Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)
Кафедра комп'ютерних наук
(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

магістр
(назва освітнього ступеня)

на тему: Інформаційні системи з використанням релевантного пошуку на основі Elasticsearch та Apache Solr

Виконав: студент VI курсу, групи СНнм-61
спеціальності 122 Комп'ютерні науки
(шифр і назва спеціальності)

_____ Дзядик Т.О.
(підпис) (прізвище та ініціали)

Керівник _____ Фриз М.Є.
(підпис) (прізвище та ініціали)

Нормоконтроль _____ Дуда О.М.
(підпис) (прізвище та ініціали)

Завідувач кафедри _____ Боднарчук І.О.
(підпис) (прізвище та ініціали)

Рецензент _____ Гащин Н.Б.
(підпис) (прізвище та ініціали)

Тернопіль
2024

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра комп'ютерних наук
(повна назва кафедри)

ЗАТВЕРДЖУЮ
Завідувач кафедри
_____ Боднарчук І.О.
(підпис) (прізвище та ініціали)
«_23_» _травня_ 2024 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

на здобуття освітнього ступеня Магістр
(назва освітнього ступеня)

за спеціальністю 122 Комп'ютерні науки
(шифр і назва спеціальності)

Студенту Дзядику Тарасу Орестовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Інформаційні системи з використанням релевантного пошуку на основі Elasticsearch та Apache Solr

Керівник роботи Фриз Михайло Євгенович, кандидат технічних наук, доцент кафедри КН
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від «_29_» _листопада_ 2024 року № _4/7-1139_

2. Термін подання студентом завершеної роботи 22 травня 2024р.

3. Вихідні дані до роботи Наукові публікації про інформаційні системи для релевантного пошуку, інформаційні технології, «Elasticsearch» та «Apache Solr»

4. Зміст роботи (перелік питань, які потрібно розробити)
Вступ. 1 Аналіз предметної області, пошукові системи, ключові підходи та алгоритми в інформаційних системах пошуку та індексації. 2 Провідні системи, технології пошуку та нейро-символьні підходи до пошуку інформації. 3 Інформаційна система з використанням релевантного пошуку. 4 Охорона праці та безпека в надзвичайних ситуаціях. Висновки. Додатки

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)
1 Титьна сторінка. 2 Тема, Мета, Об'єкт, Предмет дослідження. 3 Завдання дослідження. 4 Актуальність дослідження. 5 Порівняння рішень для пошуку зображень та їх технічних підходів. 6 Ключові компоненти «Elastic Stack». 7 Переваги Elasticsearch у пошуку даних. 8 Архітектура інформаційної системи з використанням релевантного пошуку. 9 Інтеграція Solr. 10 Індексація «Apache Solr». 11 Сумісність з різними форматами даних. 12 Перспективні напрямки використання. 13. Участь в конференціях. 14 Висновки. 17 Завершальний слайд.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Охорона праці	Сенчишин В.С., доцент		
Безпека в надзвичайних ситуаціях	Клепчик В.М., ст. викладач		

7. Дата видачі завдання 29 листопада 2024 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Ознайомлення з завданням до кваліфікаційної роботи	29.11.2024	
2.	Підбір наукових джерел про інформаційні системи	02.12.2024-15.12.2024	
	для релевантного пошуку, інформаційні технології,		
	«Elasticsearch» та «Apache Solr»		
3.	Опрацювання наукових публікацій та збір даних по	16.12.2024-29.12.2024	
	темі роботи		
4.	Виконання дослідження згідно мети кваліфікаційної	13.01.2025-06.04.2025	
	роботи		
5.	Оформлення розділу «Аналіз предметної області,	14.04.2025-20.04.2025	
	пошукові системи, ключові підходи та алгоритми в		
	інформаційних системах пошуку та індексації»		
6.	Оформлення розділу «Провідні системи, технології	21.04.2025-27.04.2025	
	пошуку та нейро-символьні підходи до пошуку		
	інформації»		
7.	Оформлення розділу «Інформаційна система з	28.04.2025-04.05.2025	
	використанням релевантного пошуку»		
8.	Виконання завдання до підрозділу «Охорона праці»	05.05.2025-07.05.2025	
9.	Виконання завдання до підрозділу «Безпека в	08.05.2025-11.05.2025	
	надзвичайних ситуаціях»		
10.	Оформлення кваліфікаційної роботи	12.05.2025-18.05.2025	
11.	Нормоконтроль	19.05.2025-20.05.2025	
12.	Перевірка на плагіат	21.05.2025	
13.	Попередній захист кваліфікаційної роботи	22.05.2025	
14.	Захист кваліфікаційної роботи	26.05.2025	

Студент _____ Дзядик Т.О.
(підпис) (прізвище та ініціали)

Керівник роботи _____ Фриз М.Є.
(підпис) (прізвище та ініціали)

АНОТАЦІЯ

Інформаційні системи з використанням релевантного пошуку на основі Elasticsearch та Apache Solr // Кваліфікаційна робота освітнього рівня «Магістр» // Дзядик Тарас Орестович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем і програмної інженерії, кафедра комп'ютерних наук, група СНнм-61 // Тернопіль, 2025 // С. 81, рис. – 3, табл. – 8, кресл. – 15, додат. – 1, бібліогр. – 72.

Ключові слова: пошук, інформаційна система, зображення, нейро-символічний, обробка природної мови, Elasticsearch, Apache Solr.

Кваліфікаційна робота присв'ячена розробці інформаційної системи з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr». В першому розділі кваліфікаційної роботи розглянуто контекст та актуальність розробки інформаційної системи з використанням релевантного пошуку. Проведено аналіз предметної області. Описано популярні пошукові системи зображень системи пошуку та індексації зображень. Проаналізовано процес пошуку зображень людиною. Досліджено ключові підходи та алгоритми в системах пошуку та індексації зображень. В другому розділі кваліфікаційної роботи описано провідні системи пошуку та індексації зображень. Описано існуючі інформаційні технології пошуку та індексації зображень, зокрема технології від Google. Розглянуто нейро-символьні підходи до пошуку інформації за зображеннями. Проведено огляд дата-пайплайнів. Також було провеено огляд принципів ключових принципів проєктування пайплайнів для обробки даних нового покоління. В третьому розділі кваліфікаційної роботи спроєктовано архітектуру інформаційної системи з використанням релевантного пошуку. Описано платформи та технологічний стек. Розглянуто інтеграцію та взаємодію інформаційної системи з використанням релевантного пошуку. Висвітлено безпекові аспекти використання інформаційної системи з використанням релевантного пошуку та відповідність вимогам. Подано особливості застосування інформаційної системи.

ANNOTATION

Information Systems Using Relevant Search Based on Elasticsearch and Apache Solr // The educational level "Master" qualification work // Dziadyk Taras // Ternopil Ivan Pulyuy National Technical University, Faculty of Computer Information Systems and Software Engineering, Department of Computer Science, SNnm-61 group // Ternopil, 2025 // P. 81, fig. - 3, tables - 8, posters - 15, annexes - 1, ref. - 72.

Key words: search, information system, image, neuro-symbolic, natural language processing, Elasticsearch, Apache Solr

The qualification thesis is dedicated to the development of an information system utilizing relevant search based on "Elasticsearch" and "Apache Solr". The first chapter of the qualification work discusses the context and relevance of developing an information system with the use of relevant search. An analysis of the subject area is provided. Popular image search systems and image indexing systems are described. The process of image search by humans is analyzed. Key approaches and algorithms in image search and indexing systems are explored.

The second chapter of the qualification work describes leading image search and indexing systems. Existing information technologies for image search and indexing, particularly those by Google, are outlined. Neuro-symbolic approaches to image-based information retrieval are discussed. An overview of data pipelines is provided. Additionally, an overview of the key principles of designing next-generation data processing pipelines is presented.

The third chapter of the qualification work outlines the architecture of the information system using relevant search. The platforms and technological stack are described. The integration and interaction of the information system with the relevant search system are discussed. Security aspects of using the information system with relevant search and its compliance with requirements are highlighted. The specifics of applying the information system are presented.

ПЕРЕЛІК СКОРОЧЕНЬ І ТЕРМІНІВ

API (англ. Application Programming Interface) – це набір інструкцій, які дозволяють різним програмам спілкуватись між собою та обмінюватись даними.

AWS (англ. Amazon Web Services) – дочірня компанія Amazon.com, що надає платформу хмарних обчислень в оренду приватним особам, компаніям та урядам на основі платної підписки.

CSV (англ. Comma-Separated Values) – файл даних із роздільниками-комами.

ELT (англ. Extract, Load, Transform) – або Витяг, Перетворення та Завантаження – процес, який використовується в базах даних та, особливо, у сховищах даних та у засобах Business Intelligence для забезпечення їх роботи для підтримки прийняття рішень.

ETL (англ. Extract, Transform, Load) – ELT.

GCP (англ. Google Cloud Platform) – запропонований компанією Google набір хмарних служб, які виконуються на тій же самій інфраструктурі, яку Google використовує для своїх продуктів призначених для кінцевих споживачів, таких як Google Search та YouTube.

IAM (англ. Identity and Access Management) – частина Amazon Web Services що дозволяє керувати доступом до ресурсів AWS.

JSON (англ. JavaScript Object Notation) – текстовий формат обміну даними між комп'ютерами.

KMS (англ. Key Management Service) – безпечний та надійний хмарний сервіс, за допомогою якого можна централізовано керувати ключами.

ML (англ. Machine Learning) – машинне навчання.

XML (англ. EXtensible Markup Language) – запропонований консорціумом World Wide Web Consortium (W3C) стандарт побудови мов розмітки ієрархічно структурованих даних для обміну між різними застосунками, зокрема, через Інтернет.

ШІ – штучний інтелект.

ЗМІСТ

ВСТУП

Актуальність теми. В умовах стрімкого зростання обсягів інформації та ускладнення цифрових сервісів, створення сучасних інформаційних систем з підтримкою релевантного пошуку стає надзвичайно актуальним. Традиційні підходи до зберігання, обробки та пошуку даних більше не відповідають вимогам часу, оскільки не здатні ефективно реагувати на високі швидкості, великі обсяги та різноманітність даних, які генеруються в реальному часі. Це призводить до затримок у доступі до важливої інформації, погіршує якість прийняття рішень та знижує загальну продуктивність бізнес-процесів. Водночас користувачі очікують миттєвого та точного пошуку інформації, що є ключовим чинником успішної взаємодії з будь-якою інформаційною системою.

Інформаційна система, побудована з урахуванням потреб релевантного пошуку, має забезпечувати не лише швидкий доступ до даних, а й інтелектуальне ранжування результатів відповідно до змісту запиту, контексту користувача та семантичних зв’язків між даними. Це особливо важливо в умовах хмарних середовищ, де дані постійно змінюються та оновлюються. Для цього необхідне впровадження сучасних технологій — обробки потоків даних у режимі реального часу, безсерверної інфраструктури, інтелектуальної маршрутизації та оркестрації даних, а також безперервної інтеграції з моделями машинного навчання для адаптивного аналізу та класифікації інформації.

Сучасні архітектурні підходи передбачають використання таких інструментів, як Apache Kafka, Apache Flink, Elasticsearch, Solr та AWS Kinesis, які спільно дозволяють створити гнучкий і масштабований дата-пайплайн. Така система забезпечує автоматизоване збирання, очищення, збагачення та індексацію даних з подальшим виконанням релевантного пошуку у реальному часі. Це дозволяє не тільки зменшити затримки в обробці інформації, а й підвищити точність та контекстуальну відповідність результатів пошуку.

У результаті, побудова інформаційної системи нового покоління з фокусом на релевантному пошуку є відповіддю на сучасні виклики цифрової епохи. Така система здатна задовольнити вимоги бізнесу, науки, медицини, безпеки та інших

галузей, де швидкий і точний доступ до релевантної інформації визначає ефективність роботи. Це робить тему розробки подібних систем надзвичайно актуальною як для дослідників, так і для розробників практичних рішень у сфері обробки даних та штучного інтелекту.

Мета і задачі дослідження. Метою даної кваліфікаційної роботи освітнього рівня «Магістр» є проєктування інформаційної системи з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr». Для досягнення поставленої мети потрібно виконати ряд завдань, зокрема:

– Проаналізувати стан досліджень в області інформаційних систем для пошуку зображень.

– Виконати порівняння існуючих систем пошуку та індексації зображень.

– Дослідити існуючі на даний час ключові підходи та алгоритми в системах пошуку та індексації зображень.

–Проаналізувати провідні системи пошуку та індексації зображень.

– Розробити проєкт архітектури інформаційної системи з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr».

Об’єкт дослідження інформаційні системи пошуку та індексації зображень.

Предмет дослідження. процеси пошуку та індексації зображень.

Наукова новизна одержаних результатів отримали подальший розвиток та застосування методи релевантного пошуку на основі «Elasticsearch» та «Apache Solr».

Практичне значення одержаних результатів. Виконано проєктування інформаційної системи з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr».

Апробація результатів магістерської роботи. Основні результати проведених досліджень обговорювались на IV Міжнародній науково-практичній конференції «Open science nowadays: main mission, trends and instruments, path and its development» (Вінниця-Відень).

Публікації. Основні результати кваліфікаційної роботи опубліковано у двох працях конференції (Див. додаток А).

Структура й обсяг кваліфікаційної роботи. Кваліфікаційна робота складається зі вступу, чотирьох розділів, висновків, списку літератури з 72 найменувань та одного додатка. Загальний обсяг кваліфікаційної роботи складає 81 сторінка, з них 53 сторінки основного тексту, який містить 3 рисунки та 8 таблиць.

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ, ПОШУКОВІ СИСТЕМИ, КЛЮЧОВІ ПІДХОДИ ТА АЛГОРИТМИ В ІНФОРМАЦІЙНИХ СИСТЕМАХ ПОШУКУ ТА ІНДЕКСАЦІЇ

1.1 Контекст та актуальність розробки інформаційної системи з використанням релевантного пошуку

Експоненційне зростання цифрових даних зумовило зростаючий попит на обробку даних у режимі реального часу в хмарних середовищах. Організації з різних галузей покладаються на аналітику в режимі реального часу для отримання практичних інсайтів, оптимізації операцій та покращення процесів прийняття рішень. Однак досягнення низької затримки та високої пропускної здатності при масштабованій обробці даних вимагає використання передових архітектур, здатних обробляти динамічні навантаження, різнорідні джерела даних і складні аналітичні запити. Традиційні дата-пайплайни часто не справляються з цими вимогами через жорстку структуру, неефективне використання ресурсів і обмеження масштабованості.

Незважаючи на досягнення в хмарних обчисленнях та технологіях великих даних, існуючі архітектури дата-пайплайнів стикаються з рядом проблем при реалізації аналітики в режимі реального часу. Масштабованість залишається ключовою проблемою, оскільки традиційні методи пакетної обробки призводять до затримок, що робить їх непридатними для застосунків із жорсткими часовими вимогами. Крім того, забезпечення відмовостійкості та стійкості системи до непередбачуваних навантажень є серйозним викликом. Поточні програмно-алгоритмічні рішення часто потребують складних налаштувань і високих експлуатаційних витрат, що обмежує їхню практичність для багатьох організацій. Отже, існує нагальна потреба у більш ефективній та адаптивній структурі для обробки даних у режимі реального часу в хмарному середовищі.

Тому проектування та розробка інформаційної системи з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» є актуальним напрямком сучасних досліджень. Для цього доцільно створити нову структуру,

розроблену для подолання обмежень існуючих дата-пайплайнів в аналітиці хмарних обчислень у режимі реального часу. Структура спрямована на покращення масштабованості, мінімізацію затримок обробки та підвищення відмовостійкості шляхом використання сучасних технологій, таких як безсерверні обчислення, обробка на периферії та інтелектуальна оркестрація даних. Такий підхід забезпечує безперебійну, високопродуктивну аналітику в режимі реального часу, даючи змогу міським установам та організаціям максимально ефективно використовувати потенціал своїх даних у хмарно-орієнтованому середовищі.

1.2 Аналіз предметної області

Пошук зображень залишається складним завданням через складну взаємодію між людським сприйняттям, пам’яттю та обчислювальними процесами. Поточні пошукові інформаційні системи з використанням релевантного пошуку для зображень часто мають труднощі з ефективним пошуком зображень на основі природних мовних описів, оскільки вони залежать від ресурсомістких етапів передобробки, тегування та машинного навчання. Тому доцільно сформувати інформаційну систему з новим підходом до пошуку зображень – пошукову систему для зображень, що орієнтована на людину, яка використовує нейро-символічну індексацію для покращення пошуку зображень, зосереджуючись на індексації, орієнтованій на людину. Інтегруючи знання з когнітивних наук із сучасними обчислювальними техніками, інформаційні системи з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» покращують процес пошуку, роблячи його узгодженішим з тим, як люди сприймають, зберігають та відтворюють візуальну інформацію. Нейро-символічна структура поєднує переваги нейронних мереж та символічного міркування, пом’якшуючи їхні окремі обмеження. Запропоновані інформаційні системи з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» оптимізують пошук зображень, пропонуючи інтуїтивно зрозуміліше та ефективніше рішення для користувачів. Дизайн і впровадження

інформаційних систем з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr», підкреслюємо його потенційні застосування для виявлення помилок у дизайні та управління знаннями, удосконалення метрик і можливостей системи.

На перетині людського досвіду та штучного інтелекту виникає основне завдання: як ефективно подолати розрив між людським знанням і обчислювальним інтелектом. Традиційні підходи символічного ШІ, які намагаються кодувати людський досвід у правилах, що базуються на системах, довго стикалися з проблемою припущення замкнутого світу – вони можуть міркувати лише в межах чітко визначених знань, що обмежує їхню здатність адаптуватися до нових ситуацій. Натомість сучасні підходи машинного навчання (ML) відзначаються виявленням закономірностей, але мають два критичних обмеження:

– упереджені результати, що відображають дані для навчання;

– відсутність пояснюваності, через що їхні процеси прийняття рішень стають непрозорими для людського розуміння.

Спочатку доцільно розглянути завдання пошуку зображень, їхні поточні підходи до розв’язання і їхні недоліки. Потім презентуємо підхід та пояснюємо, як його можна інтегрувати з пошуком зображень. На наступному етапі доцільно запропонувати алгоритм функціонування інформаційної системи з використанням релевантного пошуку на основі Elasticsearch та Apache Solr.

«Чудове життя починається з картинки, що тримається у Вашій уяві, того, чого Ви хочете досягти або бути» [1]. Ця концепція уявлення результатів резонує з інформаційною системою пошуку зображень, орієнтованою на людину, яка має на меті покращити пошук зображень, орієнтуючись на те, як люди візуалізують і описують зображення. Використовуючи нейро-символічний індексинг, інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» долає розрив між людським пізнанням та обчислювальними системами, даючи змогу користувачам отримувати зображення на основі описів природною мовою, подібно до того, як уявляється бажаний результат і втілюється в реальність.

