем. Показник Cumulative Layout Shift (CLS) дорівнює 0, що гарантує стабільність макету сторінки під час завантаження та запобігає різкій зміні розташування елементів інтерфейсу. Speed Index, що становить 1.1 с, свідчить про дуже швидке завантаження сторінки, забезпечуючи безперервний і плавний досвід для користувачів.

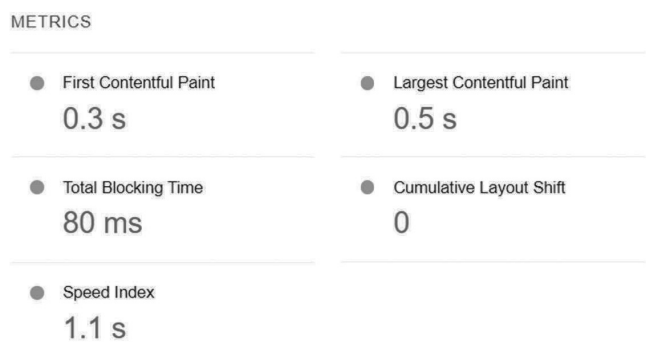


Рис. 2. Показники метрик на сайті Rozetka.ua

Тим не менш, що стосується вебдоступності та дотримання кращих практик, на сайті Rozetka.ua є певні нюанси, які потребують уваги (рис. 3). Наприклад, деякі кнопки не мають доступних імен, що ускладнює взаємодію для користувачів з обмеженими можливостями, зокрема тих, хто використовує екранні зчитувачі. Це є можливістю для покращення семантики елементів навігації. Крім того, наявність атрибуту `user-scalable="no"` в `<meta name="viewport">` та встановлення `maximum-scale` менш ніж на 5 можуть обмежувати можливість масштабування контенту,

що є важливим для користувачів з порушеннями зору. Порушення достатнього контрасту між фоном і текстом також потребує уваги, оскільки це може ускладнити читання для людей з ослабленим зором. Особливий фокус має бути спрямований на покращення доступності для людей з обмеженими можливостями через усунення цих проблем, таких як надання доступних імен для кнопок, належне використання ARIA-ролей для нестандартних елементів.

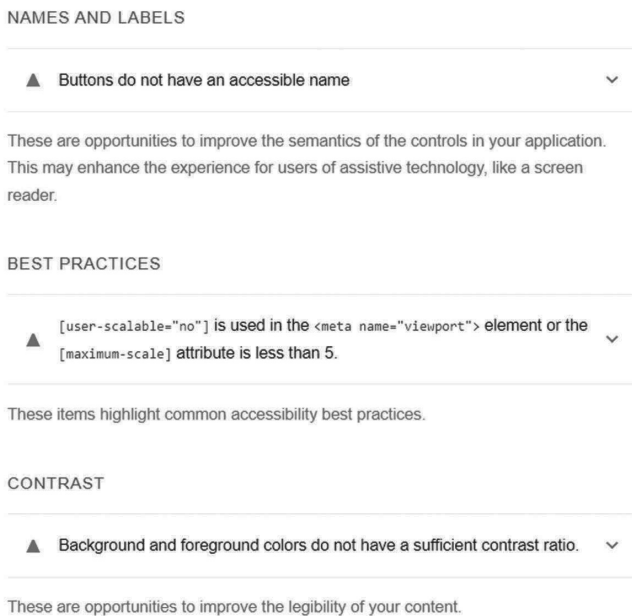


Рис. 3. Метрики вебдоступності на сайті Rozetka.ua

Інший популярний сайт, який часто використовується для того, щоб дізнатися погоду у різних регіонах є Sinoptik.ua. В даному випадку метрики показують значно кращі результати, зокрема в аспекті інклюзивності та загальної продуктивності (рис. 4).



Рис. 4. Загальні метрики на сайті Sinoptik.ua

Performance досягнув максимальної оцінки, що свідчить про надзвичайно швидке завантаження та оптимізацію контенту. SEO також отримало ідеальний результа, що означає, що сайт добре оптимізований для пошукових систем і має високі шанси на хорошу видимість в результатах пошуку.

З точки зору accessibility, сайт отримав оцінку 90/100, що вказує на те, що він надає користувачам з обмеженими можливостями високу доступність. Більшість основних аспектів доступності вже налаштовані належним чином, однак є проблеми з контрастом між фоном і текстом, що може ускладнити читання для людей з ослабленим зором. Покращення контрасту значно підвищить доступність сайту для цієї категорії користувачів (рис. 5).

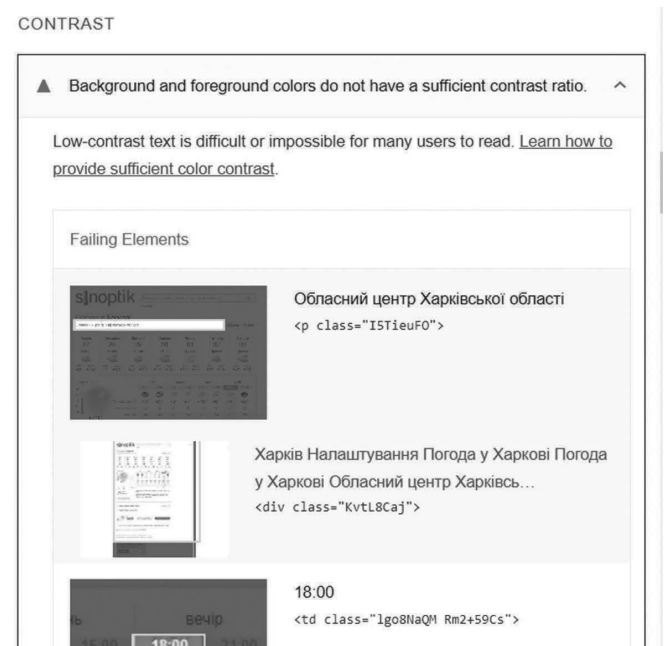


Рис. 5. Метрики вебдоступності на сайті Sinoptik.ua

Що стосується Best Practices, оцінка 83/100 свідчить про те, що сайт добре використовує сучасні підходи, хоча є місце для оптимізації. Загалом, сайт демонструє високий рівень виконання практик доступності та ефективної продуктивності і більше орієнтований на людей з вадами.

5. Роль інклюзивного вебдизайну в сучасному світі

Інклюзивний вебдизайн є критично важливою складовою сучасного цифрового середовища, оскільки він забезпечує доступність і зручність використання вебресурсів для всіх категорій користувачів. Люди з вадами можуть зіштовхуватися з різними незручностями при взаємодії з вебресурсами, оскільки при розробці показник вебдоступності не враховувався або ж йому приділялось мало уваги. Вебсайти, які підтримують інклюзивність, дозволяють залучати ширшу аудиторію, сприяючи підвищенню лояльності користувачів та покращенню загальної репутації

бренду. Окрім цього, застосування практик доступності позитивно впливає на SEO.

ARIA Authoring Practices Guide (APG) – специфікація для створення вебсайтів, які відповідають правилам інклюзивності. Вона описує, як застосовувати семантику доступності до поширених шаблонів дизайну та віджетів, надаючи дизайнерські шаблони та функціональні приклади [9]. Використання даної специфікації дозволяє забезпечити коректну взаємодію користувачів з обмеженими можливостями з інтерфейсами, а також полегшити навігацію за допомогою таких технологій як екранні зчитувачі.

6. Інструменти для оцінки вебдоступності

Існує низка інструментів для оцінки вебдоступності [10]. Це спеціалізовані програми або онлайн-сервіси, які допомагають розробникам та дизайнерам перевіряти відповідність вебсайтів стандартам інклюзивності. Вони дозволяють виявляти потенційні проблеми, які можуть ускладнити взаємодію з вебресурсами для людей з обмеженими можливостями, таких як слабкий зір, слухові порушення або обмеження в русі.

Основною метою цих інструментів є полегшення процесу тестування вебсайтів на відповідність вимогам WCAG (Web Content Accessibility Guidelines), ARIA (Accessible Rich Internet Applications) та іншим стандартам. Ці інструменти можуть автоматично перевіряти такі аспекти, як доступність контенту для екранних зчитувачів, правильне використання семантики HTML, контрастність кольорів, навігацію за допомогою клавіатури та інші важливі параметри. Окрім автоматичних перевірок, багато інструментів надають рекомендації щодо покращення доступності та пропонують розробникам практичні поради для досягнення більш інклюзивного дизайну.

Інструменти для оцінки вебдоступності не лише допомагають забезпечити відповідність стандартам, але й сприяють загальному покращенню користувацького досвіду, оскільки роблять вебресурси доступнішими для ширшої аудиторії.

Висновки

Підвищення продуктивності вебсайтів через оптимізацію швидкості завантаження є необхідною умовою для забезпечення комфортного користувацького досвіду в сучасному цифровому середовищі. Застосування таких методів, як кешування, відкладене завантаження ресурсів, оптимізація запитів до бази даних та мініфікація файлів, дозволяє значно знизити час завантаження та підвищити ефективність роботи сайтів.

Водночас, важливо пам'ятати про інклюзивність і доступність для всіх категорій користувачів. Цифрові ресурси, які враховують потреби осіб з вадами, мають потенціал до залучення більшої кількості аудиторії.

Для досягнення оптимальної доступності важливо використовувати інструменти для оцінки вебдоступності, такі як Lighthouse, який допомагає виявляти й виправляти проблеми з контрастністю, семантикою елементів та іншими критичними аспектами. Крім того, керівництва на кшталт ARIA Authoring Practices Guide надають необхідні знання та приклади для створення доступних інтерфейсів. Використання таких інструментів дозволяє покращити взаємодію користувачів з обмеженими можливостями з вебресурсами, що в результаті підвищує ефективність вебсайту.

Таким чином, інтеграція методів оптимізації продуктивності та інклюзивного дизайну є ключовим фактором для створення успішного вебресурсу, який забезпечує не лише швидкий доступ до інформації, а й рівні можливості для всіх користувачів.

Конфлікт інтересів

Автори заявляють, що не мають конфлікту інтересів.

Список джерел

- [1] Nielsen, Jakob. "User Interface Directions for the Web." Communications
- [2] Web Vitals. URL: <https://web.dev/articles/vitals?hl=en>
- [3] Largest Contentful Paint. URL: <https://developer.mozilla.org/en-US/docs/Web/API/LargestContentfulPaint>
- [4] First Input Delay. URL: https://developer.mozilla.org/en-US/docs/Glossary/First_input_delay
- [5] Cumulative Layout Shift. URL: <https://developer.mozilla.org/en-US/docs/Glossary/CLS>
- [6] Асинхронне завантаження скриптів. URL: <https://developer.mozilla.org/en-US/docs/Web/HTML/Reference/Elements/script>
- [7] WebP. URL: <https://developer.mozilla.org/en-US/docs/Glossary/WebP>
- [8] Kelly, Brian & Sloan, David & Brown, Stephen & Seale, Jane & Smith, Stuart & Lauke, Patrick & Ball, Simon. (2009). Accessibility 2.0: Next Steps For Web Accessibility. Journal of Access Services. 6. pp. 265-294.
- [9] Accessibility information for web authors. URL: https://developer.mozilla.org/en-US/docs/Web/Accessibility/Guides/Information_for_Web_authors
- [10] Web Accessibility Evaluation Tools List. URL: <https://www.w3.org/WAI/test-evaluate/tools/list/>
- [11] Rozetka.ua. URL: <https://rozetka.com.ua>
- [12] Sinoptik.ua. URL: <https://sinoptik.ua/pohoda/kharkiv>

Надійшла до редакції 07.05.2025



Н. В. Шаронова¹, Г. А. Плехова², С. М. Неронов³, М. В. Костікова⁴, Д. О. Плехов⁵

¹НТУ «ХПІ», м. Харків, Україна, nvsharonova@ukr.net, ORCID iD: 0009-0004-9878-1761

²ХНАДУ, м. Харків, Україна, plehovaanna1@gmail.com, ORCID iD: 0000-0002-6912-6520

³ХНАДУ, м. Харків, Україна, sernikner@gmail.com, ORCID iD: 0000-0003-2381-1271

⁴ХНАДУ, м. Харків, Україна, kmv_topaz@ukr.net, ORCID iD: 0000-0001-5197-7389

⁵ХНАДУ, м. Харків, Україна, plehov@gmail.com, ORCID iD: 0009-0004-7873-1716

СПОСІБ ІНТЕГРАЦІЇ РІЗНОРІДНИХ ДАНИХ В СИСТЕМІ ГЕОПРОСТОРОВОГО АНАЛІЗУ

У роботі запропонований спосіб інтеграції різномірних даних в системі геопросторового аналізу, який полягає в тому, що система геопросторового аналізу проводить об'єднання даних, що надходять на її вхід з метою отримання більш достовірних і точних даних про місце розташування об'єкта аналізу, його швидкість, атрибутивну (додаткову) інформацію та ідентичність певному класу, виду або типу, після чого проводиться моделювання зв'язків об'єкта аналізу з іншими об'єктами або подіями в контексті його оточуючого середовища, а після чого проводиться формування узагальненої оцінки стану об'єкта аналізу за рахунок відстеження телеметричних та метаданих. Особливістю способу є те, що система геопросторового аналізу на етапі формування узагальненої оцінки стану об'єкта аналізу разом з визначенням умов використовує удосконалену процедуру прогнозування зміни стану об'єкту аналізу, сутність якої полягає в тому, що спочатку система геопросторового аналізу будує авторегресивну модель прогнозування на основі інформаційного критерію Акаїке, а потім системою геопросторового аналізу відбувається порівняння отриманих значень помилки прогнозування з еталонними. Технічний ефект від використання зазначеного способу полягає в підвищенні ефективності застосування системи геопросторового аналізу за рахунок підвищення точності прогнозування зміни стану об'єкту аналізу.

ГЕОПРОСТОРОВИЙ АНАЛІЗ, РІЗНОРІДНІ ДАНІ, СПОСІБ ІНТЕГРАЦІЇ, СТАН ОБ'ЄКТУ АНАЛІЗУ, ПРОГНОЗУВАННЯ

N. V. Sharonova, G. A. Pliekhova, S. M. Neronov, M. V. Kostikova, D. O. Pliekhov. Method of integration of vari-ous data in the geospatial analysis system. The paper proposes a method of integrating disparate data in the geospatial analysis system, which consists in the fact that the geospatial analysis system combines the data received at its input in order to obtain more reliable and accurate data about the location of the object of analysis, its speed, attributive (additional) information and identity to a certain class, species or type, after which the analysis object's connections with other objects or events in the context of its surrounding environment are modeled, and then a generalized assessment of the state of the analysis object is formed by tracking telemetry and metadata. A feature of the method is that the geospatial analysis system at the stage of forming a generalized assessment of the state of the object of analysis together with the definition of conditions uses an improved procedure for forecasting changes in the state of the object of analysis, the essence of which is that the geospatial analysis system first builds an autoregressive forecasting model based on the Akaike information criterion, and then the geospatial analysis system compares the received forecasting error values with the reference values. The technical effect of using the specified method is to increase the efficiency of the geospatial analysis system by increasing the accuracy of forecasting the change in the state of the object of analysis.

GEOSPATIAL ANALYSIS, MISCELLANEOUS DATA, METHOD OF INTEGRATION, STATE OF THE OBJECT OF ANALYSIS, PROGNOSTICATION

Вступ

У сучасному світі геоінформаційні технології (ГІТ) є невід'ємною складовою багатьох галузей, що забезпечують ефективне управління просторовими даними, їх аналіз та візуалізацію. Вони широко застосовуються у сфері державного управління різними рівнями, енергетикою, транспортним сектором, аграрним виробництвом, землеустроєм, будівництвом та раціональним використанням природних ресурсів. Використання ГІТ сприяє оптимізації процесів, підвищує точність прийняття управлінських рішень, що, своєю чергою, забезпечує сталий розвиток галузей економіки та суспільства в цілому.

З кожним роком потреба в ГІТ невпинно зростає, що обумовлено розвитком цифрових технологій,

зростанням обсягів простору даних, глобальними викликами, такими як урбанізація, зміни клімату, екологічні катастрофи, демографічні зміни, а також забезпечення підвищення рівня безпеки в умовах політичної нестабільності. Геоінформаційні системи (ГІС) забезпечують інтеграцію та аналіз великої кількості даних, що дозволяє оперативно реагувати на критичні ситуації, планувати та прогнозувати майбутні зміни в різних сферах діяльності.

Особливі актуальності ГІТ набувають у сфері національної безпеки та оборони, де вони використовують для оперативного аналізу ситуації, картографування загроз, моделювання сценаріїв розвитку подій, моніторингу території, управління військовими операціями та планування оборонних заходів.

Автоматизовані системи геопросторового аналізу позбавляють від комплексної оцінки оперативної обстановки, що є ключовим фактором у процесі прийняття стратегічних та тактичних рішень.

В Україні вже впроваджуються ГІТ у структурах, що забезпечують завдання у сфері національної безпеки та оборони. Таким, вони застосовуються у військових операціях, розвідці, управлінні кризовими ситуаціями, забезпеченні кібербезпеки, прикордонному контролі та боротьбі з терористичними загрозами. Завдяки впровадженню сучасних геоінформаційних розробок вдосконалені процеси управління оборонними ресурсами, створений ефективний моніторинг території, аналізуються ризики та моделюється можливість сценаріїв розвитку подій у реальному масштабі.

У сфері національної безпеки ГІТ також виконують важливу роль у забезпеченні екологічної безпеки, управління природними ресурсами, протидії техногенним катастрофам та стихійним лихам. Геопросторові дані застосовуються для оцінки ризиків, розробки превентивних заходів та координації дій державних органів у разі виникнення надзвичайних ситуацій.

В Україні існує нормативно-правова база, що регулює використання ГІТ та геопросторових даних. Однак, у контексті стрімкого розвитку технологій та змін у безпековому середовищі, необхідним є подальше вдосконалення законодавства, розширення можливостей використання ГІТ на державному рівні, забезпечення інтеграції різних систем та розширення міжнародного співробітництва у сфері геоінформаційних технологій.

Таким чином, геоінформаційні технології відіграють стратегічну роль у багатьох сферах суспільного життя, а їх впровадження у сферу національної безпеки та оборони є вкрай актуальним у сучасних умовах [1-5]. Розвиток та удосконалення ГІТ дозволяють підвищити ефективність управлінських рішень, підвищити рівень ситуаційної обізнаності [6-10].

Виклад основного матеріалу

Відомий спосіб обробки різнорідних даних в системі геопросторового аналізу, сутність якого полягає в послідовному виконанні процедур попередньої обробки інформації, уточнення стану об'єкту, який підлягає аналізу, уточнення оперативної обстановки, уточнення загроз та уточнення процесу обробки різнорідних даних в системі геопросторового аналізу [11].

Недоліком зазначеного способу є відсутність можливості прогнозування стану об'єкту аналізу.

Найбільш близьким до способу, що пропонується, є спосіб прогнозування сигнальної обстановки [12], при якому пристроєм обробки сигналів засобу радіозв'язку вибирається мінімально необхідний

масив вибірових даних, після чого визначається порядок авторегресивної моделі, далі по навчальній частині вибірки методом Берга настроюються коефіцієнти авторегресії, після чого обчислюється значення величини питомої інформаційної неузгодженості та відбувається обчислення дисперсій і величини інформаційної неузгодженості.

Недоліком способу-прототипу є неможливість проводити прогнозування динаміки зміни різнотипних даних, що не дозволяє проводити багатопараметричну оцінку стану об'єкту аналізу.

Тому технічним завданням, що вирішує запропонований спосіб інтеграції різнорідних даних в системі геопросторового аналізу, є поєднання переваг прототипу та аналогів, з усуненням їхніх недоліків.

Спосіб інтеграції різнорідних даних в системі геопросторового аналізу передбачає обробку та синтез інформації, що надходить з різних джерел, з метою підвищення достовірності, точності та інформативності отриманих результатів. Основною метою способу є формування комплексної оцінки стану об'єкта аналізу на основі просторових, часових та атрибутивних характеристик.

Вхідні дані надходять до системи геопросторового аналізу з різнорідних джерел, таких як супутникові знімки, дані від GPS-трекерів, сенсорні системи, бази даних обліку, відомості з картографічних сервісів, результати дистанційного зондування Землі тощо. З огляду на неоднорідність цих даних, проводиться їх попередня обробка, що включає нормалізацію, фільтрацію шумів та корекцію похибок.

Оскільки дані можуть мати різні формати, структуру та рівні точності, система виконує їх гармонізацію, що включає приведення їх до єдиної системи координат, усунення дублювання та синхронізацію часових параметрів. В результаті формується єдиний узгоджений набір даних, який можна використовувати для подальшого аналізу.

На етапі ідентифікації та класифікації об'єкта здійснюється аналіз характеристик об'єкта, включаючи його місце розташування, швидкість руху (у разі динамічного об'єкта), функціональне призначення та додаткові параметри, отримані з атрибутивної інформації. Визначається клас, вид або тип об'єкта відповідно до заданих критеріїв та довідкових даних.

Система геопросторового аналізу виконує просторове моделювання, визначаючи взаємозв'язки об'єкта з іншими об'єктами, подіями чи природними факторами. Це дозволяє оцінити його роль у загальній структурі аналізованої території або системи та прогнозувати можливі зміни або ризики.

Для оцінки поточного стану об'єкта система аналізує сукупність телеметричних показників, метаданих та часових рядів, що дозволяє отримати всебічну характеристику об'єкта в динаміці.

Відмінною особливістю способу є застосування удосконаленої процедури прогнозування змін стану об'єкта аналізу. Спочатку система геопросторового аналізу будує авторегресивну модель прогнозування, використовуючи інформаційний критерій Акаїке (AIC) для вибору оптимальної структури моделі. Це дозволяє мінімізувати похибку прогнозування та забезпечити найбільш адекватне відображення трендів та циклічних змін.

Після побудови моделі система проводить порівняння отриманих значень прогнозних помилок з еталонними показниками, що дає змогу оцінити точність прогнозу та при необхідності скоригувати модель або використати додаткові методи регресії та машинного навчання.

На фінальному етапі проводиться аналіз отриманих прогнозних даних та формуються рекомендації щодо управлінських рішень. Ця інформація може використовуватися для моніторингу природних та техногенних процесів, оптимізації логістичних маршрутів, оцінки екологічного стану територій, управління міською інфраструктурою та інших завдань.

Пояснимо більш докладно сутність заявленого способу.

На етапі проведення збору відомостей від різних джерел.

Вхідними даними (джерелами) для зазначеного способу є зображення, що отримані в різних діапазонах електромагнітного спектру і з різних платформ, геопросторова інформація, тематичні бази даних, апріорна інформація, а також інші допоміжні відомості.

Отримані відомості фільтруються та проводиться їх попередня обробка, до якої належать процедури підвищення відношення сигнал/шум, виявлення та виділення характерних ознак об'єкта аналізу або явища, яке становить інтерес, а також алгоритмів їх просторово-часового покращення. Крім того, під час зазначеного етапу визначається тип даних і пріоритет їх (порядок) подальшої обробки, виконується класифікація і групування даних за певними критеріями та формуються сигнатури об'єкта аналізу.

Під час проведення етапу уточнення об'єкта аналізу виконуються операції приведення даних до єдиної системи координат та виміру, об'єднання (використовуються кореляційні методи), відстеження попереднього та прогнозування наступного місцеположення об'єкта аналізу та його ідентифікація на базі класифікаційних методів. Даний етап характеризується застосуванням чисельних методів обробки даних, до яких можна віднести методи прогнозування та розпізнавання образів. Головною задачею є уточнена оцінка типу або виду об'єкта аналізу, його місця розташування, швидкості, прискорення ракет, кораблів, літаків, живої та рухомої сили противника,

за рахунок поєднання та сумісного використання параметричної, ідентифікаційної та координатної інформації.

Впродовж етапу уточнення об'єкта аналізу виконуються чотири основних функції:

- трансформація даних до єдиної системи координат та виміру;
- деталізація та забезпечення додаткових можливостей щодо дослідження місцеположення об'єкта аналізу, його кінематичних та динамічних характеристик, а також атрибутивних даних;
- визначення даних для проведення статистичних розрахунків;
- проведення класифікації або ідентифікації об'єкта аналізу.

Під час проведення етапу уточнення обстановки проводиться пошук концептуального опису взаємозв'язків між об'єктами та подіями в районі їх розташування. Головною метою даного етапу є оновлення інформації про об'єкти аналізу для визначення цілевказівок військам у випадку проведення військової операції.

Під час проведення етапу формування узагальненої оцінки стану об'єкта аналізу на основі апріорної інформації проводиться прогнозування розвитку стану об'єкта аналізу, надання висновків стосовно поточного стану об'єкта аналізу, середовища функціонування об'єкта аналізу, а також кількості об'єктів, з якими встановлена взаємодія.

