

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**  
**Західноукраїнський національний університет**  
**Факультет комп'ютерних інформаційних технологій**  
Кафедра інформаційно-обчислювальних систем і управління

**КУРПЕЛА Павло Петрович**

**Інтелектуальний метод формування рекламного контенту  
на основі семантичного аналізу / Intelligent Method of  
Creating Advertising Content Based on Semantic Analysis**

спеціальність: 122 - Комп'ютерні науки  
освітньо-професійна програма - Комп'ютерні науки

Кваліфікаційна робота

Виконав студент групи КНм-21  
П. П. Курпела

---

Науковий керівник:  
к.т.н., доцент, Ліп'яніна-  
Гончаренко Х.В.

---

Кваліфікаційну роботу  
допущено до захисту:  
«\_\_\_» \_\_\_\_\_ 20\_\_\_ р.  
Завідувач кафедри  
\_\_\_\_\_ М.П. Комар

**ТЕРНОПІЛЬ - 2023**

## Факультет комп'ютерних інформаційних технологій

Кафедра інформаційно-обчислювальних систем і управління

Освітній ступінь «магістр»

спеціальність: 122 – Комп'ютерні науки

освітньо-професійна програма – Комп'ютерні науки

ЗАТВЕРДЖУЮ

Завідувач кафедри

М.П. Комар

« \_\_\_\_ » \_\_\_\_\_ 20\_\_ р.

### ЗАВДАННЯ

### НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ

Курпелі Павлу Петровичу

(прізвище, ім'я, по батькові)

#### 1. Тема кваліфікаційної роботи

Інтелектуальний метод формування рекламного контенту на основі семантичного аналізу / Intelligent Method of Creating Advertising Content Based on Semantic Analysis

керівник роботи к.т.н., доцент, Ліп'яніна-Гончаренко Х.В.

затверджені наказом по університету від 8 грудня 2022 року № 491.

2. Строк подання студентом закінченої кваліфікаційної роботи 1 грудня 2023 р.

3. Вихідні дані до кваліфікаційної роботи: завдання на кваліфікаційну роботу студента, наукові статті, технічна література.

#### 4. Основні питання, які потрібно розробити

1. Провести аналіз поняття рекламного контенту, визначити його ключові характеристики та особливості в контексті закладів вищої освіти.

2. Розглянути та проаналізувати різні підходи до формування рекламного контенту для освітніх установ.

3. Вивчити роль обробки текстів та семантичного аналізу в ефективному створенні рекламних повідомлень.

4. Використати попередню обробку тексту для підготовки даних до наступних етапів аналізу.

5. Застосувати методи векторизації документів для перетворення текстових даних у числові вектори.

6. Провести дослідження та порівняти методи декомпозиції єдиного значення (SVD, LSA) та латентного розмірностного розподілу (LDA) для семантичного аналізу текстових даних.

7. Розробити інтелектуальний метод формування рекламного контенту на основі семантичного аналізу, використовуючи результати досліджень.

8. Провести аналіз ефективності рекламних кампаній, розроблених на основі LSA та LDA, та порівняти їх результативність.

9. Оцінити можливість використання отриманих результатів для підвищення ефективності рекламних стратегій закладів вищої освіти.

Об'єктом дослідження є процес формування рекламного контенту

#### 5. Перелік графічного матеріалу у роботі

Структура інтелектуального методу формування рекламного контенту на основі семантичного аналізу

Порівняння ефективності сформованого рекламного контенту

## 6. Консультанти розділів кваліфікаційної роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		Завдання видав	Завдання прийняв

## 7. Дата видачі завдання 8 грудня 2022 р.

## КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів кваліфікаційної роботи	Строк виконання етапів кваліфікаційної роботи	Примітка
1	Сучасний стан предметної області і постановка задачі дослідження	12.2022 р. – 03.2023 р.	
2	Інтелектуальний метод формування рекламного контенту на основі семантичного аналізу	03.2023 р. – 05.2023 р.	
3	Реалізація інтелектуального методу	05.2023 р. – 11.2023 р.	
4	Повне завершення та представлення кваліфікаційної роботи на кафедрі	01.12.2023 р.	

Студент \_\_\_\_\_ П. П. Курпела  
підпис

Керівник роботи \_\_\_\_\_ к.т.н., доцент, Ліп'яніна-Гончаренко Х.В.  
підпис

## РЕЗЮМЕ

Кваліфікаційна робота на тему "Інтелектуальний метод формування рекламного контенту на основі семантичного аналізу" у напрямку отримання ступеня вищої освіти "Магістр" за спеціальністю 122 "Комп'ютерні науки" включає 75 сторінок тексту, 9 рисунків, 1 таблиць та список з 29 джерел.

Мета дослідження полягала у розробці та застосуванні інтелектуального підходу до формування рекламного контенту на основі семантичного аналізу для підвищення ефективності рекламних кампаній у цифрових медіа.

У роботі використано методи аналізу текстів, машинного навчання та оцінки ефективності рекламних стратегій на основі обробки великих обсягів даних.

Отримані результати показують значний потенціал інтелектуального підходу до формування рекламного контенту з використанням семантичного аналізу для підвищення взаємодії з аудиторією та покращення ефективності рекламних кампаній.

Розроблені методи мають практичне застосування в галузі маркетингу для оптимізації рекламних стратегій та підвищення впливу реклами в цифрових каналах комунікацій.

Ключові слова: РЕКЛАМНИЙ КОНТЕНТ, СЕМАНТИЧНИЙ АНАЛІЗ, МАРКЕТИНГ, МАШИННЕ НАВЧАННЯ.

## RESUME

The thesis on " Intelligent Method of Creating Advertising Content Based on Semantic Analysis" aimed at obtaining a Master's degree in the field of Computer Science comprises 75 pages of text, 9 figures, 1 table, and a bibliography of 29 sources.

The research objective was to develop and implement an intelligent approach for crafting advertising content using semantic analysis to enhance the effectiveness of advertising campaigns in digital media.

The work employed methods involving text analysis, machine learning, and assessing advertising strategy effectiveness based on processing extensive datasets.

The obtained results demonstrate the substantial potential of an intelligent approach in forming advertising content through semantic analysis to augment audience interaction and improve advertising campaign efficacy.

The developed methodologies find practical applications in the marketing domain, optimizing advertising strategies and amplifying advertising influence across digital communication channels.

Keywords: ADVERTISING CONTENT, SEMANTIC ANALYSIS, MARKETING, MACHINE LEARNING.

## ЗМІСТ

ВСТУП .....	7
1 СУЧАСНИЙ СТАН ПРЕДМЕТНОЇ ОБЛАСТІ І ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ.....	11
1.1 Суть та особливості поняття рекламного контенту .....	11
1.2 Роль обробки тексту та семантики в ефективному рекламному контексті .....	18
1.3 Аналіз існуючих підходів.....	20
1.4 Постановка задачі.....	23
Висновки до розділу 1.....	26
2 ІНТЕЛЕКТУАЛЬНИЙ МЕТОД ФОРМУВАННЯ РЕКЛАМНОГО КОНТЕНТУ НА ОСНОВІ СЕМАНТИЧНОГО АНАЛІЗУ.....	28
2.1 Попередня обробка тексту .....	28
2.2 Векторизація документу.....	32
2.3 Метод декомпозиції єдиного значення .....	34
2.4 Семантичний аналіз LSA та LDA .....	35
2.5 Інтелектуальний метод формування рекламного контенту на основі семантичного аналізу .....	38
Висновки до розділу 2.....	42
3 РЕАЛІЗАЦІЯ ІНТЕЛЕКТУАЛЬНОГО МЕТОДУ.....	45
3.1 Попередня обробка даних .....	45
3.2 Результати застосування методів LSA та LDA.....	48
3.3 Порівняльний аналіз ефективності рекламних кампаній ....	51
Висновки до розділу 3.....	57
ВИСНОВКИ .....	59
СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ .....	61
ДОДАТОК А Апробація отриманих результатів .....	66

## ВСТУП

Актуальність дослідження про інтелектуальний метод формування рекламного контенту на основі семантичного аналізу надзвичайно важлива у контексті сучасного ринку вищої освіти та маркетингу. Зростаюча конкуренція серед навчальних закладів вимагає розробки нових підходів до залучення студентів. В умовах цифрової епохи, де соціальні мережі та онлайн-платформи стають ключовими каналами комунікації, оптимізація рекламного контенту на основі семантичного аналізу має величезний потенціал.

Сучасні споживачі висувають вимоги до персоналізованої та значущої реклами. Це вимагає від маркетологів не лише розуміння цільової аудиторії, але й здатності використовувати інтелектуальні алгоритми для створення контенту, який буде здатен залучити, зацікавити та вплинути на цю аудиторію. Технології семантичного аналізу тексту відкривають можливості для створення такого контенту, який буде максимально адаптованим до потреб студентської аудиторії.

Підвищення конкуренції на ринку вищої освіти, розвиток онлайн-освіти та зміни в уявленнях молоді про процес навчання вимагають вдосконалення рекламних стратегій. Інтелектуальні методи аналізу тексту, такі як LSA та LDA, стають інструментами, що допомагають маркетологам зрозуміти та відповідати на змінні потреби цільової аудиторії, що може забезпечити конкурентну перевагу у сфері просування вищої освіти.

Мета дослідження полягає в розробці та аналізі інтелектуального методу формування рекламного контенту для закладів вищої освіти на основі семантичного аналізу текстів. Зокрема, мета включає в себе визначення ролі обробки тексту та семантики в контексті ефективної рекламної стратегії, а також порівняння результатів застосування методів латентного семантичного аналізу (LSA) та латентного розмірностного розподілу (LDA) для створення рекламного контенту.

Досягнення цієї мети зумовило потребу теоретичних розробок, визначення та послідовного вирішення таких завдань:

1. Провести аналіз поняття рекламного контенту, визначити його ключові характеристики та особливості в контексті закладів вищої освіти.
2. Розглянути та проаналізувати різні підходи до формування рекламного контенту для освітніх установ.
3. Вивчити роль обробки текстів та семантичного аналізу в ефективному створенні рекламних повідомлень.
4. Використати попередню обробку тексту для підготовки даних до наступних етапів аналізу.
5. Застосувати методи векторизації документів для перетворення текстових даних у числові вектори.
6. Провести дослідження та порівняти методи декомпозиції єдиного значення (SVD, LSA) та латентного розмірностного розподілу (LDA) для семантичного аналізу текстових даних.
7. Розробити інтелектуальний метод формування рекламного контенту на основі семантичного аналізу, використовуючи результати досліджень.
8. Провести аналіз ефективності рекламних кампаній, розроблених на основі LSA та LDA, та порівняти їх результативність.
9. Оцінити можливість використання отриманих результатів для підвищення ефективності рекламних стратегій закладів вищої освіти.

**Об'єктом дослідження** є процес формування рекламного контенту для закладів вищої освіти на основі семантичного аналізу. Цей процес включає в себе створення та оптимізацію текстового та візуального матеріалу для ефективного просування освітніх послуг та привертання уваги потенційних студентів через різні канали комунікації, зокрема, соціальні мережі, онлайн-платформи та веб-сайти навчальних закладів.



**Предметом дослідження** є методи аналізу текстового матеріалу та його семантичного відтворення для оптимізації рекламного контенту. Це включає в себе використання технік обробки природної мови, векторизації документів, усіченого розкладання особливих значень (SVD), а також методи LSA (латентно-семантичний аналіз) та LDA (латентна диріклетовська алокація), спрямовані на виявлення семантичної структури та ключових понять у текстах.

**Методи дослідження** включають в себе аналіз даних опитувань студентів, застосування машинного навчання для обробки та відтворення текстового матеріалу, а також порівняльний аналіз результатів застосування різних моделей для створення рекламного контенту. Такий підхід дозволяє оцінити ефективність методів, виявити переваги та недоліки кожного підходу та розробити оптимальну стратегію формування контенту для привертання цільової аудиторії.