Запропонований алгоритм «HORSE» пропонує нове рішення через унікальний підхід інтеграції нейро-символічного підходу. На відміну від традиційних нейро-символічних систем, де процеси міркування просто накладаються на виходи нейронних мереж, «HORSE» починає з витягування відповідних людських знань і потім впроваджує процеси ШІ, що спеціально узгоджуються з цією основою знань. Така орієнтованість на людину забезпечує сумісність обчислювального міркування з людською логікою, одночасно використовуючи переваги розпізнавання закономірностей нейронними мережами. Пошук візуальної інформації є прикладом проблем, які цей підхід вирішує. Поточні системи пошуку зображень мають:

– значні обмеження з точки зору зручності для користувачів;

– часто використовуючи складні процеси попередньої обробки, тегування та алгоритми машинного навчання, які створюють надмірне зберігання ознак, водночас не оптимізуючи для зображень, які вже переглядалися.

Ці системи часто повертають результати, які, хоча й алгоритмічно релевантні, але не відповідають людському сприйняттю та процесам пам’яті.

Це дослідження вивчає альтернативну парадигму, за якою користувачі можуть знаходити раніше переглянуті зображення, описуючи їх природною мовою, створюючи інтуїтивно зрозумілішу модель взаємодії людини з комп’ютером. Оптимізуючи процеси пошуку, орієнтуючись на людську пам’ять та можливості опису, прагнемо розробити системи, які працюють у гармонії з когнітивними процесами людини, а не вимагають від людини адаптуватися до обмежень машини. В процесі дослідження потрібно:

– Краще розуміння процесу пошуку зображень з обох точок зору – людини та комп’ютера.

– Оцінка взаємодії між людською пам’яттю, зберіганням, описом та пошуком.

– Огляд метрик для взаємодії людини з комп’ютером для оцінки процесу пошуку від початку до кінця.

– Оптимізоване рішення для пошуку зображень, яке покращує доступність та ефективність пошуку.

1.3 Пошукові системи зображень системи пошуку та індексації зображень

Область виявлення та управління візуальною інформацією в основному поділяється на два типи систем, кожна з яких служить різним цілям та потребам користувачів. Пошукові системи зображень «Google Images» [2] та «Bing Images» [3], призначені для загальних користувачів, які шукають зображення в Інтернет за допомогою текстових запитів. Ці орієнтовані на споживачів платформи пропонують прості інтерфейси з основними параметрами фільтрації за характеристиками, такими як розмір, колір і тип зображення, при цьому надаючи перевагу доступності та широкому покриттю над технічною складністю. На противагу цьому, системи пошуку та індексації зображень служать для технічних і корпоративних застосунків, надаючи спеціалізовані можливості для керування курованими колекціями зображень. Системи «Apache Solr» з розширеннями для зображень [4] та «Elasticsearch» з плагінами для зображень [5] підтримують як текстові, так і контентні запити, що дає можливість точніше отримувати доступ до візуальних активів.

Основні відмінності між цими типами систем полягають у їхньому обсязі, що можна побачити в таблиці 1.1.

Таблиця 1.1 – Порівняння між пошуковими системами зображень та системами пошуку та індексації зображень

Характеристика	**Пошукові системи зображень**	**Механізми пошуку та індексації зображень**
1	**2**	**3**
Цільові користувачі	Звичайні споживачі	Технічні фахівці та підприємства
Основна мета	Пошук зображень в Інтернеті	Керування курованими колекціями зображень

Продовження таблиці 1.1

1	2	3
Приклади	Зображення «Google», зображення «Bing»	Apache Solr з розширеннями зображень, Elasticsearch з плагінами зображень
Методологія індексування	Сканування Інтернету з індексацією на основі навколишнього тексту, імен файлів та основних візуальних функцій	Складні багатовимірні структури індексації (k-d дерева, R-дерева, локально-чутливе хешування)

Ці платформи включають передові функції порівняння візуальної схожості зображень [6] або індекс структурної подібності (SSIM), автоматичне видобування ознак, класифікацію контенту та всебічне індексація метаданих – можливості, які підтримують професійні робочі процеси в сферах управління цифровими активами та медичне зображення. Основні відмінності між цими типами систем у методах запитів, що можна побачити в таблиці 1.2.

Таблиця 1.2 – Порівняння запитів між пошуковими системами зображень та системами пошуку та індексації зображень

Характеристика	Пошукові системи зображень	Механізми пошуку та індексації зображень
1	2	3
Типи запитів	Переважно текстові запити	Кілька парадигм запитів: текстові, запити за прикладами, запити за ескізами, гібридні підходи

Продовження таблиці 1.2

1	2	3
Обробка запитів	Зіставлення ключових слів з індексованим текстом	Складний пошук за подібністю у векторах ознак із семантичним розумінням
Вилучення ознак	Основні функції та метадані зображень	Розширене візуальне вилучення ознак з доменно-специфічними оптимізаціями
Джерела даних	Загальнодоступний веб-контент	Кураторські колекції зі структурованими метаданими
Керування даними	Безперервне виявлення через сканування Інтернету з обмеженим контролем набору даних	Ретельно підтримувані колекції з контролем версій та правами доступу

Основні відмінності між цими типами систем у аналітичних можливостях, що можна побачити в таблиці 1.3.

Таблиця 1.3 – Порівняння аналітичних можливостей між пошуковими системами зображень та системами пошуку та індексації зображень

Характеристика	**Пошукові системи зображень**	**Механізми пошуку та індексації зображень**
Аналітичні можливості	Базова фільтрація (розмір, колір, тип)	Розширені можливості (виявлення об’єктів, розуміння сцен, розпізнавання облич)
Обчислювальний фокус	Швидкість та релевантність для величезних наборів даних	Точність у визначених доменах з доменно-специфічними алгоритмами

Основні відмінності між цими типами систем у їх характеристиках, що можна побачити в таблиці 1.4.

Таблиця 1.4 – Порівняння аналітичних можливостей між пошуковими системами зображень та системами пошуку та індексації зображень

Характеристика	**Пошукові системи зображень**	**Механізми пошуку та індексації зображень**
Варіанти інтеграції	Обмежений доступ до API	Надійні API та інтеграційні фреймворки для корпоративних застосунків
Налаштування	Узагальнені алгоритми для всього контенту	Можливість адаптації до доменно-специфічних візуальних шаблонів та випадків використання
Технічна складність	Розроблено для зручності використання	Вища складність з досконалішим керуванням
Обсяг	Ширше охоплення в загальнодоступному Інтернеті	Глибший аналіз у визначених колекціях
Фокус на масштабованості	Горизонтальне масштабування для мільярдів зображень	Точне індексування та пошук для доменно-специфічних колекцій

В той час як пошукові системи охоплюють ширшу аудиторію в публічному Інтернет, здебільшого з текстовими запитами, системи пошуку пропонують глибший аналіз у межах визначених колекцій, використовуючи різноманітні методи запитів. Ця відмінність відображає їхні різні цілі: пошукові системи з’єднують користувачів із раніше невідомими зображеннями, тоді як системи пошуку допомагають користувачам ефективно знаходити та використовувати відомі візуальні активи в керованих сховищах. Технічна складність систем пошуку супроводжується збільшеною складністю, але дає змогу досягти

точності та аналітичних можливостей, необхідних для професійних та корпоративних застосунків.

1.4 Аналіз процесу пошуку зображень людиною

Проведемо вивчення процесу пошуку зображень людиною. Людська пам’ять здатна швидко кодувати візуальну інформацію та компактно зберігати її для довгострокового відновлення [7], характеристика, яку наш підхід прагне використовувати для ефективного індексування та зберігання в комп’ютерних системах. Новизна цього полягає в інтеграції когнітивних факторів людини в пошук зображень, що, наскільки на даний час відомо, не було широко досліджено.

«HORSE» транслює ці характеристики в нейро-символічні правила. Ці правила використовуються як людські знання, що покращує подальші кроки машинного навчання та пошуку. Наскільки нам відомо, така інтеграція не була здійснена раніше. Крім того, переклад людської пам’яті в нейро-символічні правила є унікальним у цій галузі.

Алгоритм вирішує зростаючу потребу в доступних системах пошуку зображень у контексті збільшення обсягу зображень, що зберігаються та обмінюються на платформах соціальних мереж та персональних пристроях. Результати дослідження матимуть значний вплив на поліпшення доступності зображень, управління знаннями та допомогу професіоналам, таким як дизайнери та креслярі, у виявленні помилок у дизайні.

1.5 Дослідження ключових підходів та алгоритми в системах пошуку та індексації зображень

Системи пошуку та індексації зображень спираються на різні підходи та алгоритми, які допомагають ефективно керувати та шукати візуальні дані. Однією з ключових областей є видобування ознак [8], яке включає отримання інформації з зображень у різних формах. Наприклад, ознаки кольору часто

використовуються для відстеження розподілу кольорів на зображенні за допомогою методів на базі гістограм кольору, моментів кольору та кольорових корелограм [9]. Гістограми кольору відстежують розподіл кольорів, моменти кольору фіксують середнє значення, дисперсію та асиметрію розподілу кольорів, а корелограма представляє просторову кореляцію кольорів в межах зображення. Описники домінуючого кольору також важливі для ідентифікації ключових кольорів, що характеризують загальний вигляд зображення. Ознаки текстури зосереджуються на аналізі поверхневих властивостей зображень [10], із методами на базі матриці співвідношення сірого рівня («GLCM»), яка вивчає просторові відносини між інтенсивностями пікселів [11]. Фільтри Габора виявляють специфічний частотний вміст у різних напрямках [12], а локальні бінарні шаблони (LBP) захоплюють локальні текстурні шаблони [13]. Перетворення вейвлет дає змогу здійснювати багаторозрізний аналіз текстури, що є важливим у багатьох застосунках [14].

Ознаки форми – це ще один важливий аспект, і методи виявлення країв, включаючи оператори Кані та Собеля, підкреслюють краї зображення [15]. Представлення контурів захоплюють контури форми [16], тоді як моментні інваріанти забезпечують спосіб опису форм незалежно від обертання та масштабу [17]. Також використовуються контексти форми для захоплення просторових відносин між точками на межі форми, що пропонує відмінне подання об’єкта.

Підходи на основі глибокого навчання здобули значну увагу, зокрема за допомогою згорткових нейронних мереж («CNN»), які часто використовуються для видобування ознак. Попередньо навчені моделі «Visual Geometry Group» («VGG»), «Residual Network» («ResNet») та «Inception», часто використовуються для створення векторів ознак, які служать ефективними дескрипторами зображень [18]. Сіамські мережі стали популярними для навчання метрик подібності між парами зображень, що покращує здатність розрізняти схожі та несхожі зображення [19]. Автокодери застосовуються для зменшення розмірності, спрощуючи подання даних, зберігаючи при цьому ключові ознаки. Методи самонавчання, які дають можливість отримати кращі описи зображень

без потреби в мічених даних, також довели свою корисність у підвищенні ефективності пошуку.

Індексаційні структури є необхідними для ефективного пошуку та отримання зображень. Методи на основі дерев, зокрема, KD-дерева (K-вимірні) для низькорозмірних ознак та R-дерева часто використовуються для просторового індексування [20]. M-дерева підходять для індексації в метричних просторах [21], а VP-дерева («Vantage-Point») є особливо ефективними для високорозмірних даних [22]. Методи хешування, зокрема, «Locality Sensitive Hashing» («LSH»), «Semantic Hashing», «Spectral Hashing» та «Product Quantization», також відіграють важливу роль у зменшенні складності завдань пошуку, даючи змогу проводити швидші пошуки шляхом кодування даних у компактніші подання [23].

Методи вимірювання подібності використовуються для оцінки близькості між зображеннями на основі їх ознак [24]. Евклідові відстані зазвичай використовуються для порівняння векторів ознак, тоді як косинусна подібність часто застосовується у високорозмірних просторах. Відстань земельного переміщика корисна для порівняння гістограм, вимірюючи вартість трансформації одного розподілу в інший [25]. Відстань Геммінга застосовується до бінарних ознак [26], а відстань Махаланобіса використовується, коли дані містять корельовані ознаки, що дає змогу отримати точнішу міру подібності в таких випадках [27].

Пошук зображень на основі їх вмісту («CBIR») є однією з найбільш популярних технік пошуку, що дає можливість системам отримувати зображення на основі їх вмісту [28]. Запит за зразком («QBE») та механізми зворотного зв’язку з релевантністю часто використовуються для уточнення пошуків, а стратегії мульти-особливостей поєднують кілька типів ознак для покращення точності пошуку [29]. Перехресний модальний пошук, що включає техніки за текстом до зображення та відображення зображення в текст, дає змогу знаходити зображення на основі текстових описів і навпаки [30]. Ці техніки спираються на спільні простори вбудовування, які стирають межу між різними модальностями.

Сучасні методи оптимізації, зокрема, пошук найближчого сусіда «ANN», широко використовуються для покращення ефективності пошуку [31]. Інструменти «FAISS» – «Facebook AI Similarity Search», «Annoy» – бібліотека «ANN» від «Spotify» та «HNSW» – ієрархічний навігаційний малий світ, дають змогу проводити швидші пошуки найближчого сусіда, зменшуючи обчислювальні витрати при роботі з великими наборами даних [32]. При оцінці систем пошуку зазвичай використовуються метрики, зокрема, точність і відгук, середня середньоквадратична точність («MAP») та нормалізоване знижене кумулятивне підвищення («NDCG»), для оцінки релевантності та ранжування результатів [33]. Інші важливі фактори включають час пошуку та ефективність пам’яті, які є критичними у реальних застосуваннях.

Для реального впровадження необхідно вирішити кілька важливих аспектів. Масштабованість є критично важливою для обробки великих наборів даних, а механізми оновлення потрібні для адаптації до динамічних колекцій. Оптимізація зберігання гарантує ефективне використання даних, тоді як оптимізація запитів покращує швидкість обробки запитів. Балансування навантаження в розподілених системах також необхідне для обробки змінюваних вимог і забезпечення безперебійної роботи.

Покращені методи пошуку використання кількох зображень для запитів та розширення запитів можуть допомогти підвищити точність пошуку [34]. Семантичні методи пошуку покращують результати, розуміючи сенс пошукових запитів [35], а фільтрація за атрибутами дає можливість здійснювати детальніший пошук на основі конкретних характеристик зображень [36]. Просторова верифікація – це ще одна техніка, яка гарантує, що повернуті результати відповідають просторовим характеристикам запитуваного зображення [37]. Нарешті, до передових тем пошуку та індексації зображень відносяться тонко настроєний пошук зображень, який зосереджується на пошуку зображень з тонкими відмінностями, та пошук на рівні екземплярів, який орієнтований на конкретні екземпляри об’єктів. Перехресне порівняння зображень вирішує проблеми зі співвідношенням зображень одного об’єкта з різних кутів зору [38], а пошук зображень без прикладів («zero-shot image

retrieval») дає змогу здійснювати пошук зображень без попередніх прикладів чи навчальних даних [39]. Системи безперервного навчання, які адаптуються та покращуються з часом завдяки новим даним, є ще однією новою перспективною областю, яка гарантує, що системи пошуку зображень залишаються ефективними в процесі їх еволюції [40].

1.6 Висновок до першого розділу

В першому розділі кваліфікаційної роботи освітнього рівня «Магістр» розглянуто контекст та актуальність розробки інформаційної системи з використанням релевантного пошуку. Крім того проведено аналіз предметної області. Описано популярні пошукові системи зображень системи пошуку та індексації зображень. Проаналізовано процес пошуку зображень людиною. Досліджено ключові підходи та алгоритми в системах пошуку та індексації зображень.

2 ПРОВІДНІ СИСТЕМИ, ТЕХНОЛОГІЇ ПОШУКУ ТА НЕЙРО-СИМВОЛЬНІ ПІДХОДИ ДО ПОШУКУ ІНФОРМАЦІЇ

2.1 Провідні системи пошуку та індексації зображень

Існує декілька популярних систем пошуку та індексації зображень, які можна поділити на комерційні, відкриті та засновані на глибинному навчанні. До комерційних рішень відносяться «Elastic Image Search» [41], що надає можливості візуального пошуку в екосистемі «Elasticsearch», і «Amazon Rekognition» [42], сервіс «AWS» для аналізу зображень та пошуку за подібністю. «Google Cloud Vision Product Search» дає можливість здійснювати візуальний пошук продуктів і каталогізацію [43], а «Microsoft Azure Computer Vision» забезпечує можливості аналізу зображень і візуального пошуку [44]. Порівняння рішень для пошуку зображень і їхніх технічних підходів наведено в таблиці 2.1, яка підсумовує їхні основні підходи та ключові алгоритми.

Таблиця 2.1 – Порівняння рішень для пошуку зображень та їх технічних підходів

Рішення	**Ключові підходи**	**Основні алгоритми**
1	**2**	**3**
Еластичний пошук зображень	Пошук подібності на основі векторів з вилученням ознак CNN	Графіки HNSW Приблизна k-NN Оцінка щільних векторів Обчислення косинусної подібності
Розпізнавання «Amazon»	Багатомодельний підхід, що поєднує виявлення та розпізнавання	Глибокі CNN Каскадні класифікатори Сіамські мережі Варіанти YOLO

Продовження таблиці 2.1

1	2	3
«Google Cloud Vision»	Навчання користувацьких моделей на основі AutoML з ефективним вилученням ознак	Магістраль EfficientNet Контрастне навчання Глибоке метричне навчання Зіставлення тексту та зображень BERT
«Microsoft Azure Vision»	Архітектура на основі трансформатора з багатозадачними можливостями	Вилучення ResNet Генерація графів сцени Швидша R-CNN Навчання за допомогою кількох кадрів

Також «Algolia Visual Search» служить доповненням для їхньої пошукової платформи [45]. Відкриті рішення включають «Milvus» – розподілену векторну базу даних, яка обробляє вектори ознак зображень [46], та «FAISS» («Facebook AI Similarity Search») – бібліотеку високопродуктивного пошуку за подібністю [47]. «Qdrant» – ще один пошуковий рушій за векторною подібністю, який пропонує розширену підтримку фільтрації [48], а «Vearch» – високопродуктивний пошуковий рушій за векторною подібністю, розроблений «Jina AI» [49]. «Vespa» – це система обробки великих даних у режимі реального часу, яка також підтримує пошук зображень [48], а бібліотека «ImageHash» на Python дає змогу здійснювати просте зіставлення зображень за допомогою перцепційних хешів [50].

Порівняння «Milvus», «FAISS», «Qdrant» та «CLIP» для пошуку зображень і їхніх технічних підходів подано в таблиці 2.2. «LIRE» («Lucene Image Retrieval») – це Java-бібліотека для пошуку зображень, заснована на «Lucene» [51]. У галузі глибинного навчання модель «CLIP» («Contrastive Language Image Pre-Training») розроблена «OpenAI» для пошуку тексту за зображеннями [52].