Розглянемо детальніше сутність прогнозування стану об'єкта аналізу:

- системою геопросторового аналізу обирається мінімально необхідний масив вибірових даних;
- визначається порядок авторегресивної моделі;
- по навчальній частині вибірки методом Берга настроюються коефіцієнти авторегресії.

Коротко опишемо сутність методу Бергу (метод максимальної ентропії). Його центральним ланцюгом є решітчастий фільтр лінійного передбачення M -го порядку. Динаміка фільтру описується наступним рекурентним виразом

$$\begin{aligned} z_m^f(n) &= z_{m-1}^f(n) + \rho_m z_{m-1}^b(n-1), \\ z_m^b(n) &= z_{m-1}^b(n) + \rho_m z_{m-1}^f(n), \quad m=1, \dots, M, \end{aligned} \quad (1)$$

де ρ_1, \dots, ρ_M – коефіцієнти відбиття решітчастого фільтру; $z_i^f(n), z_i^b(n), i=1, \overline{M}$ – значення помилок лінійного передбачення вперед та назад відповідно. При цьому початкові значення помилок лінійного передбачення (помилки нульового порядку) прирівнюються самому сигналу $z_0^f(n) = z_0^b(n) = x(n)$.

До переваг решітчастого фільтру можна віднести малу чутливість до шуму округлення та їх флуктуаціям (випадковим збуренням) значень коефіцієнтів.

Крім того, помилки передбачення назад на виході кожної ступені взаємно ортогональні. Ідея алгоритму

Берга зводиться до мінімізації арифметичного середнього квадрату помилок лінійного передбачення вперед та назад при кожному значенні порядку m .

$$\sigma_m^2 = \frac{1}{2} \left[\frac{1}{N} \sum_{n=m+1}^N |z_m^f(n)|^2 + \frac{1}{N} \sum_{n=m+1}^N |z_m^b(n)|^2 \right]. \quad (2)$$

Результатом мінімізації є наступний вираз для оцінки коефіцієнтів відбиття по методу найменших квадратів:

$$\rho_m = \frac{-2 \sum_{n=m+1}^N z_{m-1}^f(n) z_{m-1}^b(n-1)}{\sum_{n=m+1}^N |z_{m-1}^f(n)|^2 + \sum_{n=m+1}^N |z_{m-1}^b(n-1)|^2}. \quad (3)$$

Дана оцінка коефіцієнта відбиття представляє собою гармонічне середнє коефіцієнтів частотної кореляції помилок передбачення вперед та назад.

Після визначення коефіцієнтів відбиття використовуємо алгоритм Левінсона, що зв'язує коефіцієнти авторегресії порядку m з коефіцієнтами авторегресії порядку $m-1$:

$$a_m(k) = a_{m-1}(k) + \rho_m a_{m-1}(m-k), \quad 1 \leq k \leq m-1,$$

причому, по властивостям коефіцієнтів авторегресії нульовий коефіцієнт завжди рівний одиниці. Як висновок, можемо записати формулу узгодження дисперсії вхідного сигналу $X(t)$ та породжуючого шуму $\eta(t)$

$$\sigma_\eta^2 = \prod_{m=1}^M (1 - \rho_m^2) \sigma_x^2. \quad (4)$$

Сукупність виразів зазначених виразів складають метод Берга.

— обчислюється значення величини питомої інформаційної неузгодженості та відбувається обчислення дисперсій і величини інформаційної неузгодженості.

Впродовж зазначеного етапу проводиться всебічний аналіз оперативної обстановки, розглядається та оцінюється рівень загроз, які можуть вплинути на успішність виконання поставлених завдань. Це включає виявлення потенційних загроз з боку противника, оцінку можливих ризиків для власних сил, а також визначення заходів із їхнього попередження. Одночасно з цим здійснюється розробка превентивних заходів, що сприяють зниженню ризиків і мінімізації можливих втрат.

Ключовим завданням цього етапу є своєчасне виявлення та ідентифікація загроз, що можуть вплинути на результат проведення військової операції. Це передбачає оцінку противника, його можливих намірів, аналіз слабких місць у власних силах та розробку стратегії протидії. Важливим аспектом є визначення пріоритетних цілей для засобів вогневого ураження, що забезпечує ефективне використання наявних ресурсів для нейтралізації загроз у критичних точках операційного простору.

Наступним важливим кроком є етап уточнення процесу обробки, який спрямований на підвищення ефективності збору та аналізу інформації. Він передбачає безперервне спостереження за продуктивністю системи в реальному масштабі часу, що дає змогу оперативно реагувати на зміни в бойовій обстановці. У рамках цього процесу здійснюється перерозподіл джерел відомостей для забезпечення своєчасного надходження актуальної інформації, необхідної для ухвалення рішень.

Даний етап має на меті:

- оптимізацію процесів обробки даних шляхом адаптації алгоритмів аналізу до поточної ситуації;
- зниження затримок у передачі критичної інформації та її достовірну оцінку;
- удосконалення методів обміну розвідувальними даними між підрозділами та інтеграцію різних систем спостереження;
- визначення можливих точок перегрупування або коригування напрямку дій залежно від змін у бойовій ситуації.

Таким чином, даний етап відіграє вирішальну роль у забезпеченні ефективності управління військовими діями, оскільки дозволяє своєчасно виявляти загрози, коригувати плани та забезпечувати адекватну реакцію на зміну обстановки.

Отже, якщо узагальнити наведений вище матеріал, то необхідно відмітити, що використання технологічного ланцюга, який притаманний для геопросторового аналізу, а також даних у існуючому загальному алгоритмі оцінки об'єкту аналізу є доцільним і актуальним.

Висновки

Технічний ефект від використання зазначеного способу полягає у підвищенні ефективності застосування системи геопросторового аналізу за рахунок підвищення точності прогнозування зміни стану об'єкту аналізу. Зокрема:

- підвищена точність аналізу — за рахунок комплексного об'єднання різномірних даних та використання математичних моделей прогнозування;
- гнучкість та адаптивність — можливість використання способу в різних сферах, таких як логістика, екологія, транспорт, безпека, містобудування тощо;
- автоматизоване моделювання — мінімізація впливу людського фактору при прийнятті рішень завдяки алгоритмічному аналізу та прогнозуванню;
- зниження ризиків та витрат — можливість своєчасного реагування на потенційні зміни в середовищі або стані об'єкта.

Список літератури

- [1] Романюк В. А. Архітектура системи оперативного управління тактичними радіомережами / Романюк В. А. // Збірник наукових праць ВІТІ НТУУ «КПІ». — 2009. — № 3. — С. 70–76.
- [2] Shyshatskyi, A., Dmytriieva, O., Lytvynenko, O., Borysov, I., Vakulenko, Y., Mukashev, T., Mordovtsev, O., Kashkevich, S., Lyashenko, A., Velychko, V. Development of a method for assessing the state of dynamic objects using a combined swarm algorithm. *Eastern-European Journal of Enterprise Technologies*, 2024, Vol. 3, No. 4 (129), pp. 44–54. <https://doi.org/10.15587/1729-4061.2024.304131>.
- [3] Mahdi, Q. A., Shyshatskyi, A., Voznytsia, A., Plekhova, G., Shostak, S., Tulenko, I., Semko, R., Zheliezniak, D., Momit, A., Sova, M. (2025). Development of a method for increasing the efficiency of processing different types of data in organizational and technical systems. *Eastern-European Journal of Enterprise Technologies*, 2 (4 (134)), 23–31. <https://doi.org/10.15587/1729-4061.2025.325102>.
- [4] A. Koshlan, O. Salnikova, M. Chekhovska, R. Zhyvotovskiy, Y. Prokopenko, T. Hurskyi, A. Yefymenko, Y. Kalashnikov, S. Petruk, A. Shyshatskyi. Development of an algorithm for complex processing of geospatial data in the special-purpose geoinformation system in conditions of diversity and uncertainty of data. *Eastern-European Journal of Enterprise Technologies*. Vol. 5. No. 9 (101). 2019. pp. 35–45. DOI: <https://doi.org/10.15587/1729-4061.2019.180197>.
- [5] Сніцаренко, П. М., Саричев, Ю. А., Зубков, В. П., Піщанський, Ю. А. (2022). Методичний підхід до управління ризиками безпеки інформації як складової забезпечення інформаційної безпеки держави. Збірник наукових праць Центру воєнно-стратегічних досліджень Національного університету оборони України імені Івана Черняхівського, 2 (75), 47–55. URL: <http://znp-cvnsd.nuou.org.ua/article/view/266779/262799>.
- [6] Плехова Г. А., Неронов С. М., Костікова М. В., Кашкевич С. О. Удосконалення моделі безпечної маршрутизації в програмно-конфігурованих мережах // Г. А. Плехова, С. М. Неронов, М. В. Костікова, С. О. Кашкевич / Біоніка інтелекту, 2024, № 1 (100), С. 50–57. DOI: 10.30837/bi.2024.1(100).07. URL: <http://bionics.nure.ua/article/view/314707/305583>.
- [7] Плехова Г. А., Костікова М. В., Неронов С. М., Багмут Р. Б., Яценко О. О. Пристрій утворення маршрутів передачі інформації в радіомережах спеціального призначення із можливістю самоорганізації // Г. А. Плехова, М. В. Костікова, С. М. Неронов, Р. Б. Багмут, О. О. Яценко / Біоніка інтелекту, 2024, № 2 (101), С. 30–33.
- [8] Kashkevich, S., Litvinenko, O., Shyshatskyi, A., Salnyk, S., & Velychko, V. (2024). The method of self-organization of information networks in the conditions of the complex influence of destabilizing factors. *Advanced Information Systems*, 8(3), 59–71. <https://doi.org/10.20998/2522-9052.2024.3.07>.
- [9] Плехова Г. А., Костікова М. В., Неронов С. М., Кашкевич С. О. Обробка різнотипних даних в геоінформаційних системах за допомогою засобу ультракороткохвильового радіозв'язку // Г. А. Плехова, М. В. Костікова, С. М. Неронов, С. О. Кашкевич / Біоніка інтелекту, 2024, № 2 (101), С. 52–55. DOI: 10.30837/bi.2024.2(101).08. URL: <http://bionics.nure.ua/article/view/322707/313061>.
- [10] Шаронова Н. В., Плехова Г. А., Костікова М. В., Неронов С. М., Кашкевич С. О. Пристрій управління ризиками інформаційної безпеки в інформаційних системах // Н. В. Шаронова, Г. А. Плехова, М. В. Костікова, С. М. Неронов, С. О. Кашкевич / Біоніка інтелекту, 2024, № 2 (101), С. 48–51. DOI: 10.30837/bi.2024.2(101).07. URL: <http://bionics.nure.ua/article/view/322702/313058>.
- [11] Подліпаєв В. О. Геопросторова розвідка, як шлях реалізації геоінформаційного підходу у комплексній обробці розвідувальної інформації. Системи обробки інформації: збірник наукових праць. — Х.: Харківський університет Повітряних Сил імені Івана Кожедуба, 2013. — Вип. 5 (112). — С. 53–55. Режим доступу: http://nbuv.gov.ua/UJRN/soi_2013_5_13.
- [12] I. Alieinykov, K. A. Thamer, Y. Zhuravskiy, O. Sova, N. Smirnova, R. Zhyvotovskiy, S. Hatsenko, S. Petruk, R. Pikul, A. Shyshatskyi. Development of a method of fuzzy evaluation of information and analytical support of strategic management. *Eastern-European Journal of Enterprise Technologies*. Vol. 6. No. 2 (102). 2019. pp. 16–27. DOI: <https://doi.org/10.15587/1729-4061.2019.184394>.

Надійшла до редколегії 05.03.2025

УДК 004:681.518:623.74::623-1/-8

DOI 10.30837/bi.2025.1(102).10



- Г. А. Плехова¹, С. М. Неронов², М. В. Костікова³, Ю. С. Асаєнко⁴, Д. О. Плехов⁵
¹ХНАДУ, м. Харків, Україна, plehovaanna1@gmail.com, ORCID iD: 0000-0002-6912-6520
²ХНАДУ, м. Харків, Україна, sernikner@gmail.com, ORCID iD: 0000-0003-2381-1271
³ХНАДУ, м. Харків, Україна, kmv_topaz@ukr.net, ORCID iD: 0000-0001-5197-7389
⁴ХНАДУ, м. Харків, Україна, yzx321123@tutanota.com, ORCID iD: 0009-0002-8181-2846
⁵ХНАДУ, м. Харків, Україна, plehov@gmail.com, ORCID iD: 0009-0004-7873-1716

ПРИСТРІЙ ОБРОБКИ РІЗНОТИПНИХ ГЕТЕРОГЕННИХ ДАНИХ В СИСТЕМАХ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

У роботі запропоновано метод створення пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень, що містить передавальну частину пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень та приймальну частину пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень, причому передавальна частина пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень та приймальна частина пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень з'єднані між собою послідовно, який відрізняється тим, що до складу пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень додатково введено блок нейро-нечіткого прийняття рішень, блок прогнозування та блок навчання баз знань.

ОБРОБКА ГЕТЕРОГЕННИХ ДАНИХ, МОДЕЛЬ, СИСТЕМА ПРИЙНЯТТЯ РІШЕНЬ, ПРОГНОЗУВАННЯ

G. A. Pliekhova, S. M. Neronov, M. V. Kostikova, Yu. S. Asaienko, D. O. Pliekhov. Device for processing various types of heterogeneous data in decision support systems. The paper proposes a method for creating a device for processing heterogeneous data in decision support systems, comprising a transmitting part of the device for processing heterogeneous data in decision support systems and a receiving part of the device for processing heterogeneous data in decision support systems, wherein the transmitting part of the device for processing heterogeneous data in decision support systems and the receiving part of the device for processing heterogeneous data in decision support systems are interconnected in series, which differs in that the device for processing various types of heterogeneous data in decision support systems additionally includes a neuro-fuzzy decision-making unit, a forecasting unit, and a knowledge base training unit.

HETEROGENOUS DATA PROCESSING, MODEL, DESICION MAKING SYSTEM, FORECASTING

Вступ

Запропонований пристрій належить до галузі спеціальної техніки, зокрема, до систем підтримки прийняття рішень. Для системи управління різноманітного призначення є актуальною задачею забезпечення необхідної оперативності та достовірності прийняття рішень.

Відомий пристрій прийняття рішень, що містить блок аналізу та блок прийняття рішень, зв'язані між собою лінією зворотнього зв'язку [1]. До недоліків відомого пристрою прийняття рішень відноситься низька оперативність прийняття рішень.

Виклад основного матеріалу

Найбільш близьким технічним рішенням, як за суттю, так і задачею, що вирішується, яке обрано за найближчий аналог (прототип), є пристрій обробки різнотипних даних в системах підтримки прийняття рішень, що містить передавальну частину пристрою обробки різнотипних даних в системах підтримки прийняття рішень та приймальну частину пристрою обробки різнотипних даних в системах підтримки прийняття рішень, що з'єднані між собою послідовно [2]. Недоліком пристрою обробки різнотипних

даних в системах підтримки прийняття рішень, який обрано за найближчий аналог (прототип), є низька ефективність адаптації до динамічної зміни навколишньої обстановки та вхідних даних.

В основу розробки покладено задачу шляхом додаткового введення до складу пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень блоку нейро-нечіткого прийняття рішень, блоку прогнозування та блоку навчання баз знань забезпечити в комплексі підвищення оперативності та достовірності прийняття рішень, можливості адаптації параметрів пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень до динамічної зміни навколишньої обстановки, забезпечити можливість прогнозування змін навколишньої обстановки та здійснити корегування бази знань систем підтримки прийняття рішень. Суть моделі пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень, що містить передавальну частину пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень та приймальну частину пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень, причому передавальна

частина пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень та приймальна частина пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень з'єднані між собою послідовно, який відрізняється тим, що до складу пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень додатково введено блок нейронечіткого прийняття рішень, блок прогнозування та блок навчання баз знань, при цьому приймальна частина пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень з'єднана послідовно з першим входом блоку нейронечіткого прийняття рішень, вихід блоку прогнозування з'єднано з другим входом блоку нейронечіткого прийняття рішень, вихід блоку навчання баз знань з'єднано з третім входом блоку нейронечіткого прийняття рішень, а вихід блоку нейронечіткого прийняття рішень з'єднано з входом передавальної частини пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень.

Рішення технічної задачі в пристрої обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень (що заявляється), дійсно можливе тому, що:

- введення блоку нейронечіткого прийняття рішень дозволяє забезпечити прийняття достовірних рішень з необхідною оперативністю за рахунок зменшення обчислювальної складності;

- введення блоку прогнозування дозволяє забезпечити короткострокове прогнозування змін вихідної обстановки з необхідною оперативністю за рахунок представлення зміни оточуючої обстановки в вигляді часових рядів;

- введення блоку навчання баз знань дозволить проводити корегування параметрів баз знань та забезпечити необхідну достовірність прийнятих рішень, проводити корегування помилок, що виникають в ході роботи системи підтримки прийняття рішень.

Суть моделі пояснюється за допомогою креслень, де на рис. 1 показано блок-схему пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень, що заявляється. Пристрій 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень, що заявляється, містить (див. блок-схему на рис. 1) приймальну 2 частину пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень, блок 3 нейронечіткого прийняття рішень пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень, передавальну 4 частину пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень та блок 5 прогнозування пристрою 1 обробки різнотипних

гетерогенних даних в системах підтримки прийняття рішень та блок 6 навчання баз знань пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень.

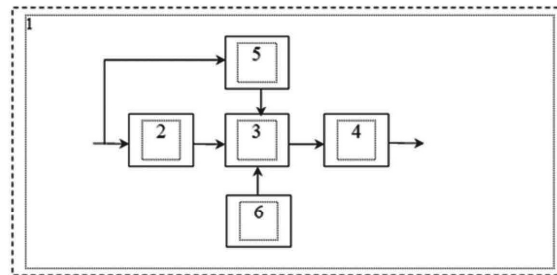


Рис. 1. Функціональна схема запропонованої системи

При цьому в пристрої 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень конструктивні елементи з'єднано між собою таким чином:

- вихід приймальної 2 частини пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень з'єднано з першим входом блоку 3 нейронечіткого прийняття рішень пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень;

- вихід блоку 5 прогнозування пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень з'єднано з другим входом блоку 3 нейронечіткого прийняття рішень пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень;

- вихід блоку 3 нейронечіткого прийняття рішень пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень з'єднано з входом передавальної 4 частини пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень;

- вихід блоку 6 навчання баз знань пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень з'єднано з третім входом блоку 3 нейронечіткого прийняття рішень пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень.

Пристрій 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень, (що заявляється) працює таким чином.

На вхід приймальної 2 частини пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень надходить інформація, яку необхідно обробити, а також ця інформація надходить на вхід блоку 5 прогнозування пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень. З виходу приймальної 2 частини пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень інформація для обробки надходить на перший вхід

блоку 3 нейро-нечіткого прийняття рішень пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень, а з виходу блоку 5 прогнозування пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень результати прогнозу надходять на другий вхід блоку 3 нейро-нечіткого прийняття рішень пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень.

Сутність прогнозування, що відбувається в блоці 5 прогнозування пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень полягає в представленні динаміки зміни оточуючої обстановки в вигляді часових рядів, а також відповідної екстраполяції зміни оточуючої обстановки на певну кількість часових інтервалів вперед.

Основою для роботи блоку 3 нейро-нечіткого прийняття рішень пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень обрано метод Rete. Основним недоліком методу Rete є те його робота тільки з чіткими продукціями, що не дозволяє його використовувати при обробці різнотипних гетерогенних даних.

Обробка інформації в блоці 3 нейро-нечіткого прийняття рішень пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень має наступну послідовність.

Дія 1 Введення вихідних даних.

Дія 2 Формування бази знань (БЗ) з урахуванням невизначеності.

Формальна модель нейро-нечіткої бази правил буде мати вигляд (1)

$$\{P_n\} = \{Rule\}, \quad (1)$$

де *Rule* – правило нейро-нечіткої експертної системи. Кожне правило визначається наступним чином (2):

$$Rule = \langle C \rightarrow S \rangle, \quad (2)$$

де *C* – умова правила; *S* – наслідок правила.

Оскільки модель повинна забезпечувати подання граматичної структури правил з різного виду вкладеними умовами, буде використаний рекурсивний механізм опису вузлів і кінцевих вершин дерева умови правила. Параметр *C* визначається наступним чином (3):

$$C = \langle C_l, R, C_r \rangle, \quad (3)$$

де *C_l* – лівий вузол умови правила; *R* – відношення між вузлами правил; *C_r* – правий вузол умови правила.

Далі розглянемо наведені параметри.

$$C_l = FC_l \parallel Null \parallel C, \quad (4)$$

$$C_r = FC_r \parallel Null \parallel C, \quad (5)$$

де *FC_l* – ліва кінцева трійка умови правила; *FC_r* – права кінцева трійка умови правила.

Формули (4) та (5) дозволяють описати умови з різним ступенем вкладеності.

$$FC_l = \langle L, Z, W \rangle, \quad (6)$$

$$FC_r = \langle L, Z, W \rangle, \quad (7)$$

де *L* – лінгвістична змінна; *Z* – знак умови, $Z = \{<, >, <=, >=, =, !=\}$; *W* – значення умови, яке визначається наступним чином (8):

$$W = L \parallel V, \quad (8)$$

де *L* – лінгвістична змінна; *V* – фіксоване значення (9).

$$V = T_i \parallel const, \quad (9)$$

де *T_i* – значення нечіткої змінної з терм-множин лінгвістичної змінної; *const* – константа. Зазначена модель допускає використання не тільки лінгвістичних змінних, але й класичних змінних. В цьому випадку їх значення може порівнюватися також з константами.

R – множина відношень між вузловими вершинами. $R \subset (C_l \times C_r)$ або $R: C_l \rightarrow C_r$.

Аналогічно параметру *C* визначається параметр *S* – наслідок правила.

$$S = \langle S_l, R, S_r \rangle, \quad (10)$$

де *S_l* – лівий вузол наслідку правила; *R* – відношення між вузлами наслідку правила; *S_r* – правий вузол наслідку правила.

$$S_l = FS_l \parallel Null \parallel S, \quad (11)$$

$$S_r = FS_r \parallel Null \parallel S, \quad (12)$$

де *FS_l* – ліва кінцева трійка наслідку правила, *FS_r* – права кінцева трійка наслідку правила. Формули (11) та (12) дозволяють описати наслідки з різним ступенем вкладеності.

$$FS_l = \langle L, Op, W \rangle, \quad (13)$$

$$FS_r = \langle L, Op, W \rangle, \quad (14)$$

де *L* – лінгвістична змінна; *Op* – операція, $Op = \{:=\}$; *W* – значення наслідку.

Дія 3. Пошук кінцевих трійок та навчання штучних нейронних мереж (ШНМ).

На даному етапі роботи по методу Rete виконується пошук близьких кінцевих трійок у всіх правилах продукційної бази знань. Знайдені відповідності між кінцевими трійками позначаються. У правилах встановлюються посилання такі кінцеві трійки для забезпечення їх одноразової обробки. На відміну від класичних нейро-нечітких експертних систем, в зазначеній нейро-нечіткій експертній системі в якості штучної нейронної мережі пропонується використовувати нейро-нечітку еволюційну мережу. Також на зазначеному етапі відбувається навчання параметрів та архітектури штучної нейронної мережі.

Розглянемо алгоритм пошуку відповідностей кінцевих трійок дерева рішення.

Вхідні дані: *Rule* – база правил, представлена у вигляді дерева рішень.

Вихідні дані: $Rule'$ – скорочена база правил, представлена у вигляді дерева рішень. Проміжні дані: FC_i та FC_j – поточні кінцеві трійки.

Крок 3.1. Спочатку роботи алгоритму всі кінцеві трійки не помічені (не перевірені), m – кількість кінцевих трійок. Встановити початкове значення $i=1$.

Крок 3.2. Якщо $i > m$, то до кроку 3.11.

Крок 3.3. Якщо FC_i помічена, то $i=i+1$ та до кроку 3.2.

Крок 3.4. Вибрати FC_i . Встановити $j=1$.

Крок 3.5. Якщо $j > m$, то помітити FC_i , як проглянуту кінцеву трійку та перейти до кроку 3.2.

Крок 3.6. Якщо FC_j помічена, то $j=j+1$ та до кроку 3.5.

Крок 3.7. Обрати FC_j . Виконати процедуру перевірки близькості кінцевих вузлів та кінцевих трійок FC_i та FC_j .

Крок 3.8. Якщо результат успішний, то додати FC_j в список відповідностей для FC_i , FC_j , кінцеву трійку, що була перевірена.