**Наукова новизна отриманих результатів** полягає в розробці методу створення рекламного контенту для закладів вищої освіти, що ґрунтується на комплексному аналізі текстових даних з використанням сучасних методів обробки природної мови. Особливість полягає у використанні не лише базових алгоритмів аналізу тексту, але й комбінації векторизації документів, усіченого розкладання особливих значень, а також методів LSA та LDA для отримання ключових понять та семантичної структури тексту. Отримані методи дозволяють не лише підвищити ефективність рекламного контенту, а й визначити більш точні підходи до спілкування з цільовою аудиторією у сучасному онлайн-середовищі.

**Практичне значення отриманих результатів** полягає в можливості покращення стратегій маркетингу для закладів вищої освіти через удосконалення методів створення рекламного контенту. Це дозволить ефективніше спілкуватися з аудиторією, адаптувати повідомлення та пропонувати зміст, що краще відповідає її потребам. Розроблені підходи

дають можливість збільшити якість та різноманітність рекламного контенту, спрямованого на студентів, що може призвести до підвищення зацікавленості та залученості цільової аудиторії, а також покращити результативність маркетингових кампаній закладів вищої освіти в цілому.

**Структура та обсяг роботи.** Кваліфікаційна робота включає вступ, три головні розділи, висновки, список літератури та додатки. Загалом, текст роботи нараховує 75 сторінки, в якому представлено 9 рисунків та 1 таблицю. У списку літератури міститься 29 джерел.

**Апробація результатів дослідження.** Основні теоретичні положення роботи й практичні результати дослідження доповідалися й обговорювалися: IV Всеукраїнської мультидисциплінарної студентської наукової конференції «Експериментальні та теоретичні дослідження в контексті сучасної науки», яка відбулася 29 вересня 2023 року у місті Чернігів, Україна; IV Міжнародна студентська конференція «Тренди та перспективи розвитку мультидисциплінарних досліджень» 01.12.2023 м. Луцьк, Україна.

# 1 СУЧАСНИЙ СТАН ПРЕДМЕТНОЇ ОБЛАСТІ І ПОСТАНОВКА ЗАДАЧІ ДОСЛІДЖЕННЯ

## 1.1 Суть та особливості поняття рекламного контенту

Рекламний контент представляє собою інформацію чи матеріали, спрямовані на просування продукту, послуги, бренду чи ідеї з метою залучення уваги, створення інтересу та стимулювання певної дії аудиторії.

Особливості рекламного контенту:

1. Стратегічне спрямування: Рекламний контент планується з урахуванням конкретної мети: залучення уваги, підвищення продажів, збільшення свідомості про бренд тощо. Він підбирається таким чином, щоб досягти конкретних цілей в маркетинговій стратегії.
2. Креативність і естетика: Ефективний рекламний контент вимагає оригінальності, вражаючого дизайну, привабливих фотографій чи відеороликів, які здатні зацікавити аудиторію та викликати емоційну відгуку.
3. Цільова аудиторія: Рекламний контент створюється з урахуванням потенційних споживачів. Він повинен звертатися до конкретної аудиторії з врахуванням її потреб, інтересів та характеристик.
4. Мультимедійність: Сучасний рекламний контент використовує різноманітні медіаформати - відео, зображення, тексти, аудіо, анімацію тощо, для досягнення більшого впливу та привертання уваги аудиторії.
5. Call-to-Action: Ефективний рекламний контент часто включає прямі виклики до дії (СТА - Call-to-Action), щоб спонукати аудиторію до конкретних кроків, таких як покупка, підписка, перехід на веб-сайт тощо.

6. Адаптація до платформи: Успішний рекламний контент враховує особливості та формати різних медіа-платформ (соціальні мережі, телебачення, Інтернет), щоб максимально ефективно досягти своїх цілей.

Стратегічне спрямування рекламного контенту полягає в тому, щоб максимально спрямувати його зусилля для досягнення конкретних цілей маркетингової стратегії. Кожен елемент контенту - будь то текст, зображення, відео чи аудіо - планується та створюється з урахуванням цілей, які він повинен досягти в рамках маркетингової кампанії. Ось детальніше про це:

1. Цілі кампанії: Рекламний контент формулюється з урахуванням конкретних цілей кампанії. Наприклад, це може бути збільшення уваги до продукту, підвищення обізнаності про бренд, залучення нових клієнтів, збільшення продажів, підвищення лояльності тощо.
2. Аудиторія: Рекламний контент адаптується до цільової аудиторії. Зрозуміння інтересів, потреб та поведінки аудиторії допомагає створити контент, який буде приваблювати та залучати саме цю групу людей.
3. Креативність та розробка: Контент розробляється з урахуванням креативних ідей, які будуть відповідати цілям кампанії. Він повинен бути цікавим, інноваційним та здатним вражати аудиторію.
4. Платформи розміщення: Рекламний контент планується з урахуванням платформ, на яких він буде розміщуватися (соціальні мережі, телебачення, Інтернет), і відповідно адаптується до форматів кожної платформи для максимального ефекту.
5. Аналіз та вимірювання: Після запуску кампанії рекламний контент аналізується та вимірюється з метою оцінки ефективності в досягненні поставлених цілей. Це дозволяє коригувати стратегію в разі необхідності.

6. Контроль бюджету: Стратегічне спрямування також включає контроль над витратами на створення та розміщення рекламного контенту, щоб забезпечити оптимальне використання бюджету і максимальний результат.

Стратегічне планування рекламного контенту є важливим етапом в маркетинговій стратегії, оскільки від нього залежить успіх кампанії та досягнення її цілей.

Креативність і естетика в рекламному контенті відіграють ключову роль у привертанні уваги та створенні позитивного враження у аудиторії. Ось детальніше про ці аспекти:

1. Оригінальність: Рекламний контент повинен бути унікальним, відрізнятися від інших подібних матеріалів. Це може бути новаторська ідея, нестандартний підхід до презентації продукту чи послуги, що привертає увагу та залишає враження.
2. Дизайн: Ефективний дизайн графіки чи відеоряду має важливе значення. Привабливість, чіткість, спрощення інформації - всі ці аспекти можуть викликати позитивну реакцію у споживача.
3. Фотографії та відео: Використання вражаючих фотографій чи відеороликів може бути потужним засобом привернення уваги. Вони здатні викликати емоційну відгуку, створювати атмосферу та відображати сутність продукту чи послуги.
4. Емоційна зв'язок: Контент, який викликає емоційну реакцію у споживачів, має більший вплив. Він може викликати радість, зацікавленість, співчуття або будь-яку іншу емоцію, яка стимулює споживачів до взаємодії з рекламою.
5. Спілкування бренду: Ефективний контент також допомагає у спілкуванні та позиціонуванні бренду. Він передає основні цінності та ідеї компанії, роблячи акцент на те, що робить її особливою та унікальною.

Креативність і естетика в рекламі створюють змогу не лише привернути увагу аудиторії, а й залишити сильне враження, що збудує позитивний імідж та сприятиме побудові відносин з клієнтами.

Цільова аудиторія визначається як конкретна група осіб чи сегмент ринку, до якої спрямована рекламна кампанія чи контент. Основні аспекти, які важливо враховувати при створенні рекламного контенту для цільової аудиторії:

1. Демографічні дані: Вік, стать, місце проживання, освіта, заробітна плата - це лише деякі з характеристик, що визначають демографічну складову цільової аудиторії.
2. Психографічні аспекти: Інтереси, уподобання, стиль життя, цінності, переконання - це важливі фактори, які допомагають краще зрозуміти аудиторію та створити контент, який буде відповідати їхнім потребам та бажанням.
3. Поведінка споживачів: Розуміння того, як аудиторія взаємодіє з продуктами чи послугами, їхні покупні звички, медіа-використання, способи витрат - це допомагає створювати контент, що більш точно відповідає їхнім потребам.
4. Контекст споживання: Розуміння того, коли та де аудиторія може зустріти контент - це важливо. Наприклад, реклама одягу може бути вдалою, якщо вона представлена в спеціалізованих журналах чи соціальних медіа, які використовуються цільовою аудиторією.
5. Персоналізація та адаптація: Здатність адаптувати контент до конкретних вимог та характеристик аудиторії, а також персоналізувати його для кожного сегмента, може значно підвищити його ефективність.

Розуміння цільової аудиторії є ключовим для успішної рекламної стратегії, оскільки це дозволяє створювати контент, який реально

відповідає потребам та очікуванням споживачів, забезпечуючи більшу ефективність і результативність рекламної кампанії.

Використання різних медіаформатів у рекламному контенті є ключовим для створення сильного впливу на аудиторію. Ось деякі аспекти мультимедійності в рекламному контенті:

1. Відеоформат: Відеоролики здатні захопити увагу та передати повідомлення більш ефективно через звук, візуальний контент та емоції. Це може бути короткий кліп, реклама на YouTube або навіть вебінар, що дозволяє взаємодіяти з аудиторією.
2. Зображення та Графіка: Фотографії, ілюстрації, графічні малюнки - це важливий елемент, який може швидко захопити увагу та сприйнятися легше за текст. Графіка також допомагає змістовно передавати ідеї та інформацію.
3. Аудіоформат: Подкасти, аудіоголошення, музичні фони - це формат, який дозволяє споживачеві взаємодіяти з контентом без необхідності активного перегляду.
4. Анімація та інтерактивність: Використання анімації може надати контенту динаміки та привабливості. Технології, що дозволяють інтерактивність (наприклад, AR та VR), роблять рекламу більш захопливою та цікавою.
5. Текстовий контент: Навіть у мультимедійному контенті текст є важливим елементом. Підписи, описи, вказівки та інші текстові елементи можуть доповнювати або пояснювати відео чи зображення.
6. Соціальні медіа та інтернет-ресурси: Використання різних медіаформатів відображається у всіх аспектах соціальних медіа та інтернет-ресурсів: від постів на Facebook до інтерактивних інструментів на веб-сайтах.

Мультимедійний підхід дає можливість різноманітним способами досягти аудиторію, оскільки різні люди сприймають інформацію по-

різному. Це дозволяє створювати рекламний контент, який краще відповідає потребам і уподобанням різних груп споживачів.

Call-to-Action (СТА) - це справжня пікова точка рекламного контенту, що надає аудиторії чіткі вказівки чи запрошення на подальші дії.

Ось кілька ключових аспектів СТА:

1. Ясність і привабливість: СТА повинен бути простим, лаконічним та відразу викликати реакцію. Фрази типу "Купи зараз!", "Підпишись!", "Дізнайся більше" працюють ефективно.
2. Створення невідкладності: Слова чи фрази, які підсилюють відчуття брошеного шансу або обмеженості пропозиції, часто стимулюють людей до дії. "Тільки сьогодні!", "Обмежений час!", "Лише 24 години!" - ці фрази збільшують вірогідність реакції.
3. Підказки для дії: Вказівки, що робити далі, є важливою частиною СТА. Кнопки "Купити зараз", "Зареєструватись", "Дізнатися більше", "Перейти" - вони спрямовують аудиторію.
4. Співвідношення з контентом: СТА повинен бути злагодженим зі змістом контенту та місцем, де він розміщується. Він має створювати продовження або завершення того, що було запропоновано раніше в контенті.
5. Вимірювання ефективності: Слід вимірювати ефективність СТА через метрики, такі як конверсія, кількість кліків, відвідуваність сторінки тощо, щоб аналізувати його вплив та вдосконалювати.

СТА важливий не лише для привертання уваги, але й для перетворення цієї уваги в конкретні дії. Грамотно сформульований та розташований СТА може значно підвищити ефективність рекламного контенту.