Таблиця 2.2 – Порівняння «Milvus», «FAISS», «Qdrant» та «CLIP» для пошуку зображень та їх технічних підходів

Рішення	**Ключові підходи**	**Основні алгоритми**
«Milvus»	Розподілений векторний пошук з підтримкою кількох індексів	Індексування IVF/HNSW/ANNOY Прискорення GPU Динамічне квантування Оптимізація SIMD
«FAISS»	Високопродуктивний пошук подібності зі стисненням	Квантування продукту Інвертований файловий індекс Багатозондовий LSH Індексація на основі кластерів
«Qdrant»	Векторний пошук на основі графів з фільтрацією	Індексування HNSW Фільтрація корисного навантаження Сегментне зберігання Оптимізація запитів
«CLIP»	Контрастне навчання між зображенням і текстом	Трансформатор зору Класифікація нульового кадру Крос-модальна увага Масштабування температури

«DupDetector» використовує нейронні мережі для виявлення дублікатів або схожих зображень [53]. «DeepSight» – ще один пошуковий рушій для візуального пошуку, заснований на глибинному навчанні [54].

Порівняння «Milvus», «FAISS», «Qdrant» та «CLIP» для пошуку зображень і їхніх технічних підходів подано в таблиці 2.3.

Таблиця 2.3 – Порівняння «Milvus», «FAISS», «Qdrant» та «CLIP» для пошуку зображень та їх технічних підходів

Рішення	**Ключові підходи**	**Основні алгоритми**
«ImageHash»	Зіставлення зображень на основі перцептивного хешування	Хешування середніх значень Хешування різниць Вейвлет-хешування Хешування колірних моментів
«LIRE»	Традиційні функції комп’ютерного зору	Дескриптори SIFT/SURF Виявлення кольору/краю Текстури Габора Гістограми країв

Ці рішення зазвичай надають кілька ключових функцій, зокрема, «CBIR» («Content-Based Image Retrieval»), зворотний пошук зображень, виявлення майже однакових зображень, індексація векторів ознак, перцепційне хешування та мультимодальний пошук, який поєднує текст і зображення, при цьому мають масштабовані розподілені можливості індексації.

2.2 Існуючі інформаційні технології

Було розроблено низку відкритих та хмарно-орієнтованих інформаційних систем для підтримки обробки даних у режимі реального часу. Порівняння ключових технологій подано в таблиці 2.4.

Ці інформаційні технології є основою сучасних пайплайнів обробки даних у режимі реального часу, однак вони мають різні компроміси щодо простоти розгортання, вартості та продуктивності.

Потреба у новій інформаційній системі з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» зумовлена цими обмеженнями та прагненням створити більш зручний, масштабований і адаптивний підхід до хмарної аналітики в режимі реального часу.

Таблиця 2.4 – Порівняння ключових технологій

Технологія	Переваги	Недоліки
«Apache Kafka»	Висока пропускна здатність, обробка подій у пайплайні	Складне налаштування, потребує додаткових компонентів для аналітики
«Apache Flink»	Станова обробка потоків, низька затримка	Високий поріг входу, ресурсоємність
«Google Dataflow»	Безсерверне середовище, автоматичне масштабування, базується на Apache Beam	Тісна інтеграція з Google Cloud, високі витрати
«Apache Spark Streaming»	Об’єднана пакетна й потокова обробка, потужна екосистема	Вища затримка порівняно з Flink, потребує багато пам’яті

Інновації Google в технологіях пошуку та індексації зображень відіграють важливу роль у змінах того, як користувачі взаємодіють із візуальною інформацією та отримують доступ до неї онлайн. Google розробила різноманітні передові технології пошуку та індексації зображень.

«Google Images», стандартна пошукова система для споживачів, підтримує зворотний пошук за зображеннями, використовуючи передові технології комп’ютерного зору та машинного навчання, а також інтегрується з технологією «Google Lens» [55].

«Google Lens» – це мобільний інструмент для візуального пошуку, який здатний ідентифікувати об’єкти, текст і пам’ятки в режимі реального часу, а також надає можливості для пошуку товарів за зображеннями, з функціями витягування тексту та перекладу [56]. Він доступний як окрема програма та інтегрований в інші продукти Google. Для корпоративних застосувань «Google Cloud Vision AI» пропонує кілька сервісів, зокрема:

– «Product Search» для створення систем пошуку в роздрібних каталогах;
– «Vision AI» для загального аналізу та класифікації зображень;
– «AutoML Vision» для навчання кастомізованих моделей [57];
– «Video Intelligence API» для аналізу відеоконтенту [58].

Внутрішні технології Google:

– «PlaNet» – модель для географічної оцінки місцезнаходження за зображеннями [59];

– «DELF» («DEep Local Features») для розпізнавання пам’яток [60];

– «SwAV» («Swapping Assignments between Views») для самонавчання в розумінні зображень [61];

– «MUM» («Multimodal Unified Model»), який обробляє одночасно текст та зображення [62].

2.3 Нейро-символьні підходи до пошуку інформації за зображеннями

Оскільки інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» орієнтована на пошук графічної інформації на основі символьних описів, то доцільно провести огляд нейро-символьні підходи до пошуку інформації за зображеннями.

Нейро-символьні підходи до пошуку інформації за зображеннями поєднують сильні сторони нейронних мереж та символьного міркування для підвищення ефективності систем пошуку [63]. Традиційні моделі пошуку інформації зазвичай покладаються або на нейронні мережі, або на символьні підходи, кожен з яких має свої переваги та обмеження. Нейронні мережі чудово справляються з обробкою високорозмірних даних і розпізнаванням патернів у візуальному контенті, тоді як символьне міркування пропонує інтерпретованість, логічну послідовність і інтеграцію структурованих знань людини. Інтеграція цих доповнюючих переваг у нейро-символічних системах створює перспективну основу для вирішення складних завдань пошуку зображень, усуваючи семантичний розрив між низькорівневими ознаками зображень і високорівневим розумінням людиною.

Нейро-символьні підходи пропонують декілька ключових переваг для систем пошуку зображень [64].

Однією з основних переваг є можливість зберігати інтерпретованість, використовуючи здатності нейронних мереж до розпізнавання патернів [65]. На

відміну від чистих нейронних моделей, які часто функціонують як «чорні ящики», нейро-символьні системи зберігають прозорість завдяки своїм символьним компонентам, що робить міркування, яке стоїть за їхніми рішеннями, зрозумілішим.

Крім того, ці підходи можуть явно інтегрувати знання з певної області, людську експертизу та когнітивні принципи, які важко відобразити лише за допомогою даних. Інтеграція цього знання є особливо цінною в спеціалізованих галузях, де позначені дані можуть бути обмеженими або неповними. Більше того, включення символьного міркування дає можливість нейро-символічним системам виконувати логічні висновки про просторові взаємовідносини, властивості об’єктів і семантичні контексти – ключові аспекти розуміння зображень, подібного до людського. Ця здатність міркувати про зображення додає рівень глибини, якого не можуть досягти чисті нейронні системи.

Включення символьних знань також зменшує залежність від великих наборів даних, що робить нейро-символічні підходи життєздатнішими в сценаріях, коли дані для навчання обмежені. Нарешті, символьний компонент виступає як захист, забезпечуючи надійність проти типових проблем, що виникають у чисто нейронних підходах, і покращує загальну послідовність та надійність системи.

Унікальні виклики пошуку зображень роблять нейро-символьні підходи особливо придатними для цього завдання. Візуальна інформація є ієрархічною та взаємозалежною за своєю природою, і значення виникає не лише з окремих об’єктів, а й з їхнього просторового розташування, контексту та взаємовідносин [66]. Ця складність добре відповідає можливостям нейро-символьних систем. Одним з ключових викликів пошуку зображень є значний розрив між представленими на рівні пікселів даними і семантичним розумінням.

Хоча нейронні мережі можуть ефективно обробляти сирі візуальні дані, вони часто не можуть вловити абстрактні взаємовідносини та контекстуальні знання, на які покладаються люди при описі або пошуку зображень. Символьні компоненти допомагають подолати цей розрив, явно представляючи вищі рівні

концепцій просторові взаємовідносини або властивості об’єктів, які є важливими для розуміння та інтерпретації зображень.

Більше того, запити природною мовою для пошуку зображень часто включають неточні, суб’єктивні або контекстно залежні терміни, які потребують інтерпретації, що виходить за межі простого зіставлення ключових слів [67]. Нейро-символьні системи особливо ефективно справляються з такими запитами, мапуючи лінгвістичні описи на візуальні особливості через символьні уявлення таких концепцій, як «вищий за», «більший за» або «схожий на». Ця здатність обробляти нюанси мови робить нейро-символічні системи ефективнішими в реальних застосунках, де запити рідко бувають прямолінійними.

Ще одна причина, чому нейро-символьні підходи добре підходять для пошуку зображень, полягає в тому, що людська пам’ять про зображення працює на кількох рівнях абстракції – від загальних вражень до конкретних деталей. Це відображає доповнюючу обробку нейронних і символьних компонентів у нейро-символьних системах, що робить їх узгодженішими з когнітивними процесами людини [68]. Така узгодженість дає змогу нейро-символьним системам створювати механізми пошуку, які здаються інтуїтивно зрозумілими і природними для користувачів, що ще більше покращує користувацький досвід.

Крім того, нейро-символьні системи можуть використовувати попередні знання про типові взаємовідносини об’єктів і просторові конфігурації для зроблення висновків про зображення [69]. Ця здатність знижує потребу в великих обсягах навчальних даних, які часто необхідні для чисто нейронних підходів для вивчення цих взаємовідносин.

В спеціалізованих областях, де навчальні дані можуть бути обмеженими, але знання в галузі є багатим, ця перевага стає особливо важливою. Комбінуючи сильні сторони нейронних і символьних підходів, ці системи пропонують потужне рішення для подолання викликів пошуку зображень, що робить їх ідеальним вибором для застосунків, які вимагають як високої точності, так і інтерпретованості.

2.4 Огляд дата-пайплайнів

Оскільки інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» орієнтована на дані, то доцільно провести огляд дата-пайплайнів.

Дата-пайплайни є основою сучасних застосунків, орієнтованих на дані, забезпечуючи ефективний потік інформації від «сирих» джерел до практичних інсайтів. Традиційно обробку даних поділяють на два основні типи: пакетну («batch») та обробку в режимі реального часу:

– Пакетна обробка – дані збираються, зберігаються та обробляються з певною періодичністю. Такий підхід підходить для великомасштабної, але не критичної за часом аналітики. Приклади включають архітектури на основі «Hadoop» та робочі процеси ETL.

– Обробка в режимі реального часу – дані обробляються одразу після надходження, що дає змогу здійснювати миттєвий аналіз і реагування. Такий підхід є критично важливим для застосунків, що потребують прийняття рішень із мінімальною затримкою, як-от виявлення шахрайства, моніторинг IoT-систем та динамічне ціноутворення.

Типовий дата-пайплайн складається з чотирьох ключових компонентів:

– Інтеграція даних («ingestion») – збір даних із різноманітних джерел, сенсорів, журналів подій або баз даних.

– Трансформація даних («transformation») – очищення, фільтрація та збагачення даних перед зберіганням або аналізом.

– Зберігання даних («storage») – збереження даних у структурованих (SQL), напівструктурованих (NoSQL) або розподілених системах зберігання.

– Аналіз даних («analysis») – отримання інсайтів за допомогою машинного навчання, бізнес-аналітики або систем на основі правил.

Аналітика в режимі реального часу в хмарному середовищі. Перехід до хмарних обчислень відкрив нові можливості та виклики для аналітики в режимі реального часу. Хмарні платформи забезпечують масштабованість, ефективність

витрат і глобальну доступність, але водночас створюють проблеми, пов’язані з мережею, керуванням ресурсами та безпекою.

Типові виклики та рішення у хмарному середовищі:

– Масштабованість – управління динамічним навантаженням через автоштабування та розподілені інформаційні системи.

– Затримки – оптимізація за допомогою обчислень на периферії («edge computing») та обробки в оперативній пам’яті («in-memory processing»).

– Відмовостійкість – використання механізмів збереження контрольних точок («checkpointing»), реплікації та аварійного перемикання («failover»).

Приклади застосування аналітики в режимі реального часу:

– Інтернет речей («IoT») – давачі «розумного міста» та промислові IoT-системи використовують потоки даних у режимі реального часу для моніторингу трафіку, енергоспоживання та стану обладнання.

– Фінансові транзакції – системи виявлення шахрайства аналізують транзакційні шаблони в режимі реального часу, щоб виявити аномалії та запобігти порушенням безпеки.

– Потокові дані – соціальні мережі та онлайн-рекомендаційні системи миттєво обробляють взаємодії користувачів, щоб надавати персоналізований контент.

2.5 Принципи проєктування пайплайнів для обробки даних нового покоління

При побудові інформаційної системи з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» доцільно провести огляд принципів проєктування пайплайнів для обробки даних нового покоління.

Масштабованість є основною вимогою до сучасних дата-пайплайнів, оскільки обсяги потокових даних стрімко зростають. Пайплайн нового покоління має бути здатним обробляти великі обсяги даних без зниження продуктивності.

Ключові стратегії масштабування:

– Горизонтальне масштабування – замість використання одного потужного вузла система має розподіляти навантаження між кількома вузлами для забезпечення плавної масштабованості. Хмарні рішення, зокрема, «Kubernetes» і контейнеризовані мікросервіси, ефективно керують розподіленою обробкою.

– Динамічний розподіл ресурсів – хмарні середовища пропонують еластичне надання ресурсів, що дає можливість системам масштабуватись вгору або вниз залежно від потреб. Безсерверні обчислення та автозростаючі кластери, наприклад, «AWS Lambda», «Google Cloud Functions», забезпечують масштабування з оптимізацією витрат.

– Балансування навантаження – розподіл вхідних потоків даних між обчислювальними одиницями для запобігання вузьким місцям і рівномірного завантаження системи.

Аналітика в режимі реального часу потребує надзвичайно низької затримки для забезпечення своєчасних інсайтів. Зменшення затримок обробки є критично важливим для застосунків, таких як виявлення шахрайства, моніторинг IoT і рекомендаційні системи в режимі реального часу.

Техніки зниження затримки:

– Потокова обробка на основі подій – на відміну від пакетної обробки, яка виконується з фіксованими інтервалами, архітектури, що керуються подіями, наприклад, «Apache Kafka Streams», «Apache Flink», забезпечують безперервний потік даних з мінімальними затримками.

– Обчислення в оперативній пам’яті – зберігання та обробка даних у пам’яті, наприклад, «Apache Ignite», «Redis», суттєво знижує витрати на доступ до диска і прискорює виконання запитів.

– Edge-обчислення (обробка даних на периферії) – попередня обробка безпосередньо на джерелі даних, наприклад, IoT-пристрої на периферії, мінімізує мережеву затримку та зменшує навантаження на центральні хмарні сервери.

– Ефективна серіалізація даних – використання легких форматів, таких як Apache Avro або Protobuf, знижує накладні витрати на передавання та десеріалізацію даних.

Надійний пайплайн обробки даних повинен зберігати працездатність у разі збоїв – як через порушення в мережі, так і через апаратні чи програмні збої.

Стратегії забезпечення стійкості:

– Надмірність і реплікація – зберігання копій даних на кількох вузлах дає можливість системі продовжувати роботу навіть у разі відмови окремих елементів.

– Збереження стану та відновлення – періодичне створення контрольних точок дає змогу інформаційній системі з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr»потокової обробки зі збереженням стану, наприклад, «Apache Flink», відновлювати роботу після збою без втрати даних.

– Механізми автоматичного перемикання – автоматичне перенаправлення навантаження на справні вузли у випадку відмови забезпечує високу доступність сервісів.

– Самовідновлювані архітектури – хмарно-орієнтовані системи використовують інструменти автоматичного моніторингу, наприклад, перевірки стану в «Kubernetes», для виявлення збоїв і автоматичного перезапуску несправних компонентів.

Пайплайни обробки даних мають бути адаптивними до змін бізнес-вимог та технологічних нововведень. Жорстка архітектура може швидко застаріти з появою нових джерел даних, аналітичних інструментів і цілей бізнесу.

Основні аспекти проєктування для гнучкості:

– Модульна архітектура – розділення компонентів пайплайну, наприклад, отримання, трансформація, зберігання, дає змогу незалежно оновлювати і налаштовувати частини системи. Архітектура на базі мікросервісів спрощує інтеграцію з новими сервісами.

– Підключення розширень («pluggable connectors») – підтримка широкого спектра джерел даних, наприклад, IoT-пристроїв, реляційних баз даних, API, та

цільових систем, зокрема, інформаційних панелей, моделей машинного навчання, підвищує сумісність.

– Підтримка мультихмарних і гібридних середовищ – можливість працювати на різних хмарних платформах (AWS, Azure, GCP) та в локальних середовищах забезпечує портативність і дає можливість уникнути залежності від конкретного провайдера.

– Обробка еволюції схем – підтримка гнучких схем даних за допомогою реєстрів схем, наприклад, «Apache Avro Schema Registry», допомагає уникнути проблем сумісності при додаванні нових форматів даних.

Дотримання цих принципів дає змогу пайплайнам нового покоління ефективно обробляти дані в режимі реального часу, гарантуючи масштабованість, низьку затримку, відмовостійкість і адаптивність у хмарному середовищі.

2.6 Висновок до другого розділу

В другому розділі кваліфікаційної роботи описано провідні системи пошуку та індексації зображень. Описано існуючі інформаційні технології пошуку та індексації зображень, зокрема технології від Google. Розглянуто нейро-символьні підходи до пошуку інформації за зображеннями. Оскільки інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» орієнтована на дані, то в даному розділі проведено огляд дата-пайплайнів. Також було провеено огляд принципів ключових принципів проєктування пайплайнів для обробки даних нового покоління.

3 ІНФОРМАЦІЙНА СИСТЕМА З ВИКОРИСТАННЯМ РЕЛЕВАНТНОГО ПОШУКУ

3.1 Архітектура інформаційної системи з використанням релевантного пошуку

Запропонована архітектура інформаційної системи з використанням релевантного пошуку розроблена для забезпечення обробки даних у режимі реального часу в хмарних середовищах, із високою продуктивністю, масштабованістю та гнучкістю. Вона базується на модульному та розподіленому підході, що інтегрує різні технології для забезпечення безперервного прийому, трансформації, зберігання та аналітики великих потоків даних. Розглянемо загальну схему архітектури, яка включає основні компоненти:

– Отримання даних («Data Ingestion») – захоплює вхідні дані з різноманітних джерел.

– Потокова обробка («Stream Processing») – виконує обробку, трансформацію і збагачення даних у режимі реального часу.

– Сховище режимі реального часу («Real-Time Storage») – зберігає дані для швидкого доступу та подальшого аналізу.