Крок 3.9. Визначення помилки навчання. Прийняття рішення щодо навчання ШНМ з урахуванням типу невизначеності.

Крок 3.10. Перейти до кроку 3.2.

Крок 3.11. Кінець.

З виходу блоку 6 навчання баз знань пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень на третій вхід блоку 3 нейро-нечіткого прийняття рішень пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень надходить корегувальна послідовність. Основним функціональним призначенням блоку 6 навчання баз знань пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень є корегування синаптичних ваг ШНМ, виду та параметрів функції належності, а також архітектури ШНМ, що покладена в основу функціонування блоку 3 нейро-нечіткого прийняття рішень пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень. Оброблена інформація з виходу блоку 3

нейро-нечіткого прийняття рішень пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень надходить на вхід передавальної 4 частини пристрою 1 обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень для подальшої передачі споживачу.

Підвищення оперативності та достовірності прийняття рішень, можливість прогнозування змін оточуючої обстановки досягається за рахунок введення блоку нейро-нечіткого прийняття рішень пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень, блоку навчання баз знань пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень та блоку прогнозування пристрою обробки різнотипних гетерогенних даних в системах підтримки прийняття рішень.

Висновки

Технічний результат: Забезпечення в комплексі підвищення оперативності та достовірності прийняття рішень, можливості адаптації параметрів до динамічної зміни навколишньої обстановки та забезпечення можливості прогнозування змін навколишньої обстановки.

Список літератури

- [1] Shyshatskyi, A., Dmytriieva, O., Lytvynenko, O., Borysov, I., Vakulenko, Y., Mukashev, T., Mordovtsev, O., Kashkevich, S., Lyashenko, A., Velychko, V. Development of a method for assessing the state of dynamic objects using a combined swarm algorithm. *Eastern-European Journal of Enterprise Technologies*, 2024, Vol. 3, No. 4 (129), pp. 44–54. DOI: <https://doi.org/10.15587/1729-4061.2024.304131>.
- [2] Koshlan, A., Salnikova, O., Chekhovska, M., Zhyvotovskiy, R., Prokopenko, Y., Hurskyi, T., Yefymenko, A., Kalashnikov, Y., Petruk, S., Shyshatskyi, A. Development of an algorithm for complex processing of geospatial data in the special-purpose geoinformation system in conditions of diversity and uncertainty of data. *Eastern-European Journal of Enterprise Technologies*, 2019, Vol. 5, No. 9 (101), pp. 35–45. DOI: <https://doi.org/10.15587/1729-4061.2019.180197>.

Надійшла до редколегії 12.03.2025



О. В. Золотухін¹, М. С. Кудрявцева², Є. Ю. Єремєєв³

¹ХНУРЕ, м. Харків, Україна, oleg.zolotukhin@nure.ua,
ORCID iD: 0000-0002-0152-7600

²ХНУРЕ, м. Харків, Україна, maryna.kudryavtseva@nure.ua,
ORCID iD: 0000-0003-0524-5528

³ХНУРЕ, м. Харків, Україна, yevhenii.yeremieiev@nure.ua,
ORCID iD: 0009-0000-9752-1774

СТОХАСТИЧНА ІНІЦІАЛІЗАЦІЯ НЕЙРОННИХ МЕРЕЖ НА ОСНОВІ АНАЛІЗУ БІОЛОГІЧНИХ СИСТЕМ

Штучні нейронні мережі, як правило, ініціалізуються за допомогою математично визначених методів, які не враховують структурного та функціонального різноманіття біологічних систем. Хоча традиційні підходи забезпечують стабільність навчання, вони ігнорують природні механізми формування синаптичних зв'язків. У цьому дослідженні запропоновано біологічно натхненний підхід до ініціалізації вагових коефіцієнтів, заснований на стохастичних закономірностях, виявлених в емпіричних траєкторіях руху, зафіксованих у контрольованому біологічному середовищі. Дані проходять попередню обробку, що включає згладжування, нормалізацію та масштабування, для генерації вагових значень, які надалі використовуються для ініціалізації нейронної мережі. Ефективність запропонованого методу оцінюється у порівнянні зі стандартними стратегіями ініціалізації на основі трьох тестових наборів даних: MNIST, Fashion-MNIST і Gas Sensor Array Drift. Експериментальні результати демонструють, що біологічно натхненний підхід забезпечує порівнянну ефективність за всіма оціночними критеріями, включаючи точність на тестовій і валідаційній вибірках, кількість епох до збіжності, чутливість за класами та макроусереднений показник F1-метрики. У ряді випадків метод сприяв швидшій збіжності без зниження якості класифікації. Попри те, що запропонована стратегія не завжди перевершує стандартні методи, вона вводить структуровану стохастичність у процес навчання, спираючись на біологічні принципи, та формує перспективне підґрунтя для подальших досліджень у сфері складніших архітектур і біологічно натхнених моделей навчання.

ШТУЧНІ НЕЙРОННІ МЕРЕЖІ, ІНІЦІАЛІЗАЦІЯ ВАГОВИХ КОЕФІЦІЄНТІВ, СТОХАСТИЧНЕ ЗРОСТАННЯ, МАШИННЕ НАВЧАННЯ, БЕГГІНГ, СТЕКІНГ, АНСАМБЛЕВЕ НАВЧАННЯ, ШВИДКІСТЬ ЗБІЖНОСТІ, ЕФЕКТИВНІСТЬ НАВЧАННЯ, ТОЧНІСТЬ КЛАСИФІКАЦІЇ

O. V. Zolotukhin, M. S. Kudryavtseva, Y. Y. Yeremieiev. Stochastic initialization for neural networks based on the analysis of biological systems. Artificial neural networks are typically initialized using mathematically defined methods that do not account for the structural and functional diversity of biological systems. While traditional approaches ensure training stability, they overlook natural mechanisms of synaptic connection formation. This study proposes a biologically inspired approach to weight initialization based on stochastic patterns identified in empirical movement trajectories recorded in a controlled biological environment. The data undergo preprocessing, including smoothing, normalization, and scaling, to generate weight values subsequently used for neural network initialization. The effectiveness of the proposed method is evaluated against standard initialization strategies using three benchmark datasets: MNIST, Fashion-MNIST, and the Gas Sensor Array Drift dataset. Experimental results demonstrate that the biologically inspired approach achieves comparable performance across all evaluation criteria, including test and validation accuracy, the number of epochs required for convergence, class-wise sensitivity, and the macro-averaged F1-score. In several cases, the method facilitates faster convergence without compromising classification accuracy. Although the proposed strategy does not consistently outperform conventional methods, it introduces structured stochasticity into the training process based on biological principles and provides a promising foundation for further research into more complex architectures and biologically inspired learning models.

ARTIFICIAL NEURAL NETWORKS, WEIGHT INITIALIZATION, STOCHASTIC GROWTH, MACHINE LEARNING, BAGGING, STACKING, ENSEMBLE LEARNING, CONVERGENCE SPEED, TRAINING PERFORMANCE, CLASSIFICATION ACCURACY

Вступ

Ефективність навчання штучних нейронних мереж значною мірою залежить від того, як обрано початкові значення вагових коефіцієнтів. Саме початкова ініціалізація визначає характер поширення градієнтів, впливає на збіжність алгоритму та формує загальну динаміку перебігу навчання. Чутливість до початкових значень ваг може зумовити зникнення або вибухове зростання градієнтів, що ускладнює процес оптимізації та знижує здатність моделі до

узагальнення. У відповідь на ці виклики було розроблено низку стратегій ініціалізації, орієнтованих на збереження стабільного поширення сигналу між шарами та ефективне функціонування нейромережі [1].

Традиційні підходи до ініціалізації зазвичай ґрунтуються на заздалегідь заданих математичних розподілах, а також на емпіричних припущеннях щодо глибини мережі та властивостей функцій активації. Хоча такі стратегії значно покращують стабільність і швидкість навчання, вони практично не враховують

принципи, притаманні біологічним нейронним системам [2]. Зокрема, у стратегіях ініціалізації штучних нейронних мереж часто ігноруються просторові закономірності, природна розрідженість і стохастичні механізми зростання, притаманні формуванню синаптичних зв'язків у біологічних нейронних системах.

На відміну від цього, розвиток нейронних структур у біологічних системах відбувається через механізми зростання, що поєднують випадкові процеси та впливи зовнішнього середовища. Синаптичні з'єднання виникають у результаті аксонального росту та просторово спрямованого формування мереж, що призводить до розріджених і функціонально адаптивних схем з'єднань [3]. Такі біологічні особливості можуть слугувати основою для переосмислення принципів ініціалізації ваг у штучних мережах, замінюючи суто формальні підходи на більш природні, з елементами структурного різноманіття та стохастичної організації.

Останні дослідження демонструють, що біологічно натхненні моделі здатні підвищити ефективність нейронних обчислень, зокрема за рахунок відтворення природних механізмів пластичності та навчання [4]. Такі підходи не лише сприяють зростанню продуктивності, а й покращують інтерпретованість і надійність моделей, оскільки ґрунтуються на нейробіологічних закономірностях.

У цьому дослідженні розглядається метод стохастичної ініціалізації штучних нейронних мереж, що базується на даних, отриманих із біологічних систем. Запропонований підхід базується на концепції формування зв'язків у процесі росту та моделює стохастичні закономірності їх розвитку для генерації початкових вагових коефіцієнтів. На відміну від класичних повнозв'язних архітектур, такий метод забезпечує формування розрідженої, структуровано впорядкованої початкової конфігурації, що може сприяти покращенню навчальної динаміки й обчислювальної ефективності.

Об'єктом дослідження є процес ініціалізації вагових коефіцієнтів у штучних нейронних мережах із використанням біологічно натхненого підходу, заснованого на стохастичних закономірностях, виявлених у реальних траєкторіях руху біологічних систем. Предметом дослідження є метод ініціалізації, побудований на моделі стохастичного зростання, що відтворює ключові механізми формування синаптичних зв'язків у біологічних нейронних структурах. У роботі проаналізовано теоретичні основи, реалізовано алгоритм методу та проведено його оцінювання щодо впливу на динаміку навчання й ефективність моделей у порівнянні зі стандартними підходами.

1. Постановка проблеми

Попри суттєвий прогрес у методах навчання штучних нейронних мереж, проблема ефективної та

біологічно обґрунтованої ініціалізації вагових коефіцієнтів залишається відкритою. Сучасні стратегії ініціалізації здебільшого базуються на математичних моделях, розроблених для забезпечення статистичної стабільності активацій і градієнтів між шарами мережі. Вони демонструють високу ефективність у запобіганні градієнтній нестабільності та покращенні динаміки навчання, проте спираються на абстрактні ймовірнісні припущення і не враховують ані структурних, ані функціональних особливостей біологічних систем.

Натомість нейронні зв'язки в біологічних мережах не формуються з використанням рівномірних або симетричних розподілів. Вони виникають у результаті стохастичних, просторово обмежених процесів, зумовлених локальними взаємодіями, динамікою розвитку та адаптивною реакцією на зовнішнє середовище. Формування синаптичних зв'язків у біологічних системах визначається такими принципами, як розрідженість, локальність та пластичність, що спричиняє появу різноманітних і неоднорідних закономірностей зв'язків, які зазвичай не відтворюються в штучних моделях. Дані нейронауки свідчать про те, що така варіативність відіграє ключову роль у підвищенні ефективності навчання, забезпеченні різноманіття сигналів та адаптивності системи загалом.

Однак у більшості штучних нейронних мереж біологічно обґрунтована варіативність майже не враховується під час ініціалізації. Існуючі методи зазвичай розглядають цей процес як суто математичну процедуру, відокремлену від емпіричних біологічних даних. Як наслідок, моделі втрачають потенційні переваги структурованої природної випадковості, серед яких – покращена динаміка навчання, краща здатність до узагальнення та вища стійкість до зміни даних або зовнішніх умов.

Центральною проблемою, що розглядається у цьому дослідженні, є відсутність практичного та відтворюваного методу впровадження біологічно обґрунтованої варіативності в процес ініціалізації ваг штучних нейронних мереж. Хоча ідеї біологічного натхнення широко визнаються в машинному навчанні, більшість підходів не використовує реальні біологічні процеси як структуроване джерело для формування початкових розподілів ваг.

Метою дослідження є розв'язання актуального питання, пов'язаного з відсутністю біологічно обґрунтованих методів ініціалізації, шляхом використання патернів руху, виявлених на основі аналізу біологічних систем. Перетворюючи природні траєкторії руху на структуровані вагові коефіцієнти, запропонована методика має на меті підвищити стабільність навчання, забезпечити ефективну збіжність і з'ясувати, чи може біологічна варіативність позитивно впливати на продуктивність нейронної мережі. Такий підхід

особливо актуальний для задач, пов'язаних із нерівномірним розподілом вхідних даних, динамічними середовищами або сферами, у яких перевагу мають моделі з біологічною інтерпретованістю. У ширшому контексті запропонований підхід сприяє подоланню концептуального розриву між моделями навчання штучних нейронних мереж і принципами, характерними для біологічних нейронних систем.

2. Огляд літератури

Ініціалізація ваг залишається важливою складовою навчання штучних нейронних мереж, особливо у глибоких архітектурах, де вибір початкових значень може суттєво впливати на стабільність градієнтів і загальну динаміку оптимізації. Важливим кроком у цьому напрямі стало запровадження ініціалізації Ксав'єра (Xavier initialization), цей підхід полягає в масштабуванні вагових коефіцієнтів відповідно до кількості нейронів у вхідному та вихідному шарах, що дозволяє зберігати дисперсію активацій і градієнтів стабільною як під час прямого поширення сигналу, так і на етапі зворотного поширення помилки [2]. Ініціалізація Ксав'єра продемонструвала ефективність у прискоренні збіжності та забезпеченні стабільного навчання, особливо у випадку використання симетричних активаційних функцій, таких як гіперболічний тангенс.

Одним з альтернативних широко вживаних підходів є ортогональна ініціалізація (orthogonal initialization), яка зберігає норму вхідного сигналу та виявилася ефективною для стабілізації градієнтного потоку між шарами. Цей метод, окрім математичної обґрунтованості, викликає інтерес і з біологічної точки зору, оскільки орієнтовані схеми з'єднань розглядаються як потенційний механізм стабільного передавання сигналу в біологічних системах [5].

Останніми роками з'явилися альтернативні стратегії, які прагнуть запровадити структуровану випадковість у процес ініціалізації. Зокрема, методи ініціалізації на основі хаотичних відображень (chaos-based initialization) використовують детерміновані хаотичні функції для генерації різноманітних вагових коефіцієнтів, порушуючи симетрію на початкових етапах навчання та сприяючи зростанню репрезентативного різноманіття. Такі підходи демонструють покращену збіжність і вищу точність класифікації, особливо в умовах неопуклої оптимізації, де початкова ініціалізація суттєво впливає на траєкторію навчання [6].

На відміну від штучних моделей, які здебільшого ґрунтуються на абстрактних математичних розподілах, біологічні системи формуються через стохастичні, проте функціонально структуровані процеси росту. У таких системах синаптичні зв'язки виникають завдяки просторовій організації, локальній конкуренції та адаптивним механізмам, залежним від активності.

Дослідження в галузі нейронауки свідчать, що ці процеси формують розріджені, модульні та високоадаптивні мережі, що, своєю чергою, дає підстави вважати, що біологічно натхненні методи (bio-inspired initialization) ініціалізації здатні суттєво покращити характеристики штучних нейронних моделей [7].

Одним із таких підходів є ініціалізація на основі флуктуацій (fluctuation-based initialization), яка імітує природні коливання, зумовлені синаптичним шумом і балансом збуджувальної та гальмівної активності. Цей метод відображає біологічні механізми шумової модуляції, що забезпечують адаптивність нейронних систем. Ініціалізація з флуктуаційною варіативністю асоціюється з покращеною збіжністю та підвищеною здатністю моделі до навчання узагальнених уявлень [8].

Окрім біологічно натхнених схем, активну увагу дослідників привертають оптимізаційно-орієнтовані методи ініціалізації. Зокрема, було запропоновано застосування еволюційних алгоритмів та евристик, натхнених природними процесами, для налаштування початкового розподілу ваг ще до початку навчання. Ці підходи забезпечують гнучкість і адаптивність, використовують стохастичні варіації для пошуку ефективних параметрів, і особливо добре працюють у задачах регресії та підкріплювального навчання [9].

Архітектурні рішення в нейромережевому проектуванні також включають елементи біологічного моделювання — такі як синаптичний шум і стохастична активація. Прикладом є нейросемплінгова машина, яка інтегрує мультиплікативний синаптичний шум у ініціалізацію та обчислення, реалізуючи ймовірнісне навчання, наближене до біологічного [10].

Теоретичні основи навчання нейронних мереж підкреслюють важливість узгодження стратегій ініціалізації з особливостями оптимізаційного процесу. Ефективність зворотного поширення помилки залежить не лише від властивостей градієнтного потоку, а й від масштабу та розподілу початкових вагових коефіцієнтів. Неузгоджена ініціалізація може уповільнити збіжність або спрямувати модель до субоптимальних рішень. Відтак, структуровано побудована схема ініціалізації залишається ключовим чинником для досягнення стабільного та ефективного навчання [11].

Дослідження в галузі моделювання складних систем підтверджують вирішальну роль ініціалізації у формуванні поведінки моделі. При виконанні аналітичних або прогностичних завдань у структурованих середовищах початкові параметри здатні довгостроково впливати на поведінку моделі, динаміку збіжності та рівень інтерпретованості результатів. Інтелектуальний аналіз таких систем показує, що початкові структурні умови часто визначають успішність подальших етапів навчання [12]. Разом із цим, семантичне моделювання предметних галузей

акцентує увагу на необхідності логічно вивіреного, формалізованого представлення знань, що, своєю чергою, сприяє точнішій ініціалізації параметрів завдяки чіткому визначенню структури та залежностей у моделі даних [13].

Сучасні дослідження, орієнтовані на біологічно правдоподібну динаміку навчання, доповнюють це розуміння через запровадження диференційованих навчальних правил, що базуються на нейрофізіологічних спостереженнях. Наприклад, у стохастичних нейронних мережах досліджуються правила навчання, засновані на кореляції активності із затримкою, які відтворюють механізми хеббівської пластичності та синаптичної адаптації, залежної від часу спайкування. Такі підходи демонструють значущість впровадження біологічних принципів у алгоритми навчання та процеси ініціалізації [14].

Ще одним перспективним напрямом є ініціалізація на основі інтервалів, де окремим нейронам або шарам призначаються специфічні діапазони початкових значень. Це сприяє контрольованому внесенню варіативності, порушенню симетрії та покращенню збіжності, особливо в нейронних мережах прямого поширення сигналу [15].

У прикладних сферах значення оптимальної ініціалізації ще більше зростає, зокрема, в комп'ютерному зорі, обробці природної мови та аналізі біомедичних сигналів. У глибоких згорткових мережах початковий розподіл ваг суттєво впливає на якість навчання, особливо у випадках обмеженої кількості навчальних прикладів. У медичній візуалізації та дистанційному зондуванні метод ініціалізації визначає чутливість і надійність виявлення структур, що робить його критично важливим етапом побудови моделей [16].

Моделі адаптивної поведінки в інтелектуальних системах також підтверджують користь структурованої випадковості під час ініціалізації. Така варіативність забезпечує чутливість до контексту та дозволяє агентам змінювати стратегії відповідно до змінних зовнішніх умов [17]. Аналогічні принципи застосовуються в нейромережах, де стохастична ініціалізація сприяє різноманітності поведінки та підвищує ефективність у динамічних і непередбачуваних середовищах. Подібні закономірності виявлено і в дослідженнях прогнозування для систем Інтернету речей, де початкові сигнальні параметри суттєво впливають на точність передбачення екологічних показників у «розумних» будинках [18].

Окрім біологічно обґрунтованих патернів і евристичних методів, значну роль відіграє й попередня обробка даних, орієнтована на прикладні потреби. Так, у біомедичних системах, наприклад, при аналізі ринанометричних сигналів, згорткові нейронні мережі використовуються для автоматичної ідентифікації та виправлення аномалій, що дозволяє

підвищити якість даних ще до етапу навчання [19]. У тривимірному аналізі ефективною виявилася нечітка фільтрація хмар точок на основі F-перетворення, що забезпечує збереження геометричної структури даних і підвищення точності сигналу, що також важливо для ініціалізації мереж, що працюють з просторовими вхідними даними [20].

Підсумовуючи, слід зазначити, що ініціалізація є не просто підготовчим етапом, а фундаментальним компонентом, який визначає якість і ефективність усього процесу навчання. Попри досягнутий прогрес, більшість сучасних методів залишаються в межах абстрактних математичних моделей і не враховують природну варіативність, характерну для біологічних систем. Запропонований підхід спрямований на подолання цього обмеження шляхом використання траєкторій природного руху для генерації початкових ваг.

3. Матеріали та методи

Запропонований метод ініціалізації вагових коефіцієнтів у штучних нейронних мережах ґрунтується на використанні траєкторій руху, отриманих із біологічних систем. Дані про переміщення були зібрані шляхом відеоспостереження за біологічними зразками у контрольованому водному середовищі. Запис здійснювався за допомогою стаціонарної камери, розміщеної над експериментальним об'єктом, з роздільною здатністю 1080p та частотою 30 кадрів на секунду.

Відстеження руху біологічних зразків виконувалося покадрово з використанням алгоритмів виявлення контурів на основі бібліотеки OpenCV. У результаті було сформовано послідовність двовимірних координат (x_t, y_t) , що відповідають положенню об'єкта на кожному часовому кроці t .

Для кількісної оцінки інтенсивності руху розраховувалося переміщення між сусідніми кадрами за допомогою евклідової відстані:

$$d_t = \sqrt{(x_t - x_{t-1})^2 + (y_t - y_{t-1})^2}, \quad (1)$$

де d_t — величина переміщення на часовому кроці t , x_t та y_t — координати об'єкта у поточному кадрі, а x_{t-1} та y_{t-1} — координати об'єкта у попередньому кадрі.

Для зменшення високочастотного шуму та пригнічення короткочасних флуктуацій було застосовано двоетапну процедуру згладжування сигналу. На першому етапі використовувався фільтр ковзного середнього з розміром вікна п'ять кадрів. На другому етапі використовувався фільтр Гауса зі стандартним відхиленням $\sigma = 1$. Обрані параметри були підібрані з метою досягнення оптимального балансу між ефективним усуненням шуму та збереженням природної варіативності у траєкторіях руху біологічних об'єктів.

Після згладжування сигнал переміщення було нормалізовано за допомогою Z-нормалізації, яка обчислюється за формулою:

$$z_t = \frac{d_t - \mu}{\sigma}, \quad (2)$$

де z_t — нормалізоване значення переміщення на часовому кроці t , μ — середнє значення всіх вимірних переміщень, а σ — стандартне відхилення. Така трансформація забезпечує нульове середнє значення та одиничну дисперсію нормалізованого сигналу, що запобігає впливу екстремальних значень на процес ініціалізації.

Для адаптації нормалізованих значень до задачі ініціалізації ваг у нейронній мережі, а також з метою забезпечення стабільної дисперсії активацій, виконувалося масштабування за формулою:

$$w_{scaled} = z_t \times \frac{1}{\sqrt{n_{in} + n_{out}}}, \quad (3)$$

де w_{scaled} — остаточне масштабоване значення ваги, n_{in} — кількість нейронів на вході шару, а n_{out} — кількість нейронів на виході. Масштабування здійснювалося з урахуванням структури шару для підтримання стабільного розповсюдження сигналу та уникнення перенасичення або згасання активацій.

Отриманий одномірний масив вагових коефіцієнтів було перетворено у двовимірну матрицю ваг шляхом позиційного відображення, яке відповідало топології з'єднань між нейронами:

$$w_{i,j} = w_{scaled} [i \times n_{out} + j], \quad (4)$$

де $w_{i,j}$ — вага з'єднання між вхідним нейроном i та вихідним нейроном j . Ця процедура виконується у рядковому порядку, що забезпечує відповідність структури матриці типовим реалізаціям нейронних мереж.

Для інтеграції запропонованого методу ініціалізації у нейромережеву архітектуру було визначено модель, що складається з вхідного шару, одного прихованого шару та вихідного шару. Ілюстративна конфігурація містила 100 вхідних нейронів, 64 нейрони у прихованому шарі та 10 нейронів на виході, що відповідали категоріям класифікації. Сформована матриця ваг була використана для ініціалізації з'єднань між вхідним і прихованим шарами. Ваги інших шарів були ініціалізовані стандартними методами з метою збереження сумісності та забезпечення можливості порівняльного аналізу.