Адаптація до платформ - це ключовий аспект ефективності рекламного контенту. Ось кілька основних принципів адаптації:



1. Розуміння аудиторії платформи: Кожна платформа має свою унікальну аудиторію з особливими потребами та очікуваннями. Розуміння цих відмінностей допомагає створити контент, який краще відповідає споживачам цієї платформи.
2. Формат та контекст: Контент, який працює на одній платформі, може не підійти для іншої. Наприклад, короткі, динамічні відео можуть привертати увагу на соціальних мережах, але на телебаченні можуть потребувати адаптації до формату програм.
3. Звернення до особливостей платформи: Використання унікальних можливостей кожної платформи - таких як хештеги на Twitter, інтерактивні функції на Instagram чи використання ключових слів у пошукових системах - може підвищити ефективність реклами.
4. Контент, що відповідає контексту: Реклама має бути адаптована до того, як користувачі споживають контент на конкретній платформі. Наприклад, споживачі можуть швидше сприймати короткі повідомлення на Twitter, а довші форми контенту - на YouTube.
5. Технічні обмеження та можливості: Різні платформи можуть мати обмеження щодо довжини відео, розміру зображень чи форматів файлів. Адаптація до цих технічних обмежень важлива для оптимальної презентації контенту.

Адаптація до різних медіа-платформ відіграє ключову роль у створенні рекламного контенту, який привертає увагу та досягає поставлених цілей, враховуючи специфіку кожної платформи та її аудиторії.

Отже, рекламний контент є фундаментальною частиною маркетингової стратегії, яка визначає не тільки як бренд сприймається аудиторією, але й безпосередньо впливає на успіх рекламної кампанії. Важливо враховувати стратегічне спрямування, креативність, естетику, визначення цільової аудиторії, мультимедійність, наявність чітких

викликів до дії (Call-to-Action) та адаптацію до специфіки різних медіа-платформ. Ефективний рекламний контент не лише приваблює увагу, але й стимулює взаємодію з брендом, створюючи довготривалі та позитивні асоціації. Використання даних, інноваційних підходів та технологій дозволяє розробляти рекламні матеріали, які не тільки захоплюють, але й ефективно досягають цілей маркетингової кампанії, забезпечуючи оптимальне використання ресурсів і підвищуючи рентабельність інвестицій..

## **1.2 Роль обробки тексту та семантики в ефективному рекламному контексті**

Рекламний ландшафт постійно змінюється, змушуючи маркетологів шукати нові підходи для ефективного спілкування з аудиторією. Серед найважливіших засобів у цьому процесі - обробка тексту та семантики, які стали ключовими інструментами для розуміння та аналізу потреб споживачів, їхніх емоцій та реакцій на рекламу.

Перше, що робить ці підходи надзвичайно важливими, - це їх здатність розуміти контекст споживача. Аналіз відгуків, коментарів у соціальних мережах та інших джерелах дозволяє належним чином налаштовувати рекламні стратегії для більш ефективного залучення аудиторії.

Крім того, обробка семантики дозволяє створювати персоналізований контент, що відповідає інтересам та уподобанням споживачів, що в свою чергу поліпшує зв'язок між брендом і аудиторією, роблячи його більш релевантним.

Окрім адаптації до поточних потреб споживачів, аналітика тексту також допомагає прогнозувати майбутні тренди, забезпечуючи здатність швидко адаптуватися до змін, зокрема в споживчому попиті.

Важливим аспектом є також розуміння емоційної реакції споживачів на рекламний контент. Аналіз емоцій допомагає створювати більш ефективні та залучаючі рекламні повідомлення, спрямовані на конкретні реакції аудиторії.

Завершальним фактором є можливість точно визначати метрики успішності рекламних кампаній. Це дозволяє удосконалювати стратегії та максимізувати результативність, що є критично важливим у конкурентному світі реклами.

Підходи обробки тексту та семантики в рекламному контексті є ключовими, оскільки вони дозволяють краще розуміти споживачів, їхні потреби та емоції. Ось декілька аргументів, що підтверджують їх ефективність:

- Розуміння контексту споживача: Обробка тексту дозволяє аналізувати відгуки, коментарі, відгуки у соціальних мережах та інші джерела для зрозуміння реакції споживачів на рекламний контент. Це дозволяє ліпше адаптувати рекламні стратегії до їхніх потреб.
- Персоналізація реклами: Аналіз семантики тексту допомагає створювати персоналізований контент, що краще відповідає інтересам та уподобанням споживачів. Це дозволяє більш ефективно спілкуватися з аудиторією.
- Аналіз та прогнозування трендів: Обробка тексту може допомогти виявити наслідки рекламних кампаній та визначити тенденції в споживчому поведінці. Це дозволяє швидко реагувати на зміни та підлаштовувати рекламні стратегії під нові умови.
- Розуміння емоційної реакції: Аналіз семантики тексту допомагає виявляти емоції, які викликає реклама у споживачів. Це дозволяє створювати контент, який легше сприймається та здатний викликати певну реакцію.

- Підвищення ефективності рекламних кампаній: Аналіз тексту та семантики дозволяє точніше визначати метрики успішності рекламних кампаній, такі як конверсія, взаємодія, відкриття. Це дає змогу покращувати стратегії та максимізувати результативність.

Ці підходи дозволяють більш точно аналізувати та розуміти потреби, поведінку та реакції споживачів, що дозволяє створювати більш ефективні та спрямовані рекламні кампанії.

Отже, обробка тексту та семантика в рекламному контексті відіграють критичну роль у сучасному маркетингу, дозволяючи брендам глибше розуміти та взаємодіяти зі своєю аудиторією. Вони надають можливість не лише адаптувати контент до індивідуальних інтересів та уподобань споживачів, але й прогнозувати майбутні тренди та швидко реагувати на зміни в споживчому попиті. Це забезпечує більшу релевантність та особистий зв'язок між брендом і споживачем, підвищуючи ефективність рекламних кампаній. Аналіз емоційних реакцій та відгуків споживачів дозволяє створювати більш залучаючі та впливові рекламні повідомлення. Крім того, точне визначення метрик успішності рекламних кампаній допомагає в удосконаленні стратегій та максимізації результативності, що є важливим у конкурентоспроможному світі реклами. Таким чином, обробка тексту та семантика стають необхідними інструментами для ефективного комунікаційного підходу в ері цифрового маркетингу.

### **1.3 Аналіз існуючих підходів**

Інтернет-реклама, має велику кількість площадок для реклами [8], а саме пошукові системи [18], соціальних мережах [19] та медійна реклама [20]. З огляду на складність сучасної екосистеми цифрової реклами, є багато досліджень, що описують вплив рекламного вмісту в соціальних

мережах на залучення клієнтів [11], використовуючи дані з Facebook в: медицині [12], психології [13], соціології [14], політиці [15], харчовій промисловості [16], туризмі [17]. Дослідження [9] дає розуміння поточного ландшафту соціальних мереж для вищих навчальних закладів.

У роботі [2] пропонується онтологічна система рекомендацій щодо реклами, яка використовує дані, отримані користувачами на сайтах соціальних мереж, і цей підхід обґрунтовується спільною моделлю онтології, за допомогою якої можна представити як профілі користувачів, так і зміст реклами. І користувачі, і реклама представлені за допомогою векторів, створених із використанням методів обробки природними мовами, які збирають онтологічні сутності з текстового вмісту. Система рекомендацій щодо реклами широко перевірена в модельованому середовищі, MAP@3 та рівна 85,6%.

У роботі [21] запропоновано вдосконалену семантикою та контекстну гібридну спільну фільтрацію для рекомендацій подій, яка поєднує семантичний аналіз вмісту та вплив контекстуальної події на вибір сусідства користувача.

У [10] роботі представлена інтелектуальна система управління рекламою в соціальних мережах, на основі методів аналізу даних для автоматичного створення оголошень. У роботі [1] для класифікації рекламного відео запропоновано представлення відео, яке має на меті охопити приховану семантику рекламного відео без нагляду, експерименти над реальними рекламними відео, демонструють, що запропонований метод може ефективно диференціювати рекламні відео. У статті [3] пропонується підхід довіри та семантичної соціальної рекомендації для усунення проблем старту реклами.

Запропоновано метод [4], що допомагає перетворити нечитабельні дані в читабельний структурований формат за допомогою класифікації машинного навчання, а кластеризація відіграє вирішальну роль у

перетворенні операційних даних у моделі даних та візуалізації обробленої інформації для кінцевого користувача. Оскільки організації мають конкретні вимоги, розглядаючи їх, впроваджено прихований розподіл Діріхле (LDA) та прихований семантичний аналіз (LSA), які змогли обробляти дискретні дані. У документі [5] використовується підхід LDA для визначення зворотного зв'язку для аналізу настроїв інтернет-покупок, оскільки підхід LDA призначений для вирішення питань LSA та PLSA. З метою підвищення ефективності аналізу громадських настроїв в Інтернеті пропонується [6] метод аналізу настроїв тексту, що поєднує подання тексту прихованого розподілу Діріхле (LDA) та згорткової нейронної мережі (CNN). У роботі [7] представлена емпірична оцінка трьох найважливіших методів моделювання тем - прихованого семантичного аналізу, прихованого розподілу Діріхле та корелюючого моделювання тем.

Треба відмітити, що згадані вище роботи з одної сторони здебільшого аналізують дії користувачів на інтернет-рекламу. І з іншої сторони ряд робіт - представляють дослідження семантичного аналізу тексту також є роботи, що використовують семантичний аналіз для інтернет-реклами (аналоги).

У зв'язку із цим, метою даної статті є розробка інтелектуального методу формування рекламного контенту закладів вищої освіти на основі семантичного аналізу.

На відміну від аналогів [2, 21] розроблений інтелектуальний метод формування рекламного контенту закладів вищої освіти на основі семантичного аналізу, дозволить виділити ключові слова на основі, яких можна сформувавши рекламний текст оголошення.

Отже, аналіз існуючих підходів до інтернет-реклами і семантичного аналізу демонструє значний прогрес у розумінні та використанні даних для оптимізації рекламних стратегій. Сучасні методи, такі як онтологічні системи, гібридні фільтри, аналіз настроїв і моделювання тем, вже довели

свою ефективність у різних сферах, включаючи медицину, політику, туризм та інші. Однак, існує потреба у подальшому розвитку та адаптації цих технологій, особливо в контексті формування рекламного контенту для закладів вищої освіти. Розробка інтелектуального методу на основі семантичного аналізу, що фокусується на виокремленні ключових слів для формування рекламних текстів, може запропонувати новітній підхід у цій сфері. Це дозволить не лише підвищити ефективність рекламних кампаній, але й забезпечити більшу релевантність та персоналізацію контенту, що є критично важливим для залучення потенційних студентів та підвищення репутації освітніх установ.

#### **1.4 Постановка задачі**

Інтелектуальний метод формування рекламного контенту на основі семантичного аналізу в сучасному світі має величезну актуальність. Перш за все, це пов'язано з постійною зміною підходів до споживання інформації. Сучасні користувачі медіа вимагають контенту, який буде не лише цікавим, але й особисто спрямованим на їхні потреби та інтереси. Інтелектуальний аналіз тексту та семантики дозволяє створювати контент, який більш точно відповідає цілям та очікуванням аудиторії.

Далі, зростає конкуренція в цифровому просторі. Компанії постійно змагаються за увагу споживачів у світі перенасиченому рекламою. Використання інтелектуальних методів дає можливість створити рекламний контент, який привертає увагу через персоналізацію, відповідність інтересам та змістовність.

Також, швидкий та точний аналіз великих обсягів даних - це один із основних плюсів використання інтелектуальних методів у маркетингу.

Завдяки такому аналізу можна швидко ідентифікувати тенденції, реагувати на зміни в споживчому попиті та ефективно коригувати рекламні стратегії.