– Аналітика і візуалізація («Analytics and Visualization») – надає аналітичні результати у режимі реального часу за допомогою розширених алгоритмів та інтерактивних інформаційних панелей.

Ця модульна структура (див. рисунок 3.1) дає можливість масштабувати кожен компонент окремо, оптимізуючи використання ресурсів і підвищуючи ефективність обробки.

Ефективне отримання та передача даних є основою для обробки великих обсягів інформації в режимі реального часу. Інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» повинна підтримувати високу пропускну здатність і низьку затримку, щоб встигати за безперервними потоками даних із різних джерел.

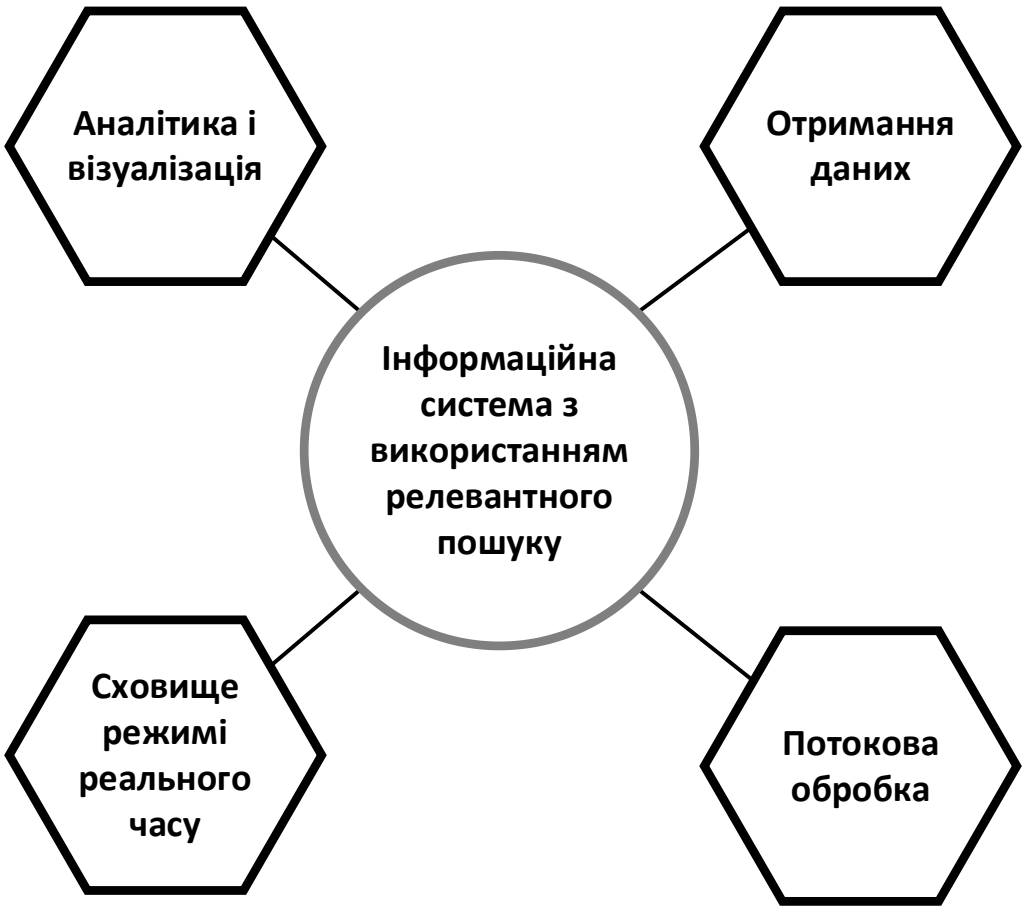

Рисунок 3.1 – Архітектура інформаційної системи з використанням релевантного пошуку

Методи отримання та потокової обробки великих наборів даних:

– Озера даних («Data Lakes») – хмарні озера даних, наприклад, «AWS S3», «Google Cloud Storage», слугують центральними сховищами для великих обсягів сирих даних. Вони забезпечують гнучке та масштабоване середовище зберігання як структурованих, так і неструктурованих даних.

– Черги повідомлень – технології «Apache Kafka» або «RabbitMQ» забезпечують надійні та відмовостійкі брокери повідомлень для обробки високонавантажених потоків подій. Такі системи гарантують, що дані надходять у режимі реального часу та можуть оброблятися послідовно.

– Потоки даних у режимі реального часу – інтеграція з джерелами даних у режимі реального часу, такими як пристрої IoT, API або транзакційні бази даних, є критично важливою для своєчасної доступності даних. Для захоплення і передачі даних у компоненти обробки використовуються потокові протоколи «Apache Kafka Streams» або «AWS Kinesis».

Потокова обробка є ядром аналітики в режимі реального часу, перетворюючи «сирі» дані на придатну для дій інформацію. Обробка даних у пам’яті для зниження затримки. Для мінімізації затримки у інформаційній системі з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» використовується обробка даних у оперативній пам’яті. Технології «Apache Flink» і «Apache Spark Streaming» виконують обчислення безпосередньо в пам’яті, що значно зменшує кількість операцій введення/виведення та підвищує пропускну здатність.

До трансформацій даних належать фільтрація, агрегація та збагачення, що дає змогу перетворити сирі дані у корисні для аналізу формати:

– Фільтрація – видалення нерелевантних або помилкових даних у режимі реального часу для зменшення навантаження на систему.

– Агрегація – узагальнення даних у значущі метрики або групи, наприклад, сума, середнє значення, кількість, для отримання швидких інсайтів.

– Збагачення – додавання контекстної інформації, наприклад, геолокаційних даних або відомостей про профіль користувача, для підвищення цінності вхідних даних.

Запропонована інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» використовує хмарні інструменти для зберігання та аналітики, щоб забезпечити швидке отримання даних і безперебійну інтеграцію з аналітичними системами.

Хмарні системи зберігання:

– Розподілені бази даних – бази даних «Google Bigtable», «AWS DynamoDB» та «Cassandra», оптимізовані для високої пропускної здатності та низької затримки, забезпечуючи швидкий доступ до даних для аналізу та запитів.

– Бази даних часових рядів – спеціалізовані бази даних «InfluxDB» або «Prometheus» ефективно зберігають та обробляють дані з часовими мітками, що робить їх ідеальними для моніторингу та аналітики в режимі реального часу.

– Озера та сховища даних – озера даних можуть запитуватись за допомогою сервісів «AWS Athena» або «Google BigQuery», що забезпечує швидкий доступ до великих обсягів даних із мінімальним навантаженням.

Платформи аналітики та інформаційні панелі в режимі реального часу:

– Інтерактивні панелі – інструменти візуалізації даних у режимі реального часу, наприклад, «Grafana», «Tableau», «Power BI», забезпечують інтерактивні панелі для відстеження ключових метрик та інсайтів. Ці платформи підтримують безперервне оновлення даних та дають користувачам можливість досліджувати потоки даних у режимі реального часу.

– Розширена аналітика – хмарні моделі машинного навчання, наприклад, «AWS SageMaker» та «Google AI Platform», дають змогу реалізовувати передову аналітику в режимі реального часу, включаючи прогностичне моделювання, виявлення аномалій і автоматизоване прийняття рішень.

Для забезпечення ефективності та стійкості пайплайна даних необхідна автоматизація та постійний моніторинг.

Автоматизоване управління пайплайном:

– Розгортання – використання інструментів інфраструктури як коду (IaC) «Terraform» або «AWS CloudFormation», автоматизує розгортання компонентів пайплайна, забезпечуючи узгодженість і зменшуючи кількість помилок людини.

– Масштабування – система повинна автоматично масштабувати компоненти відповідно до навантаження, використовуючи хмарні можливості автоскейлінгу «Kubernetes» або серверлес-архітектури, наприклад, «AWS Lambda» або «Azure Functions».

– Обслуговування – пайплайни CI/CD, безперервної інтеграції та доставки, автоматизують тестування, інтеграцію та впровадження оновлень, зменшуючи час простою та забезпечуючи стабільність системи.

Моніторинг і сповіщення в режимі реального часу:

– Моніторинг якості даних – автоматичні перевірки на узгодженість, точність і повноту даних. Інструменти «Prometheus» та «Datadog» забезпечують моніторингові метрики для візуалізації якості та продуктивності даних у режимі реального часу.

– Моніторинг продуктивності – відстеження стану пайплайна, включаючи час обробки, пропускну здатність та використання системних ресурсів, щоб виявити вузькі місця й оптимізувати роботу.

– Виявлення збоїв і сповіщення – налаштування автоматичних сповіщень за допомогою систем, таких як PagerDuty або Slack, дає змогу оперативно інформувати відповідальні команди про збої чи аномальну поведінку системи, забезпечуючи швидке реагування на проблеми.

Запропонована інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» забезпечує безперервний потік даних – від отримання до аналітики – з низькою затримкою, високою доступністю та відмовостійкістю. Завдяки використанню хмарних технологій і інтеграції автоматизації, ця система є гнучким, масштабованим і надійним рішенням для аналітики в режимі реального часу.

3.2 Платформи та технологічний стек

В процесі проєктування інформаційної системи з використанням релевантного пошуку доцільно провести вибір платформи та технологічного стеку. Успіх аналітичного пайплайна в режимі реального часу значною мірою залежить від вибору відповідних хмарних платформ і технологій. Ці інструменти мають бути масштабованими, надійними та здатними обробляти різноманітні запити до обробки даних та аналітики. Проведемо огляд хмарних платформ.

«AWS» пропонує широкий спектр інструментів для аналітики в режимі реального часу, зокрема, «AWS Kinesis» для потокової обробки, «AWS Lambda» для серверлес-обчислень і «Amazon Redshift» для сховищ даних. «AWS» особливо підходить для підприємств з існуючою інфраструктурою в «AWS» або для тих, хто шукає глобальну доступність і підтримку зрілої екосистеми.

«GCP» забезпечує потужні сервіси, зокрема, «Google BigQuery» для аналітики великих обсягів даних, «Google Cloud Pub/Sub» для обміну повідомленнями в режимі реального часу та «Google Cloud Dataflow» для потокової обробки. «GCP» відзначається високою продуктивністю в режимі реального часу та особливо корисний для інтеграції з машинним навчанням і розширеної аналітики.

«Microsoft Azure» – «Azure» надає комплексні інструменти «Azure Stream Analytics», «Azure Event Hubs» і «Azure Synapse Analytics». «Azure» добре інтегрується з екосистемою Microsoft, що робить його чудовим вибором для підприємств, які покладаються на сервіси Microsoft або потребують безшовної інтеграції з корпоративними інструментами «Microsoft Office 365», «Dynamics» або «SharePoint».

Проведемо описовий огляд технологій. «Apache Kafka» – це розподілена стрімінгова платформа, яка використовується для побудови пайплайнів обробки даних у режимі реального часу. Вона забезпечує високу пропускну здатність, низьку затримку та широко застосовується в подієво-орієнтованих архітектурах. Kafka підтримує масштабованість і відмовостійкість, що робить її ідеальною для обробки великих потоків даних.

«Apache Spark» – «Spark» підтримує як пакетну, так і потокову обробку, забезпечуючи обробку в режимі реального часу за допомогою «Spark Streaming». Його архітектура з обробкою в пам’яті забезпечує низьку затримку та високу продуктивність, що робить «Spark» придатним для аналітики великих обсягів даних і навантажень з машинного навчання.

«Apache Flink» – це інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr»для потокової обробки, оптимізований для роботи з даними в режимі реального часу та збереженням стану. Він підтримує обробку за подієвим часом і семантику «точно один раз», що робить його особливо корисним для складної обробки подій і розширеної аналітики.

Серверлес-фреймворки, наприклад, «AWS Lambda», «Google Cloud Functions» – серверлес-обчислення дають можливість автоматично масштабуватися без необхідності керувати інфраструктурою. Це ідеально підходить для подієво-орієнтованих застосунків, де обробка ініціюється потоками даних або подіями без постійного виділення обчислювальних ресурсів.

Обґрунтування вибору інструментів і технологій:

– «Apache Kafka» і «Apache Flink» обрані через їх здатність обробляти потоки даних з високою пропускною здатністю, низькою затримкою та

відмовостійкістю, що є критично важливим для аналітики в режимі реального часу.

– Apache Spark є ідеальним рішенням для середовищ, де потрібна як потокова, так і пакетна обробка, що робить його універсальним інструментом для складної аналітики.

– Хмарні сервіси «AWS Lambda» і «Google Cloud Pub/Sub» забезпечують простоту використання й автоматичне масштабування, знижуючи операційні витрати та підвищуючи гнучкість системи.

«Elasticsearch» – це розподілена пошукова та аналітична система з відкритим кодом, створена для зберігання, пошуку та аналізу великих обсягів даних майже в реальному часі. Вона забезпечує надзвичайно швидкий пошук навіть у великих масивах даних завдяки використанню індексації документів у форматі JSON, де кожне поле автоматично стає доступним для пошуку.

Система є розподіленою за своєю природою, тому легко масштабується горизонтально, дозволяючи будувати кластери з великою кількістю вузлів. Окрім пошуку, «Elasticsearch» також підтримує потужну аналітику через агрегації та статистичні запити, що дає змогу аналізувати великі обсяги інформації в режимі, наближеному до реального часу. Вона також забезпечує пошук за релевантністю, тобто з урахуванням значущості результатів для користувача.

Можливості «Elasticsearch» можна розширювати за допомогою численних плагінів та інтеграцій. Найпоширенішою є інтеграція з «Kibana» – інструментом для візуалізації даних. «Elasticsearch» часто використовується для пошуку товарів на вебсайтах, зберігання і аналізу логів у поєднанні з «Logstash» і «Beats», аналізу даних із соціальних мереж, пошуку за зображеннями з використанням плагінів та реалізації рекомендаційних систем.

«Elasticsearch» є ключовим компонентом «Elastic Stack» (див. рисунок 3.2), раніше відомого як «ELK Stack», що включає «Elasticsearch», «Logstash», «Kibana», а також «Beats» – легкі агенти для збору даних.

Рисунок 3.2 – Ключові компоненти «Elastic Stack»

Використання «Elasticsearch» для інформаційної системи з реалізацією релевантного пошуку є цілком обґрунтованим з огляду на технічні можливості цієї платформи. «Elasticsearch» дає змогу виконувати потужний повнотекстовий пошук із високою швидкістю та точністю, що особливо важливо для систем, у яких користувачі вводять запити природною мовою.

Його основна перевага – це індексація всіх текстових полів документа, що забезпечує гнучке ранжування результатів за релевантністю, а також можливість враховувати синоніми, вагу полів, частоту слів та інші параметри для точнішого відповідника запиту.

«Elasticsearch» підтримує аналіз тексту, включаючи морфологічну нормалізацію, токенізацію, фільтрацію стоп-слів, що особливо корисно для обробки мов із розвиненою граматикою, таких як українська. Завдяки цьому система може краще зрозуміти зміст запиту та знайти найбільш відповідні документи, навіть якщо точне формулювання не збігається з текстом у базі даних. Це значно покращує зручність користувача та ефективність взаємодії з системою.

Окрім цього, «Elasticsearch» дає можливість реалізувати складні запити, які комбінують повнотекстовий пошук із фільтрами за метаданими, такими як дата, категорія, автор тощо. Це робить його особливо корисним у тих випадках, коли інформаційна система обслуговує великий масив структурованих і неструктурованих даних, наприклад – у бібліотечних, медіа- або юридичних системах.

Ще однією важливою перевагою є масштабованість «Elasticsearch», що дає змогу ефективно обробляти великі обсяги даних у розподіленому середовищі. Це гарантує стабільну роботу системи навіть при високому навантаженні або

зростанні кількості користувачів. Також варто відзначити можливість реального часу, яка дає можливість оновлювати дані в індексі майже миттєво, що є критичним для систем, які працюють із динамічною інформацією.

У підсумку, «Elasticsearch» є оптимальним інструментом для побудови інформаційної системи з релевантним пошуком, оскільки поєднує гнучкість, швидкість, аналітичні можливості та підтримку багатих пошукових сценаріїв, що разом значно покращують точність і ефективність пошуку.

«Apache Solr» – це потужна, масштабована платформа для реалізації пошукових систем, розроблена на основі бібліотеки «Apache Lucene». Вона спеціалізується на повнотекстовому пошуку та аналізі даних у великих масивах інформації. «Solr» підтримує індексацію (див. рисунок 3.3), пошук та керування документами, забезпечуючи при цьому високу продуктивність і гнучкість у налаштуванні.

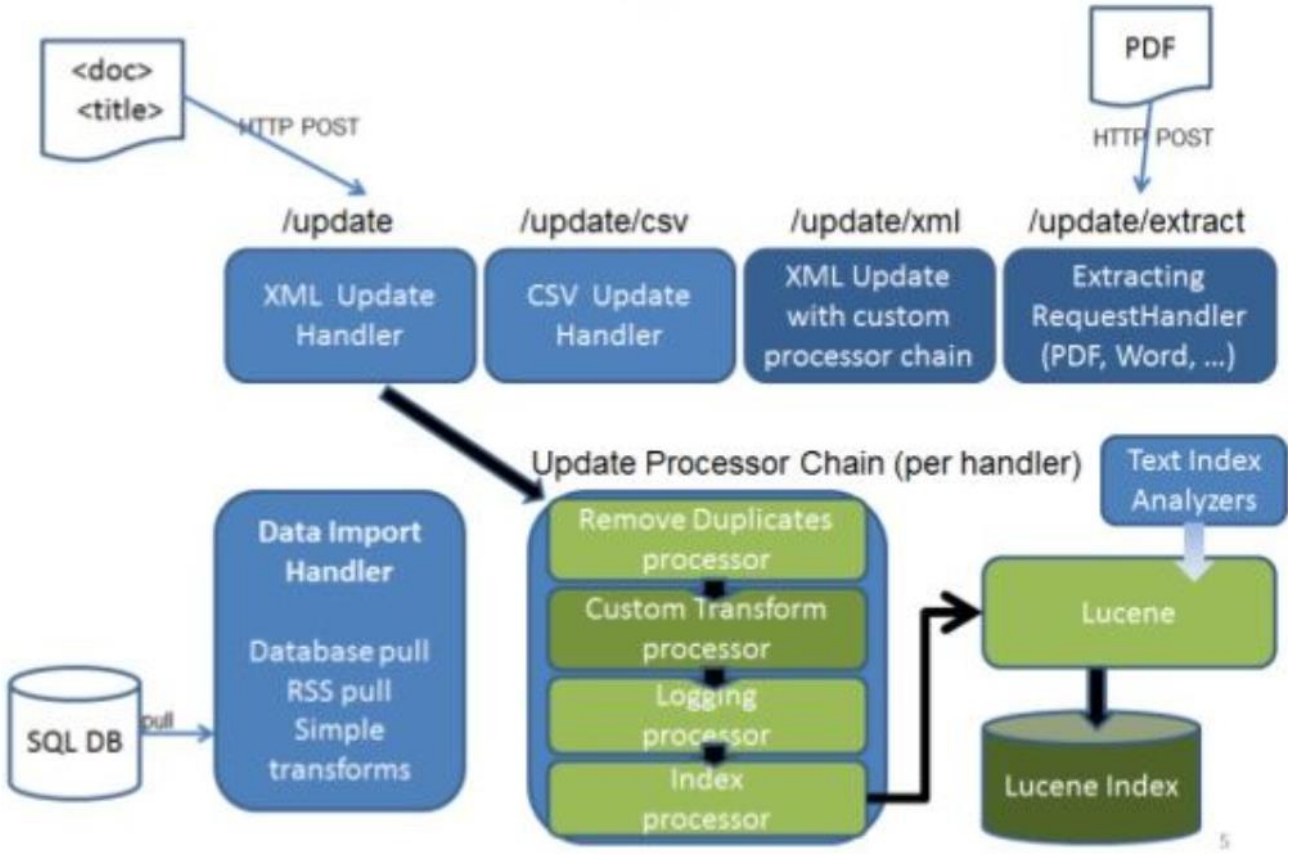

Рисунок 3.3 – Індексація «Apache Solr»

Його функціонал дає змогу реалізувати складні запити, релевантний пошук, фільтрацію результатів та кластеризацію, що робить його зручним

інструментом для побудови інформаційних систем, де критично важлива якість пошуку.