Функції активації, застосовані в нейронній мережі, були підібрані з метою запровадження нелінійності, забезпечення ефективного перетворення сигналу та підтримки стабільного поширення градієнтів. У прихованому шарі використовувалася функція активації ReLU, яка визначається за формулою:

$$f(x) = \max(0, x), \quad (5)$$

де x — вхідне значення нейрона. Вибір ReLU зумовлений її обчислювальною ефективністю, здатністю

сприяти формуванню розріджених активацій та стійкістю до проблеми зникнення градієнтів.

У вихідному шарі застосовувалася сигмоїдна функція активації, яка відображає вихідні значення у проміжок $[0,1]$, що дозволяє інтерпретувати результати як ймовірності у задачах класифікації. Вона визначається наступним чином:

$$f(x) = \frac{1}{1 + e^{-x}}, \quad (6)$$

де e^{-x} — експонента від від'ємного значення вхідного сигналу.

Описана методологія ґрунтується на аналізі біологічних систем і забезпечує статистично коректну ініціалізацію штучних нейронних мереж. Інтеграція природної варіативності в процес генерації початкових параметрів сприяє підвищенню структурного різноманіття мережі та відкриває нові перспективи для розробки біологічно натхненних рішень у сфері машинного навчання. Ефективність запропонованого підходу розглядається в наступному розділі, присвяченому експериментальній оцінці.

4. Експерименти

Експериментальна перевірка запропонованого методу була проведена із застосуванням повнозв'язної нейронної мережі прямого поширення, реалізованої у середовищі PyTorch. Метою експериментів було аналізування динаміки навчання та якісних характеристик моделі під час використання біологічно натхненного методу ініціалізації в порівнянні з іншими відомими підходами. Дослідження здійснювалося в контрольованих і відтворюваних умовах, що дозволяло ізолювати вплив саме ініціалізації на результати навчання та оцінити доцільність включення стохастичних патернів, характерних для біологічних систем, у процес навчання штучної нейронної мережі.

Навчання моделей проводилося з використанням алгоритму оптимізації Adam, який поєднує переваги адаптивного регулювання швидкості навчання та імпульсного прискорення. Швидкість навчання було встановлено на рівні 0.001 як компроміс між швидкою збіжністю та стабільністю, тоді як розмір міні-вибірки, що становив 64, забезпечував ефективне оновлення ваг без надмірного обчислювального навантаження. Кожен сеанс навчання складався з 30 епох, що надавало моделі достатньо часу для стабілізації та досягнення високої точності класифікації. Функцією втрат обрано CrossEntropyLoss, яка є стандартом для багатокласової класифікації й дозволяє безпосередньо інтерпретувати точність моделі у ймовірнісних термінах. Для запобігання перенавчанню та покращення здатності до узагальнення після кожного прихованого шару було додано шари випадкового вимикання із ймовірністю 0.25.

Усі архітектурні, навчальні та оптимізаційні параметри залишалися незмінними протягом

експериментів; варіювався лише метод ініціалізації ваг. Запропонований біологічно натхненний підхід було порівняно з такими стратегіями, як ініціалізація Ксав'єра, ортогональна ініціалізація, ініціалізація на основі хаотичних відображень та ініціалізація на основі флуктуацій. Щоб забезпечити об'єктивність порівняння, кожна конфігурація тренувалася з однаковими випадковими початковими значеннями у стабільному обчислювальному середовищі. Це гарантувало, що всі відмінності в результатах навчання були зумовлені виключно обраним методом ініціалізації.

Для експериментальної перевірки використано три тестові набори даних, що відрізняються типами вхідної інформації та складністю. Набір MNIST [21], що містить 70 000 зображень рукописних цифр у відтінках сірого, слугував для оцінки ефективності моделі у задачах візуальної класифікації зі структурованими та чітко розділеними класами. Цей набір використовувався як базовий для аналізу збіжності та ефективності навчання в умовах спрощеного середовища. Другим був набір Fashion-MNIST [22], який має ту саму структуру, але представляє складніше завдання класифікації одягу, де візуальна подібність між класами значно вища, що ускладнює навчання та вимагає кращого розпізнавання тонких відмінностей. Третім набором був Gas Sensor Array Drift [23], який складається з часових рядів, отриманих від газових сенсорів протягом кількох місяців. Через старіння сенсорів і зміну середовища вхідні розподіли піддаються поступовому дрейфу, що робить цей набір особливо цінним для перевірки стійкості моделей до нестабільності даних і концептуального дрейфу.

Поєднання статичних зображень і динамічних часових рядів формує різноманітне середовище тестування, що дозволяє оцінити як здатність моделі до короткотривалого розпізнавання шаблонів, так і до довгострокової адаптації. Така експериментальна конструкція забезпечує глибший аналіз впливу стратегії ініціалізації за різних рівнів складності та стаціонарності даних.

Кожну конфігурацію експерименту було повторено п'ять разів для врахування стохастичної природи навчального процесу та зменшення варіації результатів. Усі показники усереднювались для отримання стабільних і статистично достовірних висновків. Для оцінювання ефективності використовувався набір ключових метрик, що відображають різні аспекти якості моделі. Точність на тестовій вибірці характеризувала загальну продуктивність моделі на нових даних, тоді як точність на валідаційній вибірці відображала її здатність до узагальнення під час навчання. Кількість епох до збіжності, визначена як момент досягнення 95% точності на навчальному наборі, використовувалася для оцінки ефективності початкової ініціалізації. Чутливість за класами дозволяла проаналізувати здатність моделі виявляти приклади

кожної окремої категорії, а макроусереднений показник F1-метрики, що є гармонічним середнім між точністю та повнотою по класах, слугував інтегральним критерієм збалансованості класифікації, особливо в умовах нерівномірного розподілу даних.

Усі експерименти були проведені в однаковому апаратному середовищі із використанням стандартних програмних бібліотек, щоб уникнути впливу сторонніх факторів, пов'язаних із конфігурацією обладнання чи програмного забезпечення. Експериментальний протокол було розроблено з урахуванням повної відтворюваності та можливості безпосереднього порівняння результатів з майбутніми дослідженнями, що базуватимуться на подібних стратегіях ініціалізації.

Ця розширена експериментальна структура дозволяє не лише оцінити безпосередню ефективність біологічно натхненого методу ініціалізації, а й закласти підґрунтя для його інтеграції в ширші машинно-навчальні процеси.

5. Результати

У цьому розділі подано результати експериментів, які відображають ефективність кожного методу ініціалізації ваг за різними критеріями оцінювання. Аналіз охоплює такі показники: точність на тестовій вибірці, точність на валідаційній вибірці, кількість епох до збіжності, чутливість за класами та макроусереднений показник F1-метрики для трьох різних наборів даних. Це дозволяє здійснити комплексне порівняння стратегій ініціалізації з урахуванням точності класифікації та динаміки навчання в умовах різноманітних інформаційних доменів.

Огляд точності на тестовій вибірці демонструє, наскільки ефективно кожна модель навчалася на відповідній вибірці даних. Усі методи досягли високої точності на наборі даних MNIST, тоді як на складніших наборах даних Fashion-MNIST та Gas Sensor Drift спостерігалася дещо більша варіативність результатів. Біологічно натхненна ініціалізація показала стабільну ефективність, зіставну з іншими методами ініціалізації на всіх розглянутих наборах даних. Результати продемонстровано в табл. 1.

Таблиця 1

Точність на тестовій вибірці (%)

Метод ініціалізації	MNIST	Fashion-MNIST	Gas Sensor Drift
Xavier	99.2	94.8	91.4
Orthogonal	99.1	95.0	91.7
Chaos-Based	99.0	94.4	90.5
Fluctuation-Driven	98.9	94.6	90.9
Bio-Inspired	99.2	94.9	91.6

Точність на валідаційній вибірці відображає здатність моделі узагальнювати знання на нові дані, які

не використовувалися під час навчання. Цей показник є критично важливим для оцінки реальної працездатності моделі за межами тренувального середовища. Отримані результати свідчать про те, що всі методи зберігали високу здатність до узагальнення з незначними відмінностями між ними, що свідчить про їхню надійність. Біологічно натхненна ініціалізація демонструвала стабільну ефективність, результати якої були близькими до показників ініціалізації Ксав'єра та ортогональної ініціалізації. Це вказує на те, що використання стохастичної ініціалізації на основі аналізу біологічних систем не лише не знижує здатність моделі до узагальнення, а навпаки – може сприяти зменшенню перенавчання завдяки введенню початкової структурної неоднорідності. Така властивість є особливо важливою у випадках, коли навчальна вибірка обмежена або містить внутрішній шум, що робить стабільне узагальнення пріоритетом. Результати продемонстровано в табл. 2.

Таблиця 2

Точність на валідаційній вибірці (%)

Метод ініціалізації	MNIST	Fashion-MNIST	Gas Sensor Drift
Xavier	97.8	89.9	88.3
Orthogonal	97.7	90.1	88.6
Chaos-Based	97.6	89.5	87.7
Fluctuation-Driven	97.5	89.7	88.0
Bio-Inspired	97.7	90.1	88.5

Швидкість збіжності визначається кількістю епох, необхідних для досягнення 95% точності на навчальній вибірці, що є важливим критерієм ефективності початкової ініціалізації. Менша кількість епох свідчить про здатність моделі швидше входити в режим стабільного навчання з мінімальними обчислювальними витратами. Біологічно натхненна ініціалізація, відповідно до експериментальних результатів, у більшості випадків забезпечувала таку ж або меншу кількість епох до збіжності порівняно з іншими методами, особливо на наборах MNIST і Fashion-MNIST. Така динаміка може бути зумовлена тим, що стохастичні вагові розподіли, сформовані на основі біологічних траєкторій, сприяють формуванню початкової конфігурації з вищою чутливістю до градієнтів, що зменшує ризик затримки на плато помилок і пришвидшує оптимізацію. Результати подано в табл. 3.

Таблиця 3

Кількість епох до збіжності

Метод ініціалізації	MNIST	Fashion-MNIST	Gas Sensor Drift
Xavier	13	17	21
Orthogonal	14	16	20
Chaos-Based	15	18	23
Fluctuation-Driven	16	17	22
Bio-Inspired	14	16	21

Показник чутливості за класами характеризує здатність моделі коректно виявляти об'єкти, що належать до кожного окремого класу, і є важливим критерієм оцінювання якості класифікації в умовах нерівномірного розподілу класів або присутності класів із низькою частотою появи. Аналіз отриманих результатів свідчить про те, що біологічно натхненна ініціалізація забезпечила стабільну ефективність моделі у розрізі класів. За рівнем чутливості вона не поступалася класичним методам ініціалізації і не спричиняла зниження надійності класифікації на жодному з досліджуваних класів. Таким чином, можна зробити висновок про її здатність підтримувати баланс якості розпізнавання в межах усіх класів, що є критично важливим для задач, де помилка в окремому класі має високу вагу. Результати продемонстровано в табл. 4.

Таблиця 4

Чутливість за класами

Метод ініціалізації	MNIST	Fashion-MNIST	Gas Sensor Drift
Xavier	0.980	0.891	0.873
Orthogonal	0.978	0.894	0.875
Chaos-Based	0.976	0.885	0.865
Fluctuation-Driven	0.975	0.888	0.870
Bio-Inspired	0.979	0.892	0.874

Більш узагальнене уявлення про якість класифікації дає макроусереднений показник F1-метрики. Цей показник поєднує точність і чутливість, забезпечуючи збалансовану оцінку ефективності моделі по всіх класах. Він є особливо інформативним у випадках дисбалансу класів або різної складності їх розпізнавання. Біологічно натхненна ініціалізація знову продемонструвала порівнювані результати, підтвердивши свою надійність у різних задачах та здатність зберігати стабільну якість класифікації без упередженості до окремих категорій. Результати продемонстровано в табл. 5.

Таблиця 5

Макроусереднений показник F1-метрики

Метод ініціалізації	MNIST	Fashion-MNIST	Gas Sensor Drift
Xavier	0.978	0.889	0.871
Orthogonal	0.977	0.892	0.874
Chaos-Based	0.976	0.884	0.864
Fluctuation-Driven	0.975	0.887	0.868
Bio-Inspired	0.977	0.890	0.873

Отримані результати підтверджують, що біологічно натхненна ініціалізація демонструє стабільну ефективність за всіма метриками та на всіх наборах

даних. У низці випадків цей підхід сприяв швидшому навчанню без зниження точності класифікації, що свідчить про його потенціал як повноцінної альтернативи традиційним стратегіям ініціалізації.

6. Обговорення

Результати експериментів дозволяють глибше зрозуміти практичну поведінку біологічно натхненого методу ініціалізації у порівнянні з усталеними підходами. У межах усіх наборів даних і за всіма критеріями оцінювання запропонований підхід продемонстрував ефективність, не нижчу за показники загальноновизнаних методів ініціалізації. Така стабільність свідчить про те, що впровадження варіативності, отриманої з біологічних систем, не має негативного впливу на процес навчання моделі або якість класифікації. Це свідчить про те, що структурована випадковість, сформована на основі аналізу біологічних систем, може слугувати ефективною основою для ініціалізації нейронних мереж.

Щодо точності на навчальній та валідаційній вибірках, запропонований метод забезпечив результати, практично ідентичні тим, що були досягнуті за допомогою інших стратегій ініціалізації. Це вказує на те, що стохастичні патерни, отримані на основі аналізу біологічних систем, є цілком достатніми для підтримки стабільного процесу навчання та ефективного узагальнення. Хоча самі дані були отримані з емпіричних спостережень, процес ініціалізації на їх основі виявився сумісним із статистичною динамікою навчання в штучних системах. Це підтверджує гіпотезу про те, що варіативність, запозичена з біологічних систем, може функціонально передаватися у моделі штучних нейронних мереж без втрати ефективності навчання.

Одним із найпомітніших спостережень стала швидкість збіжності. Біологічно натхненна ініціалізація в багатьох випадках потребувала меншої кількості епох для досягнення порогу точності в 95%, зокрема на задачах MNIST і Fashion-MNIST. Хоча різниця не була значною, вона може свідчити про те, що такий розподіл початкових ваг створює більш сприятливі умови для швидкого входження моделі в ефективний навчальний режим. Згладжені криві збіжності, зафіксовані в окремих експериментах, також вказують на те, що природна варіативність, закладена у ініціалізацію, може допомагати уникати плато помилок на ранніх етапах навчання.

Метод також продемонстрував стабільну поведінку за показниками чутливості по класах та макроусередненим значенням F1-метрики, що підтверджує його відсутність упередженості до окремих класів і надійність навіть за умов дисбалансу даних. Такі результати мають особливу практичну цінність для реальних застосувань, де важливо забезпечити рівномірну якість

класифікації для всіх категорій. У прикладних задачах віддається перевага моделям, що забезпечують узгоджену продуктивність на різномірних вхідних даних без домінування окремих класів, тоді як запропонований метод задовольняє цю вимогу.

Ще одним важливим аспектом є потенційна стійкість методу до зсувів розподілу даних та навчання в нестабільних середовищах. Оскільки біологічні траєкторії руху, що використовуються для генерації ваг, містять у собі структуровану стохастичність та часову динаміку, така ініціалізація може виявитися більш придатною для задач, у яких розподіл вхідних даних змінюється з часом. Хоча ця гіпотеза не перевірялася безпосередньо у межах даного дослідження, вона становить перспективний напрям подальших робіт.

У теоретичному контексті запропонований метод розглядається як внесок у розвиток напрямку машинного навчання, що орієнтується на біологічні принципи. На відміну від більшості традиційних стратегій ініціалізації, які ґрунтуються на формальних положеннях оптимізаційної теорії, запропонований підхід використовує стохастичні закономірності, виявлені в поведінці біологічних систем. Отримані вагові коефіцієнти відображають характерні властивості природних нейронних структур — зокрема, розрідженість з'єднань, присутність шуму та просторову варіативність. Попри те, що навчання штучної нейронної мережі реалізується у вигляді формального алгоритму, впровадження біологічно обґрунтованої структури ще на етапі ініціалізації дозволяє частково відтворити функціональні особливості природних нейронних систем.

Крім того, запропонований підхід може слугувати основою для створення гібридних архітектур, які поєднують стохастичну біологічну ініціалізацію з іншими елементами біологічного наслідування, зокрема регуляризацією типу випадкового вимикання, правилами навчання, керованим шумом, або обчисленням, що ґрунтується на подіях. Така багаторівнева інтеграція біологічних компонентів потенційно дозволяє підвищити як ефективність моделей, так і рівень їх інтерпретованості.

Хоча біологічно натхненна ініціалізація не продемонструвала істотного перевищення показників традиційних методів, вона підтвердила свою стабільність, відтворюваність та практичну придатність. Найважливішим її внеском є новий погляд на те, як біологічні джерела випадковості можуть бути інтегровані у процес навчання без шкоди для якості моделі. Цінність цього підходу полягає не стільки в перевершенні існуючих технік на окремих тестових наборах, скільки в розширенні концептуального інструментарію для побудови нейронних систем з вищим рівнем структурної реалістичності та функціонального різноманіття.

З огляду на це, метод біологічно натхненної ініціалізації може бути особливо корисним для адаптивних або онлайн-навчальних сценаріїв, де початкові умови повинні швидко налаштуватися під динамічні потоки даних. У таких задачах класичні фіксовані методи ініціалізації можуть виявитися неефективними або надто інертними до змін, тоді як ініціалізація на основі біологічних патернів відкриває можливість для створення систем з підвищеною гнучкістю та контекстною чутливістю. Подальше дослідження таких властивостей могло б розширити сферу застосування запропонованого підходу на задачі адаптивного контролю, безперервного навчання та роботи з потоковими даними.

Висновки

У представленому дослідженні запропоновано новий біологічно натхнений метод ініціалізації вагових коефіцієнтів штучних нейронних мереж, який базується на аналізі стохастичних закономірностей, виявлених у траєкторіях руху біологічних об'єктів у контрольованому середовищі. Такий підхід реалізує концепцію перенесення природної варіативності до математичних моделей, спрямовану на подолання обмежень, властивих традиційним стратегіям ініціалізації, що базуються виключно на абстрактних статистичних розподілах.

Результати експериментів підтверджують, що запропонований підхід забезпечує ефективність, сумірну з усталеними методами, такими як ініціалізація Ксав'єра, ортогональна ініціалізація, ініціалізація на основі хаотичних відображень та ініціалізація на основі флуктуацій, за всіма ключовими метриками: точністю класифікації на тестовій і валідаційній вибірках, кількістю епох до збіжності, чутливістю за класами та макроусередненою F1-метрикою. Стабільність запропонованого методу, відсутність деградації навчання й неупередженість до окремих класів свідчать про його надійність, незалежно від природи вхідних даних чи особливостей предметної області.

Особливо примітним є те, що біологічно натхненна ініціалізація в окремих випадках сприяла швидшому досягненню бажаного рівня точності, демонструючи потенціал до покращення динаміки навчання. Враховуючи, що початковий розподіл ваг значною мірою впливає на траєкторію оптимізації у багатовимірному просторі параметрів, можна припустити, що природна структурована варіативність забезпечує сприятливіші умови для ефективного навчання, зменшуючи ймовірність застрягання в локальних мінімумах.

Ще одним важливим внеском є можливість поєднання даного методу з іншими біологічно натхненими компонентами: шумовими регуляризаторами, стохастичними правилами навчання,

моделями спайкової активності або подією орієнтованими обчисленнями. Біологічна сумісність стохастичної ініціалізації закладає основу для створення мультикомпонентних архітектур, які відтворюють не лише функціональні, а й структурні аспекти роботи природної нервової системи.

Наукове значення дослідження полягає не лише у підтвердженні працездатності методу, але й у демонстрації концептуальної життєздатності підходу, що ґрунтується на емпіричних біологічних процесах як джерелі ініціалізації. Це означає, що у майбутньому вагові конфігурації можуть не просто оптимізуватися за допомогою стохастичних функцій, а формуватися на основі природно отриманих динамічних сигналів, адаптованих до конкретного типу задачі або середовища.

Практична цінність підходу підтверджується його сумісністю з існуючими архітектурами, простотою інтеграції в типові фреймворки та відсутністю потреби в модифікації основних елементів навчального процесу. Це дозволяє впроваджувати метод у широке коло застосувань без ризику втрати продуктивності. Зокрема, він є перспективним у тих контекстах, де моделі працюють з динамічними, нестабільними або біологічно релевантними даними — наприклад, у біомедичних системах, робототехніці, когнітивному моделюванні та обробці сенсорних потоків у реальному часі.

Водночас результати дослідження відкривають кілька перспективних напрямів подальших наукових робіт. По-перше, доцільно розширити метод для використання в складніших нейромережевих архітектурах: глибоких згорткових мережах, рекурентних моделях і трансформерах. Це дозволить оцінити, чи зберігається перевага запропонованої стохастичної ініціалізації в умовах значної глибини або часової залежності сигналів. По-друге, варто дослідити адаптивні стратегії побудови вагових розподілів, у яких ініціалізація буде динамічно коригуватися залежно від типу даних, початкового розміру вибірки або змінних зовнішніх умов. По-третє, перспективним є вивчення впливу різних джерел біологічного руху — не лише лінійних траєкторій у просторі, а й багатовимірних патернів активності, зокрема нейрофізіологічних сигналів.

Крім того, важливо дослідити вплив даної ініціалізації на інтерпретованість і надійність моделей у критично важливих додатках, таких як медицина або автономні системи, де неконтрольована варіативність може призвести до небажаних наслідків. У цьому контексті можливим є інтегрування біологічно натхненого методу з механізмами пояснення рішень, що дозволить не лише покращити ефективність, а й підвищити довіру до моделей у чутливих доменах.

Підсумовуючи, варто зазначити, що запропонована стратегія стохастичної ініціалізації, розроблена

на основі аналізу біологічних систем, формує основу для подальшого розвитку адаптивних, стійких та структурно оптимізованих систем штучного інтелекту. Її застосування не лише забезпечує конкурентоспроможну продуктивність у розв'язанні стандартних задач класифікації, а й створює передумови для побудови моделей навчання з високим рівнем гнучкості, інтерпретованості та здатності до адаптації.