Реклама, яка глибоко вникає у потреби аудиторії та комунікує з нею, має більший успіх. Використання семантичного аналізу та обробки тексту дозволяє краще розуміти споживачів, їхні уподобання, побажання та реакції на рекламні послання.

Нарешті, в умовах стрімкого розвитку технологій та штучного інтелекту, застосування інтелектуальних методів у формуванні рекламного контенту є необхідністю для компаній, які прагнуть бути конкурентоспроможними і успішними на ринку.

Мета дослідження полягає в розробці та аналізі інтелектуального методу формування рекламного контенту для закладів вищої освіти на основі семантичного аналізу текстів. Зокрема, мета включає в себе визначення ролі обробки тексту та семантики в контексті ефективної рекламної стратегії, а також порівняння результатів застосування методів латентного семантичного аналізу (LSA) та латентного розмірностного розподілу (LDA) для створення рекламного контенту.

Досягнення цієї мети зумовило потребу теоретичних розробок, визначення та послідовного вирішення таких завдань:

10. Провести аналіз поняття рекламного контенту, визначити його ключові характеристики та особливості в контексті закладів вищої освіти.
11. Розглянути та проаналізувати різні підходи до формування рекламного контенту для освітніх установ.
12. Вивчити роль обробки текстів та семантичного аналізу в ефективному створенні рекламних повідомлень.
13. Використати попередню обробку тексту для підготовки даних до наступних етапів аналізу.



14. Застосувати методи векторизації документів для перетворення текстових даних у числові вектори.
15. Провести дослідження та порівняти методи декомпозиції єдиного значення (SVD, LSA) та латентного розмірностного розподілу (LDA) для семантичного аналізу текстових даних.
16. Розробити інтелектуальний метод формування рекламного контенту на основі семантичного аналізу, використовуючи результати досліджень.
17. Провести аналіз ефективності рекламних кампаній, розроблених на основі LSA та LDA, та порівняти їх результативність.
18. Оцінити можливість використання отриманих результатів для підвищення ефективності рекламних стратегій закладів вищої освіти.

Завдання 1-3 розглянуто в параграфах 1.1, 1.2 та 1.3 відповідно, а виконання завдань 4-9 представлені у параграфах 2.1-3.3.

## Висновки до розділу 1

1. Рекламний контент є фундаментальним елементом маркетингової стратегії, який виконує ключову роль у взаємодії з цільовою аудиторією. Ефективний рекламний контент повинен бути стратегічно спрямованим, креативним, естетично привабливим, а також повинен враховувати потреби та характеристики цільової аудиторії. Використання різноманітних медіаформатів, таких як відео, зображення, текст та аудіо, а також адаптація контенту до особливостей різних платформ, є важливими для досягнення максимальної ефективності. Включення чітких закликів до дії (СТА) та аналіз ефективності контенту дозволяють оптимізувати взаємодію з аудиторією та досягнути конкретних цілей маркетингової кампанії. Таким чином, ретельно спланований та ефективно реалізований рекламний контент може істотно підвищити впізнаваність бренду, залучити нових клієнтів та сприяти загальному успіху рекламної кампанії.
2. Обробка тексту та семантичний аналіз відіграють вирішальну роль у сучасному рекламному контексті, надаючи маркетологам могутні інструменти для ефективної комунікації з аудиторією. Ці технології не лише поліпшують розуміння контексту та потреб споживачів, але й дозволяють створювати персоналізований та емоційно резонансний контент, що забезпечує більшу залученість і взаємодію з брендом. Аналіз тексту в реальному часі дає можливість прогнозувати тренди та швидко адаптуватися до змін на ринку, а також точно визначати метрики успішності рекламних кампаній для їх оптимізації та покращення результативності. Використання цих підходів є критично важливим для успіху в конкурентному світі реклами, дозволяючи брендам більш точно й ефективно досягати своєї цільової аудиторії.

3. Аналіз існуючих підходів до інтернет-реклами та семантичного аналізу в різних сферах, включаючи медицину, психологію, політику, туризм, та інші, підкреслює важливість інтеграції цих технік в маркетинг вищої освіти. Зокрема, використання онтологічних систем, гібридних фільтрацій та інтелектуальних систем управління рекламою в соціальних мережах відкриває нові можливості для більш точної та ефективної цільової реклами. Водночас, зосередження на семантичному аналізі зокрема дозволяє глибше зрозуміти потреби та інтереси цільової аудиторії, що, у свою чергу, сприяє розробці більш персоналізованого та впливового рекламного контенту. Таким чином, інтеграція цих сучасних підходів у створення рекламного контенту для закладів вищої освіти може значно підвищити ефективність та результативність рекламних кампаній, виходячи за рамки традиційних методів та забезпечуючи більш цільове та стратегічно обгрунтоване просування.

## **2 ІНТЕЛЕКТУАЛЬНИЙ МЕТОД ФОРМУВАННЯ РЕКЛАМНОГО КОНТЕНТУ НА ОСНОВІ СЕМАНТИЧНОГО АНАЛІЗУ**

### **2.1 Попередня обробка тексту**

Розглядаючи обробку тексту, токенізація, лематизація та ігнорування символів відіграють важливу роль у підготовці даних для подальшого аналізу та роботи з текстом. Токенізація полягає в розділенні тексту на менші одиниці, які називаються токенами, і може використовувати різні правила та методи. Лематизація відповідає за перетворення слів до їх базової форми, враховуючи граматичні особливості мови, що спрощує подальший аналіз. Ігнорування символів стосується вилучення з тексту знаків пунктуації та інших символів, що не несуть суттєвої інформації.

Зазвичай ці процеси використовуються для побудови моделей машинного навчання, аналізу синтаксичних структур, пошуку ключових слів та іншого виду аналізу семантики тексту. Такий підхід є ключовим у сфері обробки природної мови.

Через математичний підхід можна виразити ці процеси за допомогою функцій, що оперують текстом, розділяючи його на певні частини або токени, або ж приводячи слова до їх базової форми з урахуванням граматичних особливостей.

Токенізація документу - це процес розбиття тексту на менші одиниці, які називаються токенами. Це може бути розбиття на окремі слова, фрази, речення або навіть символи, залежно від конкретної потреби аналізу. Токени створюють базові елементи тексту, з якими можна працювати для подальшої обробки або аналізу.

Зазвичай процес токенізації використовує різні правила та методи, наприклад:

1. Розбиття за пробілами: Одним із найпростіших методів є розділення тексту на токени за допомогою пробілів.
2. Розділення за розділовими знаками: Текст може бути розділений за розділові знаки, такі як коми, крапки, апострофи тощо.
3. Токенізація на рівні слів або фраз: Вона може враховувати спеціальні випадки, такі як розпізнавання складних фраз або врахування контексту для правильного розбиття.
4. Лексичний аналіз: Деякі методи враховують особливості мови, наприклад, враховують форми слів (наприклад, різні часи дієслів або форми іменників) і розбивають їх на специфічні токени.

Після токенизації текст може стати основою для подальших аналітичних операцій, таких як аналіз синтаксичних структур, побудова моделей машинного навчання, пошук ключових слів або аналіз семантики тексту. Токенізація є ключовим етапом обробки при роботі з текстовими даними для багатьох завдань у сфері обробки природної мови.

У математичному контексті токенизація тексту може бути виражена через різні методи, які враховують специфіку обробки природної мови. Один з методів можна виразити так:

Нехай  $D$  - текстовий документ, який потрібно токенизувати, а  $T$  - множина токенів, отриманих в результаті токенизації.

Ми можемо описати токенизацію як функцію  $Tokenize(D)$ , яка призводить до отримання множини токенів  $T$ :

$$T = Tokenize(D)$$

Математично, функція токенизації може використовувати різні правила, символи, регулярні вирази або моделі мови для розділення тексту на токени. Наприклад, для поділу тексту на окремі слова можна

використовувати функцію  $\text{TokenizeByWhitespace}(D)$ , яка розділяє текст за допомогою пробілів:

$$T = \text{TokenizeByWhitespace}(D)$$

де  $T$  - це множина токенів, які отримані розділенням тексту  $D$  за пробілами.

Це лише загальний опис, але реалізація самої токенизації може включати багато додаткових етапів та алгоритмів, які розбивають текст на менші одиниці для подальшої обробки та аналізу.

Лематизація - це процес приведення слова до його базової форми (леми) з урахуванням його граматичної основи та контексту. Лематизація враховує словникову форму слова, що дозволяє співставити його з найбільш базовим виглядом цього слова.

Математично цей процес можна виразити через функцію лематизації  $\text{Lemmatize}(W)$ , яка перетворює слово  $W$  у його базову форму (лему):

$$\text{Lemmatize}(W) = \text{lemma}$$

де  $\text{lemma}$  - це базова форма слова  $W$ , яка може бути отримана за допомогою лематизаційного процесу.

Алгоритми лематизації враховують граматичні правила мови для знаходження базової форми слова. Наприклад, слово "бігаючий" може бути зведене до базової форми "бігати". Цей процес важливий для стандартизації слів і дозволяє враховувати різні форми того ж слова як один токен під час аналізу тексту, що спрощує обробку природної мови.

Точність лематизації може залежати від мови тексту, а також від алгоритму, використаного для проведення лематизації.

Процес ігнорування символів в тексті часто пов'язаний з видаленням певних символів або знаків пунктуації, які не несуть значущої інформації

для аналізу або обробки тексту. Це можуть бути розділові знаки, символи пунктуації, або навіть окремі літери, які не впливають на сутність тексту.

Математично ігнорування символів може бути виражене через функцію:

$$\text{IgnoreSymbols}(T) = T'$$

де  $T$  - вихідний набір токенів або текст, який містить символи для ігнорування, а  $T'$  - новий набір токенів без цих символів.

Для цього можна використовувати певний набір стоп-символів або пунктуаційних знаків, які будуть вилучені з тексту. Наприклад, символи крапка, кома, апостроф, круглі дужки тощо можуть бути вилучені через функцію:

$$\begin{aligned} &\text{IgnoreSymbols}(\text{"Текстізрозділовимизнаками!"}) \\ &= \text{"Текстізрозділовимизнаками"} \end{aligned}$$

Цей процес допомагає зменшити шум у тексті, спрощує обробку, аналіз та зберігання даних. Вилучення символів може бути корисним при побудові моделей обробки природної мови або для виокремлення важливої інформації з текстових даних.

Отже, в розглянутому параграфі описано процес токенізації тексту, що включає розділення тексту на менші одиниці, так звані токени. Ці токени можуть представляти окремі слова, фрази, речення або навіть символи, залежно від потреби аналізу тексту. Токенізація є ключовим етапом обробки природної мови та може використовувати різні правила та методи для поділу тексту. Після токенізації текст стає основою для подальшої аналітики, включаючи аналіз синтаксичних структур, побудову моделей машинного навчання та аналіз семантики тексту. Паралельно з цим, лематизація дозволяє привести слова до їх базових форм для

спрощення аналізу, а процес ігнорування символів допомагає у вилученні зайвих елементів, що не несуть значення для аналізу тексту, сприяючи підвищенню ефективності обробки даних.

## 2.2 Векторизація документу

Векторизація документу - це процес перетворення текстового документу в числовий вектор, який можна використовувати для подальшого аналізу та обробки за допомогою алгоритмів машинного навчання або аналізу даних.

Щоб векторизувати текстовий документ, спочатку необхідно виконати передобробку, таку як токенізація (розбиття на токени), лематизація (перетворення слів до їх базових форм) та видалення стоп-слів (зайвих слів, які не несуть суттєвої інформації).