Однією з ключових переваг «Apache Solr» є його розширювана архітектура та підтримка великої кількості форматів даних, таких як XML, JSON, CSV та інші. «Solr» забезпечує підтримку реплікації, балансування навантаження та масштабованості, що дає йому можливість працювати у розподіленому середовищі з великим обсягом запитів. Це робить його особливо привабливим для корпоративних застосунків, систем електронної комерції, бібліотек, наукових порталів та інших систем, які працюють з великим обсягом текстових даних.

«Solr» також включає модулі для аналізу мови, автоматичної корекції помилок у запитах, підказок («autocomplete») та сортування результатів за складними критеріями. Його відкритий код і активна спільнота розробників сприяють постійному розвитку проєкту, що дозволяє системам, побудованим на «Apache Solr», залишатися актуальними та ефективними в умовах зростаючих вимог до обробки інформації. Завдяки цим характеристикам, «Solr» є одним з найпопулярніших рішень для реалізації високоякісного текстового пошуку у професійних та наукових інформаційних системах.

Використання «Apache Solr» для інформаційної системи з релевантним пошуком є обґрунтованим рішенням, оскільки ця платформа забезпечує високу точність та швидкість обробки текстових запитів. Завдяки своїй архітектурі, побудованій на базі «Apache Lucene», «Solr» дає змогу здійснювати повнотекстовий пошук із підтримкою розширених механізмів ранжування результатів за ступенем відповідності, що критично важливо для систем, орієнтованих на надання релевантної інформації користувачу.

«Solr» забезпечує підтримку морфологічного аналізу тексту, нормалізації термінів, стемінгу, синонімів та автозаповнення, що значно покращує користувацький досвід. Його алгоритми дозволяють обробляти запити навіть із помилками, неточностями чи варіативними формами слів, що збільшує ймовірність отримання коректних результатів.

Крім того, завдяки гнучкому налаштуванню фільтрів і фасетного пошуку, «Solr» дає можливість користувачам звузити результати за змістовими ознаками, підвищуючи якість навігації по великих обсягах даних.

«Solr» також є ефективним інструментом у системах, де необхідна висока масштабованість, адже підтримує розподілену архітектуру, реплікацію даних і балансування навантаження. Це забезпечує стабільну роботу навіть при великій кількості одночасних користувачів та великих обсягах індексованої інформації. Враховуючи ці переваги, «Apache Solr» є раціональним вибором для побудови інформаційних систем з фокусом на релевантний пошук, де важливо швидко і точно відповідати на запити користувачів.

3.3 Інтеграція та взаємодія інформаційної системи з використанням релевантного пошуку

Ключовим аспектом впровадження інформаційної системи з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr»для аналітики в режимі реального часу в хмарі є забезпечення безшовної інтеграції з існуючими корпоративними системами та підтримка сумісності з різноманітними форматами даних.

Забезпечення безшовної інтеграції з існуючими корпоративними системами. API та конектори – пайплайни даних у режимі реального часу повинні інтегруватися з різними корпоративними системами, наприклад, «CRM», «ERP», транзакційні бази даних. Використання стандартизованих API та попередньо створених конекторів, наприклад, «Kafka Connect», «AWS Glue» полегшує інтеграцію з системами, такими як «Salesforce», «SAP» або «Oracle».

Інструменти ETL/ELT – для імпорту даних з застарілих систем інтеграція з інструментами ETL або ELT, такими як «Talend» або «Apache Nifi», може автоматизувати процеси витягування та трансформації даних.

Інтеграція API в режимі реального часу – API «REST» або «gRPC» повинні використовуватися для потокової передачі даних з різних джерел, що дає змогу

здійснювати імпорт та обробку даних у режимі реального часу через різні платформи.

Сумісність з різними форматами даних:

– Інформаційна система з використанням релевантного пошуку повинна бути здатна обробляти структуровані, напівструктуровані та неструктуровані формати даних, щоб врахувати різноманітність даних, що генеруються в різних галузях.

– Структуровані дані, наприклад, SQL-бази даних – звичні формати CSV, «Parquet» та «Avro», підтримуються більшістю платформ для обробки даних і інформаційних систем, що забезпечує безперебійну інтеграцію з реляційними базами даних.

– Напівструктуровані дані, наприклад, JSON, XML – формати даних, як JSON і XML, часто використовуються в API та логах. Інструменти «Apache Avro» та «Google Protocol Buffers» можуть стандартизувати ці формати і полегшити ефективне зберігання та обробку.

– Неструктуровані дані, наприклад, текст, зображення, логи – джерела неструктурованих даних, зокрема, веб-логи або мультимедійний контент, можна імпортувати за допомогою рішень для зберігання об’єктів, наприклад, «AWS S3», «Google Cloud Storage», та обробляти через спеціалізовані фреймворки «Apache Solr» або «Elasticsearch» для пошуку та індексації.

3.4 Безпека інформаційної системи з використанням релевантного пошуку та відповідність вимогам

Конфіденційність даних, шифрування та контроль доступу є важливими аспектами при роботі інформаційної системи з використанням релевантного пошуку з даними в режимі реального часу, особливо коли йдеться про чутливу або регульовану інформацію. Тому доцільно розглянути конфіденційність даних та шифрування.

Шифрування під час передачі та зберігання – для захисту цілісності та конфіденційності даних у інформаційній системі з використанням релевантного

пошуку повинно бути реалізовано шифрування TLS/SSL для даних під час передачі та шифрування AES-256 для даних, що зберігаються. Постачальники хмарних послуг надають вбудовані можливості шифрування «AWS KMS» та «Google Cloud KMS» для безпечного управління ключами шифрування.

Маскування даних – для чутливих даних, таких як персонально ідентифікована інформація (PII), повинні використовуватися техніки маскування даних, щоб лише авторизовані користувачі мали доступ або могли переглядати чутливу інформацію.

Токенізація – методи токенізації можуть замінювати чутливі дані на неконфіденційні токени для використання в обробці або зберіганні, забезпечуючи, що оригінальні дані не будуть невиправдано відкриті.

Контроль доступу на основі ролей – потрібно забезпечити, щоб користувачам і сервісам надавались мінімально необхідні права доступу до ресурсів. Інструменти, як «AWS IAM» та «Azure Active Directory», допомагають управляти ролями користувачів і впроваджувати політики доступу.

Протоколи автентифікації – необхідно впровадити «OAuth 2.0» або «OpenID Connect» для безпечної автентифікації та авторизації користувачів, особливо в багатокористувацьких хмарних середовищах.

Аудиторські журнали – потрібно вести детальні аудиторські журнали для відстеження діяльності користувачів, змін у системі та можливих порушень безпеки. Більшість хмарних сервісів, наприклад, «AWS CloudTrail», «Azure Monitor», надають вбудовані можливості для ведення журналів і моніторингу, що полегшує проведення аудитів безпеки.

Розглянемо відповідність інформаційної системи з використанням релевантного пошуку вимогам галузевих стандартів.

«GDPR» – загальний регламент захисту даних – інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» повинна бути розроблена відповідно до вимог GDPR, щоб забезпечити обробку персональних даних законним, прозорим та відповідальним способом.

«HIPAA» – закон про доступність та переносимість медичного страхування – для медичних даних система повинна відповідати вимогам

«HIPAA», щоб забезпечити конфіденційність, цілісність і безпеку медичної інформації.

«SOC 2»/«ISO 27001» – для компаній, що обробляють чутливі дані, відповідність стандартам «SOC 2» або «ISO 27001» гарантує, що система відповідає суворим вимогам щодо безпеки, доступності та конфіденційності.

Дотримуючись цих вимог до платформи, технологічного стеку, інтеграційних вимог та відповідності стандартам безпеки, ця інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» забезпечує надійне, масштабоване і безпечне рішення для аналітики в режимі реального часу в хмарі.

3.5 Особливості застосування інформаційної системи з використанням релевантного пошуку

Запропонована інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» для аналітики в режимі реального часу може бути застосована у різних галузях для вирішення конкретних бізнес-завдань. Розглянемо декілька прикладів використання, де інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» має значний потенціал.

Аналіз IoT-даних для «розумних міст» або промислового моніторингу. «Розумні міста» та системи промислового моніторингу активно використовують IoT-давачі для збору даних про рух транспорту, якість повітря, споживання енергії, стан обладнання та інше. Обробка даних в режимі реального часу дає можливість отримувати миттєві висновки та приймати рішення, що підвищує ефективність, безпеку та сталий розвиток.

IoT-давачі та потоки даних – пристрої «розумні» лічильники, камери та давачі руху, постійно генерують величезну кількість даних. Запропонована інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» може отримувати та обробляти ці дані в режимі реального часу, даючи змогу здійснювати моніторинг та управління міськими

системами, такими як громадський транспорт, управління відходами чи енергомережі.

Аналітика в режимі реального часу та прийняття рішень – наприклад, дані про потік транспорту можуть бути проаналізовані для оптимізації часу роботи світлофорів та зменшення заторів, а дані про якість повітря можуть спонукати до автоматичних сповіщень про рекомендації для здоров'я.

Технологічний стек – поєднання «Elasticsearch», «Apache Solr» пошукові рушії на базі Apache Lucene для індексації та реалізації повнотекстового релевантного пошуку, «Apache Kafka» для обробки повідомлень з високою пропускною здатністю, «Apache Flink» для обробки потоків та «AWS Kinesis» для отримання даних в режимі реального часу забезпечує те, що дані отримуються, обробляються та зберігаються без затримок.

Виявлення шахрайства в режимі реального часу під час фінансових транзакцій. У фінансовому секторі виявлення шахрайських операцій в режимі реального часу є критично важливим. Аналітика в режимі реального часу може виявити незвичайні патерни в фінансових транзакціях і позначити їх для негайного розслідування, що знижує ризик втрат і підвищує довіру клієнтів.

Потоки транзакцій – дані про транзакції з кредитних карток, банківські перекази та онлайн-платежі передаються в режимі реального часу та обробляються за допомогою моделей машинного навчання для виявлення підозрілих дій, таких як незвичайні витрати або транзакції з незнайомих локацій.

Запобігання шахрайству – інформаційна система може надавати миттєві сповіщення, що дає змогу вжити заходів, таких як блокування транзакції або повідомлення користувача, таким чином запобігаючи потенційному шахрайству.

Технологічний стек – «Apache Spark» для обробки даних та виконання моделей машинного навчання, «Kafka» для потокових даних, а також «AWS Lambda» для безсерверного виконання детекції шахрайства за запитом роблять це рішення висококоскалованим та ефективним.

Прогнозне обслуговування в виробництві. Прогнозне обслуговування використовує дані з давачів обладнання та історичні записи про обслуговування для прогнозування, коли техніка або обладнання ймовірно вийде з ладу. Обробка

цих даних в режимі реального часу даючи можливість зменшити час простою, знизити витрати на обслуговування та підвищити ефективність роботи.

Дані з давачів в режимі реального часу – давачі, встановлені на машинах та обладнанні, генерують безперервні потоки даних, зокрема вібрації, температуру та тиск. Ці дані отримуються та обробляються в режимі реального часу для прогнозування потенційних поломок або потреби в обслуговуванні.

Прогнозні висновки – система може ініціювати автоматичні дії щодо планування роботи обслуговуючої команди або надсилання сповіщень операторам, коли машина ймовірно зазнає поломки.

Технологічний стек – «Apache Flink» для обробки даних в режимі реального часу та «Google Cloud Pub/Sub» для отримання потокових даних, разом з «Google BigQuery» для зберігання та аналізу, забезпечують швидкі та точні прогнози з низькою затримкою.

Оцінка продуктивності є критично важливою для розуміння того, наскільки добре запропонована інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» відповідає вимогам до обробки даних у режимі реального часу для великих масштабних застосунків. Нижче наведено ключові метрики продуктивності для оцінки ефективності інформаційної системи з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr».

Аналіз затримки. Затримка – це час, який проходить від моменту, коли дані потрапляють у інформаційну систему, до моменту їх обробки та надання результату, який можна використовувати. Для систем аналітики в режимі реального часу низька затримка є критично важливою. Бенчмаркінг затримки обробки в режимі реального часу – інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» повинна бути оцінена за галузевими стандартами щодо затримки, з метою досягнення субсекундної затримки при обробці даних. Наприклад, у типовій системі виявлення шахрайства в фінансових транзакціях затримка має бути нижчою за 200 мілісекунд, щоб сповіщення були надіслані в розумні терміни.

Техніки для зменшення затримки – оптимізації процесів обробки даних в пам’яті, наприклад, «Apache Ignite», «Apache Flink» та периферійні обчислення («edge computing») для початкового фільтрування даних, можуть значно знизити затримку.

Аналіз пропускної здатності. Пропускна здатність визначає здатність інформаційної системи обробляти великі обсяги даних без погіршення продуктивності. Висока пропускна здатність необхідна для того, щоб інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» могла обробляти великомасштабні потоки даних.

Бенчмаркінг пропускної здатності – пропускну здатність зазвичай вимірюють у транзакціях на секунду (TPS) або подіях на секунду. Інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» повинна бути здатною масштабуватися для обробки мільйонів подій на секунду, особливо в випадках використання IoT або моніторингу соціальних мереж.

Оптимізації для високої пропускної здатності – технології «Apache Kafka» забезпечують відмовостійкість і горизонтальне масштабування для обробки великих потоків даних. Кластерні конфігурації та балансування навантаження можуть додатково підвищити пропускну здатність.

Очікувані результати – у системі моніторингу IoT-трафіку «розумного міста» інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» повинна мати здатність обробляти понад десятки мільйонів подій на секунду, зберігаючи цілісність даних та забезпечуючи аналітику в режимі реального часу.

Аналіз масштабованості. Масштабованість визначає здатність інформаційної системи з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» адаптуватися до зростаючих обсягів даних без втрати продуктивності. Інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» повинна масштабуватися

горизонтально через розподілені хмарні ресурси, щоб відповідати зростаючим вимогам до обробки даних в режимі реального часу.

Еластичне масштабування – використання хмарних рішень, таких як Kubernetes для оркестрації контейнерів та безсерверні обчислення, наприклад, «AWS Lambda», «Google Cloud Functions», даючи системі можливість масштабуватися вгору чи вниз в залежності від обсягу даних, забезпечуючи оптимальне використання ресурсів.

Бенчмаркінг горизонтального масштабування – система повинна бути здатна масштабуватися безшовно, від обробки сотень тисяч подій до мільйонів подій на секунду, без значного збільшення часу обробки чи затримки.

Очікувані результати – у випадку з прогнозним обслуговуванням інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» повинна продемонструвати здатність масштабуватися від сотень пристроїв до тисяч пристроїв, зберігаючи стабільну продуктивність за пропускною здатністю та затримкою.

Відмовостійкість і відновлення. Відмовостійкість гарантує, що система залишатиметься працездатною навіть у разі відмови обладнання, проблем з мережею чи збоїв програмного забезпечення. Аналітика в режимі реального часу на основі хмари повинна бути стійкою, щоб мінімізувати час простою та запобігти втраті даних.

Тестування відмовостійкості – система повинна бути оцінена на здатність відновлюватися після збоїв, таких як аварія вузла, без втрати даних. Техніки, зокрема, контрольні точки та реплікація в фреймворках «Apache Flink» або «Kafka» можуть забезпечити високу доступність. Очікувані результати – інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» повинна продемонструвати автоматичне відновлення після збоїв вузлів протягом кількох секунд, без втрати даних або значних збоїв у обробці даних в режимі реального часу.

Ефективність системи та використання ресурсів. Ефективність системи оцінюється за тим, як ефективно вона використовує хмарні ресурси, зокрема,

обчислювальна потужність, сховище та пам’ять. Оптимізація використання ресурсів є важливою для масштабування, яке даючи змогу зменшити витрати.

Моніторинг споживання ресурсів – інструменти «Google Stackdriver» і «AWS CloudWatch», можуть використовуватися для моніторингу використання ресурсів під час пікових навантажень. Очікувані результати – інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» повинна бути здатною динамічно масштабувати ресурси залежно від навантаження, одночасно мінімізуючи неактивні ресурси, що веде до економії витрат і ефективної роботи.

3.6 Перспективні напрямки розвитку інформаційна система з використанням релевантного пошуку

Попри досягнення в галузі аналітики в режимі реального часу на основі хмари, існують ряд викликів, які потрібно вирішити, щоб забезпечити оптимальну продуктивність і масштабованість цих систем. Розглянемо їх. Якість даних, точність обробки та складність системи. Якість даних – одним з головних викликів в аналітиці в режимі реального часу є забезпечення якості даних. Дані з різних джерел, таких як IoT-давачі, API або транзакційні системи, можуть бути шумними, непослідовними або неповними. Забезпечення точності збору, очищення та попередньої обробки даних є критично важливим для надійної аналітики. Точність обробки – у середовищах з високою швидкістю даних утримання точності обробки стає складним через постійний потік даних. Невеликі помилки на ранніх етапах конвеєра можуть поширюватися і впливати на подальший аналіз, призводячи до неправильних висновків.

Складність системи – системи аналітики в режимі реального часу на основі хмари часто включають складні архітектури з багатьма компонентами, що ускладнює їхнє управління, моніторинг і усунення несправностей. Ця складність зростає з масштабуванням систем, що вимагає ретельної оркестрації та оптимізації для забезпечення безперебійної роботи.

Управління та аналіз великих обсягів гетерогенних даних у режимі реального часу. Гетерогенність даних – системи аналітики в режимі реального часу зазвичай потребують обробки різних типів даних, наприклад, структурованих, напівструктурованих і неструктурованих. Обробка та інтеграція цих різноманітних форматів даних у єдиний конвеєр вимагає спеціалізованих технік і інструментів.