Список літератури:

- [1] Dai, K., Barati Farimani, A., Webster-Wood, V.: Bio-inspired stochastic growth and initialization for artificial neural networks. In: Proceedings of the 17th International Conference on Artificial Intelligence Applications and Innovations (IAAI), pp. 88–100. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24741-6_8
- [2] Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS), pp. 249–256. PMLR, Sardinia (2010). <https://proceedings.mlr.press/v9/glorot10a/glorot10a.pdf>
- [3] Goodfellow, I., Bengio, Y., Courville, A.: Deep Learning. MIT Press, Cambridge (2016).
- [4] Squadrani, L., Curti, N., Giampieri, E., Remondini, D., Blais, B., Castellani, G.: Effectiveness of biologically inspired neural network models in learning and patterns memorization. Entropy 24(5), 682 (2022). <https://doi.org/10.3390/e24050682>
- [5] Manchev, N., Spratling, M.: On the biological plausibility of orthogonal initialization for solving gradient instability in deep neural networks. In: Proceedings of the 2022 International Symposium on Computational and Business Intelligence (ISCMI), pp. 47–55. IEEE, Shanghai (2022). <https://doi.org/10.1109/ISCMI56532.2022.10068489>
- [6] Masood, S., Doja, M.N., Chandra, P.: Chaos-based network initialization approach for feedforward artificial neural networks. J. Comput. Theor. Nanosci. 17(4), 418–424 (2020). <https://doi.org/10.1166/jctn.2020.8684>
- [7] Pircher, T., Schlücker, E., Pircher, B., Feigenspan, A.: The structure dilemma in biological and artificial neural networks. Sci. Rep. 11, 5621 (2021). <https://doi.org/10.1038/s41598-021-84813-6>
- [8] Rossbroich, J., Gygax, J., Zenke, F.: Fluctuation-driven initialization for spiking neural network training. Neuroinform. Comput. Eng. 2, 044016 (2022). <https://doi.org/10.1088/2634-4386/ac97bb>
- [9] Demidova, L.A., Gorchakov, A.V.: Biology-inspired optimization algorithms applied to intelligent input weights selection of an extreme learning machine in regression problems. AIP Conf. Proc. 2700(1), 030003 (2023). <https://doi.org/10.1063/5.0124917>
- [10] Dutta, S., Detorakis, G., Khanna, A., Grisafe, B., Neftci, E., Datta, S.: Neural sampling machine with stochastic synapse allows brain-like learning and inference. Nat. Commun. 13, 2571 (2022). <https://doi.org/10.1038/s41467-022-30305-8>
- [11] LeCun, Y., Bottou, L., Orr, G.B., Müller, K.-R.: Efficient backprop. In: Orr, G.B., Müller, K.-R. (eds.) Neural Networks: Tricks of the Trade, pp. 9–50. Springer, Berlin (2000). https://doi.org/10.1007/3-540-49430-8_2
- [12] Filatov, V., Yerokhin, A., Zolotukhin, O., Kudryavtseva, M.: Methods of intellectual analysis of processes in medical information systems. Inf. Extr. Process. 48(124), 92–98 (2020). <https://doi.org/10.15407/vidbir2020.48.092>
- [13] Filatov, V., Semenets, V., Zolotukhin, O.: Synthesis of semantic model of subject area at integration of relational databases. In: Proceedings of the IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL), pp. 598–601. IEEE, Sozopol (2019). <https://doi.org/10.1109/CAOL46282.2019.9019532>
- [14] Saglietti, L., Gerace, F., Ingrosso, A., Baldassi, C., Zecchina, R.: From statistical inference to a differential learning rule for stochastic neural networks. Interface Focus 8, 20180033 (2018). <https://doi.org/10.1098/rsfs.2018.0033>
- [15] Sodhi, S., Chandra, P.: Interval-based weight initialization method for sigmoidal feedforward artificial neural networks. AASRI Procedia 6, 19–25 (2014). <https://doi.org/10.1016/j.aasri.2014.05.004>
- [16] Zangana, H., Mohammed, A., Mustafa Alfaqhi, F.: Advances and applications of convolutional neural networks in image analysis: A comprehensive review. J. Intell. Comput. Syst. 3(1), 16–29 (2024). <https://doi.org/10.58602/jics.v3i1.30>
- [17] Zolotukhin, O., Filatov, V., Yerokhin, A., Lanovyy, O., Kudryavtseva, M., Semenets, V.: An approach to the selection of behavior patterns autonomous intelligent mobile systems. In: Proceedings of the IEEE International Conference on Problems of Infocommunications Science and Technology (PIC S&T), pp. 349–352. IEEE, Kyiv (2021). <https://doi.org/10.1109/PICST54195.2021.9772110>
- [18] Zolotukhin, O., Filatov, V., Yerokhin, A., Kudryavtseva, M.: The methods for the prediction of climate control indicators in the Internet of Things systems. CEUR Workshop Proc. (2021). <https://doi.org/10.5281/zenodo.14526027>
- [19] Yerokhin, A., Nechyporenko, A., Babii, A., Turuta, O.: A new intelligence-based approach for rhinomanometric data processing. In: 2016 IEEE 36th International Conference on Electronics and Nanotechnology (ELNANO), pp. 198–201. IEEE, Kyiv (2016). <https://doi.org/10.1109/ELNANO.2016.7493047>
- [20] Yerokhin, A., Semenets, V., Nechyporenko, A., Turuta, O., Babii, A.: F-transform 3D point cloud filtering algorithm. In: 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), pp. 524–527. IEEE, Lviv (2018). <https://doi.org/10.1109/DSMP.2018.8478581>
- [21] LeCun, Y., Bottou, L., Bengio, Y., Haffner, P.: Gradient-based learning applied to document recognition. Proc. IEEE 86(11), 2278–2324 (1998). <https://doi.org/10.1109/5.726791>
- [22] Xiao, H., Rasul, K., Vollgraf, R.: Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. arXiv:1708.07747 (2017). <https://doi.org/10.48550/arXiv.1708.07747>
- [23] Vergara, A., Vembu, S., Ayhan, T., Ryan, M.A., Homer, M.L., Huerta, R.: Chemical gas sensor drift compensation using classifier ensembles. Sens. Actuators B Chem. 166–167, 320–329 (2012). <https://doi.org/10.1016/j.snb.2012.01.074>



Н. В. Шаронова¹, Г. А. Плехова², С. М. Неронов³, М. В. Костікова⁴, Д. О. Плехов⁵

¹НТУ «ХПІ», м. Харків, Україна, nvsharonova@ukr.net, ORCID iD: 0009-0004-9878-1761

²ХНАДУ, м. Харків, Україна, plehovaanna11@gmail.com, ORCID iD: 0000-0002-6912-6520

³ХНАДУ, м. Харків, Україна, sernikner@gmail.com, ORCID iD: 0000-0003-2381-1271

⁴ХНАДУ, м. Харків, Україна, kmv_topaz@ukr.net, ORCID iD: 0000-0001-5197-7389

⁵ХНАДУ, м. Харків, Україна, plehov@gmail.com, ORCID iD: 0009-0004-7873-1716

СПОСІБ ФОРМУВАННЯ МАРШРУТІВ ПЕРЕДАЧІ ДАНИХ В МОБІЛЬНИХ ІНФОРМАЦІЙНИХ МЕРЕЖАХ

У роботі запропонований спосіб формування маршрутів передачі даних в мобільних інформаційних мережах, який належить до галузі спеціальної техніки зв'язку, зокрема, до систем зв'язку, а саме, до інформаційних мереж спеціального призначення із можливістю самоорганізації, у яких застосовуються спеціальні методи передачі даних. Спосіб відрізняється тим, що на етапі формування множини маршрутів передачі інформації в інформаційній мережі вузлами a та b шляхом передачі службових повідомлень в інформаційній мережі з можливістю до самоорганізації відбувається прогнозування стану маршрутів передачі інформації за критерієм максимуму пропускної спроможності шляхом порівняння отриманих значень пропускної спроможності з граничними показниками мережі. Для інформаційних мереж є актуальним завданням забезпечення заданої пропускної спроможності, забезпечення завадозахищеності та скритності. Технічний результат від застосування зазначеного способу полягає у зменшенні обчислювальних затрат, підвищенні швидкості передачі інформації, забезпечення можливості здійснювати прогнозування стану маршрутів передачі інформації в умовах високої мобільності мобільних інформаційних мереж спеціального призначення.

МОБІЛЬНІ ІНФОРМАЦІЙНІ МЕРЕЖІ, ПРОГНОЗУВАННЯ, МАРШРУТ ПЕРЕДАЧІ ДАНИХ, НЕЙРОННА МЕРЕЖА, АЛГОРИТМ НАВЧАННЯ

N. V. Sharonova, G. A. Pliekhova, S. M. Neronov, M. V. Kostikova, D. O. Pliekhov. Method of forming data transmission routes in mobile information networks. The paper proposes a method for forming data transmission routes in mobile information networks, which belongs to the field of special communication technology, in particular, to communication systems, namely, to special-purpose information networks with the ability to self-organize, in which special data transmission methods are used. The method differs in that at the stage of forming a set of information transmission routes in the information network by nodes a and b by transmitting service messages in the information network with the ability to self-organize, the state of information transmission routes is predicted using the maximum throughput criterion by comparing the obtained throughput values with the network's limit indicators. For information networks, it is an urgent task to ensure a given bandwidth, ensure interference immunity, and secrecy. The technical result of using the specified method consists in reducing computational costs, increasing the speed of information transmission, and providing the ability to predict the state of information transmission routes in conditions of high mobility of special-purpose mobile information networks.

MOBILE INFORMATION NETWORKS, FORECASTING, DATA TRANSMISSION ROUTE, NEURAL NETWORK, LEARNING ALGORITHM

Вступ

Збройні сили сучасних держав при веденні бойових дій мають відповідати високим вимогам технологічності та гнучкості, що передбачає територіальний розподіл підрозділів у бойовому просторі й інтеграцію їхніх дій завдяки єдиному інформаційно-командному середовищу. Важливою умовою ефективного застосування таких збройних сил є використання мережецентричних систем управління, що забезпечують інтегроване та синхронізоване залучення військових сил і засобів одночасно у різних просторових і часових вимірах бойових операцій у режимі реального часу. Це дозволяє підвищити ефективність оперативного реагування, скоротити цикл ухвалення рішень та досягти переваги над противником завдяки інформаційній перевазі.

Особливий акцент у сучасних умовах робиться на створення перспективних тактичних радіомереж, які

стають ядром мережецентричних концепцій [1-10]. Ключову роль у цьому контексті відіграють мобільні радіомережі, що характеризуються низкою переваг, серед яких варто виокремити здатність до самоорганізації, що мінімізує або повністю виключає етап попереднього планування та значно спрощує масштабування мережі залежно від конкретних оперативних завдань. Такі мережі вирізняються оперативністю розгортання, високою стійкістю до пошкоджень або втрат окремих вузлів, а також забезпечують безперервну та стабільну роботу навіть за умов інтенсивного маневрування всіх їхніх елементів.

Важливим аспектом при формуванні таких інформаційних мереж є забезпечення високої пропускної спроможності каналів передачі даних, їхньої завадозахищеності та скритності зв'язку. З цією метою активно використовуються сучасні технології шифрування, адаптивного управління частотними

ресурсами, цифрові методи обробки сигналів та інші передові рішення, що дозволяють гарантувати захищеність інформації та стабільність функціонування мережі навіть за умов активного застосування проти-вником радіоелектронної боротьби.

Виклад основного матеріалу

Відомий спосіб передачі інформації в інформаційних мережах з можливістю до самоорганізації, при якому передача інформації між вузлами інформаційної мережі на інформаційному напрямку $a-b$ відбувається наступним чином:

- вузлом a та вузлом b відбувається формування загальної топології інформаційної мережі шляхом передачі службових повідомлень в інформаційній мережі з можливістю до самоорганізації;

- вузлом a та вузлом b відбувається формування множини маршрутів передачі інформації в інформаційній мережі шляхом передачі службових повідомлень в інформаційній мережі з можливістю до самоорганізації;

- перед початком передачі інформації між вузлом a та вузлом b відбувається визначення якості каналів для передачі інформації шляхом передачі тестових повідомлень в інформаційній мережі з можливістю до самоорганізації;

- передача інформації між вузлом a та вузлом b по найкращому маршруту передачі інформації;

- підтвердження отримання пакетів або в разі неотримання пакетів запит на повторну передачу між вузлом a та вузлом b [11].

Недоліками зазначеного способу передачі є відсутність можливості здійснення прогнозування стану каналів передачі інформації між вузлами a та b мереж з можливістю до самоорганізації з урахуванням альтернатив можливих маршрутів, інтенсивності трафіку та впливу навмисних завад.

Найбільш близьким до способу, що пропонується, є спосіб прогнозування стану каналів інформаційних мереж з можливістю до самоорганізації. Сутність зазначеного методу полягає в тому, що вузол-відправник в інформаційній мережі з можливістю до самоорганізації, перед здійсненням передачі інформації до вузла-кореспондента здійснює послідовність наступних дій шляхом розсилки службових повідомлень в мережі: контроль завантаженості каналів зв'язку в інформаційній мережі; виявлення перевантажень; виявлення виходу з ладу програмної та апаратної складової; виявлення активності користувачів та подій; виявлення вразливостей інформаційної мережі та порушень в налаштуваннях та прогнозування стану інформаційної мережі. На підставі зазначеної послідовності перетворень вузол-відправник визначає раціональний маршрут передачі інформації до вузла кореспондента в інформаційній мережі з можливістю до самоорганізації [12].

Недоліком способу-прототипу є неможливість використання в умовах високої мобільності об'єктів та не врахування впливу навмисних завад на процес передачі інформації між вузлами зв'язку інформаційних мереж з можливістю до самоорганізації.

Тому технічним завданням, що вирішує заявлений спосіб формування топології мережі, є поєднання переваг прототипу та аналогу, з усуненням їхніх недоліків.

Спосіб формування маршрутів передачі даних в мобільних інформаційних мережах, при якому передача інформації між вузлами інформаційної мережі на інформаційному напрямку $a-b$, відбувається наступним чином: вузлом a та вузлом b відбувається формування загальної топології інформаційної мережі шляхом передачі службових повідомлень в інформаційній мережі з можливістю до самоорганізації, після чого вузлом a та вузлом b відбувається формування множини маршрутів передачі інформації в інформаційній мережі шляхом передачі службових повідомлень в інформаційній мережі з можливістю до самоорганізації, після чого перед початком передачі інформації між вузлом a та вузлом b відбувається визначення якості каналів для передачі інформації шляхом передачі тестових повідомлень в інформаційній мережі з можливістю до самоорганізації, після чого відбувається передача інформації між вузлом a та вузлом b по найкращому маршруту передачі інформації, після чого відбувається підтвердження отримання пакетів або в разі неотримання пакетів запит на повторну передачу між вузлом a та вузлом b . У запропонованому посібі на етапі формування множини маршрутів передачі інформації в інформаційній мережі вузлами a та b шляхом передачі службових повідомлень в інформаційній мережі з можливістю до самоорганізації відбувається прогнозування стану маршрутів передачі інформації за критерієм максимуму пропускної спроможності шляхом порівняння отриманих значень пропускної спроможності з граничними показниками мережі.

Пояснимо більш докладно сутність заявленого способу, а саме процес прогнозування. У зв'язку з неможливістю збору в режимі реального часу інформації про стан мобільної інформаційної мережі (МІМ), будемо розглядати процес маршрутизації потоків даних на інформаційному напрямку $a-b$, який складається з кінцевих вузлів a та b (відповідно відправник та адресат), а також множини вузлів, які формують канали передачі між a та b .

Припустимо, що між вузлами a та b існує маршрут m , в якому загальна кількість вузлів рівна k . Вузли можуть змінювати потужність передавача $p_i(t) \leq p_{i_{\max}}$. Параметри вузлів: $e_i(t)$ — це шлях, по якому інформація передається від початкового вузла i до кінцевого вузла j , $i = \overline{1, N}$, N — це кількість

всіх вузлів, через які проходить маршрут (включаючи вузли-відправники та вузли-одержувачі); $T_i(t)$ – час життя i -го вузла. Тип інформації – $\xi = 1-3$ (мова, відео, передача даних); кількість адресатів при кожній сесії $|b|=1$ (однокористувальницька передача). Час існування діючого маршруту $T_m(t)$ визначається мінімальним часом «життя» i -го вузла $T_i(t)$ на маршруті m : $T_m(t) = \min(T_1(t), T_2(t), \dots, T_i(t))$, $i=1, k$. Параметри стану мережі: x_1 – тип трафіка, x_2 – об'єм інформації, x_3 – кількість адресатів, x_4 – розмір черги, x_5 – час існування маршруту, x_6 – швидкість зміни розміру черги, x_7 – коефіцієнт втрати пакетів, x_8 – затримка передачі пакетів в мережі, x_9 – пропускна здатність каналів.

Необхідно: спрогнозувати час перевантаження маршрутів передачі даних в МІМ.

Суть запропонованої у способі формування маршрутів процедури прогнозування полягає у прийнятті рішення щодо пошуку нових маршрутів на основі прогнозованого часу перевантаження маршрутів передачі, з метою задоволення вимог користувальницької оптимізації.

Слід врахувати те, що прогнозування є окремим випадком завдання регресії, тобто залежність залежної змінної від незалежних за заданих умов, то варіантом вирішення цього питання може бути застосування нейронних мереж (НМ), а саме: багат шарового перцептрону, радіально-базисної мережі, узагальнено-регресійної мережі, мережі Вольтеррі та мережі Елмана.

Проведений аналіз застосування НМ при вирішенні завдань прогнозування подій вказує на доцільність застосування обчислення часових рядів в основі яких буде покладена нейронна мережа Елмана, яка представляє собою один з видів рекурентної мережі. Мережа Елмана складається з багат шарового перцептрону з зворотнім зв'язком. Дана функція дозволяє враховувати попередні дії та накопичувати інформацію для підтримки прийняття управлінських рішень на основі прогнозування часового ряду. Прогнозування часового ряду зводиться до задачі інтерполяції (визначення проміжних значень величини) функції багатьох змінних та вирішення задачі апроксимації (приведення до спрощеного вигляду) багатовимірної функції, що невід'ємно впливає на якість прогнозування.

У свою чергу мережа Елмана складається з трьох шарів: вхідного, прихованого та вихідного. При цьому прихований шар має зворотний зв'язок сам на себе, в наслідок чого відбувається навчання нейронної мережі, аналіз подій що сталися та як наслідок прогнозування майбутніх.

На відміну від звичайної мережі прямого поширення, вхідний образ рекурентної мережі це не один вектор, а послідовність векторів $\{x(1), x(2), \dots, x(n)\}$

вхідного образу, що подаються на вхід в заданому порядку, при цьому новий стан прихованого шару залежить від його попередніх станів. Мережу Елмана можливо описати наступними співвідношеннями:

$$u(t) = f(w \cdot x(t) + z \cdot u(t-1) + b_h), \quad (1)$$

$$y(t) = g(v \cdot u(t) + b_y), \quad (2)$$

де: $x(t)$ – вхідний вектор номер t ; $u(t)$ – стан прихованого шару для входу $x(t)$, ($u(0) = 0$); $y(t)$ – вихід мережі для входу $x(t)$; z – вагова матриця розподільного шару; v – вагова (квадратна) матриця зворотного зв'язку прихованого шару; b_h – вектор зрушень прихованого шару; b_y – вектор зрушень вихідного шару; f – функція активації прихованого шару; g – функція активації вихідного шару.

Для навчання мережі Елмана застосовуються градієнтні методи в наслідок чого нейронна мережа обчислюється за допомогою методу зворотного поширення з розгортанням мережі в часі.

Використання мережі передбачає, що процес прогнозування імітується вихідним сигналом деякою нелінійною динамічною системою, яка залежить від множини факторів, у тому числі включаючи і минулі стани системи. В мережі введено шар зворотного зв'язку. Цей шар отримує сигнали з виходу прихованого шару і через елементи затримки z подає їх на вхідний шар, зберігаючи таким чином оброблювану інформацію з попередніх тактів мережі.

Розглядаючи прогнозування перевантаження маршрутів в мережі зазвичай застосовують стандартні нейрони з активаційними функціями, елементи затримки z та блоки фазифікації, призначені для перетворення вхідних порядкових і номінальних змінних, що характеризують вплив мережі, в кількісну форму.

Мережа Елмана для прогнозування часу перевантаження маршрутів з декількома входами, де число нейронів у шарі введення m і прихований шар n і один вихідний блок. Нехай x_{it} ($i=1, 2, \dots, m$) позначають набір вхідних векторів нейронів в момент часу t , y_{it+1} позначає вихід мережі в момент часу $t+1$, u_{jt} ($j=1, 2, \dots, n$) позначають висновок нейронів прихованого шару в часі t та z_{jt} ($j=1, 2, \dots, n$) позначають нейрони рекурентного шару, де w_{ij} – ваги, які з'єднують вузол V у вхідному шарі нейронів до вузла j у прихованому шарі; c_j, v_j – ваги, які з'єднують вузол j в нейронах прихованого шару з вузлом в рекурентному шарі.

Входи нейронів прихованого шару:

$$NET_{ji}(k) = \sum_{i=1}^m w_{ij} x_{it}(k-1) + \sum_{j=1}^n c_{ij} z_{it}(k), \quad (3)$$

де $z_{it}(k) = u_{jt}(k-1)$, $i=1, 2, \dots, n$, $j=1, 2, \dots, m$.

Виходи нейронів прихованого шару:

$$u_{ji}(k) = f_H \left(\sum_{i=1}^m w_{ij} x_{it}(k) + \sum_{j=1}^n c_{ij} z_{it}(k) \right), \quad (4)$$

де сигмоїдальна функція в прихованому шарі обрана як функція активації: $f_H(x) = 1 / (1 + e^{-x})$.

Алгоритм мережі Елмана зі стохастичною ефективністю часу. Алгоритм зворотного поширення є контрольованим алгоритм навчання, який мінімізує глобальну помилку E з використанням методу градієнтного спуску. Для моделі стохастичної ефективності часу мережі Елмана, ми припускаємо, що отримана помилка виходу $\varepsilon_{e_n} = d_{t_n} - y_{t_n}$ та помилка вибірки n визначається як:

$$E(t_n) = 0,5\varphi(t_n)(d_{t_n} - y_{t_n})^2, \quad (5)$$

де t_n – час відгуку n ($n=1, 2, \dots, N$); d_{t_n} – фактичне значення; y_{t_n} – вихід в момент часу t_n ; $\varphi(t_n)$ – ефективна функція стохастичного часу. Ефективна функція часу даних розглядається як функція змінної часу:

$$\varphi(t_n) = \frac{1}{\beta} \exp\left\{\int_{t_0}^{t_n} \mu(t)dt + \int_{t_0}^{t_n} \sigma(t)B(t)\right\}, \quad (6)$$

де $\varphi(t_n)$ – функція змінної часу даних у момент часу відгуку; β – коефіцієнт нормування загального рівня величини функції часу; $\mu(t)$ – функція інтенсивності процесу у момент стохастичного часу t ; $\sigma(t)$ – функція волатильності невизначеності процесу у часі t ; $B(t)$ – шумовий процес, який вносить стохастичний характер у поведінку системи.

Помилка даних мережі визначається як:

$$E = \frac{1}{N} \sum_{n=1}^N E(t_n) = \frac{1}{2N} \sum_{n=1}^N \varphi(t_n) \cdot (d_{t_n} - y_{t_n})^2. \quad (7)$$

Основним завданням алгоритму навчання є мінімізація значення функції стану мережі E до тих пір, поки воно не досягне заданого мінімального значення ξ шляхом повторного навчання. При кожному повторенні, висновок розраховується і виходить глобальна помилка. Градієнт функції стану мережі визначається $\Delta E = \partial E / \partial W$. Для вузлів ваги у вхідному шарі градієнт з'єднувального ваги w_{ij} задається формулою:

$$\Delta w_{ij} = -\eta \frac{\partial E(t_n)}{\partial \omega_{ij}} = \eta \varepsilon_{t_n} v_j \varphi(t_n) f'_H(NET_{j_n}) x_{i_n}, \quad (8)$$

де ω_{ij} – сила (значимість) зв'язку між двома нейронами; η – швидкість навчання; $f'_H(NET_{j_n})$ – похідна функції активації.

Для вузлів ваги в рекурентному шарі задається формулою:

$$\Delta c_j = -\eta \frac{\partial E(t_n)}{\partial c_j} = \eta \varepsilon_{t_n} v_j \varphi(t_n) f'_H(NET_{j_n}) z_{i_n}. \quad (9)$$

Для вузлів ваги в прихованому шарі задається формулою:

$$\Delta v_j = -\eta \frac{\partial E(t_n)}{\partial v_j} = \eta \varepsilon_{t_n} v_j \varphi(t_n) f'_H(NET_{j_n}). \quad (10)$$

Правила оновлення для ваг w_{ij} , A_j та v_j задаються як:

$$w_{ij}^{k+1} = w_{ij}^k + \Delta w_{ij}^k, \quad (11)$$

$$c_j^{k+1} = c_j^k + \Delta c_j^k, \quad (12)$$

$$v_j^{k+1} = v_j^k + \Delta v_j^k. \quad (13)$$

Нейронна мережа Елмана повинна змінювати ваги, щоб мінімізувати помилку між передбаченням мережі та мети передбачення.

Висновки

Технічний результат від застосування зазначеного способу полягає у зменшенні обчислювальних затрат, підвищенні швидкості передачі інформації, забезпечення можливості здійснювати прогнозування стану маршрутів передачі інформації в умовах високої мобільності мобільних інформаційних мереж спеціального призначення.

Список літератури

- [1] Shyshatskiy, A., Dmytriieva, O., Lytvynenko, O., Borysov, I., Vakulenko, Y., Mukashev, T., Mordovtsev, O., Kashkevich, S., Lyashenko, A., Velychko, V. Development of a method for assessing the state of dynamic objects using a combined swarm algorithm. Eastern-European Journal of Enterprise Technologies, 2024, Vol. 3, No. 4 (129), pp. 44-54. <https://doi.org/10.15587/1729-4061.2024.304131>.
- [2] Mahdi, Q. A., Shyshatskiy, A., Voznytsia, A., Plekhova, G., Shostak, S., Tulenko, I., Semko, R., Zheliezniak, D., Momit, A., Sova, M. (2025). Development of a method for increasing the efficiency of processing different types of data in organizational and technical systems. Eastern-European Journal of Enterprise Technologies, 2 (4 (134)), 23-31. <https://doi.org/10.15587/1729-4061.2025.325102>.
- [3] A. Koshlan, O. Salnikova, M. Chekhovska, R. Zhyvotovskiy, Y. Prokopenko, T. Hurskiy, A. Yefymenko, Y. Kalashnikov, S. Petruk, A. Shyshatskiy. Development of an algorithm for complex processing of geospatial data in the special-purpose geoinformation system in conditions of diversity and uncertainty of data. Eastern-European Journal of Enterprise Technologies. Vol. 5. No. 9 (101). 2019. pp. 35-45. DOI: <https://doi.org/10.15587/1729-4061.2019.180197>.
- [4] Сніцаренко, П. М., Саричев, Ю. А., Зубков, В. П., Піщанський, Ю. А. (2022). Методичний підхід до управління ризиками безпеки інформації як складової забезпечення інформаційної безпеки держави. Збірник наукових праць Центру воєнно-стратегічних досліджень Національного університету оборони України імені Івана Черняхівського, 2 (75), 47-55. URL: <http://znp-cvds.nuou.org.ua/article/view/266779/262799>.
- [5] Плехова Г. А., Неронов С. М., Костікова М. В., Кашкевич С. О. Удосконалення моделі безпечної маршрутизації в програмно-конфігурованих мережах // Г. А. Плехова, С. М. Неронов, М. В. Костікова, С. О. Кашкевич / Біоніка інтелекту, 2024, № 1 (100), С. 50-57. DOI: 10.30837/bi.2024.1(100).07. URL: <http://bionics.nure.ua/article/view/314707/305583>.
- [6] Плехова Г. А., Костікова М. В., Неронов С. М., Багмут Р. Б., Яценко О. О. Пристрій утворення маршрутів передачі інформації в радіомережах спеціального призначення із можливістю самоорганізації // Г. А. Плехова, М. В. Костікова, С. М. Неронов, Р. Б. Багмут, О. О. Яценко / Біоніка інтелекту, 2024, № 2 (101), С. 30-33.