TF (Term Frequency - частота терміну): Це відносна частота, з якою термін (слово) зустрічається у документі. Формула TF визначає, наскільки часто слово зустрічається у документі в порівнянні з загальною кількістю слів у цьому документі. Формула для TF може бути виражена як:

$$TF(t, D) = \frac{\text{кількість входжень терміну } t \text{ в документ } D}{\text{загальна кількість термінів у документі } D}$$

IDF (Inverse Document Frequency - інверсія частоти документа): Це міра, яка визначає, наскільки інформативний є термін шляхом обчислення зворотного значення частоти входження терміну у всіх документах колекції. Терміни, які зустрічаються рідко, отримують вищий показник IDF, тоді як терміни, що зустрічаються частіше, мають менший показник IDF. Формула для IDF може бути виражена як:



$$IDF(t, \text{колекція документів}) = \log \left( \frac{\text{кількість документів у колекції}}{\text{кількість документів, що містять термін } t+1} \right)$$

TF (term frequency) - IDF (inverse document frequency) — статистичний показник, що використовується для оцінки важливості слів у контексті документа, що є частиною колекції документів чи корпусу. Вага (значимість) слова пропорційна кількості вживань цього слова у документі, і обернено пропорційна частоті вживання слова у інших документах колекції.

$$TF-IDF = TF \times IDF, \quad (1)$$

Векторизація тексту дозволяє використовувати методи аналізу даних, які працюють з числовими даними, для обробки та розуміння текстової інформації. Це важливий етап у використанні тексту для різних завдань машинного навчання, класифікації тексту, пошуку інформації тощо.

Отже, векторизація документу є ключовим етапом у перетворенні текстових даних на формат, придатний для обробки алгоритмами машинного навчання та аналізу даних. Через використання таких метрик як частота терміну (TF) та інверсія частоти документа (IDF), текст перетворюється в числовий вектор, що відображає важливість окремих слів у контексті документа та колекції документів загалом. Таке представлення тексту в числовій формі дозволяє застосовувати складні аналітичні методи для виявлення ключових патернів, класифікації тексту, а також для підвищення точності і глибини аналізу текстових даних. Векторизація є незамінною у сфері аналізу великих обсягів даних, дозволяючи використовувати потужність машинного навчання для розуміння та обробки природної мови.

### 2.3 Метод декомпозиції єдиного значення

Метод декомпозиції єдиного значення (SVD - Singular Value Decomposition) є одним із методів зменшення розмірності даних. SVD є потужним математичним інструментом, що застосовується до матриць з метою розкладання матриці на більш прості компоненти, що дозволяє зберегти важливу інформацію та зменшити розмірність даних.

При застосуванні SVD до матриці  $A$  розмірності  $m \times n$ , де  $m$  - кількість рядків (спостережень), а  $n$  - кількість стовпчиків (ознак), матриця  $A$  розкладається на добуток трьох матриць:  $U$ ,  $S$  та  $VT$ :

$$A = U \cdot S \cdot VT$$

де:

$U$  - ортогональна матриця розмірності  $m \times m$ , яка містить лінійно незалежні власні вектори матриці  $AA^T$ .

$S$  - діагональна матриця розмірності  $m \times n$ , де на діагоналі розташовані сингулярні числа, які відображають важливість кожної компоненти.

$VT$  - транспонована ортогональна матриця розмірності  $n \times n$ , яка містить лінійно незалежні власні вектори матриці  $ATA$ .

Суть зменшення розмірності полягає у тому, що лише перші  $k$  сингулярних чисел та відповідні власні вектори  $U$  та  $V$  зберігаються, тобто використовуються лише перші  $k$  компонент для побудови нової апроксимованої матриці  $A_k$ .

$$A_k \approx U_k \cdot S_k \cdot V_k^T$$

Цей підхід дозволяє зберегти найбільш важливу інформацію, а використання лише частини сингулярних чисел і відповідних власних

векторів допомагає зменшити розмірність даних, зберігаючи їхню сутність. SVD широко застосовується у сферах комп'ютерного зору, обробки сигналів, рекомендаційних системах, зображеннях та інших галузях, де необхідно зменшити розмірність даних без втрати важливої інформації.

Метод декомпозиції єдиного значення (SVD) є ефективним інструментом у процесі зменшення розмірності даних, дозволяючи істотно спростити обробку великих матриць, зберігаючи при цьому ключову інформацію. Цей метод розкладає матрицю на три складові (U, S, VT), виділяючи найважливіші компоненти даних. Використання лише частини сингулярних чисел і відповідних векторів дозволяє зменшити обсяг даних без суттєвої втрати інформації. Широке застосування SVD у різних галузях, включаючи комп'ютерний зір, обробку сигналів, рекомендаційні системи та обробку зображень, підкреслює його значимість і ефективність як засобу оптимізації та аналізу великих об'ємів даних.

## **2.4 Семантичний аналіз LSA та LDA**

Зниження розмірності є процесом скорочення кількості випадкових змінних шляхом отримання множини головних змінних. Для семантичного аналізу (LSA) методом зменшення розмірності є метод декомпозиції єдиного значення (SVD), який обчислює лише  $k$  найбільших одиничних значень, де  $k$  є заданим користувачем параметром.

Моделювання тем в тексті на основі LSA методу. Коли SVD застосовується до матриць термінових документів, це перетворення відоме як прихований семантичний аналіз (LSA), оскільки воно перетворює такі матриці у „семантичний” простір низької розмірності. Зокрема, відомо, що LSA бореться з ефектами синонімії та полісемії (обидва з яких приблизно означають, що в слові є кілька значень), що спричиняє надмірну

розрідженість матриць термінових документів та виявляє низьку схожість за такими показниками, як подібність косинусів.

Відповідно до теореми про сингулярному розкладанні, будь-яка матеріальна прямокутна матриця може бути розкладена на твір трьох матриць:

$$A = U S V^T, \quad (2)$$

де матриці  $U$  і  $V$  – ортогональні, а  $S$  – діагональна матриця, значення на діагоналі якої називаються сингулярними значеннями матриці  $A$ . Буква  $T$  в вираженні  $V^T$  означає транспонування матриці. Таке розкладання володіє чудовою особливістю: якщо в матриці  $S$  залишити тільки  $k$  найбільших сингулярних значень, а в матрицях  $U$  і  $V$  – тільки відповідні цим значенням стовпці, то твір одержані матриць  $S$ ,  $U$  і  $V$  буде найкращим наближенням початкової матриці  $A$  до матриці  $\hat{A}$  рангу  $k$ :

$$\hat{A} \approx A =, \quad (3)$$

Основна ідея латентно-семантичного аналізу полягає в тому, що якщо в якості матриці  $A$  використовувалася матриця терми-на-документи, то матриця  $\hat{A}$ , Що містить тільки  $k$  перших лінійно незалежних компонент  $A$ , Відображає основну структуру різних залежностей, присутніх у вихідній матриці. Структура залежностей визначається ваговими функціями термів.

Latent Dirichlet Allocation (LDA) - - це метод розкладання матриць. У векторному просторі будь-який корпус (колекція документів) може бути представлений у вигляді матриці термінів документа. Наступна матриця (4) показує корпус із  $N$  документів  $D_1, D_2, D_3 \dots D_n$  та розмір словника з  $M$  слів  $W_1, W_2 \dots W_n$ . Значення комірки  $i, j$  дає підрахунок частоти слова  $W_j$  у документі  $D_i$ .

$$\begin{pmatrix} D_1 W_1 & D_1 W_2 & \dots & D_1 W_n \\ D_2 W_1 & D_2 W_2 & \dots & D_2 W_n \\ \dots & \dots & \dots & \dots \\ D_n W_1 & D_n W_2 & \dots & D_n W_n \end{pmatrix} \quad (4)$$

LDA перетворює (4) матрицю терміну документа у дві матриці нижнього розміру -  $M_1$  і  $M_2$ .

$M_1$  - матриця тематик документів (5), а  $M_2$  - матриця тем (6) - термінів із розмірами  $(N, K)$  та  $(K, M)$  відповідно, де  $N$  - кількість документів,  $K$  - кількість тем, а  $M$  - розмір словникового запасу.

$$M_1 = \begin{pmatrix} D_1 K_1 & D_1 K_2 & \dots & D_1 K_n \\ D_2 K_1 & D_2 K_2 & \dots & D_2 K_n \\ \dots & \dots & \dots & \dots \\ D_n K_1 & D_n K_2 & \dots & D_n K_n \end{pmatrix} \quad (5)$$

$$M_2 = \begin{pmatrix} K_1 W_1 & K_1 W_2 & \dots & K_1 W_n \\ K_2 W_1 & K_2 W_2 & \dots & K_2 W_n \\ \dots & \dots & \dots & \dots \\ K_n W_1 & K_n W_2 & \dots & K_n W_n \end{pmatrix} \quad (6)$$

Ці (5, 6) матриці вже забезпечують розподіл тем за темами та словами, однак, цей розподіл потребує вдосконалення, що є основною метою LDA. LDA використовує методи вибірки для покращення цих матриць. Після ряду ітерацій досягається стійкий стан, коли розподіл теми документа та термінів теми досить хороший. Це точка збіжності LDA.

Отже, латентний семантичний аналіз (LSA) та Latent Dirichlet Allocation (LDA) є важливими методами для семантичного аналізу тексту та зниження розмірності даних. LSA використовує метод декомпозиції єдиного значення для перетворення матриць термін-документ у семантичний простір низької розмірності, ефективно виокремлюючи важливі залежності та структури з даних. З іншого боку, LDA працює на принципі векторного простору, представляючи корпус як матрицю термін-документ і перетворюючи її у дві матриці нижньої розмірності, які відображають розподіл тем у документах та розподіл слів у темах. Обидва методи використовують складні математичні операції для виявлення

глибоких семантичних зв'язків у текстових даних, що є ключовим для розуміння, класифікації та аналізу великих обсягів текстової інформації.

## **2.5 Інтелектуальний метод формування рекламного контенту на основі семантичного аналізу**

Розвиток технологій обробки природної мови та семантичного аналізу відкриває нові можливості для створення інноваційних методів у сфері рекламного контенту. Особливо це стосується рекламних стратегій, де потрібно враховувати специфічні потреби та переваги цільової аудиторії. У цьому контексті розробка інтелектуального методу формування рекламного контенту на основі семантичного аналізу відіграє ключову роль, оскільки цей підхід дозволяє глибоко аналізувати великі обсяги текстових даних, виявляти основні тенденції та інтереси потенційних студентів. Застосування методів як LSA (Латентний Семантичний Аналіз), так і LDA (Латентне Розподілення Діріхле) в контексті зменшення розмірності даних та моделювання тематичних структур відкриває нові перспективи для створення ефективного та персоналізованого рекламного контенту. Цей підхід не тільки спрощує процес аналізу великих обсягів даних, але й забезпечує більш точне відображення інтересів та потреб аудиторії, що є вирішальним для розробки ефективних рекламних стратегій.