Обсяг і швидкість – зростаючий обсяг даних, разом із швидкістю, з якою їх потрібно обробляти, накладає величезне навантаження на систему. Забезпечення масштабування системи по горизонталі без впливу на продуктивність є постійним викликом. Крім того, забезпечення балансу між пропускною здатністю даних і обробкою з низькою затримкою залишається ключовою проблемою в багатьох випадках використання.

Майбутнє аналітики в режимі реального часу на основі хмари буде визначатися кількома ключовими тенденціями та інноваціями в технологіях. Ці тенденції покращать можливості та застосування обробки даних у режимі реального часу. Інтеграція машинного навчання та ШІ в конвеєри даних у режимі реального часу. Машинне навчання в режимі реального часу – з розвитком моделей машинного навчання, їх інтеграція безпосередньо в конвеєри даних у режимі реального часу дозволить здійснювати динамічне прийняття рішень. Моделі для виявлення аномалій, прогнозної аналітики та розпізнавання шаблонів будуть застосовуватися в режимі реального часу, даючи системам можливість автоматично адаптуватися до нових шаблонів даних без необхідності втручання людини. Інсайти на основі ШІ – використання ШІ для виявлення тенденцій, аномалій та кореляцій в даних у режимі реального часу може забезпечити більш точні та своєчасні інсайти. Інтеграція ШІ на периферійних пристроях IoT, наприклад, дозволить ще швидше приймати рішення, зменшуючи залежність від хмарних ресурсів для певних завдань аналітики.

Автоматизована підготовка даних. Алгоритми машинного навчання можуть використовуватися для автоматичного очищення та попередньої обробки даних під час їх збору, що підвищує точність і ефективність аналітики в режимі реального часу.

Розвиток периферійних обчислень і мереж 5G для розподіленої обробки даних. Периферійні обчислення – периферійні обчислення означають обробку даних ближче до джерела їх генерації, наприклад, пристроїв IoT, давачів, а не передачу всіх даних до хмари для обробки. Це значно зменшить затримку, покращить аналітику в режимі реального часу і мінімізує навантаження на ширину смуги пропускання хмарної інфраструктури. З ростом використання периферійних пристроїв, аналітика в режимі реального часу на основі хмари повинна та інформаційна система з використанням релевантного пошуку адаптуватися для обробки розподілених даних як в хмарі, так і на периферійних вузлах.

Запуск мереж 5G значно покращить швидкість передачі даних в інформаційній системі з використанням релевантного пошуку, зменшить затримку і дозволить ефективніше передавати великі набори даних. Це підтримуватиме реальні застосування в таких галузях, як автономне водіння, смарт-міста та доповнена реальність, де обробка з низькою затримкою та високою пропускною здатністю є критично важливою.

Покращені хмарні технології для оптимізації операцій. Безсерверні обчислення – зростаюче впровадження безсерверних архітектур, таких як «AWS Lambda» та «Google Cloud Functions», спростить розгортання та масштабування конвеєрів даних в режимі реального часу. Безсерверні обчислення зменшують складність управління інфраструктурою та дає розробникам змогу зосередитися на написанні бізнес-логіки, а не на обслуговуванні серверів.

Контейнеризація та «Kubernetes» – контейнеризовані сервіси, оркестровані за допомогою «Kubernetes», ще більше оптимізують управління розподіленими системами обробки даних, що дає можливість швидко масштабувати та розгортати конвеєри даних для потреб інформаційної системи з використанням релевантного пошуку.

3.7 Потенційні напрямки досліджень

Існує кілька напрямків досліджень, які мають великий потенціал для розвитку галузі аналітики в режимі реального часу на основі хмари та вирішення існуючих проблем інформаційної системи з використанням релевантного пошуку. Екологічні обчислення – з ростом систем аналітики в режимі реального часу енергоспоживання в хмарних дата-центрах стає важливою проблемою. Дослідження з оптимізації розподілу ресурсів для зменшення вуглецевого сліду при збереженні продуктивності системи є критичним. Це включає розробку алгоритмів, які динамічно розподіляють ресурси залежно від навантаження та мінімізують час простою для зменшення енергоспоживання.

Енергоефективні алгоритми – оптимізація алгоритмів для підвищення енергоефективності є ще одним напрямком досліджень при подальному розвитку інформаційної системи з використанням релевантного пошуку. Це включає зменшення обчислювальної складності аналітики в режимі реального часу та покращення ефективності апаратного забезпечення, що використовується для обробки даних.

Покращення прийняття рішень в режимі реального часу за допомогою передової аналітики для потреб інформаційної системи з використанням релевантного пошуку. Розширене злиття даних – інтеграція даних з кількох джерел, наприклад, давачів, історичних даних, зовнішніх API, в режимі реального часу є важливою для покращення прийняття рішень. Дослідження передових методів злиття даних дозволить системам ефективніше об’єднувати різні потоки даних, надаючи глибші інсайти та точніші прогнози.

Аналітика графів в режимі реального часу – обробка графів в режимі реального часу стає все важливішою, особливо в таких сферах, як виявлення шахрайства, аналітика соціальних мереж та рекомендаційні системи. Розробка нових графових алгоритмів, здатних обробляти великі, динамічні графи в режимі реального часу, дозволить швидше виявляти зв’язки та шаблони в складних наборах даних інформаційної системи з використанням релевантного пошуку.

Пояснювальний ШІ в системах в режимі реального часу – оскільки моделі ШІ інтегруються в конвеєри даних в режимі реального часу, важливим стає забезпечення пояснюваності їхніх рішень. Дослідження в галузі пояснювального

ШІ (XAI) дасть змогу забезпечити прозорість прийняття рішень в системах, що використовують моделі машинного навчання, а саме в інформаційній системі з використанням релевантного пошуку, що дасть можливість зацікавленим сторонам довіряти та перевіряти інсайти і прогнози в режимі реального часу.

Розподілене та федеративне навчання для аналітики в режимі реального часу для потреб інформаційної системи з використанням релевантного пошуку. Федеративне навчання дає можливість тренувати моделі на децентралізованих пристроях без необхідності передачі «сирих» даних з пристрою, що зберігає конфіденційність. Ця технологія є особливо перспективною в середовищах периферійних обчислень та IoT, де конфіденційність даних є важливою. Дослідження моделей федеративного навчання, які можуть працювати в конвеєрах даних в режимі реального часу, буде критичним для забезпечення безпечного та масштабованого машинного навчання на периферії.

Розподілені інформаційні системи з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr»обробки даних – оскільки дані все більше розподіляються між різними хмарними та периферійними середовищами, розробка розподілених інформаційних систем з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» обробки даних, які безшовно агрегують та аналізують дані з кількох джерел, стане необхідною. Потрібні дослідження для ефективнішого поділу даних, агрегації та обробки даних в режимі реального часу на розподілених системах.

Запропонована інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» для аналітики в режимі реального часу на основі хмари пропонує гнучке, масштабоване та малозатратне рішення для різних галузей. Завдяки практичним випадкам використання, таким як аналіз IoT та прогнозне обслуговування, інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» демонструє здатність надавати дієві аналітичні дані в режимі реального часу. Оцінка продуктивності підтверджує, що інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr» здатна обробляти дані з високою пропускною здатністю та низькою затримкою, має

надійну відмовостійкість і масштабованість, що робить його цінним інструментом для сучасних орієнтованих на дані застосунків.

Галузь аналітики в режимі реального часу на основі хмари швидко розвивається, і проблеми якості даних, складності систем та масштабованості залишаються важливими напрямками для покращення. Майбутні тенденції, зокрема, інтеграція машинного навчання, ШІ, периферійних обчислень та 5G мереж, обіцяють ще більше покращити можливості систем аналітики в режимі реального часу. Дослідження в таких галузях, як оптимізація ресурсів, передова аналітика та федеративне навчання, продовжать сприяти інноваціям та допомагатимуть вирішувати існуючі проблеми в обробці даних в режимі реального часу. Завдяки впровадженню цих розробок організації будуть краще підготовлені для ефективного використання всього потенціалу даних в режимі реального часу для прийняття обґрунтованих рішень та підвищення операційної ефективності інформаційної системи з використанням релевантного пошуку.

Запропонована архітектура інформаційної системи з використанням релевантного пошуку для аналітики в режимі реального часу на основі хмари пропонує всебічне рішення для вирішення проблем, з якими стикаються існуючі архітектури даних у процесінгу великих обсягів потокових даних з низькою затримкою та високою стійкістю до відмов. Зосередивши увагу на ключових принципах проєктування, таких як масштабованість, обробка з низькою затримкою, стійкість до відмов, гнучкість і розширюваність, ця архітектура забезпечує надійний фундамент для побудови потоків даних нового покоління, які можуть задовольняти вимоги сучасних хмарних середовищ. Основні внески в архітектуру включають:

– Масштабована архітектура – горизонтальне масштабування забезпечує здатність обробляти зростаючі обсяги даних без шкоди для продуктивності.

– Обробка з низькою затримкою – інтеграція даних, трансформація та аналітика в режимі реального часу досягаються за допомогою обробки в пам’яті та технік обробки потоків.

– Стійкість до відмов і висока доступність – резервування, контрольні точки та стратегії відновлення забезпечують стійкість і надійність системи в розподілених середовищах.

– Розширюваність – модульна архітектура дає змогу розвивати систему відповідно до змінних потреб бізнесу, підтримуючи різні джерела та призначення даних.

Впровадження цієї новаторської архітектури може значно вплинути на бізнеси, надаючи їм можливість обробляти та аналізувати дані в режимі реального часу, що відкриває численні переваги:

– Поліпшення прийняття рішень – аналітика в режимі реального часу дає змогу компаніям приймати своєчасні, обґрунтовані даними рішення, що може покращити операційну ефективність, зменшити ризики та стимулювати інновації. Наприклад, компанії в фінансовому секторі можуть миттєво виявляти шахрайство, а виробничі підприємства можуть виконувати передбачуване технічне обслуговування для уникнення дорогих простоїв.

– Поліпшення досвіду клієнтів – аналізуючи поведінку користувачів і дані про взаємодію в режимі реального часу, компанії можуть персоналізувати свої послуги та оптимізувати взаємодії з клієнтами, що призводить до покращення задоволеності та лояльності. Наприклад, платформи електронної комерції можуть динамічно коригувати ціни або рекомендувати продукти в залежності від реальних умов на ринку.

– Конкурентна перевага – аналітика даних в режимі реального часу може надати значну конкурентну перевагу інформаційним системам з використанням релевантного пошуку, даючи можливість організаціям швидше реагувати на зміни на ринку, виявляти нові тенденції та приймати проактивні рішення. В таких галузях, як IoT, фінансові послуги та охорона здоров’я, де своєчасні дані критично важливі, впровадження трубопроводів даних нового покоління може стати ключовим моментом.

– Економічна ефективність – автоматизуючи управління трубопроводами даних та використовуючи ресурси, оптимізовані для хмари, компанії можуть оптимізувати використання інфраструктури, що призводить до заощаджень.

Гнучкість архітектури в обробці різних форматів і джерел даних також знижує складність управління трубопроводами даних, що в свою чергу знижує операційні витрати.

Еволюція аналітики даних на основі хмари змінює спосіб, яким компанії обробляють і використовують дані. Зі зростанням обсягів, різноманітності та швидкості даних традиційні методи пакетної обробки більше не здатні задовольнити вимоги до аналітики в режимі реального часу. Запропонована архітектура пропонує шлях уперед, надаючи високомасштабоване, з низькою затримкою та стійке до відмов рішення, здатне справлятися з викликами сучасної хмарної аналітики.

Якщо галузі продовжують впроваджувати хмарні технології, аналітика даних у режимі реального часу відіграватиме ключову роль у наданні організаціям можливості отримувати цінні інсайти з їхніх даних швидше та ефективніше. Майбутні досягнення в таких сферах, як машинне навчання, периферійні обчислення та мережі 5G, ще більше покращать можливості аналітики в режимі реального часу, прокладаючи шлях для більш інтелектуальних і адаптивних інформаційних систем, які стимулюватимуть інновації та успіх бізнесу.

На завершення, запропонована архітектура інформаційної системи з використанням релевантного пошуку є значним кроком до того, щоб бізнес структури, установи та організації могли повною мірою реалізувати потенціал аналітики даних у режимі реального часу на основі хмари, сприяючи новій епосі більш розумного, орієнтованого на дані прийняття рішень у різних галузях.

3.8 Висновок до третього розділу

В третьому розділі кваліфікаційної роботи спроєктовано архітектуру інформаційної системи з використанням релевантного пошуку. Описано платформи та технологічний стек. Розглянуто інтеграцію та взаємодію інформаційної системи з використанням релевантного пошуку. Висвітлено безпекові аспекти використання інформаційної системи з використанням

релевантного пошуку та відповідність вимогам. Подано особливості застосування інформаційної системи з використанням релевантного пошуку. Описано перспективні напрямки розвитку інформаційна система з використанням релевантного пошуку. Проаналізовано потенційні напрямки подальших досліджень.

4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1 Контроль за станом охорони праці

Кваліфікаційна робота освітнього рівня «Магістр» присвячена створенню інформаційної системи з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr». Оскільки інформаційно-технологічні проекти такого класу потребують активної інтеграції інформаційних та комунікаційних технологій в різнотипову за своєю природою обчислювальну інфраструктуру, то актуальним постає контроль за станом охорони праці.

Контроль у сфері охорони праці є ключовим інструментом для забезпечення безпечних умов роботи та запобігання нещасним випадкам. Його суть полягає в тому, що відповідальний суб’єкт здійснює моніторинг і перевірку того, як об’єкт контролю дотримується встановлених вимог, виконує покладені на нього обов’язки та реалізує свої функції в галузі охорони праці. Актуальність посилення контролю пояснюється тривожною тенденцією — попри реалізацію профілактичних заходів, рівень виробничого травматизму та кількість смертельних випадків не лише не зменшується, а подекуди навіть зростає. Однією з головних причин цієї ситуації є недостатня ефективність існуючих механізмів контролю за дотриманням норм та вимог законодавства з охорони праці. Здійснення контролю в цій сфері покладається як на роботодавців, так і на державні органи. Роботодавці відповідають за внутрішній контроль, що спрямований на постійне забезпечення безпечних умов праці всередині підприємства. Водночас державні органи здійснюють зовнішній контроль у межах нагляду за дотриманням нормативних актів з охорони праці на загальнодержавному рівні. Основною метою такої системи контролю є не лише виявлення та усунення порушень, а й запобігання їх виникненню шляхом впровадження превентивних заходів та зміцнення культури безпеки на робочих місцях.

– Контроль у сфері охорони праці є одним із найдієвіших засобів забезпечення безпечних умов на робочому місці, однак для досягнення бажаного

результату він має бути дійсно ефективним. Це вимагає комплексного підходу, зокрема вдосконалення чинного законодавства та нормативної бази, що регламентує вимоги до охорони праці. Також важливо сформувати дієвий державний механізм нагляду, який дозволятиме оперативно реагувати на порушення та запобігати їм. Крім того, вагому роль відіграє підвищення рівня обізнаності та відповідальності як роботодавців, так і працівників щодо дотримання правил безпеки. Формування культури охорони праці на всіх рівнях трудового процесу є необхідною умовою для зменшення виробничих ризиків і запобігання нещасним випадкам.

На практиці існує кілька основних видів контролю за станом охорони праці, які проводяться на підприємствах. Одним із них є трьохступеневий контроль, який закріплений нормативними актами, зокрема розділом 3 Галузевої Угоди між Міністерством регіонального розвитку та будівництва України і Профспілкою працівників промисловості будівельних матеріалів України на 2017-2018 роки. Окрім цього, оперативний контроль здійснюють керівники робіт та інші відповідальні особи підприємства, зокрема служба охорони праці. Вищі організації також виконують контроль на більш високих рівнях, таких як четвертий чи п’ятий ступінь. Місцеві органи влади та органи самоврядування також займаються контролем за охороною праці на підприємствах.

Відомчий контроль, що проводиться роботодавцем, є важливим елементом забезпечення безпеки праці на підприємстві. Згідно з Законом України про охорону праці, роботодавець зобов’язаний створити належні умови праці відповідно до нормативних вимог та гарантувати дотримання прав працівників у сфері охорони праці. Для цього роботодавець має впровадити систему управління охороною праці, проводити інструктажі для працівників та здійснювати перевірки дотримання вимог охорони праці [70].

Відомчий контроль за охороною праці є важливою складовою механізму забезпечення безпеки праці на підприємстві. Цей контроль дає можливість виявляти та усувати порушення норм і законодавства в галузі охорони праці, а також сприяє підвищенню рівня культури безпеки серед працівників. Окрім того, важливу роль відіграють громадський контроль та контроль з боку органів

державного нагляду, які здійснюють перевірки та нагляд за дотриманням вимог охорони праці. В межах одного підприємства адміністративний контроль за станом охорони праці реалізують керівник підприємства, головні спеціалісти та інші особи, які отримали адміністративні повноваження згідно з наказом роботодавця [70].

Відомчий контроль за охороною праці – це вид контролю, який здійснюють органи управління підприємством, установою, організацією з метою забезпечення безпеки праці на підпорядкованих їм об'єктах. Згідно з Порядком проведення планових виїзних перевірок підприємств, установ та організацій з питань охорони праці, затвердженим постановою Кабінету Міністрів України від 1 червня 2013 року № 440, відомчий контроль за охороною праці здійснюється службами охорони праці вищих організацій і галузевими спеціалістами. Під час перевірок складається акт, у якому фіксуються виявлені порушення вимог законодавства в галузі охорони праці, а також заходи, що необхідно вжити для їх усунення. Основною метою відомчого контролю є забезпечення виконання вимог законодавства з охорони праці на підпорядкованих підприємствах, усунення порушень та підвищення рівня культури безпеки праці серед працівників. Під час здійснення контролю перевіряються різні аспекти, зокрема виконання планів роботи з охорони праці, використання коштів на ці потреби, розробка проектної документації, виконання обов'язків з охорони праці службовими особами та інші питання, що стосуються безпеки праці. Для того, щоб відомчий контроль був ефективним, він має проводитися регулярно і систематично.

Оперативний контроль за охороною праці має важливе значення для своєчасного виявлення порушень і недоліків, що можуть призвести до нещасних випадків або загрози безпеці працівників. Такий контроль забезпечує не лише моніторинг виконання норм і стандартів охорони праці, але й дозволяє оперативно реагувати на зміни в умовах праці, технічне обладнання або людський фактор. За допомогою постійного контролю можна вчасно виявити потенційно небезпечні ситуації та усунути їх до того, як вони переростуть у серйозні проблеми.