- [7] Kashkevich, S., Litvinenko, O., Shyshatskyi, A., Salnyk, S., & Velychko, V. (2024). The method of self-organization of information networks in the conditions of the complex influence of destabilizing factors. *Advanced Information Systems*, 8(3), 59-71. <https://doi.org/10.20998/2522-9052.2024.3.07>.
- [8] Плехова Г. А., Костікова М. В., Неронов С. М., Кашкевич С. О. Обробка різнотипних даних в геоінформаційних системах за допомогою засобу ультракороткохвильового радіозв'язку // Г. А. Плехова, М. В. Костікова, С. М. Неронов, С. О. Кашкевич / Біоніка інтелекту, 2024, № 2 (101), С. 52-55. DOI: 10.30837/bi.2024.2(101).08. URL: <http://bionics.nure.ua/article/view/322707/313061>.
- [9] Шаронова Н. В., Плехова Г. А., Костікова М. В., Неронов С. М., Кашкевич С. О. Пристрій управління ризиками інформаційної безпеки в інформаційних системах // Н. В. Шаронова, Г. А. Плехова, М. В. Костікова, С. М. Неронов, С. О. Кашкевич / Біоніка інтелекту, 2024, № 2 (101), С. 48-51. DOI: 10.30837/bi.2024.2(101).07. URL: <http://bionics.nure.ua/article/view/322702/313058>.
- [10] Подліпаєв В. О. Геопросторова розвідка, як шлях реалізації геоінформаційного підходу у комплексній обробці розвідувальної інформації. Системи обробки інформації: збірник наукових праць. — Х.: Харківський університет Повітряних Сил імені Івана Кожедуба, 2013. — Вип. 5 (112). — С. 53-55. Режим доступу: http://nbuv.gov.ua/UJRN/soi_2013_5_13.
- [11] Романюк В. А. Архітектура системи оперативного управління тактичними радіомережами / Романюк В. А. // Збірник наукових праць ВІТІ НТУУ «КПІ». — 2009. — № 3.— С. 70 — 76.
- [12] I. Alieinykov, K. A. Thamer, Yu. Zhuravskyi, O. Sova, N. Smirnova, R. Zhyvotovskiy, S. Hatsenko, S. Petruk, R. Pikul, A. Shyshatskyi. Development of a method of fuzzy evaluation of information and analytical support of strategic management. *Eastern-European Journal of Enterprise Technologies*. Vol. 6. No. 2 (102). 2019. pp. 16-27. DOI: <https://doi.org/10.15587/1729-4061.2019.184394>.

Надійшла до редколегії 12.05.2025



O. I. Ivanenko¹, Z. R. Musaiev²

¹KhNADU, Kharkiv city, Ukrain, zaur.musaiev92@gmail.com, ORCID iD: 0000-0002-5533-0897

² KhNADU, Kharkiv city, Ukrain, olehiv2@gmail.com, ORCID iD: 0000-0002-3597-5102

USE OF 3D MODEL IN THE STUDY OF TOWER CRANE STABILITY

The article presents the results of a study of the stability of a tower crane during operation, taking into account the influence of loads on the supports. The results of comparing the distribution of loads on the crane supports determined as a result of mathematical and computer modeling during the turning and lifting of the boom are presented. An analysis of tower crane designs, installation methods and operating conditions, as well as methods for calculating their stability, shows that recently there has been a need to install cranes that can handle loading and unloading operations in confined spaces. This raises the question of how important the stability of cranes under different operating conditions and different wind loads is. In this case, the correct choice of the support circuit comes to the fore, for which it is very important to determine the loads on the support elements.

RESEARCH, TOWER CRANE, STABILITY, LOAD DISTRIBUTION, MATHEMATICAL MODEL, COMPUTER MODEL, ADEQUACY

О. І. Іваненко, З. Р. Мусаєв. Використання 3D-моделі у дослідженні стійкості баштового крану. У статті представлено результати дослідження стійкості баштового крану під час експлуатації з урахуванням впливу навантажень на опори. Наведено результати порівняння розподілу навантажень на опори крану, визначених у результаті математичного та комп'ютерного моделювання під час повороту та підйому стріли. Аналіз конструкцій баштових кранів, методів їх монтажу та умов експлуатації, а також методів розрахунку їх стійкості показує, що останнім часом зросла потреба у встановленні кранів, здатних виконувати вантажно-розвантажувальні роботи в обмежених просторах. Це питання про важливість забезпечення стійкості кранів в умовах різної експлуатації та дії вітрових навантажень. У такому випадку на перший план виходить правильний вибір опорного контуру, для чого надзвичайно важливим є визначення навантажень на опорні елементи.

ДОСЛІДЖЕННЯ, БАШТОВИЙ КРАН, СТІЙКІСТЬ, РОЗПОДІЛ НАВАНТАЖЕНЬ, МАТЕМАТИЧНА МОДЕЛЬ, КОМП'ЮТЕРНА МОДЕЛЬ, АДЕКВАТНІСТЬ

Introduction

An analysis of tower crane designs, installation methods and operating conditions, as well as methods for calculating their stability, shows that recently there has been a need to install cranes that can handle loading and unloading operations in confined spaces. This raises the question of how important the stability of cranes under different operating conditions and different wind loads is. In this case, the correct choice of the support circuit comes to the fore, for which it is very important to determine the loads on the support elements.

However, sometimes cranes have accidents that lead to serious consequences for people, equipment, structures, and significant material damage, so it is necessary to comply with existing safety rules when designing and operating them [1].

Sometimes an accident can be caused by the limited space of the construction site itself. In addition, the presence of several tower cranes in central urban areas with overlapping operating cycles can lead to a collision, which significantly increases the number of accidents.

The working environment of the crane also affects the safety of its operation, for example: - dynamic loads caused by work operations and wind. The resulting vibrations affect the base of the tower crane; - the presence of trenches near the crane that change the characteristics of the soil [2-3].

The development and further improvement of domestic tower cranes today is impossible without a thorough

study of the loads acting on the crane, justification of tower crane support structures under various operating conditions, without studying the actual modes of crane use in construction, without developing advanced methods for calculating stability, calculating loads on the supports arising at the base of the crane at arbitrary boom positions, outreach and load on the hook. Fulfillment of these conditions for the development of tower cranes is of great practical importance, as ensuring safety during their operation always remains an urgent problem.

Research in the field of strength, stability and the impact on the stability of external loading is devoted to the works of A.A. Vinson, V.O. Podobed, M.F. Barshtein, L.O. Nevzorov, V.G. Krupko, O.V. Sinelshchikov, V.O. Obidenov, etc.

The analysis of scientific papers in recent years has shown that researchers' efforts are focused on improving models of external unsteady loading; searching for the most loaded elements of the metal structure; tower cranes of a certain design, making engineering and design decisions to reduce the stresses that arise, as well as developing adaptive control systems for tower cranes to reduce the impact of external loads.

Foreign scientists have studied the stability and distribution of support reactions of tower cranes [4-5].

Most authors point to the imperfection of tower crane design methods, which does not allow taking into account the flexibility of the crane's metal structure and the presence of additional loads [6-9].

The aim of this work is to develop a computer model of a tower crane during its operation and compare the results with a mathematical model.

The main part

To build a computer model of the tower crane and its subsequent analysis, the SolidWorks computer-aided design system with the Simulation application was chosen. This program is designed for use on a personal computer in the Microsoft Windows environment. SolidWorks uses the principle of three-dimensional solid and surface parametric design, which allows the designer to create three-dimensional parts and assemble assemblies in the form of three-dimensional models.

Three-dimensional product modeling offers many advantages over traditional two-dimensional design. With SolidWorks, you can see the future product from all sides in volume and give it a realistic representation in accordance with the selected material for a preliminary design assessment.

The associative relationships between parts, assemblies, and their SolidWorks drawings ensure that the model and drawing are consistent, as any changes made to a part are automatically transferred to the associated assembly and drawing.

Processing and comparison of the results with the mathematical model. The study is carried out for different angles of rotation and tilt of the crane boom. Consider the load of the support points at the boom lift angle $\alpha = 00$ crane parking angles $\gamma = 0$, $\theta = 000$. After completing

the calculation, you can estimate the load on the bearing surface in the Feature Manager study window by calling the context menu of the Results section and selecting Reaction force. By indicating the upper edges of the rails on which the crane stands in the open window and selecting the "Free body force" item, you can view their supporting reactions, and therefore the forces acting on the crane (Fig. 1).

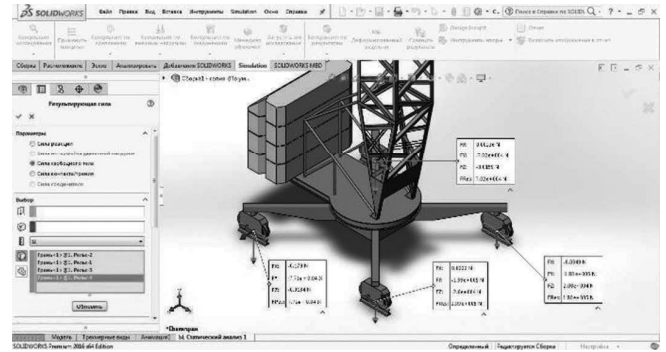


Fig. 1. Forces acting on the supports at the angle of rotation $\beta = 00$

Changing the angle of the boom, we continue to calculate the loads in increments of 450. We compare the results with the result of the mathematical study and draw a graph (Fig. 2).

Changing the angle of the boom, we continue to calculate the loads in increments of 450. For convenience, we summarize the results in Table 1, compare them with the results of the mathematical study, and draw a graph.

Comparison of loads determined in the mathematical and computer models when changing the boom angle

Table 1

			Boom angle β , deg							
			0	45	90	135	180	225	270	315
Load on the support, kN	Support 1	mate. mod.	69,5	33,7	69,5	144,0	223,4	258,2	233,4	144,0
		comp. mod.	70,2	38,2	72,3	142,6	234,6	260,8	236,8	141,1
		discrepancy	1,01	7,00	4,03	0,97	5,01	1,01	6,00	2,01
	Pillar 2	mate. mod.	191,6	114,2	35,7	2,0	35,7	114,2	191,6	223,4
		comp. mod.	180,1	115,3	38,2	1,7	36,1	113,1	201,2	212,2
		discrepancy	6,00	0,96	7,00	15,00	1,12	0,96	5,01	5,01
Load on the support, kN	Pillar 3	mate. mod.	191,6	223,4	191,6	114,2	35,7	2,0	35,7	114,2
		comp. mod.	199,3	216,7	182,0	124,5	32,8	2,1	36,8	119,9
	Pillar 4	discrepancy	4,02	3,00	5,01	9,02	8,12	5,00	3,08	4,99
		mate. mod.	69,5	144,0	223,4	258,2	233,4	144,0	69,5	33,7
		comp. mod.	77,1	131,0	234,6	250,5	230,1	152,6	72,3	34,6
	discrepancy	10,94	9,03	5,01	2,98	3,00	5,97	4,03	3,08	

The difference between the results of mathematical and computer modeling is calculated according to the following relationship:

$$\varepsilon = \frac{|P_k - P_m|}{P_m} \times 100\%$$

where P_k is the load on the support calculated by a computer model;

P_m is the load on the support calculated by the mathematical model;

Now consider the change in the load of each anchor point as the boom outreach changes from the maximum to the minimum value. Repeat the calculation for each boom position from 00 to 500 in ten-degree increments. Boom angle $\beta = 00$; crane parking angles $\gamma = 00$, $\theta = 00$. The obtained values are also entered in Table 2, evaluating the data discrepancy and building a graph (Fig. 3).

Table 2

Comparison of loads determined in the mathematical and computer models when changing the angle of the boom

			Boom lift angle α , deg					
			0	10	20	30	40	50
Load on the support, kN	Supports 1, 4	mathematical model	69,5	86,8	111,8	139,7	169,7	194,6
		computer model	66,7	85,1	107,0	145,0	196,0	189,0
		discrepancy	2,82	0,41	3,01	1,09	5,01	4,05
	Pillars 2, 3	mathematical model	191,6	170,7	142,7	109,8	77,8	46,9
		computer model	197,0	170,	147,	111,0	81,7	45,0
		discrepancy	4,03	1,96	4,29	3,79	3,71	2,88

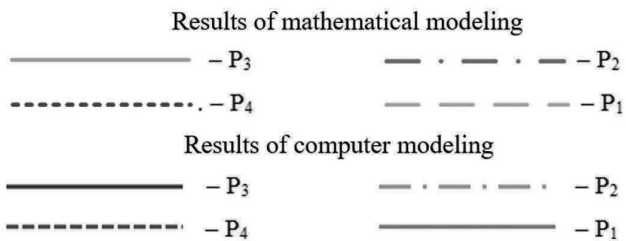
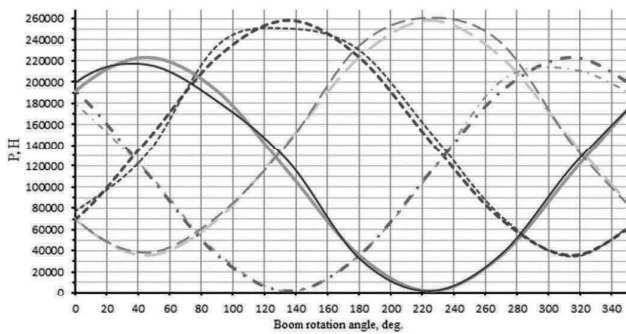


Fig. 2. Graph of comparison of mathematical and computer models

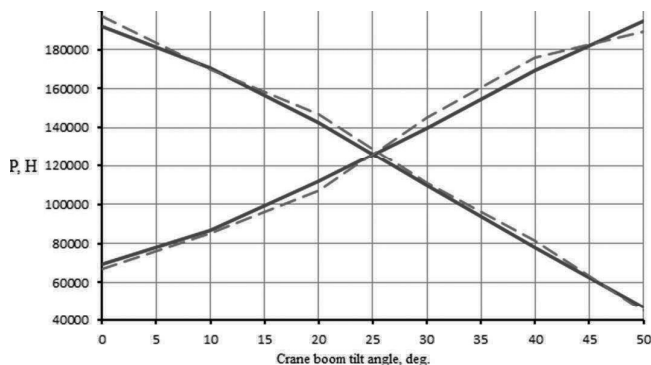


Fig. 3. Graph of comparison of mathematical and computer models

Analyzing the graphs shown in Fig. 3, it can be seen that the discrepancy between the results of the study of the mathematical and computer models varies within ten percent, averaging 4.83%. Taking into account possible inaccuracies in the construction of the computer model, this gives a fairly good result and allows us to consider the mathematical model adequate.

Conclusions

Using the developed algorithm for calculating the loads on crane supports, a computational experiment was conducted to determine the magnitude and nature of changes in the loads on crane supports under different operating conditions: with different loads and variable outreaches, wind loads.

It was found that the load of the crane support elements is oscillatory in nature with an amplitude of about 200 kN.

When the crane boom is rotated by 1350, support 4 can withstand the largest load of 250 kN, and support 2 is almost unloaded — 5 kN. At a turn of $\beta = 2250$, a similar situation occurs with supports 1 and 3, respectively.

When changing the outreach, supports 1 and 4 can withstand the same loads ranging from 68 to 190 kN, and supports 2 and 3 can withstand loads ranging from 190 to 50 kN.

At maximum outreach, the critical positions of the crane are those with boom angles of $\beta=450$ and $\beta=3150$, in which case the crane rests on almost three supports.

References:

- [1] Pravyla pobudovy i bezpechnoi ekspluatatsii vantazhopid-
iomnykh kraniv. NPDOPO.00-1.01-07. K. : Osnova, P68
2007. 312 s.
- [2] Slavomir Hric, Jozef Tkac Darina Matiskova, Milan Micko ,
Dusan Mandulak. Preliminary analysis of tower crane as a type
of truss structure. MATEC Web of Conferences 299, 03003
(2019)<https://doi.org/10.1051/mateconf/201929903003>
MTeM 2019. [https://www.matec-
conferences.org/articles/
mateconf/pdf/2019/48/mateconf_mtem2019_03003.pdf](https://www.matec-conferences.org/articles/mateconf/pdf/2019/48/mateconf_mtem2019_03003.pdf)
- [3] Ivanenko O. I., Shcherbak O. V., Liubimov Yu. Yu. Kom-
piuterne modeliuвання stiikosti na modeli bashtovoho
kranu na osnovi vyznachennia opornykh reaktsii. Nau-
kovi visti Dalivskoho universytetu. №18 2020 rik. Elek-
tronne naukove fakhove vydannia. m. Sievierodonetsk,
ISSN 2222-3428.[http://nvdu.snu.edu.ua/wp-content/
uploads/2020/02/2020-18-6.pdf](http://nvdu.snu.edu.ua/wp-content/uploads/2020/02/2020-18-6.pdf)
- [4] Krupko I., Ivanenko O., Yermakova S. Substantive provisions
of improvement of methods of calculation of loads on carry-
ing and propelling devices of lifting and transport machines /
Norwegian Journal of development of the International Sci-
ence. – VOL.1 – №57/2021. – p. 54-59. ISSN 3453-9875.
DOI: 10.24412/3453-9875-2021-57-1-54-59.
- [5] Ivanenko O.I. , Krupko I.V. , Yermakova S.O. Teoretychni
doslidzhennia stiikosti bashtovoho kranu z urakhuvanniam
rozpodilu navantazhen na opory. Pidiomno-transportna
tekhnika. Odesa 2020 3(64). S 81-95
- [6] Ivanenko O.I., Krupko I.V., Yermakova S.O. Doslidzhen-
nia stiikosti bashtovykh kraniv z urakhuvanniam rozpodilu
navantazhen na oporni elementy / Visnyk Donbaskoi der-
zhavnoi mashynobudivnoi akademii. Vypusk № 3 (47) 2019.
– Kramatorsk. – S. 46-50.
- [7] Kyselov V. V. "Doslidzhennia stiikosti bashtovykh kraniv
za dopomohoiu 3D modeliuвання". – Naukovyi zhurnal
inzhenierii ta arkhitektury, 2021. – T. 15, №3, s. 25–32.
- [8] Myronenko O. O., Petrov I. I. "Vykorystannia tsyfrovyykh
modelei pry otsintsi bezpeky bashtovykh kraniv". – Visnyk
budivelnnykh nauk, 2019. – № 4, s. 45–53.
- [9] Kumar, R., & Singh, D. "Stability Analysis of Tower Cranes
Using Finite Element Method (FEM) and 3D Modeling."
– International Journal of Structural Engineering and Me-
chanics, 2020, Vol. 13(5), pp. 231–240.

The article was delivered to editorial staff on the 20.05.2025



N. Brynza^{1,2}, V. Borodavka³, O. Teslenko¹

¹Simon K. KNUE, Kharkiv, Ukraine, natalia.brynza@hneu.net,
ORCID iD: 0000-0002-0229-2874

²KhNURE, Kharkiv, Ukraine, nataliia.brynza@nure.ua,
ORCID iD: 0000-0002-0229-2874

³V. N. Karazin Kharkiv National University, Kharkiv, Ukraine, vladyslav.borodavka@karazin.ua,
ORCID iD: 0009-0002-3885-1364

¹Simon K. KNUE, Kharkiv, Ukraine, oleg.teslenko@hneu.net,
ORCID iD: 0000-0003-3105-9323

IMPLEMENTATION OF MEASURES TO COUNTERACT CYBER THREATS AND ABUSE IN INFORMATION SECURITY BASED ON BLOCKCHAIN TECHNOLOGIES

The selection of security methods for the ISMS of blockchain systems has been researched and justified, considering their specifics and the nature of threats. The key selection factors include ensuring reliability, decentralization, confidentiality, and cyberattack resilience. Various approaches were examined, including consensus algorithms, cryptographic methods, smart contract security, and network protocols. Based on the analysis, the implementation of a monitoring and response method was deemed appropriate, specifically utilizing tools such as iptables, ipset, fail2ban, and dynamic blacklists of IP addresses. This mechanism effectively combines various tools for automation and dynamic updating of security levels. By employing iptables and ipset, the system efficiently filters network traffic, while fail2ban monitors suspicious patterns and blocks malicious IP addresses based on configurable rules. The inclusion of dynamic blacklists adds an extra layer of protection by continuously updating the database of known malicious IP addresses, allowing real-time security adjustments. Monitoring results demonstrate that the system's effectiveness can vary significantly depending on the number of blocked attacks, the cyberattacks that bypassed the system, and the false positive rate.

WEB SERVICE, BLOCKCHAIN, NETWORK, CYBERSECURITY, CYBERATTACK, INFORMATION SECURITY MANAGEMENT SYSTEM, SECURITY MECHANISM, EFFICIENCY

Н. О. Бринза, В. В. Бородавка, О. В. Тесленко. Імплементація засобів протидії кіберзагрозам і зловживанням у системах керування інформаційною безпекою на базі блокчейн-технологій. У роботі здійснено аналіз і аргументовано обрано методи захисту, що враховують особливості блокчейн-систем та специфіку потенційних кіберзагроз. Вибір підходів ґрунтувався на потребі забезпечення високої надійності, децентралізованого управління, конфіденційності даних і стійкості до атак. Розглядалися різноманітні захисні стратегії, включаючи консенсусні алгоритми, криптографічні засоби, захист смарт-контрактів і особливості мережевої взаємодії. За результатами порівняльного аналізу доцільним виявилось застосування моніторингово-реактивної моделі із використанням таких інструментів, як iptables, ipset, fail2ban і механізми динамічного формування чорних списків IP-адрес. Запропоноване рішення дозволяє поєднати ці засоби в єдину систему автоматизованого управління рівнями безпеки. Зокрема, iptables та ipset відповідають за ефективне фільтрування трафіку, fail2ban виявляє та блокує підозрілу активність, а динамічні чорні списки забезпечують оперативне оновлення бази загроз і адаптацію системи до нових викликів у реальному часі. Проведене тестування свідчить, що ефективність системи значною мірою залежить від кількості заблокованих атак, рівня проникнення загроз і частоти хибних спрацювань.

WEB-SERVICE, БЛОКЧЕЙН, МЕРЕЖА, КІБЕРЗАХИСТ, КІБЕРАТАКА, СИСТЕМА УПРАВЛІННЯ ІНФОРМАЦІЙНОЮ БЕЗПЕКОЮ, МЕХАНІЗМ ЗАХИСТУ, ЕФЕКТИВНІСТЬ

Introduction

The development of information technology (IT) has become an integral part of society. Information is one of the most valuable resources in any business process, which determines the priority of ensuring information security (IS) as a key aspect of effective management. Information security includes a set of measures aimed at preventing and eliminating the risks of unauthorised access, processing, modification, analysis, unauthorised alteration or destruction of data.

Theoretical and practical aspects of IS are actively studied by both domestic and foreign scholars, in particular, Babenko V., Boyko A., Vasylieva T., Gontareva I.,

Horbenko I., Kachynskiy A., Leonov S., Kuzmenko O., Starkova O., Anderson R., Cardholm L., Kshetri N., Stephanides G., Tsiakis T., and others. Recently, the issue of information policy and IS has become particularly relevant.

In today's world, IT development is the backbone of many industries, including financial, healthcare, logistics and others. One of the key technologies that provides new opportunities for secure storage and transmission of information is blockchain, which ensures transparency of transactions, data immutability and a decentralised management model, making it attractive to many industries.

In a blockchain, data is stored in blocks that form a connected chain. The information in the blocks is chronologically consistent, as any changes or deletions are impossible without reaching a consensus among the network participants. It is a unique data management system that has a number of advanced features. The main difference between blockchains is the decentralised nature of management, which ensures trust in data without the need for centralised control. Traditional databases (DBs) usually lack the ability to share data between different companies.

In blockchain networks, each participant has its own copy of the registry, the consistency of which is maintained automatically. In addition, in traditional databases, data can be edited or deleted, while in the blockchain, the data entered remains unchanged, as the system provides a high level of security. Once information is recorded in the blockchain, its modification becomes extremely difficult, which increases the reliability and security of the stored data [1].