Для зменшення часових витрат на вибір популярного нового товару авторами розроблено інтелектуальний метод формування рекламного контенту на основі семантичного аналізу. Запропонований метод ілюструється схематично (рис.2.1) та представлений наступними кроками:

**Крок 1.** Проведення опитування студентів (Блок 1) та перетворення у формат .csv (Блок 2).

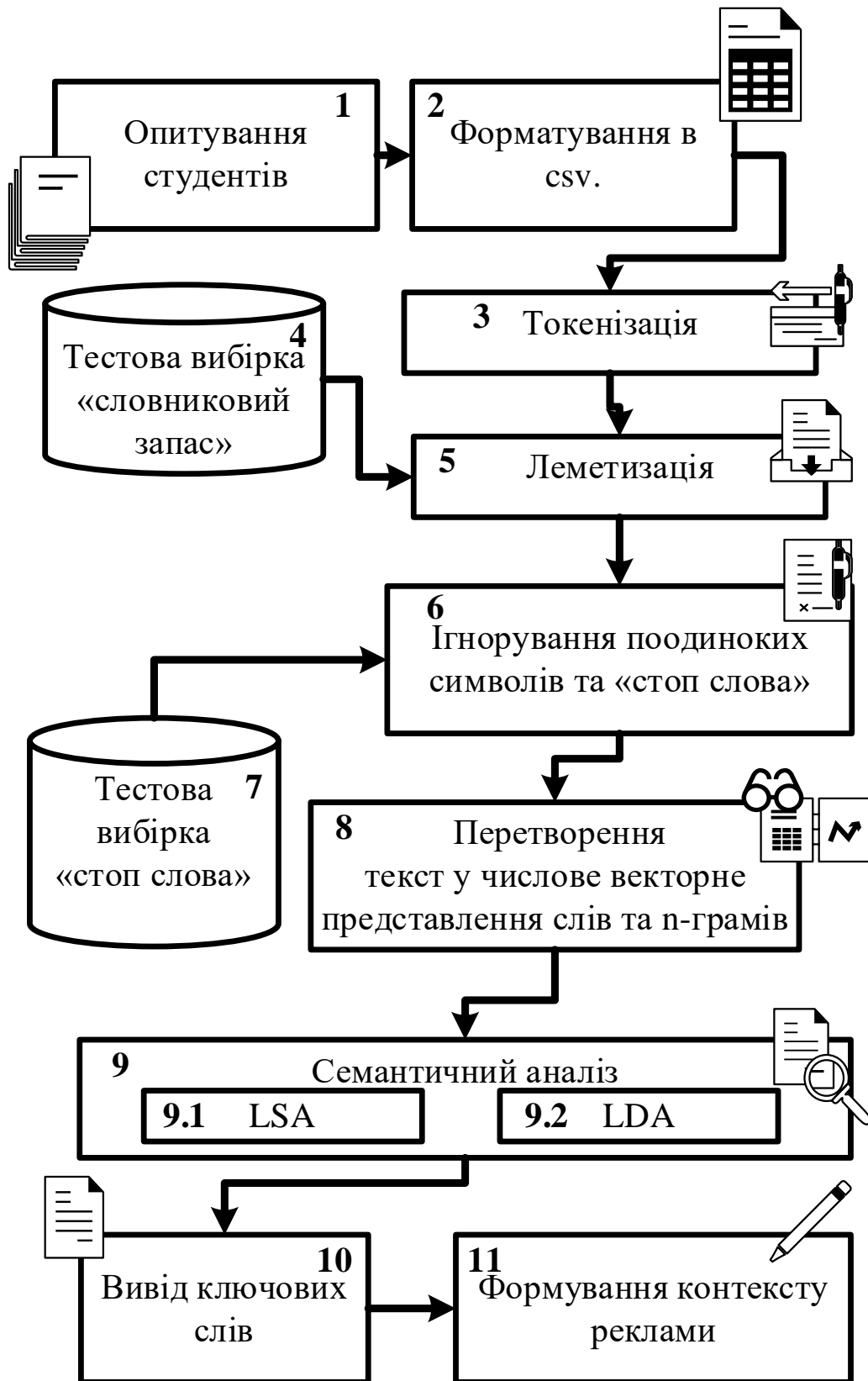


Рисунок 2.1 - Структура інтелектуального методу формування рекламного контенту на основі семантичного аналізу

**Крок 2.** Токенізацію документу (Блок 3)

**Крок 3.** Лематизація документу (Блок 5). Лематизації - це процес перетворення слова в його базову форму (Блок 4).

**Крок 4.** Процес ігнорування (Блок 6) поодиноких символів («.», «;», «:», «!», «'», «?», «,», «"», «()», «[]»), які мішають роботі програми проводиться на основі готового набору «стоп-слів».

**Крок 5.** Векторизація документу (Блок 8). Перетворення колекцію необроблених документів на матрицю функцій TF-IDF використовуючи формулу (1).

**Крок 6.** Семантичний аналіз (Блок 9).

**Крок 6.1.** Зменшення розмірності даних (Блок 9.1).

**Крок 6.2.** Моделювання тем в тексті на основі LDA методу (Блок 10.2).

**Крок 8.** Вивід ключових слів (Блок 10) по документах на основі LSA та LDA методу.

**Крок 9.** Формування контексту реклами (Блок 11), для закладів вищої освіти.

Для кращого розуміння розробленого методу варто провести експериментальне дослідження.

Інтелектуальний метод формування рекламного контенту на основі семантичного аналізу представляє собою ефективний підхід до аналізу та обробки текстових даних. Через ряд послідовних кроків, що включають опитування, токенізацію, лематизацію, видалення стоп-слів, векторизацію та семантичний аналіз за допомогою методів LSA та LDA, цей метод дозволяє глибоко аналізувати текстовий контент та виділяти ключові слова і фрази. Застосування цього методу може суттєво підвищити ефективність рекламних кампаній, допомагаючи створювати більш цілеспрямований та релевантний контент, який резонує з цільовою аудиторією. Експериментальне дослідження такого методу може надати додаткове



підтвердження його ефективності та показати практичні результати його застосування в реальних умовах.

## Висновки до розділу 2

1. Токенізація, лематизація та ігнорування символів є важливими компонентами передобробки тексту, що значно підвищують ефективність та точність подальшого аналізу та обробки тексту в сфері обробки природної мови. Токенізація допомагає розбити текст на базові елементи, що полегшує роботу з ним, лематизація забезпечує стандартизацію та спрощення словникового запасу, а ігнорування символів видаляє зайвий шум, дозволяючи зосередитись на суттєвій інформації. Ці процеси сприяють побудові більш точних та ефективних моделей машинного навчання, аналізу синтаксису та семантики, забезпечуючи глибший і більш інформативний аналіз тексту. Отже, вони є необхідними для реалізації складних завдань в області обробки тексту, від аналізу емоцій до виявлення ключових слів та фраз.
2. Векторизація документу є ключовим процесом у перетворенні текстової інформації на числові дані, що дозволяє застосувати сучасні алгоритми машинного навчання та аналізу даних до текстового контенту. Через комбінацію методів, таких як токенізація, лематизація, видалення стоп-слів та використання показників TF-IDF, векторизація дозволяє аналізувати значимість термінів у документах, визначаючи їх важливість відносно всієї колекції документів. Цей процес перетворює складні та різноманітні текстові дані на структуровану, числову форму, сприяючи глибокому аналізу, класифікації та іншим видам обробки тексту. Таким чином, векторизація документу є невід'ємною частиною багатьох сучасних технологічних підходів до обробки та аналізу текстових даних.
3. Метод декомпозиції єдиного значення (SVD) є важливим інструментом у зменшенні розмірності та обробці даних, що дозволяє ефективно розкласти великі та складні матриці на більш прості компоненти. Цей

метод використовується для збереження ключової інформації та видалення зайвих деталей, що робить його ідеальним для застосування в різноманітних сферах, таких як комп'ютерний зір, обробка сигналів, рекомендаційні системи та багато інших. Використання SVD допомагає значно знизити обсяги даних, при цьому зберігаючи їх основний зміст та структуру, що є ключовим для ефективної обробки та аналізу великих датасетів.

4. Методи Латентного Семантичного Аналізу (LSA) та Латентного Розподілу Діріхле (LDA) є важливими інструментами в аналізі семантики тексту, які використовуються для зменшення розмірності даних та виявлення прихованих зв'язків у великих колекціях текстових документів. LSA використовує метод SVD для перетворення матриць термінів у документи, ефективно борючись з проблемами синонімії та полісемії. З іншого боку, LDA використовує байєсівський підхід для моделювання розподілу тем у документах, що дозволяє аналізувати та класифікувати тексти на основі їхнього семантичного змісту. Обидва методи дозволяють виявити основні тематичні структури в текстових даних та здійснювати глибокий аналіз контенту, що є надзвичайно корисним у різних сферах застосування, від обробки природної мови до рекомендаційних систем.
5. Розроблений інтелектуальний метод формування рекламного контенту на основі семантичного аналізу демонструє ефективний підхід до обробки та аналізу текстових даних для створення цілеспрямованого рекламного контенту. Через послідовність кроків, що включають опитування студентів, токенізацію, лематизацію, видалення непотрібних символів, векторизацію та семантичний аналіз, методика дозволяє точно визначити ключові теми та слова, які резонують із цільовою аудиторією. Застосування методів LSA та LDA для моделювання тем і зменшення розмірності даних сприяє глибокому

розумінню семантичних зв'язків у текстах, що є критично важливим для створення впливового та змістовного рекламного контенту. Цей підхід відкриває широкі можливості для закладів вищої освіти щодо привертання уваги потенційних студентів, а також може бути застосований у багатьох інших сферах, де важливий глибокий аналіз текстового контенту.

## 3 РЕАЛІЗАЦІЯ ІНТЕЛЕКТУАЛЬНОГО МЕТОДУ

### 3.1 Попередня обробка даних

У цьому проекті з формування рекламного контенту для вищих навчальних закладів використовується інтелектуальний підхід, що базується на мові програмування Python. У цьому проекті для обробки та аналізу даних використовуються різноманітні бібліотеки Python, що надають широкі можливості для роботи з інформацією. Перша з них — `pandas` — це потужний інструмент для обробки та аналізу даних, який дозволяє працювати з великими наборами даних у табличному форматі.

Для математичних обчислень та маніпуляцій з числами та масивами використовується бібліотека `numpy`. Вона забезпечує високу продуктивність при виконанні операцій над числовими даними, що дозволяє ефективно працювати з числовими даними в аналізі.

Графічні візуалізації та аналітика даних реалізовані завдяки бібліотеці `matplotlib`. Ця бібліотека надає засоби для створення різноманітних графіків, діаграм та візуальних зображень, що допомагають легше сприймати та аналізувати дані.

Бібліотека `nltk` (Natural Language Toolkit) використовується для обробки тексту. Вона має ряд засобів для роботи з природним мовленням, включаючи функції токенізації, лематизації та аналізу семантики текстових даних.

Також використовується бібліотека `stop_words`, яка допомагає у вилученні стоп-слів з тексту. Ці слова не несуть значущої інформації та часто використовуються у мовленні, але не впливають на розуміння тексту.

Ці бібліотеки разом створюють потужні інструменти для обробки та аналізу різноманітних даних, роблять роботу з текстовою інформацією

більш доступною та ефективною, а також допомагають у візуалізації результатів аналізу для кращого розуміння даних.

Проект використовує дані, отримані з опитування студентів, які займаються навчанням за спеціальністю "Комп'ютерні науки". У цьому опитуванні взяли участь 152 студенти, кожен із яких відповів на 10 питань. Ці дані структуровані у форматі .csv, де кожен рядок містить відповідь одного студента на певне питання. Кожна відповідь вважається документом, що використовується для подальшого аналізу та обробки.

Ця інформація з опитування надає можливість отримати об'єктивне уявлення про думки та переконання студентів, які спеціалізуються у сфері комп'ютерних наук. Кожна відповідь студента важлива, оскільки разом вони створюють набір даних, який буде досліджуватися та аналізуватися для виявлення певних тенденцій, уподобань чи характеристик, що мають важливе значення для досліджуваної групи.

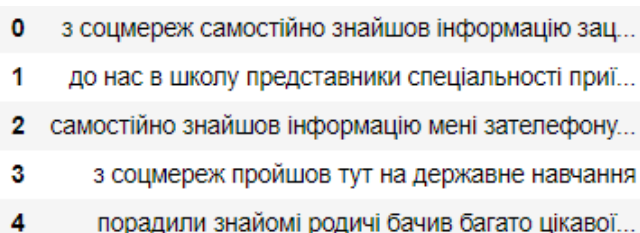
Ця інформація, структурована у вигляді .csv файлів, слугує основою для подальшого оброблення та аналізу даних з метою виявлення патернів, важливих аспектів та здійснення більш глибокого розуміння думок та ставлення студентів до певних питань у сфері комп'ютерних наук.