Процес здійснення оперативного контролю на підприємствах вимагає чіткої організації та координації між різними рівнями управління. Перший ступінь оперативного контролю, який проводять безпосередньо на місцях праці, дозволяє швидко виявляти порушення під час робочого процесу. Це дає можливість негайно вжити заходи для усунення небезпеки, не даючи їй розвинутися. Крім того, на цьому етапі робітники можуть безпосередньо повідомляти про виниклі проблеми, що також підвищує рівень безпеки.

Другий ступінь контролю здійснюється на рівні спеціалістів і керівників підрозділів, що дає змогу виявити більш системні проблеми, які не завжди можна помітити на рівні окремих працівників. Це дозволяє проводити аналіз роботи з охорони праці в контексті загальних умов виробничого процесу та інфраструктури підприємства. Під час цього етапу контролю також аналізуються витрати на охорону праці та ефективність використання ресурсів, що сприяє більш збалансованому підходу до організації безпечного середовища для працівників.

Третій ступінь контролю, на якому перевірку проводять вищі органи управління підприємства, дозволяє здійснити глибшу оцінку виконання стандартів охорони праці на підприємстві, зокрема, у більш широкому контексті — з урахуванням галузевих особливостей і загальних вимог державного регулювання. Це дає можливість виявити більш серйозні структурні проблеми в організації охорони праці, зокрема, в процесах управління, підготовки персоналу та організації робочих місць.

Четвертий ступінь оперативного контролю, який здійснюється на рівні нарад і стратегічних рішень, є критично важливим для забезпечення постійного вдосконалення системи охорони праці. На цьому етапі аналізуються результати попередніх перевірок та вжиті заходи для усунення порушень. Крім того, приймаються рішення про необхідність впровадження нових методів або технологій для покращення умов праці і зменшення ризику виникнення небезпечних ситуацій.

Підвищення ефективності оперативного контролю також пов'язане з інтеграцією новітніх технологій, таких як Інтернет речей (IoT) і штучний

інтелект, що дає змогу автоматизувати процеси моніторингу і реагування на потенційні загрози. Наприклад, за допомогою IoT можна встановити датчики для постійного моніторингу стану технічного обладнання, виявлення аварійних ситуацій та автоматичного інформування керівництва або служб охорони праці про критичні ситуації. Штучний інтелект може допомогти в аналізі великої кількості даних, що збираються на підприємствах, для виявлення потенційних трендів і ризиків, а також для розробки прогностичних моделей безпеки.

Таким чином, оперативний контроль є необхідним елементом не тільки для забезпечення безпеки на виробництві, а й для підтримки високого рівня організації та готовності до запобігання аварій. Інтеграція нових технологій в цей процес дозволяє зробити його більш ефективним і своєчасним, а також значно знижує ймовірність виникнення нещасних випадків, роблячи робочі умови більш безпечними для всіх учасників процесу [70].

4.2 Психологічні чинники небезпеки

В кваліфікаційній роботі інформаційна система з використанням релевантного пошуку на основі «Elasticsearch» та «Apache Solr». Сучасні інформаційно-технологічні проєкти такого класу потребують від персоналу підвищеної уваги та концентрації. Зростає рівень стресу та психофізичного навантаження. Тому в цьому параграфі доцільно проаналізувати психологічні чинники небезпеки.

Згідно з аналізом статистичних даних та висновками фахівців у сфері безпеки життєдіяльності, переважна більшість травм на виробництві та в побуті – від 60 до 90% – трапляється з вини самих постраждалих. Така ситуація зумовлена кількома основними чинниками. Насамперед, це недостатній рівень професійної підготовки у питаннях безпеки, що призводить до неправильних дій у потенційно небезпечних умовах. До цього додається слабке виховання у сфері культури безпечної поведінки та відсутність внутрішньої мотивації дотримуватися встановлених правил безпеки. Окрему небезпеку становить допуск до виконання небезпечних робіт осіб, які мають підвищений ризик

травматизму. Крім того, чимало нещасних випадків стається через фізичну втому або психоемоційні стани, що знижують уважність і контроль над діями людини в критичних ситуаціях [71].

Існує сукупність факторів, які підвищують індивідуальну схильність людини до небезпеки. Серед них – особливості темпераменту, фізіологічні зміни в організмі, порушення роботи органів чуття та незадоволеність виконуваною діяльністю. Одночасно з цим, підвищений ризик виникає і через несприятливі умови самої роботи, зокрема надмірні фізичні та розумові навантаження, незручне положення тіла під час праці, високий ритм виконання завдань, емоційне напруження, надмірне навантаження на зір та слух, а також невідповідність між робочим місцем, інструментами і фізичними параметрами працівника. Такі обставини спричиняють значне виснаження організму, впливають на психоемоційний стан людини, уповільнюють її реакції, знижують точність орієнтації в просторі, послаблюють увагу та здатність концентруватися, а також погіршують сприйняття навколишнього середовища. Усе це значно підвищує ймовірність нещасних випадків і травмування.

У психологічній науці виокремлено спеціальний напрям — психологію безпеки, яка досліджує психологічні особливості, що впливають на поведінку людини в умовах ризику та небезпеки, особливо в процесі професійної діяльності. У межах цієї галузі вивчаються психічні властивості особистості, а також різноманітні стани – від емоційних до когнітивних, які виникають у працівника під час виконання трудових обов’язків. Ці стани можуть як позитивно, так і негативно впливати на ефективність, безпечність та стійкість поведінки в стресових або потенційно небезпечних умовах. Основу будь-якої психічної діяльності становлять психічні процеси — сприймання, мислення, пам’ять, уява, увага, мовлення тощо. Саме вони забезпечують формування нових знань, навичок, а також накопичення життєвого досвіду. Без активного функціонування цих процесів неможливий повноцінний розвиток особистості, прийняття рішень у критичних ситуаціях та ефективна адаптація до складних умов праці. Таким чином, вивчення психології безпеки є надзвичайно важливим для підвищення рівня безпеки в різних сферах людської діяльності [72]

У психології психічні процеси поділяються на пізнавальні, емоційні та вольові, і кожен з них відіграє ключову роль у регуляції поведінки та діяльності людини. Психічні властивості є відносно стабільними характеристиками особистості, які проявляються у сфері інтелекту, емоцій, волі, трудової активності тощо. Водночас психічні стани відображають тимчасові особливості психічної діяльності, що виникають у конкретних ситуаціях та безпосередньо впливають на ефективність функціонування всіх психічних процесів. На думку більшості психологів, здатність людини до продуктивної діяльності значною мірою залежить від рівня психічного напруження. У певних межах зростання напруження стимулює працездатність і покращує результати діяльності. Однак при досягненні надмірного рівня активізації відбувається зворотний ефект – продуктивність різко падає, аж до повної втрати здатності виконувати завдання.

У випадках надмірного психічного напруження можна спостерігати два його типи – гальмівний і збудливий. Гальмівна форма проявляється в уповільненні реакцій, руховій скутості, зниженні швидкості мислення та погіршенні пам’яті. Людина стає менш уважною, розсіяною, і навіть найпростіші дії починають виконуватись із затримками. Такий стан різко контрастує з її звичайною поведінкою у нормальному, спокійному стані. Збудливий тип, навпаки, супроводжується підвищеною активністю, надмірною балакучістю, тремтінням у руках та голосі. Людина починає виконувати багато зайвих дій, що не мають практичної доцільності: вона перевіряє прилади без потреби, змінює положення елементів керування, робить зайві рухи. Одночасно з цим посилюється емоційна нестабільність – з’являється дратівливість, запальність, грубість, що також не характерно для цієї особи у звичайних умовах. Такі зміни психічного стану можуть серйозно впливати на безпеку і ефективність трудової діяльності, особливо в умовах підвищеного ризику.

Надмірне психічне напруження часто стає причиною помилкових рішень та неадекватної поведінки в умовах стресу чи складної ситуації, що значно підвищує ризик виникнення травм або аварій. Психологи наголошують на важливості своєчасного розпізнавання особливих психічних станів, які мають критичне значення для забезпечення безпеки. До таких станів належать

короткочасні порушення свідомості, різкі зміни настрою, викликані психогенними факторами, а також афективні реакції, що виникають під впливом психоактивних речовин, таких як стимулятори, заспокійливі препарати, алкоголь тощо. Ці стани здатні суттєво вплинути на здатність особи адекватно оцінювати ситуацію, приймати рішення та діяти безпечно, що створює серйозні загрози у професійній діяльності та повсякденному житті [72]

Пароксизмальні стани – це вид розладів, що характеризуються короткочасною втратою свідомості, яка може тривати від кількох секунд до кількох хвилин. Такі порушення зазвичай виникають на тлі органічних захворювань головного мозку, зокрема при епілепсії. Сучасна медична діагностика дозволяє на ранньому етапі виявити осіб із прихованою схильністю до подібних станів. Ці люди не повинні займатися діяльністю, пов’язаною з підвищеним ризиком, наприклад, працювати на висоті, керувати транспортними засобами або виконувати небезпечні виробничі завдання. Психогенні зміни емоційного фону та афективні реакції зазвичай виникають під впливом сильних психічних подразників. При цьому настрій може різко знижуватися, з’являється апатія, байдужість, загальна загальмованість і млявість, що можуть тривати від кількох хвилин до кількох тижнів або навіть місяців. Такі стани часто спричиняються особистими втратами, конфліктами чи стресовими ситуаціями. Вони впливають на темп мислення, знижують увагу і самоконтроль, що значно підвищує ймовірність нещасних випадків. Афективні стани, які супроводжуються потужними емоційними вибухами, здатні різко змінити поведінку людини. Під впливом сильного роздратування або розчарування людина може втратити здатність до раціонального мислення, виявляти агресію, діяти імпульсивно і навіть небезпечно. Через це люди з вираженою схильністю до таких реакцій не повинні обіймати посади, що вимагають високого рівня відповідальності та емоційної стабільності.

Вживання психоактивних речовин, зокрема алкоголю, значно підвищує ризик виникнення травм і знижує загальний рівень безпеки при виконанні будь-якої діяльності. Хоча легкі стимулювальні засоби, як-от чай або кава, можуть тимчасово підвищити працездатність і подолати сонливість, їхній ефект є

короткочасним і обмеженим. У випадках, коли використовуються сильніші стимулятори під час виконання відповідальних завдань, це може мати протилежний ефект – погіршення фізичного та психічного стану, зниження концентрації уваги й уповільнення реакцій. Транквілізатори, хоч і мають заспокійливу дію та можуть допомогти уникнути нервового перенапруження, водночас спричиняють зниження психічної активності, викликають апатію, млявість і сонливість, що також є небезпечним у професійному середовищі. Особливо серйозний вплив на безпеку має вживання алкогольних напоїв. За результатами численних досліджень, від 40 до 60% дорожньо-транспортних пригод відбуваються саме через алкогольне сп’яніння. У виробничій сфері виявлено, що близько 64% смертельних випадків пов’язані з вживанням алкоголю та подальшими помилковими діями постраждалих.

4.3 Висновок до четвертого розділу

В четвертому розділі кваліфікаційної роботи освітнього рівня «магістр» описано контроль за станом охорони праці. Окремо розглянуто психологічні чинники небезпеки.

ВИСНОВКИ

В першому розділі кваліфікаційної роботи освітнього рівня «Магістр»:

– Розглянуто контекст та актуальність розробки інформаційної системи з використанням релевантного пошуку.

– Проведено аналіз предметної області.

– Описано популярні пошукові системи зображень системи пошуку та індексації зображень.

– Проаналізовано процес пошуку зображень людиною.

– Досліджено ключові підходи та алгоритми в системах пошуку та індексації зображень.

В другому розділі кваліфікаційної роботи:

– Описано провідні системи пошуку та індексації зображень.

– Описано існуючі інформаційні технології пошуку та індексації зображень, зокрема технології від Google.

– Розглянуто нейро-символьні підходи до пошуку інформації за зображеннями.

– Проведено огляд дата-пайплайнів.

– Також було провеено огляд принципів ключових принципів проєктування пайплайнів для обробки даних нового покоління.

В третьому розділі кваліфікаційної роботи:

– Спроєктовано архітектуру інформаційної системи з використанням релевантного пошуку.

– Описано платформи та технологічний стек.

– Розглянуто інтеграцію та взаємодію інформаційної системи з використанням релевантного пошуку.

– Висвітлено безпекові аспекти використання інформаційної системи з використанням релевантного пошуку та відповідність вимогам.

– Подано особливості застосування інформаційної системи з використанням релевантного пошуку.

– Описано перспективні напрямки розвитку інформаційна система з використанням релевантного пошуку.

– Проаналізовано потенційні напрямки подальших досліджень.

У розділі «Охорона праці та безпека в надзвичайних ситуаціях» описано контроль за станом охорони праці. Розглянуто психологічні чинники небезпеки.

ПЕРЕЛІК ДЖЕРЕЛ

1 M. Maltz, *New psycho-cybernetics*. Penguin, 2002.

2 R. Fergus, P. Perona, and A. Zisserman, “A visual category filter for google images,” in *Computer Vision-ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings, Part I 8*. Springer, 2004, pp. 242–256.

3 H. Hu, Y. Wang, L. Yang, P. Komlev, L. Huang, X. Chen, J. Huang, Y. Wu, M. Merchant, and A. Sacheti, “Web-scale responsive visual search at bing,” in *Proceedings of the 24th ACM SIGKDD international conference on knowledge discovery & data mining*, 2018, pp. 359–367.

4 D. Shahi, *Apache Solr: a practical approach to enterprise search*. Springer, 2015.

5 B. Dixit, *Elasticsearch essentials*. Packt Publishing Ltd, 2016.

6 S. Santini and R. Jain, “Similarity measures,” *IEEE Transactions on pattern analysis and machine Intelligence*, vol. 21, no. 9, pp. 871–883, 1999.

7 L. Wang, X. Zhao, Y. Si, L. Cao, and Y. Liu, “Context-associative hierarchical memory model for human activity recognition and prediction,” *IEEE Transactions on Multimedia*, vol. 19, no. 3, pp. 646–659, 2016.

8 D. M. McKeown, T. Bulwinkle, S. Cochran, W. Harvey, C. McGlone, and J. A. Shufelt, “Performance evaluation for automatic feature extraction,” *International Archives of Photogrammetry and Remote Sensing*, vol. 33, no. B2; PART 2, pp. 379–394, 2000.

9 S. M. Singh and K. Hemachandran, “Image retrieval based on the combination of color histogram and color moment,” *International journal of computer applications*, vol. 58, no. 3, 2012.

10 V. Babak, A. Zaporozhets, Y. Kuts, M. Fryz, L. Scherbak. Noise signals: Modelling and Analyses. Cham: Springer Nature Switzerland, 2025. 222 p. DOI: https://doi.org/10.1007/978-3-031-71093-3.

11 Y. Park and J.-M. Guldmann, “Measuring continuous landscape patterns with gray-level co-occurrence matrix (glcm) indices: An alternative to patch metrics?” *Ecological Indicators*, vol. 109, p. 105802, 2020.

12 J.-K. Kamarainen, V. Kyrki, and H. Kalviainen, “Invariance properties of gabor filter-based features-overview and applications,” *IEEE Transactions on image processing*, vol. 15, no. 5, pp. 1088–1099, 2006.

13 M. Pietika¨inen and G. Zhao, “Two decades of local binary patterns: A survey,” in *Advances in independent component analysis and learning machines*. Elsevier, 2015, pp. 175–210.

14 A. W. Busch, “Wavelet transform for texture analysis with application to document analysis,” Ph.D. dissertation, Queensland University of Technology, 2004.

15 J. Jing, S. Liu, G. Wang, W. Zhang, and C. Sun, “Recent advances on image edge detection: A comprehensive review,” *Neurocomputing*, vol. 503, pp. 259–271, 2022.

16 D. Yang, B. Peng, Z. Al-Huda, A. Malik, and D. Zhai, “An overview of edge and object contour detection,” *Neurocomputing*, vol. 488, pp. 470– 493, 2022.

17 S. Qi, Y. Zhang, C. Wang, J. Zhou, and X. Cao, “A survey of orthogonal moments for image representation: Theory, implementation, and evaluation,” *ACM Computing Surveys (CSUR)*, vol. 55, no. 1, pp. 1–35, 2021.

18 S. Gkelios, A. Sophokleous, S. Plakias, Y. Boutalis, and S. A. Chatzichristofis, “Deep convolutional features for image retrieval,” *Expert Systems with Applications*, vol. 177, p. 114940, 2021.

19 I. Melekhov, J. Kannala, and E. Rahtu, “Siamese network features for image matching,” in *2016 23rd international conference on pattern recognition (ICPR)*. IEEE, 2016, pp. 378–383.

20 M. Li, H. Wang, H. Dai, M. Li, C. Chai, R. Gu, F. Chen, Z. Chen, S. Li, Q. Liu *et al.*, “A survey of multi-dimensional indexes: past and future trends,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 8, pp. 3635–3655, 2024.

21 L. Chen, Y. Gao, X. Song, Z. Li, Y. Zhu, X. Miao, and C. S. Jensen, “Indexing metric spaces for exact similarity search,” *ACM Computing Surveys*, vol. 55, no. 6, pp. 1–39, 2022.

22 N. Ukey, Z. Yang, B. Li, G. Zhang, Y. Hu, and W. Zhang, “Survey on exact knn queries over high-dimensional data space,” *Sensors*, vol. 23, no. 2, p. 629, 2023.

23 L. Han, M. E. Paoletti, X. Tao, Z. Wu, J. M. Haut, P. Li, R. PastorVargas, and A. Plaza, “Hash-based remote sensing image retrieval,” *IEEE Transactions on Geoscience and Remote Sensing*, 2024.

24 L. Zhang, L. Zhang, X. Mou, and D. Zhang, “Fsim: A feature similarity index for image quality assessment,” *IEEE transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.

25 Y. Tang, L. H. U, Y. Cai, N. Mamoulis, and R. Cheng, “Earth mover’s distance based similarity search at scale,” *Proceedings of the VLDB Endowment*, vol. 7, no. 4, pp. 313–324, 2013.

26 M. Ionescu and A. Ralescu, “Fuzzy hamming distance in a content-based image retrieval system,” in *2004 IEEE International Conference on Fuzzy Systems (IEEE Cat. No. 04CH37542)*, vol. 3. IEEE, 2004, pp. 1721–1726.

27 E. Akyilmaz and U. M. Leloglu, “Similarity ratio based adaptive mahalanobis distance algorithm to generate sar superpixels,” *Canadian Journal of Remote Sensing*, vol. 43, no. 6, pp. 569–581, 2017.

28 I. M. Hameed, S. H. Abdulhussain, and B. M. Mahmmod, “Content-based image retrieval: A review of recent trends,” *Cogent Engineering*, vol. 8, no. 1, p. 1927469, 2021.

29 T. Mei, Y. Rui, S. Li, and Q. Tian, “Multimedia search reranking: A literature survey,” *ACM Computing Surveys (CSUR)*, vol. 46, no. 3, pp. 1–38, 2014.