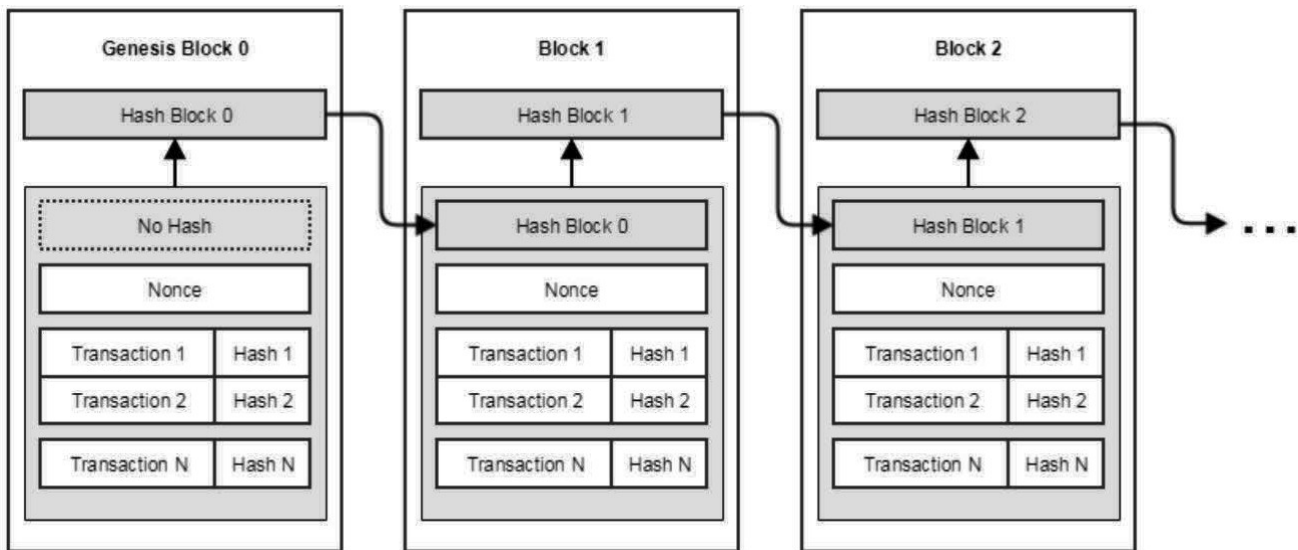


Fig. 1. The blockchain structure

Decentralisation in blockchain technology involves the transfer of control and decision-making from centralised entities (individuals, organisations or groups of them) to a distributed network. The transparency of decentralised blockchain systems reduces the dependence of participants on each other, eliminating the need for trust between them. This architecture limits the ability of one party to exert excessive influence or control while maintaining the functionality of the network. Data immutability is a key feature of the blockchain, which makes it impossible to change information once it has been entered into the register. In the event of errors in the records, a new transaction is added to correct them, while all previous transactions remain available for viewing in the history of changes. Thus, the network displays both the original and the corrected record. To confirm new transactions, the system uses consensus mechanisms to ensure that they are registered only if approved by the majority of participants.

Traditional databases pose a number of difficulties in keeping records of financial transactions. For example, when selling real estate, ownership is transferred to the buyer only after payment is made. However, the parties may register transactions separately, which creates risks of distrust: the seller may deny receiving funds, and the buyer may claim that payment has been made, even if this

is not true. To avoid such disputes, a trusted third party is usually involved to monitor and confirm the transaction. However, a centralised intermediary complicates the process by creating a single point of vulnerability, which can lead to serious consequences for both parties in the event of a system failure. Blockchain technology offers an effective solution to these problems by creating a decentralised, tamper-proof system for recording transactions. In the case of a real estate transaction, blockchain ensures that there is a single register that is synchronised between the buyer and seller in real time. All transactions have to be mutually approved, and any discrepancies in the records are instantly reflected throughout the system, increasing transparency and trust in the data. Due to its properties, blockchain technology has gained wide popularity in various industries. One of the most famous examples of its use is the creation of the digital currency Bitcoin.

However, with the growing popularity of blockchain comes new security challenges. Blockchain, like any other digital information system (IS), is becoming a target of cyber attacks and misuse. While the decentralised nature of the blockchain provides a certain level of protection, it is not completely immune to various threats. Vulnerabilities can exist at the software level, as well as at the level of access control mechanisms, cryptographic protocols, and network infrastructure.

One of the most serious threats is the ‘51%’ attack, where attackers gain control of the majority of computing power in the blockchain network, allowing them to alter transactions or even create duplicate digital assets. Other threats include phishing attacks, hacking smart contracts, and exploiting vulnerabilities in cryptographic algorithms. Insecure information security management systems (ISMS) can lead to critical data leakage, privacy breaches, and significant financial losses.

development, it is important to ensure reliable security mechanisms in the blockchain MIS. This includes constant monitoring of network activity, implementation of modern cryptographic solutions, creation of multi-level authentication systems, as well as regular software updates and improvements. Ensuring protection against blockchain-related malware and abuse is an important component of ensuring the reliable operation of systems based on this technology. ISMS should constantly adapt to new challenges in cyberspace to maintain a high level of data protection and minimise risks.

1. Analysis of mechanisms of protection against cyberattacks and abuse in information security management systems in the blockchain

One of the most well-known attacks is the 51% attack, which occurs when an attacker gains control of most of the computing resources of a blockchain network, allowing him to manipulate transactions and disrupt the chain's integrity. This allows them to control more than 50% of the mining capacity and mine new blocks faster than other participants. This advantage allows attackers to stop or change the order of transaction confirmation, as well as edit parts of the blockchain and cancel transactions that have already been carried out [2]. The 51% attack typically violates blockchain security protocols, and its consequences can range from minor to very serious, depending on the amount of hash power controlled by the attacker. The more computing resources an attacker controls, the greater the likelihood of a successful attack and the more serious the damage [3]. By controlling more than 51% of the capacity, the attacker can secretly create alternative blocks that will be considered valid because of the dominant capacity. This allows him to cancel transactions before they are confirmed, resulting in double spending of coins. In addition, legitimate miners earn less because attackers take their share of the profits from blockchain updates. Some miners, by increasing their computing power, may inadvertently cross the 50% limit of the total network capacity, but this does not pose a threat if they follow the rules and do not interfere with the normal operation of the system [4]. However, if a participant uses their advantage to act dishonestly, it can be considered an attack.

A Finney attack is a type of ‘double-spending’ where a transaction is confirmed by only one transaction confirmation [5]. In this case, the attacker creates a transaction to

pay for the goods, while simultaneously preparing a block with a transaction that transfers these funds to another own account, but does not publish this block. As soon as the payment transaction is confirmed by one of the miners and the goods are received, the attacker quickly publishes the prepared block. As a result, two blockchain branches of the same length are formed in the network. If miners start supporting the branch that contains a transaction to the attacker's account, the transaction that was supposed to transfer funds to the seller will be cancelled, and the seller will lose money because the goods have already been shipped [3]. To protect against this attack, the seller can wait for several transaction confirmations, which reduces the risk but does not guarantee complete security. If an attacker controls several nodes in the network and the seller does not wait for enough confirmations, the attacker can create a longer chain with a transaction to his account. After publishing this chain, miners will continue to work with it, maintaining a block with a transaction in favour of the attacker. If both chains have the same length, miners must choose one of them, and in this case, the probability of success of the attack is 50%.

A Race Attack occurs when an attacker makes two transactions at the same time: transaction ‘A’ to pay for the purchase and transaction ‘B’ that transfers the same funds to another account. If the seller does not wait for the transaction to be confirmed and ships the goods immediately, he or she runs a risk: with a 50% probability, transaction B can be included in the blockchain without additional actions by the attacker [6]. Even worse, an attacker can increase the probability of success by selecting specific nodes to transmit a particular transaction. The principle of this attack is similar to the Finney attack and is also a form of ‘double spending’.

Distributed denial-of-service (DDoS) attacks, although difficult to execute in a blockchain network, are still possible. In a DDoS attack, attackers seek to disable a server by overloading it with a large number of requests, which leads to the depletion of its computing resources. The main purpose of such attacks is to destabilise the functioning of mining pools, e-wallets, cryptocurrency exchanges and other financial services. In addition, the blockchain can be attacked at the application level using DDoS botnets that make massive requests to complicate the network's operation.

Timejacking exploits a potential vulnerability in the way the Bitcoin network handles timestamps. In this attack, an attacker changes the time settings of a node, forcing it to adopt an alternative blockchain. This is possible when an attacker adds several fake peers to the network with incorrect timestamps [7].

An eclipse attack assumes that the attacker controls a large number of IP addresses or uses a distributed botnet network. The attacker modifies entries in the victim's table of ‘tested’ nodes and waits for the victim's node to

restart [8]. After the restart, all outgoing connections of the victim node are directed to IP addresses controlled by the attacker. This results in the victim being unable to receive the transactions they need. At first glance, the eclipse attack may seem similar to the Sybil attack, as both involve the distribution of fake resources on the network. However, their ultimate goals are different [9]. In an eclipse attack, the attacker tries to completely isolate the victim by redirecting all of its connections to nodes that they control. The attacker creates a ring of controlled IP addresses to which the victim's node is likely to connect after the system is restarted. The restart may be forced (for example, due to a DDoS attack) or occur due to other factors that the attacker may simply wait for.

A Sybil attack is a security threat to online systems where one person attempts to take control of a network by creating multiple fake accounts, nodes, or computers. A simple example is when one person creates multiple accounts on a social network. In the context of cryptocurrencies, this could be a situation where someone runs multiple nodes on a blockchain network at the same time. The name 'Sybil' came from the case of a woman named Sybil Dorsett who suffered from dissociative identity disorder, also known as multiple personality disorder [10].

Cryptojacking is one of the types of cyber threats that has become increasingly widespread in recent years [11]. This type of cybercrime involves the use of computing resources of a user's computer, mobile phone, tablet, laptop, or server for unauthorised cryptocurrency mining without the user's consent or knowledge. The main goal of cryptojacking is to make money on cryptocurrency at the expense of other people's or organisations' resources. This threat is becoming particularly relevant due to the growing popularity of cryptocurrencies such as Monero, Bitcoin, Ethereum, and others, as well as the increasing number of devices connected to the Internet. For attackers, this is an opportunity to make significant profits by using large networks of infected devices for mining [12]. Cryptojacking is often implemented through hidden scripts embedded in websites or mobile applications.

All technologies, including blockchain, have potential attack vectors that cybercriminals can use to their advantage. In the cryptocurrency world, one of the most well-known is the Vector 76 attack. This is a form of double-spend attack that exploits a minor bug in the Bitcoin consensus system. As a result, an attacker can obtain funds and harm their victims [13]. This attack is performed when a fraudulent miner controls two complete networks of nodes. One of them (node A) is connected directly to the exchange service, and the other (node B) is connected to other key nodes in the blockchain network. For a successful attack, an attacker must monitor the transmission and propagation of transactions through different nodes to know which ones are the first to transmit transactions and thus connect to both the exchange service and the key

nodes in the network correctly.

The double-spending vulnerability is a common blockchain attack method that exploits the transaction verification process. All transactions in the blockchain must be verified by users to be valid, which takes time [14]. Attackers can take advantage of this delay to trick the system into using the same coins or tokens in multiple transactions. Other types of attacks mentioned earlier have also arisen from this vulnerability [15]. Unlike traditional financial institutions, the blockchain confirms transactions only after a consensus is reached between all network nodes. Until a block with a transaction is verified, the transaction is considered unconfirmed. However, the verification process takes some time, which creates opportunities for CAs. Similar to counterfeiting, double spending leads to inflation by increasing the amount of duplicate currency that did not previously exist. This leads to a depreciation of the currency against other currencies or goods, reduces user confidence, and disrupts the normal circulation and storage of assets.

2. Rationale for technology selection

Justification of the choice of security methods in blockchain ISMS is a key step in increasing the resilience of such systems to acts of corruption and misuse. The choice of specific technologies, mechanisms, and approaches is based on risk assessment, network features, the level of available resources, and the predicted threats that a blockchain system may face. Each security method plays a specific role in the overall security architecture, so the right choice and combination of such methods is critical to maintaining the integrity and reliability of the network.

Consensus algorithms are the basis of the blockchain, which allows decentralised nodes to reach agreement on the state of the ledger without the need for a central controller [16]. The choice of a consensus algorithm determines not only the efficiency and speed of the blockchain, but also its resistance to attacks.

The PoW algorithm provides a high level of protection against 51% attacks, since the attack requires significant computing power, which is economically unprofitable for attackers [17]. However, PoW has disadvantages, such as high power consumption and slow transaction confirmation speed. It is suitable for those blockchain projects that have a large global reach and can afford high security costs (e.g. Bitcoin).

The PoS algorithm relies on the stake that users hold in the network. This approach makes attacks less likely, as an attack on the system requires significant investment, which will be lost in case of failure [17]. PoS is less energy-intensive than PoW and offers faster transaction confirmation, which makes it an attractive choice for new blockchain systems (e.g., Ethereum 2.0). However, it is important to keep in mind that PoS can be vulnerable to a

‘trust attack’, where large validators can join together and create cartels to control the network.

The DPoS algorithm uses delegation, where participants select validators to confirm transactions. DPoS is very efficient and fast, but is vulnerable to centralisation, as several delegates can gain control of the network [18]. It is suitable for blockchains where speed and scalability are a priority (e.g., EOS).

A reasonable choice between PoW, PoS, and DPoS depends on the needs of a particular system. If the system requires maximum decentralisation and security, PoW may be the best choice. For systems that are focused on speed and environmental efficiency, PoS or DPoS are more suitable.

Cryptography is the main mechanism for ensuring confidentiality, integrity, and authenticity in a blockchain. The choice of cryptographic algorithms should take into account long-term security, including threats from quantum computing.

Asymmetric cryptography - public and private key cryptography is used to protect transactions in the blockchain. The choice of strong algorithms, such as RSA or ECDSA (elliptic curve cipher), provides protection against attackers trying to forge transactions [19-20]. However, with the development of quantum computers, these algorithms may become vulnerable, so future blockchains should start implementing post-quantum cryptographic algorithms (e.g., lattice-based algorithm).

Hashing - cryptographic hash functions such as SHA-256 (in Bitcoin) or Keccak-256 (in Ethereum) are used to ensure the integrity of data in the blockchain [19-20]. Hashing ensures that any change in the block data will be instantly detected. The choice of a strong hash function is important to prevent attacks such as hash collisions or computational hash reversal attacks.

Zero Disclosure Cryptographic Algorithms - Innovative methods such as zk-SNARKs (zero-knowledge succinct non-interactive arguments of knowledge) [21] provide privacy in blockchain transactions by allowing transactions to be confirmed without disclosing details. This is important for private blockchains or blockchains used for financial transactions where confidentiality is critical.

Smart contracts are one of the most powerful tools in blockchain systems, but they can also be a source of vulnerabilities. Therefore, to ensure their security, a number of methods should be applied.

Firstly, it is the audit of smart contracts, i.e. the external audit of the smart contract code by independent experts allows to identify possible vulnerabilities before they are used in a real network. Companies such as CertiK and OpenZeppelin specialise in smart contract security testing.

Also, smart contracts should be thoroughly tested in a sandbox environment to identify possible logical errors or security flaws. Formal verification is important for critical

applications, as it allows mathematical proof of the correctness of the contract [22].

Using a modular approach to writing smart contracts makes their code more manageable and secure. Dividing contract logic into separate components allows you to isolate potential problems and reduces the impact of code errors.

Network-level security is important for blockchain because the decentralised nature of the network makes it vulnerable to certain types of attacks. For this purpose, it is necessary to implement encryption at the P2P level - the use of secure communication protocols between nodes blocks the possibility of data interception or manipulation attempts [23]. For example, the TLS protocol ensures reliable data encryption during transmission. It is also necessary to protect against routing attacks - DHT protocols are used to organise data exchange between nodes. To ensure reliability, it is necessary to use advanced mechanisms to protect against routing attacks (e.g., Sybil attacks) to avoid misallocation of data or route falsification.

ISMS should be able to monitor activity in the blockchain network in real time [24]. This allows you to quickly detect and respond to any suspicious activities or threats, including:

- anomaly analysis: the introduction of artificial intelligence and machine learning technologies for analysing blockchain transactions allows detecting anomalies that may indicate potential attacks, for example, a sharp increase in node activity may be a sign of preparation for a DoS attack [25];

- rapid response: the system should have Incident Response Plans that will allow you to quickly isolate suspicious nodes, block malicious transactions or change consensus rules in case of a threat [25].

Social engineering is one of the most common methods of CA, so it is important to consider the human factor in security issues. Educating users on the proper use of private keys, recognising phishing attacks and other fraudulent methods is critical.

A reasonable choice of protection methods in blockchain information security management systems should be based on risk analysis, technological capabilities, and a long-term security strategy. The combination of cryptographic methods, reliable consensus algorithms, secure network protocols and continuous monitoring will minimise the risks of CI and abuse.

According to the rationale for choosing security methods, an effective approach to ensuring security in a blockchain MIS is to use a monitoring and response method. One of the most effective solutions in this direction is the implementation of a protection system based on iptables, ipset, fail2ban and dynamic blacklist IP addresses.

The combination of these tools allows you to create a multi-level protection against man-in-the-middle at the network level. The use of iptables and ipset allows

for flexible configuration of network traffic filtering and blocking of suspicious IP addresses, while fail2ban automatically detects and blocks malicious access attempts, protecting the system from brute force attacks and other hacking attempts. Dynamic blacklisted IP addresses help to keep your system up-to-date and responsive to new threats, ensuring that malicious nodes are blocked in real time.

This approach ensures constant monitoring of activity in the blockchain network and allows for a prompt response to attempts at covert action, which significantly increases the overall security of the blockchain system and reduces the risk of compromise.

This approach ensures constant monitoring of activity in the blockchain network and allows for a prompt response to attempts at CRA, which significantly increases the overall security of the blockchain system and reduces the risk of compromise. The rapid development of web services has increased their vulnerability to various forms of CRA, including DDoS, brute force attacks, SQL injections and unauthorised access attempts, and the growing prevalence of CRA on web services necessitates the development of reliable protection mechanisms to mitigate threats and preserve digital assets. Due to the growing dependence on these services for business operations, financial transactions and personal data storage, the need for reliable protection mechanisms has become more critical than ever. Traditional security measures, such as firewalls and anti-virus software, are no longer sufficient to counter the increasing sophistication of modern spacecraft. In the context of developing a mechanism for protecting web services and blockchain MIS, the key requirement is to ensure high efficiency of detecting and blocking attacks in real time with minimal impact on system performance. For this purpose, it is proposed to use a comprehensive mechanism with dynamic protection, which will consist of the following tools: iptables, ipset, fail2ban, and dynamic blacklist IP addresses. Each of these components has its own unique advantages, allowing for a multi-level security system that adapts to new threats. This approach is aimed at reducing the risk posed by intruders, automating the detection and prevention of acts of cybercrime, and adapting defences in real time to changing attack vectors.

The development will use a complex mechanism with dynamic protection. Let's take a look at its main utilities: iptables, ipset, fail2ban, and dynamic blacklist IP addresses.

Iptables is a utility for configuring and managing packet filtering rules in the network stack of Linux distributions. It allows you to effectively control incoming, outgoing, and redirected traffic by setting rules based on IP addresses, ports, protocols, and other parameters [24]. The main reasons for choosing iptables are the ability to flexibly configure rules to block malicious traffic, high performance even with a large number of rules, embeddedness in the Linux kernel, which ensures reliability and

speed, and support for integration with other security tools such as ipset and fail2ban.

Ipset is an extension to iptables that allows you to work with large sets of IP addresses or other parameters within a single rule. Instead of creating a separate rule for each IP address, you can store them in sets, which significantly speeds up the system [25]. The main reasons for choosing ipset are support for large lists of IP addresses without significantly affecting performance, the ability to quickly update and modify lists while the system is running, and increased packet filtering efficiency when using dynamic blacklists.

Fail2ban is a tool for automatically blocking IP addresses that show suspicious activity or commit acts of terrorism based on the analysis of service logs (e.g., sshd, apache, nginx). fail2ban allows you to create rules for blocking attacks at the firewall level after a certain number of failed access attempts [24]. The main reasons for choosing fail2ban are automatic detection of suspicious activity and fast response, the ability to configure it using regular expressions to analyse logs of various services, integration with iptables to create blocking rules in real time, flexibility in setting the blocking period, which allows you to dynamically control access.

The use of dynamic IP address blacklists is necessary to protect against known attacking hosts [25]. These lists can be updated both locally (based on fail2ban rules) and through integration with external sources (reputation lists, known attacks, threats from botnets, etc.). The main reasons for choosing dynamic lists are the ability to automatically update and synchronise with global sources of threat information, dynamically add new IP addresses to the blacklist without the need to restart the firewall, and ensure high request processing speeds through the use of ipset.

3. Solution implementation process

At the first stage, basic rules are created using iptables and basic traffic filtering rules are configured, including restricting access to certain ports and services, if necessary, setting rules for blocking known malicious IP addresses, and configuring logging of suspicious traffic for further analysis.

The second step is to configure the ipsets for the dynamic lists. The next step is to create sets of IP addresses using ipset that will be used in iptables rules. This will allow you to quickly update the lists of blocked addresses without rebooting the entire system. Next, sets are created to store the IP addresses of attacking hosts. The sets can be updated with data from external sources or based on local observations.

Integration of fail2ban to automatically block attacks is configured by monitoring web service logs and other critical system components. If suspicious activity is detected (for example, multiple failed authentication attempts),

the system automatically adds the offending IP address to the blacklist via ipset and updates iptables rules.

Dynamic lists of blacklisted IP addresses are integrated with external services or databases to provide up-to-date information about IP addresses that are the source of threats. These lists are regularly updated to ensure protection against new threats.

The advantages of the proposed solution:

- speed and efficiency: using ipset to manage large lists of IP addresses ensures that rules are updated quickly without affecting system performance;

- adaptability: thanks to dynamic blacklists and automatic blocking via fail2ban, the system can instantly respond to new threats and adapt to them;

- flexibility and scalability: the solution can be easily scaled up for use on large web services or blockchain systems without significant changes to the infrastructure.

Thus, the proposed solution based on iptables, ipset, fail2ban and dynamic IP address lists allows you to create an effective multi-level anti-cannonballing system that will respond to threats in real time while maintaining high performance and flexibility.

The effectiveness of the security system can be assessed using a special formula [26]:

$$E = \frac{(M - N)}{(M + V + N)} \cdot 100\% \cdot (1 - B), \quad (1)$$

where E – system efficiency; M – number of cyberattacks detected and blocked by the system; N – number of cyberattacks prevented by the system; V – number of cyberattacks that have passed the cyber defence system; B – is the percentage of false positives generated by the system. The expression makes sense only if: $M > N$ и $B \in [0;1)$.

4. Software implementation of the security system

The software implementation of the system of protection against man-in-the-middle and abuse in blockchain systems involves the integration of several key tools, such as iptables, ipset, fail2ban, and dynamic blacklist IP addresses. These solutions provide automated protection at the network level and allow blocking suspicious activities and malicious nodes.

The main goal of this implementation is to provide protection against man-in-the-middle at the network level by automatically detecting and blocking malicious actions such as brute force attempts, DoS, DDoS and other abuses.

Program Description. Iptables is the main tool for configuring a firewall on Linux, allowing you to create rules to control traffic on the network [24]. For a blockchain system, it is important to set up filtering to restrict access to certain ports used in network protocols and prevent attacks.

Basic setup steps:

- identify critical ports used by the blockchain infrastructure (e.g. for transactions and block confirmation);

- create rules to allow traffic to these ports only for certain IP addresses or geographical areas;

- use the DROP policy for all unknown or suspicious traffic: iptables -P INPUT DROP or, for example, allowing the Bitcoin port, iptables -A INPUT -p tcp --dport 8333 -j ACCEPT;

- limiting the number of connections from one IP to prevent DoS or DDoS: iptables -A INPUT -p tcp --dport 8333 -m connlimit --connlimit-above 10 -j DROP.

- ipset is a tool that allows you to efficiently manage large lists of IP addresses, making it easier to implement dynamic blacklists [25]. Using ipset in conjunction with iptables allows you to quickly update rules without the need for a full table reload.

The description of software scripts and configuration files is presented in Tables 1 and 2.