У цьому процесі текст спершу пройшов через лематизацію за допомогою функції WordNetLemmatizer, що дозволяє перетворити слова до їх базової форми або леми. Це допомагає віднайти спільні корені слів, уникнути дублювання та спростити подальший аналіз тексту. Наступним етапом було використання методу `nltk.tokenize.word_tokenize()` для розділення тексту на маркери або токени – окремі слова чи символи, що становлять основу для подальшої обробки.

Після токенізації була проведена очистка тексту від символів, які не несуть значущої інформації чи не впливають на зміст тексту. До цих символів відносяться знаки пунктуації, такі як крапка, кома, знак оклику,

апостроф, лапки та інші. Видалення цих символів допомагає зберегти лише суттєву інформацію та спростити подальший аналіз тексту.

Останнім кроком обробки стало видалення українських «стоп-слів». Це слова або терміни, які мають високу частоту вживання та незначно впливають на семантику тексту. Вони можуть бути загальними словами типу "і", "у", "на", "що" та інші, які не несуть конкретної семантичної навантаженості і можуть бути виключені з аналізу для отримання більш точних результатів (рис.3.1).



```

0   з соцмереж самостійно знайшов інформацію зац...
1   до нас в школу представники спеціальності приї...
2   самостійно знайшов інформацію мені зателефону...
3   з соцмереж пройшов тут на державне навчання
4   порадили знайомі родичі бачив багато цікавої...
  
```

Рисунок 3.2 - Лематизовані та токенізовані відзиви студентів без «стоп-слів» та символів

Розроблений інтелектуальний метод формування рекламного контенту для закладів вищої освіти на основі семантичного аналізу є важливим кроком у розвитку стратегій маркетингу та комунікації. Використання передових інструментів обробки тексту в Python, включаючи пакети `pandas`, `numpy`, `matplotlib`, `nltk` та `stop_words`, дозволяє ефективно обробляти великі обсяги текстових даних, отриманих із опитувань студентів. Через послідовність обробки, яка включає лематизацію, токенізацію, видалення непотрібних символів та стоп-слів, вдається виявити ключові теми та уподобання студентів. Цей підхід допомагає у створенні цілеспрямованого, ефективного рекламного контенту, що відповідає на попит і інтереси потенційних студентів. Це, у свою чергу, веде до підвищення ефективності рекламних кампаній та підсилює зв'язок між навчальними закладами та їх цільовою аудиторією.

### 3.2 Результати застосування методів LSA та LDA

Процес використання функції `TfidfVectorizer` полягає в перетворенні колекції необроблених текстових документів у числову матрицю за допомогою методу TF-IDF (рис.3.2), що відображає важливість кожного слова у документі та його важливість у контексті всього корпусу текстів. TF-IDF (Term Frequency-Inverse Document Frequency) оцінює вагу слова у документі шляхом порівняння його частоти в документі з його частотою у всьому корпусі текстів.

Перший етап полягає у підготовці текстів до векторизації. Документи перетворюються у вектори, де кожне слово отримує числове представлення згідно його значущості у тексті. Це дозволяє врахувати вагомість кожного слова у кожному документі порівняно з усім набором документів.

Після застосування `TfidfVectorizer` отримуємо матрицю, де кожен рядок відповідає окремому документу, а кожний стовпець представляє слово, яке зустрілося у текстах. Значення у цій матриці відображають важливість кожного слова у кожному документі, враховуючи його зважену частоту в документі та відносну інформативність у корпусі текстів.

(0, 21)	0.249923
(0, 5)	0.236487
(0, 40)	0.249923
(0, 52)	0.249923
(0, 48)	0.249923
(0, 0)	0.249923
(0, 32)	0.264944
(0, 42)	0.249923

Рисунок 3.2 - Частина TF-IDF матриці

У цьому етапі (рис.3.3) застосовується усічене SVD (Singular Value Decomposition) або LSA (Latent Semantic Analysis), яке є методом скорочення розмірності. Він дозволяє зменшити кількість вимірів даних, використовуючи усічене розкладання особливих значень (SVD). Цей метод



відрізняється від PCA (Principal Component Analysis) тим, що не центрує дані перед обчисленням декомпозиції особливих значень.

SVD використовується для зменшення розмірності даних, аналізуючи особливості матриць. У контексті аналізу тексту за допомогою методу TF-IDF, він застосовується для матриці, яка відображає відносну важливість слів у кожному документі. Цей метод вважається схованою семантичною моделлю (LSA), оскільки здатний відображати приховану семантику або відносини між словами.

Основна перевага усіченого SVD в порівнянні з PCA полягає в його здатності працювати з розрідженими матрицями, що є типовою для матриць, отриманих після векторизації тексту методом TF-IDF. Це дозволяє ефективніше враховувати важливість слів у кожному документі, створюючи компактне представлення текстової інформації з метою подальшого аналізу.

При використанні функції TruncatedSVD були визначені наступні параметри: бажана розмірність вихідних даних (документів) - `n_components=5`; алгоритм методу SVD - `algorithm='randomized'`; кількість ітерацій для рандомізованого SVD - `n_iter=10`; кількість відтворюваних результатів через кілька викликів функцій - `random_state=100`.

```
[[ 6.96660197e-01 -5.73572092e-02 -4.06682974e-01 ... -2.00406853e-18
  -7.54197595e-18  1.32222428e-17]
 [ 7.83147373e-02  2.58859227e-01 -1.10413796e-01 ... -6.93889390e-18
  -5.20417043e-18 -1.38777878e-17]
 [ 3.18620187e-01  1.45234702e-01 -6.98410495e-02 ... -7.26415456e-18
  2.25514052e-17  1.73472348e-18]
```

Рисунок 3.3 - Частина матриці LSA-методу

Далі був проведений аналіз тексту за допомогою методу LDA (рис.3.4). Функція `LatentDirichletAllocation` використовувалась з такими параметрами: бажана розмірність вихідних даних (документів) - `n_components=5`; метод оновлення `_component`, який використовується для

великих обсягів даних - `learning_method='online'`; максимальна кількість ітерацій - `max_iter=1`.

```
[[0.04455599 0.8221565 0.04445858 0.04441181 0.04441713]
 [0.7910456 0.05214722 0.05216473 0.05240864 0.05223381]
 [0.0572413 0.05729839 0.77093604 0.05725154 0.05727273]
 [0.07148326 0.07139567 0.71293007 0.07197358 0.07221742]
 [0.0459054 0.81552225 0.04597863 0.04597552 0.04661821]
 [0.04668559 0.04725528 0.04661619 0.04676363 0.81267932]]
```

Рисунок 3.4 - Частина матриці LDA-методу

LSA	LDA
Document 0 :	Document 0:
Topic 0 : 65.75057550474627 %	Topic 0 : 4.594951994120869 %
Topic 1 : 40.182712167943826 %	Topic 1 : 4.59272498367418 %
Topic 2 : 23.376891919092163 %	Topic 2 : 81.48129348616614 %
Topic 3 : -21.590881614430092 %	Topic 3 : 4.7410176809086515 %
Topic 4 : -28.60402955510097 %	Topic 4 : 4.590011855130164 %

Рисунок 3.5 - Ймовірність присутності ключових слів в окремих документах LSA та LDA-методу

З рисунку 3.4 випливає, що більшість ключових слів у документі 0 складають 66%, використовуючи LSA метод. За LDA методом, переважна кількість ключових слів представлена в документі 2, і становить 82%.

Тепер можемо отримати список важливих слів для кожного з 5 документів (див.рис.3.5) за допомогою LSA та LDA методів. Для прикладу, оберемо по 5 слів для кожної теми.

Отримані ключові слова методами LSA та LDA (див. рис.3.4 і 3.5) можна використати для створення наступного контенту для реклами закладу вищої освіти:

*«Отримай круте державне навчання. Детальніша інформація в месенджер»*

Отже, застосування методів LSA (Латентного Семантичного Аналізу) та LDA (Латентного Розподілу Діріхле) для аналізу текстових даних з опитувань студентів спеціальності «Комп'ютерні науки» виявилось

ефективним у виявленні ключових тем та слів, які відображають семантичні зв'язки та пріоритети у відповідях студентів. Використання цих методів дозволяє не лише зменшити розмірність даних, але й виявити приховані змістовні патерни, що є важливим для формування ефективного рекламного контенту. Зокрема, виявлені ключові слова можуть бути використані для створення цілеспрямованого та резонансного рекламного контенту, який відповідає інтересам та потребам потенційних студентів. Це підтверджує потенціал інтелектуальних методів обробки тексту для розробки маркетингових стратегій у освітніх установах та інших сферах, де важливо глибоко розуміти цільову аудиторію.

### **3.3 Порівняльний аналіз ефективності рекламних кампаній**

Для порівняння ефективності рекламних контентів, розроблених з використанням методів LSA та LDA, проведено експеримент на сторінці "Комп'ютерні науки ЗУНУ" у соціальній мережі Facebook (див. рис. 3.6). Перший варіант оголошення (рис. 3.6.A) був створений на основі рекомендацій, що впливають з попередніх наукових досліджень.

Цей експеримент включав у себе порівняння рекламних матеріалів, що базуються на LSA та LDA. Реклама на рисунку 3.6.A була сформована відповідно до правил, встановлених на підставі попередніх наукових робіт та досліджень.

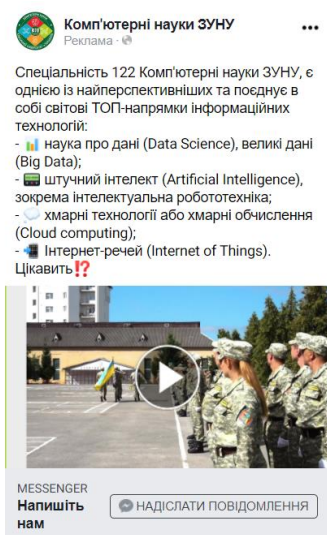
Основна мета цього порівняльного експерименту — визначити, який підхід, LSA чи LDA, ефективніший у створенні рекламного контенту для цільової аудиторії у Facebook для сторінки "Комп'ютерні науки ЗУНУ".

Ці конкретні рекомендації надавали певні напрямки для створення рекламного контенту, вказуючи на групи аудиторії, які мають найбільший потенціал взаємодії з рекламними матеріалами на сторінці "Комп'ютерні науки ЗУНУ" у Facebook.

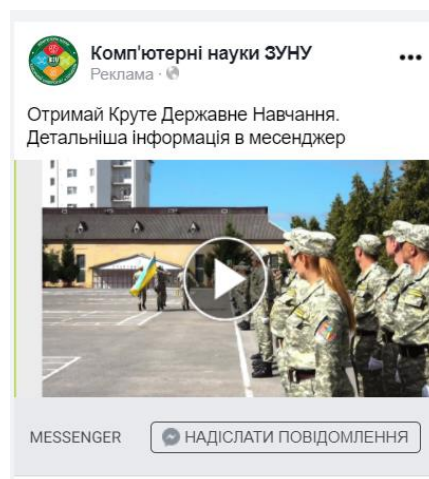
- За результатами попередніх досліджень [22], встановлено, що високий рівень зацікавленості у відеорекламі "Комп'ютерні науки ЗУНУ" спостерігається у чоловіків двох вікових категорій: 18-25 та 35-55 років. Ці групи показали найбільшу активність та взаємодію з цим типом реклами на Facebook.
- Згідно з рекомендаціями [23], найбільший інтерес до бізнес-сторінки "Комп'ютерні науки ЗУНУ" проявляють клієнти обох статей у віці 40-55 років. Ці категорії аудиторії виявили високу зацікавленість у взаємодії зі сторінкою, що створює базу для ефективної рекламної комунікації на Facebook.