30 Q. Cheng, Y. Zhou, P. Fu, Y. Xu, and L. Zhang, “A deep semantic alignment network for the cross-modal image-text retrieval in remote sensing,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4284–4297, 2021.

31 Z. Tu, W. Yang, Z. Fu, Y. Xie, L. Tan, K. Xiong, M. Li, and J. Lin, “Approximate nearest neighbor search and lightweight dense vector reranking in multi-

stage retrieval architectures," in *Proceedings of the 2020 ACM SIGIR on International Conference on Theory of Information Retrieval*, 2020, pp. 97–100.

32 M. Rahman, S. M. E. Rabbi, and M. M. Rashid, "Optimizing domainspecific image retrieval: A benchmark of faiss and annoy with fine-tuned features," *arXiv preprint arXiv:2412.01555*, 2024.

33 A. Moffat, "Computing maximized effectiveness distance for recall-based metrics," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 1, pp. 198–203, 2017.

34 H. Xie, Y. Zhang, J. Tan, L. Guo, and J. Li, "Contextual query expansion for image retrieval," *IEEE Transactions on Multimedia*, vol. 16, no. 4, pp. 1104–1114, 2014.

35 M. Ferna´ndez, I. Cantador, V. Lo´pez, D. Vallet, P. Castells, and E. Motta, "Semantically enhanced information retrieval: An ontology-based approach," *Journal of Web Semantics*, vol. 9, no. 4, pp. 434–452, 2011.

36 Бабак В.П., Куц Ю.В., Мислович М.В., Фриз М.Є., Щербак Л.М. Об'єктно-орієнтована ідентифікація стохастичних шумових сигналів. Київ: Наукова думка, 2024. 240 с. https://doi.org/10.15407/978-966-00-1883-9.

37 W. Zhou, H. Li, Y. Lu, and Q. Tian, "Sift match verification by geometric coding for large-scale partial-duplicate web image search," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 9, no. 1, pp. 1–18, 2013.

38 Y. Shi, X. Yu, D. Campbell, and H. Li, "Where am i looking at? joint location and orientation estimation by cross-view matching," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 4064–4072.

39 S. Kan, Y. Cen, Y. Cen, M. Vladimir, Y. Li, and Z. He, "Zero-shot learning to index on semantic trees for scalable image retrieval," *IEEE Transactions on Image Processing*, vol. 30, pp. 501–516, 2020.

40 R. Yang, S. Wang, H. Zhang, S. Xu, Y. Guo, X. Ye, B. Hou, and L. Jiao, "Knowledge decomposition and replay: A novel cross-modal image-text retrieval continual learning method," in *Proceedings of the 31st ACM International Conference on Multimedia*, 2023, pp. 6510–6519.

41 G. Amato, P. Bolettieri, F. Carrara, F. Falchi, and C. Gennaro, “Largescale image retrieval with elasticsearch,” in *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 2018, pp. 925–928.

42 V. Sharma, “Object detection and recognition using amazon rekognition with boto3,” in *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE, 2022, pp. 727–732.

43 Y. Zhong, P. J. Garrigues, and J. P. Bigham, “Real time object scanning using a mobile phone and cloud-based visual search engine,” in *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*, 2013, pp. 1–8.

44 M. Meena, A. R. Singh, and V. A. Bharadi, “Architecture for software as a service (saas) model of cbir on hybrid cloud of microsoft azure,” *Procedia Computer Science*, vol. 79, pp. 569–578, 2016.

45 P. Dasic, J. Dasic, and J. Crvenkovic, “Applications of the search as a service (saas),” *Bulletin of the Transilvania University of Brasov. Series I-Engineering Sciences*, pp. 91–98, 2016.

46 J. Wang, X. Yi, R. Guo, H. Jin, P. Xu, S. Li, X. Wang, X. Guo, C. Li, X. Xu *et al.*, “Milvus: A purpose-built vector data management system,” in *Proceedings of the 2021 International Conference on Management of Data*, 2021, pp. 2614–2627.

47 D. Danopoulos, C. Kachris, and D. Soudris, “Approximate similarity search with faiss framework using fpgas on the cloud,” in *International Conference on Embedded Computer Systems*. Springer, 2019, pp. 373–386.

48 P. N. Singh, S. Talasila, and S. V. Banakar, “Analyzing embedding models for embedding vectors in vector databases,” in *2023 IEEE International Conference on ICT in Business Industry & Government (ICTBIG)*. IEEE, 2023, pp. 1–7.

49 C. Sun, L. Bin Song, and L. Ying, “Product re-identification system in fully automated defect detection,” in *International Conference on Smart Multimedia*. Springer, 2022, pp. 144–156.

50 I. G. S. M. Diyasa, A. D. Alhajir, A. M. Hakim, and M. F. Rohman, “Reverse image search analysis based on pre-trained convolutional neural network model,” in *2020 6th Information Technology International Seminar (ITIS)*. IEEE, 2020, pp. 1–6.

51 M. Lux, “Lire: Open source image retrieval in java,” in *Proceedings of the 21st ACM international conference on Multimedia*, 2013, pp. 843–846.

52 P. Gao, S. Geng, R. Zhang, T. Ma, R. Fang, Y. Zhang, H. Li, and Y. Qiao, “Clip-adapter: Better vision-language models with feature adapters,” *International Journal of Computer Vision*, vol. 132, no. 2, pp. 581–595, 2024.

53 Fryz M., Scherbak L., Mlynko B., Mykhailovych T. Linear Random Process Model-Based EEG Classification Using Machine Learning Techniques. Proceedings of the 1st International Workshop on Computer Information Technologies in Industry 4.0 (CITI 2023). Ternopil, Ukraine: CEUR Workshop Proceedings, 2023. Vol. 3468. P. 126–132.

54 Y. Zhang, S. Qiao, S. Ji, and Y. Li, “Deepsite: bidirectional lstm and cnn models for predicting dna–protein binding,” *International Journal of Machine Learning and Cybernetics*, vol. 11, pp. 841–851, 2020.

55 H. Al-Lohibi, T. Alkhamisi, M. Assagran, A. Aljohani, and A. O. Aljahdali, “Awjedni: a reverse-image-search application,” *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, vol. 9, no. 3, p. 49, 2020.

56 A. Maurya, A. Kumar, and D. Alimohammadi, “Application of google lens to promote information services beyond the traditional techniques,” *Qualitative and Quantitative Methods in Libraries*, vol. 12, no. 1, pp. 111–136, 2023.

57 E. Bisong, “Google automl: cloud vision,” in *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*. Springer, 2019, pp. 581–598.

58 A. KC and S. Aravind, “Integrating google ai: Enhancing library services for the digital age,” 2024.

59 Y. Xu, L. Pan, C. Du, J. Li, N. Jing, and J. Wu, “Vision-based uavs aerial image localization: A survey,” in *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*, 2018, pp. 9–18.

60 H. Noh, A. Araujo, J. Sim, T. Weyand, and B. Han, “Large-scale image retrieval with attentive deep local features,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 3456–3465.

61 A. Jaiswal, A. R. Babu, M. Z. Zadeh, D. Banerjee, and F. Makedon, “A survey on contrastive self-supervised learning,” *Technologies*, vol. 9, no. 1, p. 2, 2020.

62 N. Fei, H. Jiang, H. Lu, J. Long, Y. Dai, T. Fan, Z. Cao, and Z. Lu, “Vemo: A versatile elastic multi-modal model for search-oriented multi-task learning,” in *European Conference on Information Retrieval*. Springer, 2024, pp. 56–72.

63 L. Dietz, H. Bast, S. Chatterjee, J. Dalton, J.-Y. Nie, and R. Nogueira, “Neuro-symbolic representations for information retrieval,” in *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2023, pp. 3436–3439.

64 D. Bouneffouf and C. C. Aggarwal, “Survey on applications of neurosymbolic artificial intelligence,” *arXiv preprint arXiv:2209.12618*, 2022.

65 A. I. Weinberg, “Neurosymbolic ai and mechanistic interpretability: Can they align in the artificial general intelligence era?” 2025.

66 I. C. S. d. Silva, “Visualization of intensional and extensional levels of ontologies,” 2014.

67 Бабак В. П., Марченко М. Є., Фриз. Б. Г. Теорія ймовірностей, випадкові процеси та математична статистика. К.: Техніка, 2004. 288 с.

68 S. Alford, “A neurosymbolic approach to abstraction and reasoning,” Ph.D. dissertation, Massachusetts Institute of Technology, 2021.

69 M. J. Khan, F. Ilievski, J. G. Breslin, and E. Curry, “A survey of neurosymbolic visual reasoning with scene graphs and common sense knowledge,” *Neurosymbolic Artificial Intelligence*, vol. 1, pp. NAI–240719, 2025.

70 Стручок В.С. Безпека в надзвичайних ситуаціях. Методичний посібник для здобувачів освітнього ступеня «магістр» всіх спеціальностей денної та заочної (дистанційної) форм навчання / В.С.Стручок. — Тернопіль: ФОП Паляниця В. А., 2022. — 156 с.

71 Ткачук, К. Н., Зацарний, В. В., Зеркалов, Д. В., Полукаров, О. І., Коз'яков, В. С., Мітюк, Л. О., ... & Луц, Т. Є. (2014). Основи охорони праці.

72 Левченко, О. Г. (2024). Охорона праці та цивільний захист.

ДОДАТКИ

Додаток А

Тези конференцій

WORLD of PAPERS

ISSN 2710–3056

INTERNATIONAL SCIENTIFIC JOURNAL

GRAIL OF SCIENCE

№ 52 (May 2025)

with the proceedings of the:

IV Correspondence International Scientific and Practical Conference

OPEN SCIENCE NOWADAYS: MAIN MISSION, TRENDS AND INSTRUMENTS, PATH AND ITS DEVELOPMENT

held on May 23th, 2025 by

NGO European Scientific Platform (Vinnytsia, Ukraine)
LLC International Centre Corporative Management (Vienna, Austria)

МІЖНАРОДНИЙ НАУКОВИЙ ЖУРНАЛ

ГРААЛЬ НАУКИ

№ 52 (червень, 2025)

за матеріалами:

IV Міжнародної науково-практичної конференції

ВІДКРИТА НАУКА СЬОГОДЕННЯ: ОСНОВНА МІСІЯ, ТЕНДЕНЦІЇ ТА ІНСТРУМЕНТИ, ШЛЯХ ТА ЇЇ РОЗВИТОК

що проводилася 23.05.2025

ГО «Європейська наукова платформа» (Вінниця, Україна)
ТОВ «International Centre Corporative Management» (Відень, Австрія)

Вінниця – Відень, 2025

International scientific journal «Grail of Science» | № 52 (May, 2025)
ISSN 2710-3056

Міжнародний науковий журнал «Грааль науки» | № 52 (травень, 2025)
ISSN 2710-3056

ABSTRACT

DOI 10.36074/grail-of-science.23.05.2025.097

АКТУАЛЬНІСТЬ СТВОРЕННЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ ДЛЯ РЕЛЕВАНТНОГО ПОШУКУ

Дзядик Тарас Орестович
магістрант кафедри комп'ютерних наук
Тернопільський національний технічний університет імені Івана Пулюя, Україна

У зв'язку з різким зростанням обсягів інформації та ускладненням цифрових сервісів, розробка сучасних інформаційних систем з підтримкою релевантного пошуку стає надзвичайно важливою. Традиційні методи зберігання, обробки та пошуку даних вже не відповідають вимогам сучасності, оскільки вони не здатні ефективно обробляти великий обсяг та різноманітність даних, що генеруються в реальному часі. Це призводить до затримок у доступі до важливої інформації, погіршення якості прийняття рішень і зниження загальної ефективності бізнес-процесів. Водночас користувачі вимагають миттєвого та точного пошуку, що є основою успішної роботи з будь-якою інформаційною системою.

Інформаційна система, що орієнтована на релевантний пошук, повинна забезпечувати не тільки швидкий доступ до даних, а й інтелектуальне ранжування результатів з урахуванням змісту запиту, контексту користувача та семантичних зв'язків між даними. Це особливо важливо в умовах хмарних середовищ, де дані постійно оновлюються. Для досягнення цієї мети необхідно використовувати сучасні технології, такі як обробка даних у реальному часі, безсерверна інфраструктура, інтелектуальна маршрутизація та оркестрація даних, а також безперервна інтеграція з моделями машинного навчання для адаптивного аналізу та класифікації. Така система автоматизує збір, очищення, збагачення та індексацію даних, забезпечуючи релевантний пошук у реальному часі. Це дозволяє зменшити затримки в обробці інформації та підвищити точність і контекстуальну відповідність результатів пошуку. Отже, розробка інформаційних систем нового покоління з акцентом на релевантний пошук є важливою відповіддю на виклики цифрової епохи.

Сфера виявлення та управління візуальною інформацією зазвичай розділяється на два типи систем, кожна з яких має своє призначення та відповідає різним потребам користувачів. Пошукові системи зображень, такі як «Google Images» [1] та «Bing Images» [2], орієнтовані на широке коло користувачів, які шукають зображення в Інтернеті за допомогою текстових запитів. Ці платформи для кінцевих споживачів мають прості інтерфейси з основними параметрами фільтрації, такими як розмір, колір і тип зображення,

International scientific journal «Grail of Science» | № 52 (May, 2025)
ISSN 2710-3056

забезпечуючи доступність та широкий охоплення замість технічної складності. В той же час, системи пошуку та індексації зображень використовуються в технічних і корпоративних цілях, надаючи спеціалізовані функції для організації і керування колекціями зображень. Системи, як «Apache Solr» [3] з розширеннями для зображень і «Elasticsearch» з плагінами для зображень, підтримують як текстові, так і контентні запити, що дозволяє здійснювати більш точний доступ до візуальних ресурсів.

Список використаних джерел:

[1] Fergus, R., Perona, P., & Zisserman, A. (2004). A visual category filter for Google Images. In T. Pajdla & J. Matas (Eds.), *Computer Vision – ECCV 2004: 8th European Conference on Computer Vision, Prague, Czech Republic, May 11–14, 2004. Proceedings, Part I* (Vol. 3021, pp. 242–256). Springer. https://doi.org/10.1007/978-3-540-24670-1_20

[2] Hu, H., Wang, Y., Yang, L., Komlev, P., Huang, L., Chen, X., Huang, J., Wu, Y., Merchant, M., & Sacheti, A. (2018). Web-scale responsive visual search at Bing. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining* (pp. 359–367). ACM. https://doi.org/10.1145/3219819.3219843

[3] Welcome to Apache Solr - Apache Solr. URL: https://solr.apache.org/.

RELEVANCE OF DEVELOPING AN INFORMATION SYSTEM FOR RELEVANT SEARCH

Dziadyk Taras
Master's student at the Department of Computer Science
Ternopil Ivan Puluj National Technical University, Ukraine

Міжнародний науковий журнал «Грааль науки» | № 52 (травень, 2025)
ISSN 2710-3056

ABSTRACT

DOI 10.36074/grail-of-science.23.05.2025.101

СИСТЕМИ ПОШУКУ ТА ІНДЕКСАЦІЇ ЗОБРАЖЕНЬ

Дзядик Тарас Орестович
магістрант кафедри комп'ютерних наук
Тернопільський національний технічний університет імені Івана Пулюя, Україна

Системи пошуку та індексації зображень ґрунтуються на різноманітних методах і алгоритмах, що дають змогу ефективно управляти візуальною інформацією та здійснювати її пошук. Однією з основних складових є екстракція ознак, яка передбачає отримання інформативних характеристик із зображень [3]. Зокрема, колірні ознаки використовуються для аналізу розподілу кольору через такі методи, як гістограми, кольорові моменти та корелограми. Гістограми надають загальний розподіл кольору, моменти передають його статистичні властивості, а корелограми описують просторову залежність між кольорами. Домінантні кольори дозволяють визначити ключові кольори, що формують візуальне сприйняття зображення.

Текстурні ознаки зосереджені на описі поверхневих характеристик за допомогою методів, як-от матриці співвідношення сірого рівня (GLCM) [1], фільтрів Габора та локальних бінарних шаблонів (LBP), які дають змогу детально аналізувати просторову структуру зображення. Перетворення вейвлет забезпечує багаторівневий аналіз текстури, що є корисним для складних візуальних задач. Формальні ознаки, пов'язані з геометрією об'єктів, включають методи виявлення контурів і країв, такі як оператори Кані та Собеля, контурні представлення, моментні інваріанти та просторові контексти, що дозволяють ідентифікувати об'єкти незалежно від їх положення чи масштабу.

Сучасні підходи активно застосовують алгоритми глибокого навчання. Згорткові нейронні мережі (CNN), зокрема архітектури VGG, ResNet і Inception, використовуються для побудови векторних подань зображень, які є ефективними дескрипторами [2]. Сіамські мережі навчаються визначати подібність між зображеннями, а автокодери дозволяють зменшити розмірність ознак без втрати змістовної інформації. Методи самонавчання також демонструють потенціал в автоматичному витягуванні ознак без потреби в ручному маркуванні даних. Інструменти індексації, такі як KD-дерева, R-дерева, M-дерева і VP-дерева, використовуються для структурованого зберігання ознак та прискорення пошуку. У високорозмірних просторах ефективними виявляються хешування на зразок LSH, Semantic Hashing, Spectral Hashing та Product Quantization, що дозволяють компактно кодувати дані для пришвидшеного доступу.

International scientific journal «Grail of Science» | № 52 (May, 2025)
ISSN 2710-3056

Оцінювання подібності зображень базується на метриках відстані, таких як евклідова, косинусна, Геммінга, Махаланобіса та земельного переміщика. Кожна з них пристосована до певного типу ознак і даних, забезпечуючи точність у різних умовах пошуку. Пошук зображень за вмістом (CBIR) широко застосовується в практиці й дозволяє системам відповідати на запити за зразком (QBE) та використовувати механізми зворотного зв'язку для уточнення результатів. Комбінування кількох типів ознак (multi-feature) сприяє підвищенню точності.

Список використаних джерел:

[1] Armi, L., & Fekri-Ershad, S. (2019). Texture image analysis and texture classification methods – A review. *arXiv preprint arXiv:1904.06554*. https://arxiv.org/abs/1904.06554

[2] Gkelios, S., Sophokleous, A., Plakias, S., Boutalis, Y., & Chatzichristofis, S. A. (2021). Deep convolutional features for image retrieval. *Expert Systems with Applications, 177*, 114940. https://doi.org/10.1016/j.eswa.2021.114940

[3] McKeown, D. M., Bulwinkle, T., Cochran, S., Harvey, W., McGlone, C., & Shufelt, J. A. (2000). Performance evaluation for automatic feature extraction. *International Archives of Photogrammetry and Remote Sensing, 33*(B2; PART 2), 379–394.

IMAGE RETRIEVAL AND INDEXING SYSTEMS

Dziadyk Taras
Master's student at the Department of Computer Science
Ternopil Ivan Puluj National Technical University, Ukraine