Table 1

Description of software scripts

Name	Description
blip.sh	BlackIP – is a project that collects and unifies public blacklists of IP addresses and subnets to make them compatible with ipset. Usage.: ./blip.sh a6o bash blip.sh.
check.sh	Checks IP addresses or subnets in the ipset lists (blnet, blip, addonnet, addonnet, addonip, geonet, geoip, f2bip) and the fail2ban ban list (/var/log/fail2ban.log). Usage.: ./check.sh IP_or_SubNET.
ipset-check.sh	Checking the ipset list, if there is nothing in the ipset, then ipset lists are created and the script is executed ipset.sh.
ipset.sh	Downloading blacklists of IP addresses and subnets from the Github repository, adding them to ipset lists (blnet, blip, addonnet, addonnet, addonip, geonet, geoip), adding IP addresses to the ipset list (f2bip), adding IP addresses and subnets to the ipset whitelist (wlip ra wlnet).
fail2ban-status.sh	Script to check the fail2ban client status in a file /etc/fail2ban/jail.d/*.conf (nginx-limit-req, nginx-conn-limit, nginx-dos, nginx-badbots, nginx-4xx, sshd).
unban.sh	A script to unban (place the banned IPs and subnets in the unban.txt file) the banned IPs and subnets from списків ipset (blnet, blip, addonnet, addonnet, addonip, geonet, geoip, f2bip), з бачу fail2ban.
wlset.sh	Script to add (place IP addresses and/or subnets in the wlset.txt file) IP addresses and subnets to the whitelist ipset (wlip ra wlnet).

Table 2

Опис конфігураційних файлів

Name	Description
/etc/fail2ban/action.d/	The directory contains action configuration files that define what fail2ban will do when it detects suspicious activity according to the filter rules. Actions can include: blocking IP addresses with a firewall, sending alerts, executing other commands or scripts. Each file is responsible for a specific action. By default, there are rules for iptables, nftables, tcpwrappers, shorewall.
/etc/fail2ban/filter.d/	This directory contains the configuration files for the filters that define which log messages should be considered suspicious or malicious. The filter files contain regular expressions that fail2ban uses to search for specific log entries. Each configuration file can correspond to a specific service (e.g. sshd, apache, etc.).
/etc/fail2ban/jail.d/	This directory contains additional or individual configuration files for 'jails'. They supplement or overwrite the basic settings from the jail.conf file. jails is a set of rules that define which filters to use and what actions to take when suspicious activity is detected. This allows you to flexibly configure protection for different services.
/etc/fail2ban/jail.conf	The main configuration file for setting up jails. It defines the general parameters for each service (or jail), such as the filter to be applied, the time to block IP addresses, the number of failed login attempts before the ban is activated, etc. However, this file is usually not edited directly, as there is a risk that changes may be overwritten when fail2ban is updated. Instead, it is recommended to create or edit separate files in the /etc/fail2ban/jail.d/.

5. Experimental testing of components

The experiment included testing of components such as iptables, ipset, fail2ban and dynamic blacklist IP addresses to assess the system's ability to detect and block malicious traffic and provide resistance to various types of CA in the IS.

For the experiment, a virtual environment with a deployed blockchain network was used to simulate various types of attacks, including brute force, DoS, DDoS, and attempts to gain unauthorised access to blockchain nodes. Test environment:

- operating system: Ubuntu 24.04 LTS;
- security tools: iptables, ipsets, fail2ban;
- type of attacks: brute force on SSH, DoS, DDoS on the blockchain port, simulation of malicious traffic through port scanning, attacks on the web server;
- testing was carried out for 7 days with different loads.

One of the system's key tasks was to detect and block brute force attacks aimed at gaining access to critical blockchain nodes via SSH, web server, and any other open port on the network. In this scenario, fail2ban monitored

system logs for failed authorisation attempts and automatically blocked the attackers' IP addresses. Results.:

- the average time to block an IP address after detecting suspicious activity was 1–2 seconds;
- fail2ban the average time to block an IP address after detecting suspicious activity was 1–2 seconds detected and blocked 95% of all attempts brute force;
- the remaining 5% are IP addresses that have carried out low-intensity attacks, but have been added to the blacklist based on the dynamic updating of blacklisted IP addresses.

These results demonstrate the high efficiency of fail2ban in detecting and blocking brute force attacks, as well as the need to use dynamic blacklisted IP addresses to handle low-intensity attacks (this percentage may increase in a real-world environment).

The experiment also tested the system's ability to protect against DoS and DDoS attacks on blockchain network ports. The attacks used a large volume of malicious requests aimed at overloading the server. Results.:

- using iptables rules and configuring the Nginx web server to limit the number of connections from one IP reduced the intensity of DDoS attacks by 85% by blocking suspicious IPs after exceeding a threshold;
- the system detected and blocked almost all malicious requests through IP address blacklists and ipset; distributed botnets, where the attack came from many IP addresses, the average blocking time for each malicious IP was 3–5 seconds.

Protection against DDoS attacks proved to be effective, but response times were slightly longer for significant distributed attacks due to the dynamic nature of updating IP blacklists.

Dynamic lists of blacklisted IP addresses, which were updated through external sources, played a key role in protecting the system from new and little-known attacks. These lists were constantly updated with the latest data on malicious IP addresses, which helped maintain a high level of security. Results.:

- automatic updating of the lists took place twice a day, which ensured timely blocking of new threats;
- automatic updates of the lists occurred twice a day, which ensured timely blocking of new threats more than 375 IP addresses from dynamic lists that were not detected during traffic monitoring by iptables or fail2ban.

The load on the system from the use of security tools was minimal. The use of iptables, ipset, and fail2ban did not cause significant delays in the operation of the blockchain network. Results.:

- the CPU load during the blocking of DDoS attacks increased by about 5%, but this did not affect the overall system performance;
- the response time of blockchain nodes to requests remained stable even during attacks, which indicates the high performance of solutions.

The operands for calculating the effectiveness of a security system are the number of attacks that were blocked after the system warned of them, the total number of attacks that were blocked, missed, or warned, and a multiplier that reduces the overall effectiveness based on the percentage of false positives. If the system frequently blocks legitimate traffic, its effectiveness decreases. The effectiveness of the system depends on the number of attacks that are successfully blocked or prevented by the system [26].

Table 3 shows the results of a week's worth of monitoring of the targeted web server. In the study, the system blocked or prevented most attacks, but false positives reduced the overall effectiveness of the system. To improve efficiency, the number of false positives can be reduced and/or the accuracy of malicious activity detection can be increased. According to the results of the weekly monitoring, 663 attacks were detected, and the final efficiency of the protection system was approximately 88%. This means that the system blocked and prevented most attacks, but some attacks (25 cases) passed through the system, there was a small number of false positives (20 cases), and 18 cases of malicious acts were detected that were prevented by the system, i.e. detected by cyber defence tools but not blocked, and the system sent notifications about such threats so that measures could be taken to further neutralise them.

Table 3

Description of the configuration files

M	N	V	B	Efficiency (%)
50	1	3	0.12	79.85
100	3	6	0.09	80.98
150	4	10	0.0667	83.09
200	7	13	0.05	83.34
250	10	15	0.04	83.78
300	11	16	0.04	84.84
350	12	18	0.04	85.39
400	14	20	0.0375	85.6
450	16	21	0.0356	85.94
500	17	22	0.032	86.74
550	17	22	0.031	87.69
600	18	25	0.031	87.71

The results demonstrate that the effectiveness of the system can vary significantly depending on the number of blocked attacks, the number of spacecraft that passed through the system, and the percentage of false positives. To achieve maximum efficiency, it is important to minimise the number of missed attacks and false positives.

Conclusions

The software implementation of the blockchain security system using iptables, ipset, fail2ban, and dynamic blacklist IP addresses provides reliable multi-level protection at the network level. It allows you to quickly detect and block suspicious activity, protecting the system from brute force attacks, DoS, DDoS, and other threats, providing continuous monitoring and automatic response to threats in real time. This implementation of a dynamic multi-level security mechanism offers an effective and scalable solution for protecting web services from man-in-the-middle attacks. The system's adaptability and the ability to integrate real-time threat intelligence increase its effectiveness in countering evolving cyber threats. The system's ability to automatically detect and block malicious IP addresses combined with low resource consumption makes it a scalable and practical solution for both small and large web applications. Future improvements could focus on further automation and incorporation of machine learning and artificial intelligence techniques for adaptive threat detection.

The study demonstrates the high efficiency and feasibility of implementing the proposed protection mechanisms in the ISMS in the blockchain and web services. The developed approaches not only provide a high level of protection against cyberattacks and abuse, but also have the potential for further development and integration with other protection systems in the information infrastructure and IS of modern enterprises.

References

- [1] Arun J. S., Gaur N., Cuomo J. Blockchain for Business. Addison-Wesley Professional: 1st edition. 2019. 224 p.
- [2] The 51% attack on blockchains: A mining behavior study / Aponte-Novoa F. A., Orozco A. L. S., Villanueva-Polanco R., Wightman P. IEEE. 2021. Vol. 9. pp. 549 – 564. DOI:10.1109/ACCESS.2021.3119291.
- [3] Anita N., Vijayalakshmi M. Blockchain security attack: A brief survey in Proc. 10th Int. Conf. Comput., Commun. Netw. Technol. (ICC-CNT). 2019. pp. 6 – 11. DOI:10.1109/ICCNT45670.2019.8944615.
- [4] A survey on the security of blockchain systems / Li X., Jiang P., Chen T., Luo X., Wen Q. Future Generation Computer Systems. 2020. Vol. 107. pp. 841 – 853. DOI:10.1016/j.future.2017.08.020.
- [5] What is a finney attack? URL: <https://bitcoin.stackexchange.com/questions/4942/what-is-a-finney-attack> (дата звернення: 04.03.2025).
- [6] Aggarwal S., Kumar N. Chapter Twenty – Attacks on blockchain. Advances in Computers. 2021. Vol. 121. pp. 399 – 410. DOI: <https://doi.org/10.1016/bs.adcom.2020.08.020>.
- [7] Vyas C. A., Lunagaría M. Security concerns and issues for bitcoin. National Conference cum Workshop on Bioinformatics and Computational Biology, NCWBCB. 2014. pp. 10 – 12. URL: <https://research.ijcaonline.org/ncwbcb/number2/ncwbcb1414.pdf> (дата звернення: 04.03.2025).

- [8] DeCusatis C., Zimmermann M., Sager A. Identity-based network security for commercial Blockchain services. IEEE 8th Annual Workshop and Conference on Computing and Communication. 2018. pp. 474 – 477. DOI:10.1109/CCWC.2018.8301713.
- [9] Sharma P. K., Moon S. Y., Park J. H. Block-VN: a distributed Blockchain based vehicular network architecture in smart city. Journal of Information Processing Systems. 2017. Vol. 13. No. 1. pp. 184 – 195. DOI: 10.3745/JIPS.03.0065.
- [10] Swathi P., Modi C., Patel D. Preventing Sybil Attack in Blockchain using Distributed Behavior Monitoring of Miners. 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT). 2019. pp. 1 – 6. DOI:10.1109/ICCCNT45670.2019.8944507.
- [11] Mining on someone else's dime: Mitigating covert mining operations in clouds and enterprises / Tahir R., Huzaifa M., Das A., Ahmad M., Gunter C. A., Zaffar F., Caesar M., Borisov N. 20th International Symposium on Research in Attacks, Intrusions and Defenses (RAID), Atlanta. 2017. pp. 287 – 310. DOI: https://doi.org/10.1007/978-3-319-66332-6_13.
- [12] Saad M., Khormali A., Mohaisen A. End-to-end analysis of in-browser cryptojacking. CoRR. 2018. Vol. abs/1809.02152. 15 p. DOI: <https://doi.org/10.48550/arXiv.1809.02152>.
- [13] Blockchain Attack Vectors: Main Vulnerabilities of Blockchain Technology. URL: <https://www.apriorit.com/dev-blog/578-blockchain-attack-vectors> (дата звернення: 04.03.2025).
- [14] Misbehavior in bitcoin: a study of double-spending and accountability / Karame G. O., Androulaki E., Roeschlin M., Gervais A., apkun S. ACM Transactions on Information and System Security. 2015. Vol. 18. No. 1. pp. 1 – 32. DOI:10.1145/2732196.
- [15] Nicolas K., Yi W. A novel double spending attack countermeasure in blockchain. IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). 2019. pp. 383 – 388. DOI:10.1109/UEMCON47517.2019.8992991.
- [16] Блокчейн інфраструктура для захисту кіберсистем / Адамов О.С., Хаханов В.І., Чумаченко С.В., Абдуллаєв В.Г. Радіоелектроніка та інформатика. 2018. №4 (83). С. 64 – 85. URL: <https://journals.indexcopernicus.com/search/article?articleId=2146726> (дата звернення: 04.03.2025).
- [17] Proof of Stake versus Proof of Work White Paper. URL: <https://bitfury.com/content/downloads/pos-vs-pow-1.0.2.pdf> (дата звернення: 04.03.2025).
- [18] Everything you Wanted to Know about the Blockchain / Puthal D., Malik N., Mohanty S. P., Kougiianos E., Das G. IEEE Consumer Electronics Magazine. 2018. Vol. 7. No. 4. pp. 06 – 14. DOI:10.1109/MCE.2018.2816299.
- [19] Analysis of cryptographic authentication and manipulation detection methods for big data / Havrylova A. A., Korol O. G., Voropay N. I., Sevriukova Y. O., Bondarenko K. O. Сучасний захист інформації. 2024. Vol. 1(57). pp. 97 – 102. DOI: <https://doi.org/10.31673/2409-7292.2024.010011>.
- [20] Горбенко І. Д. Гриненко Т. О. Захист інформації в інформаційно-телекомунікаційних системах: Навч. посібник. Ч.1. Криптографічний захист інформації. Харків : ХНУРЕ, 2004. 368 с.
- [21] Internet of Things, Blockchain and Shared Economy Applications / Huckle S., Bhattacharya R., White M., Beloff N. Procedia Computer Science. 2016. Vol. 98. pp. 461–466. DOI:10.1016/j.procs.2016.09.074.
- [22] Blockchain Technology in Financial and Banking Sector / Ragha L., Dixit A., Rodrigues B., Yadav K. International Journal of Trend in Research and Development. 2018. Vol. 1. pp. 41 – 44. URL: <http://www.ijtrd.com/papers/IJTRD15855.pdf> (дата звернення: 04.03.2025).
- [23] Курочкіна М. Г. Блокчейни – новітня технологія криптографії в цифровому світі. Світ телекомунікації та інформатизації: матеріали Міжнародної науково-технічної конференції студентства Державного університету телекомунікацій. Київ: ДУТ, 2017. С. 209–212. URL: http://www.dut.edu.ua/uploads/n_5218_58757739.pdf (дата звернення: 04.03.2025).
- [24] Bhuyan M. H., Bhattacharyya D. K., Kalita J. K. Network anomaly detection: Methods, systems and tools. Proceedings of IEEE Communications Surveys Tutorials. 2014. Vol. 16(1). pp. 303 – 336. DOI: <https://doi.org/10.1109/SURV.2013.052213.00046>.
- [25] Unsupervised anomaly detection via variational autoencoder for seasonal kpis in web applications / Xu H., Chen W., Zhao N., Li Z., Bu J., Li Z., Liu Y., Zhao Y., Pei D., Feng Y., Chen J., Wang Z., Qiao H. Proceedings of Proceedings of the 2018 World Wide Web Conference. 2018 pp. 187 – 196. DOI: <https://doi.org/10.1145/3178876.3185996>.
- [26] Савіцький Л. М., Безносенко С. Ю., Горбач Р. Я. Концептуальні погляди на побудову системи захисту від кібератак із застосуванням методів штучного інтелекту в інформаційно-комунікаційних системах. Сучасні інформаційні технології у сфері безпеки та оборони. 2024. № 1(49). С. 77 – 85. DOI: 10.33099/2311-7249/2024-49-1-77-85.

Date of submission of the article to the editorial board: 10.05.2025

ПРАВИЛА оформлення рукописів для авторів науково-технічного журналу «БІОНІКА ІНТЕЛЕКТУ»

Науково-технічний журнал «Біоніка інтелекту» приймає до друку написані спеціально для нього оригінальні рукописи, які раніше ніде не друкувались. Структура рукопису повинна бути такою: індекс УДК, відомості про авторів, заголовок, анотації (на трьох мовах), ключові слова, вступ, основний текст статті, висновки, список використаної літератури, резюме.

Відповідно до Постанови ВАК України від 15.01.2003 №7-05/1 (Бюлетень ВАК, №1, 2003, с. 2), стаття повинна мати такі необхідні елементи: постановка проблеми у загальному вигляді та її зв'язок із важливими науковими чи практичними завданнями; аналіз останніх досліджень і публікацій і виділення не вирішених раніше частин загальної проблеми в даній області; формулювання цілей та завдань дослідження; виклад основного матеріалу досліджень з повним обґрунтуванням отриманих наукових результатів; висновки з даного дослідження та перспективи подальших досліджень у даному напрямку.

Статті мають бути виконані в редакторі Microsoft Word. Формат сторінки — А4 (210×297 мм), поля: верхнє — 25 мм, нижнє — 20 мм, ліве, праве — 17 мм. Кількість колонок — 2, з інтервалом між ними 5 мм, основний шрифт Times New Roman, кегль основного тексту — 10 пунктів, міжрядковий інтервал — множник (1,1), абзацний відступ — 6 мм. Обсяг рукопису — від 6 до 12 сторінок (мови: українська, англійська, російська та мовою оригінала).

УДК друкується з першого рядка, без відступів, вирівнювання по лівому краю.

ПІБ автора (-ів), назва статті, назва та адреса учбового закладу необхідно надати повністю російською, українською та англійською мовами.

Назва статті друкується прописними літерами; шрифт прямий, напівжирний, кегль 12.

Назви розділів нумерують арабськими цифрами, виділяють жирним шрифтом. Відступи для назви статті, ініціалів та прізвищ авторів, відомостей про авторів, назв розділів, вступу та висновків, списку літератури: зверху — 6 пт, знизу — 3 пт.

Анотації (мовою статті, абзац 6—12 рядків, кегль 9) розміщують на початку статті, в ній має бути розміщена інформація про очікувані результати описаних досліджень (на трьох мовах).

Ключові слова (4—10 слів з тексту статті, які з точки зору інформаційного пошуку несуть змістовне навантаження) наводять мовою рукопису, через кому в називному відмінку, кегль 9.

Рисунки та таблиці (чорно-білі, контрастні) розміщуються у тексті після першого посилання у вигляді окремих об'єктів і нумерують арабськими цифрами наскрізною нумерацією за наявності більше ніж одного об'єкта. Невеликі схеми, що складаються з 3—4 елементів виконують, використовуючи вставку об'єкта Рисунок Microsoft Word. Більш складні виконують у графічних редакторах у вигляді чорно-білих графічних файлів форматів .tif, .jpg, .wmf, .cdr із

розділенням 300 dpi. Рисунки мають міститися у текстовому файлі й обов'язково подаватися окремими файлами з відповідними назвами (наприклад, рис1.jpg).

Усі елементи рисунка, включаючи написи, повинні бути згруповані. Усі написи в рисунках і таблицях мають бути виконані шрифтом Times New Roman, кегль у рисунках — 10, у таблицях — 9.

Рисунок повинен мати центрований підпис (поза рисунком), шрифт 9, відступи зверху і знизу по 6 пт. Ширина рисунка має відповідати ширині колонки (або ширині сторінки).

Формули, символи, змінні повинні бути набрані в редакторі формул **MathType**. Формули розміщують посередині рядка й нумерують за наявності посилань на них у рукописі. Шрифт — Times New Roman. Висота змінної — 10 пунктів, великих і малих індексів — 8 пт, основний математичний символ — 12 (10) пт. Змінні, позначені латинськими літерами, набирають курсивом, грецькі літери, скорочення російських слів і цифри — прямим написанням. Змінні, які є в тексті, також набирають у редакторі формул.

Список літератури вміщує опубліковані джерела, на які є посилання в тексті, укладені у квадратні дужки, друкують без абзацного відступу, кегль 9 пт, відступ зверху — 6 пт.

Після списку літератури з відступом зверху 6 пт зазначають *дату подання статті до редколегії*. Число та місяць задають двозначними числами через крапку. Розмір шрифту — 9 пт, курсив, вирівнювання по правому краю.

Резюме (Times New Roman, кегль — 10 пунктів,) подають англійською мовою: обсяг резюме до 2000 знаків (бажаний переклад). *Структура резюме: Background, Materials and methods, Results, Conclusion.*

Разом із рукописом (на аркушах білого паперу формату А4 щільністю 80–90 г/м², надрукований на лазерному принтері) необхідно подати такі документи:

1. Заяву, яку повинні підписати всі автори.
2. Акт експертизи про можливість опублікування матеріалів у відкритому друці (якщо потрібно).
3. Рецензію, підписану доктором чи кандидатом наук.
4. Відомості про авторів.
5. Електронний варіант рукопису, резюме та відомостей про авторів.
6. Зробити оплату публікації.

Необхідно також зазначити один з наступних тематичних розділів, якому відповідає рукопис:

1. Теоретичні основи інформатики та кібернетики. Теорія інтелекту.
2. Математичне моделювання. Системний аналіз. Прийняття рішень.
3. Інтелектуальна обробка інформації. Розпізнавання образів.
4. Інформаційні технології та програмно-технічні комплекси.
5. Структурна, прикладна та математична лінгвістика.
6. Дискусійні повідомлення.

ЗМІСТ

СТРУКТУРНА, ПРИКЛАДНА ТА МАТЕМАТИЧНА ЛІНГВІСТИКА

Dorozhynska A., Dorozhynskyy S. From dictionary to database: transformation and protection 3

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ. МАШИННЕ НАВЧАННЯ. БАЗИ ДАНИХ. РОЗПІЗНАВАННЯ ОБРАЗІВ

Лозинська О. В., Марків О. О., Висоцька В. А. Метод виявлення джерел дезінформації на основі ансамблевих моделей машинного навчання 11

Кириченко І. В., Терещенко Г. Ю., Ніколенко В. В. Дослідження методів зберігання візуальних даних 20

Линник Р. О., Висоцька В. А. Аналіз проблематики виявлення трендів громадської думки в україномовних дописах засобами кластеризації та нейронних мереж 27

Кирий В. В., Калиниченко О. В. Прогнозування фінансових показників ринку кіновиробництва засобами штучного інтелекту 37

Назаров О. С., Биковська С. А., Назарова Н. В. Дослідження методів анімації використовуючи фреймворк Flutter 46

Удовенко С. Г., Затхей В. А., Тесленко О. В., Чала Л. Е. Автоматизоване тестування програмного забезпечення модулів інформаційних систем 52

Нуралієва Л. М., Назаров О. С., Назарова Н. В. Дослідження методів програмної оптимізації завантаження сайтів на формування досвіду користувачів 64

ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ. НЕЙРОННІ МЕРЕЖІ. ГЕОПРОСТОРОВИЙ АНАЛІЗ

Шаронова Н. В., Плехова Г. А., Неронов С. М., Костікова М. В., Плехов Д. О. Спосіб інтеграції різнорідних даних в системі геопросторового аналізу 70

Плехова Г. А., Неронов С. М., Костікова М. В., Асаєнко Ю. С., Плехов Д. О. Пристрій обробки різноманітних гетерогенних даних в системах прийняття рішень 75

Золотухін О. В., Кудрявцева М. С., Єрмесєв Є. Ю. Стохастична ініціалізація нейронних мереж на основі аналізу біологічних систем 79

Шаронова Н. В., Плехова Г. А., Неронов С. М., Костікова М. В., Плехов Д. О. Спосіб формування маршрутів передачі даних в мобільних інформаційних мережах 89

Ivanenko O., Musaiev Z. Use of 3D model in the study of tower crane stability 94

ІНФОРМАЦІЙНА БЕЗПЕКА ТА НАДІЙНІСТЬ СИСТЕМ

Brynza N., Borodavka V., Teslenko O. Implementation of measures to counteract cyber threats and abuse in information security management systems based on blockchain technologies 98

ПРАВИЛА

оформлення рукописів для авторів науково-технічного журналу «БІОНІКА ІНТЕЛЕКТУ» 108

Наукове видання

БІОНІКА ІНТЕЛЕКТУ
інформація, мова, інтелект

Науково-технічний журнал

№ 1 (102)
2025

Головний редактор — *Г. Г. Четвериков*
Відповідальний редактор — *І. В. Кириченко*

Комп'ютерна верстка — *О. Б. Ісаєва*

Рекомендовано секцією № 2 «Інформаційні технології»
науково-технічної ради Харківського національного університету радіоелектроніки
(протокол № 5 від «19» червня 2025 р.)

Адреса редакції:
Україна, 61166, Харків-166, просп. Науки, 14,
Харківський національний університет радіоелектроніки, к. 127
тел. 702-14-77, факс 702-10-13,
e-mail: bionics@nure.ua

Підписано до друку 30.06.2025. Формат 60 × 84 1/8. Друк ризографічний.
Папір офсетний. Гарнітура Newton. Умов. друк. арк. 12,8. Обл.-вид. арк. 12,6.
Тираж 20 прим.

Віддруковано в редакційно-видавничому відділі ХНУРЕ
61166, Харків, просп. Науки, 14.