<b>L</b>	Торіс 0: навчання державне пройшов інформацію круто
	Торіс 1: круто соцмережах цікавої інформації бачив
	Торіс 2: спеціальності зацікавило представлення представниками представники
	Торіс 3: вчитись зателефонували переконали цікаво інформацію
<b>S</b>	Торіс 4: знайомі порадили родичі влаштувала вартість
	Торіс 0: спеціальності представлення представниками зацікавило інформацію круто
	Торіс 1: представлення інформацію спеціальності зацікавило карантинном представниками
	Торіс 2: інформації круто соцмережах бачив цікавої вирішив
<b>A</b>	Торіс 3: державне пройшов навчання спеціальності порадили родичі
	Торіс 4: знайомих навчається тернополі сфер спеціальність подобається

Рисунок 3.6 - Ключові слова для формування рекламного контенту в окремих документах LSA та LDA-методу



А. Попередня версія



В. Сформовано на основі LSA та LDA-методу

Рисунок 3.7 - Реклама «Комп'ютерні науки» Західноукраїнського національного університету в Facebook

Таблиця 3.1 наочно демонструє порівняльний аналіз результативності двох рекламних кампаній, розроблених з використанням різних методів, LSA та LDA. Перша кампанія, яка тривала з Jan 4, 2022 по Jan 31, 2022, використовувала значний обсяг текстового контенту (зображений на рис.3.7.А). У той же час, друга кампанія, яка проводилась з Feb 1, 2022 по Feb 28, 2022, була побудована на основі текстового матеріалу, сформованого за допомогою методів LSA та LDA (див. рис.3.7.В).

За першими відомостями з таблиці, можна зрозуміти, що перша кампанія використовувала більший обсяг текстового матеріалу, який, ймовірно, мав інший стиль та підхід до привертання уваги аудиторії. У той же час, друга кампанія, побудована на основі аналізу методами LSA та LDA, могла мати іншу структуру контенту та спрямованість до цільової аудиторії.

Такий порівняльний аналіз допомагає виявити різницю у результатах, отриманих за різних умов та підходів до створення рекламного контенту,

що є важливим для подальшого вдосконалення та оптимізації стратегій маркетингу.

Таблиця 3.1 - Порівняння ефективності сформованого рекламного контенту

<b>Показник</b>	<b>Jan 4, 2021 – Jan 31, 2021</b>	<b>Feb 1, 2021 – Feb 28, 2021</b>	<b>Change</b>
Coverage	6792	7350	558
Frequency	1,18	1,19	0,02
Results	9,00	13,00	4,00
Price for the result	1,42	0,98	-0,44
Amount of expenses (USD)	12,77	12,75	-0,02
CPM (Cost per 1000 views)	1,60	1,46	-0,14
Link clicks	46,00	43,00	-3,00
CPC (cost-per-click)	0,28	0,30	0,02
CTR (clickability)	0,58	0,49	-0,09

В результатах таблиці 3.1 виявлені певні позитивні зміни. Охоплення аудиторії збільшилося на 558 осіб, що є значущим показником. Також середня кількість переглядів реклами на кожного користувача зросла на 0,02. Результативність рекламних оголошень, які ведуть до взаємодії в месенджерах, підвищилася на 4. Хоча збільшення цих метрик виглядає незначним, враховуючи скромний бюджет та специфіку цільової аудиторії, можна зазначити позитивний результат. Вартість отриманого результату зменшилась на 0,44 центи, враховуючи збільшення охоплення, ця ціна знизилась на 0,02 цента. Середня вартість за кожні 1000 показів оголошень (CPM) знизилася на 0,14 центів.

Проте, кількість кліків на посилання в рекламних оголошеннях зменшилась на 3, що призвело до підвищення вартості кожного кліку (CPC)

на 0,02 центи. Це призвело до зниження частки перегляду і взаємодії з рекламою (CTR) на 0,09 відсотка. Інформація щодо порівняльного аналізу ефективності рекламного контенту представлена на рисунку 3.8 у відсотковому вираженні.

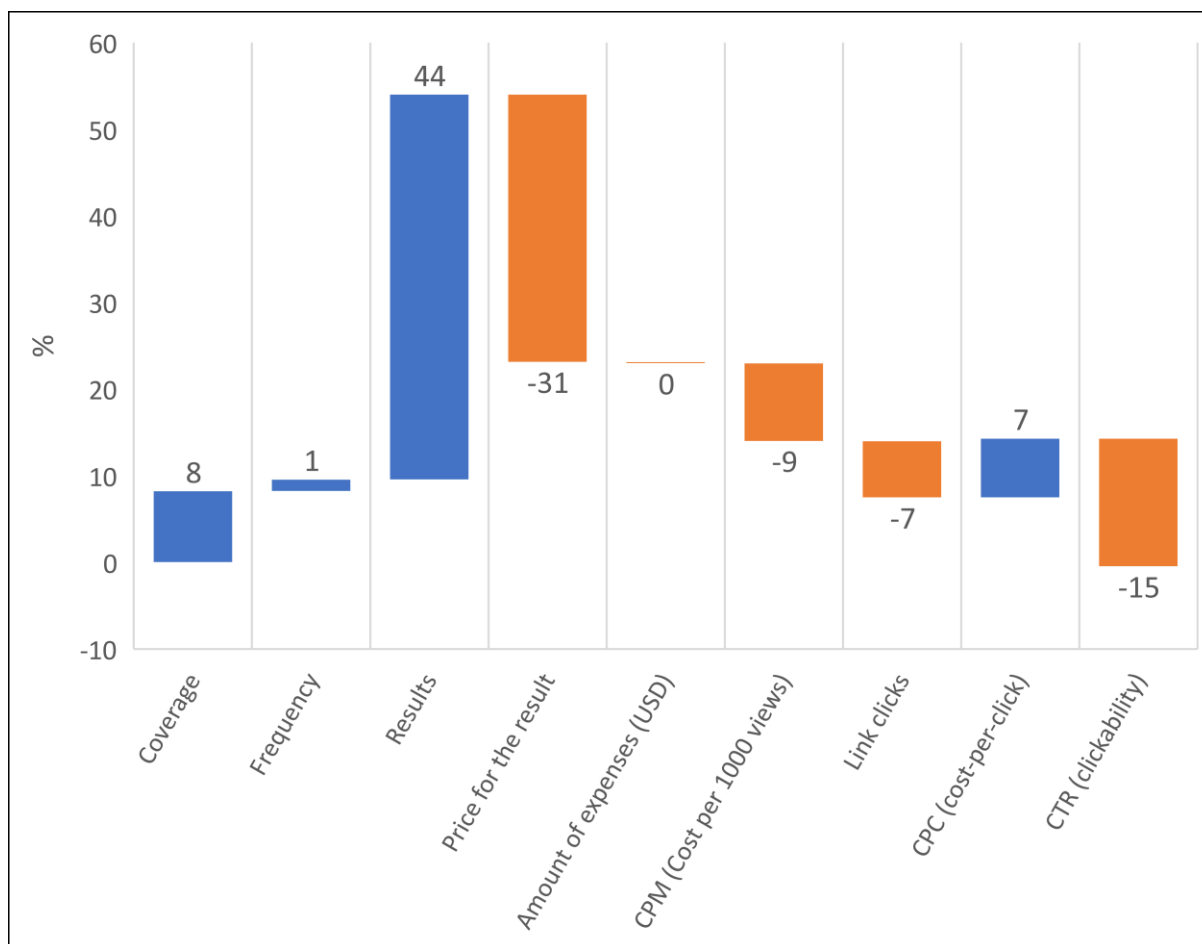


Рисунок 3.8 - Change (%) ефективності рекламної кампанії за періоди Jan 4, 2021 – Jan 31, 2021 та Feb 1, 2021 – Feb 28, 2021

З аналізу таблиці 3.1 та рисунку 3.8 можна зробити важливий висновок щодо покращення результативності рекламних оголошень. За цими показниками видно, що результативність оголошень підвищилась на 44%, що є значним поліпшенням. При цьому, вартість отриманого результату скоротилась на 31%. Результативність оголошень в цьому контексті оцінюється через частоту переходів на веб-сайт компанії або

взаємодію з рекламою, як у нашому випадку - через месенджер та задання питань.

Отже, інтелектуальний метод формування рекламного контенту для вищих навчальних закладів, базуючись на семантичному аналізі, виявляється надзвичайно ефективним, збільшуючи результативність реклами в соціальних мережах принаймні на 44%. Автор вважає, що розширення обсягу опитувань студентів сприятиме покращенню якості семантичного аналізу. Це дозволить отримати більш точні ключові слова для формування контенту та, відповідно, підвищити результативність рекламних заходів, що веде до зменшення витрат на них.

У відмінність від зазначених аналогів [2, 21], розроблений інтелектуальний підхід формування рекламного контенту для вищих навчальних закладів, ґрунтуючись на семантичному аналізі, відкриває можливість виокремлення ключових слів. Ці слова використовуються для створення тексту рекламного оголошення, результативність якого зросла на 44%. Крім того, вартість одного результату знизилась на 31%.

Отже, порівняльний аналіз ефективності рекламних кампаній, створених за допомогою методів LSA та LDA, свідчить про значний успіх у використанні інтелектуальних підходів до формування рекламного контенту. Використання семантичного аналізу в рекламних стратегіях для вищих навчальних закладів демонструє позитивний вплив на залучення цільової аудиторії та підвищення загальної ефективності кампаній. Відзначається підвищення результативності рекламних оголошень на 44% та зниження вартості на результат на 31%, що підтверджує ефективність семантичного підходу. Такий аналіз виявляє важливість точної націленості реклами та використання інноваційних технік аналізу даних для створення більш впливового та резонансного рекламного контенту. Ці результати підкреслюють значення продовження досліджень у цій сфері, а також можливості для покращення стратегій маркетингу в освітній галузі.



### Висновки до розділу 3

1. У цьому проекті використання інтелектуального методу формування рекламного контенту на основі семантичного аналізу демонструє високу ефективність у зборі та аналізі даних з опитувань студентів спеціальності «Комп'ютерні науки». Застосування різноманітних бібліотек Python, таких як pandas, numpy, matplotlib, nltk і stop\_words, дозволяє ефективно обробляти, аналізувати та візуалізувати зібрані дані. Процес включає в себе етапи лематизації, токенизації, видалення непотрібних символів і стоп-слів, що дозволяє виявити ключові теми та слова, важливі для розуміння поглядів студентів. Цей метод допомагає не лише в ефективному виявленні тенденцій та уподобань у цільовій аудиторії, але й надає можливість створення більш цілеспрямованого та резонансного рекламного контенту для закладів вищої освіти. Такий підхід відкриває широкі перспективи для покращення стратегій маркетингу та комунікацій в освітній сфері.
2. Застосування методів LSA та LDA для аналізу текстових даних, зібраних з опитувань студентів спеціальності «Комп'ютерні науки», демонструє ефективність цих методів у виявленні ключових тем і слів, які мають значення для цільової аудиторії. Використання TfidfVectorizer для перетворення тексту в числову матрицю та подальше її оброблення за допомогою LSA та LDA дозволяє здійснити глибокий семантичний аналіз, що включає зменшення розмірності даних і виділення найважливіших аспектів текстів. Такий підхід дозволяє не тільки глибше зрозуміти інтереси та потреби студентів, але й розробити більш цілеспрямований та ефективний рекламний контент для закладів вищої освіти. Вибрані ключові слова та теми, виявлені через ці методи, можуть бути використані для створення впливових

рекламних повідомлень, що відображають реальні інтереси та уподобання потенційних студентів.

3. Проведений порівняльний аналіз ефективності рекламних кампаній для закладів вищої освіти, використовуючи інтелектуальні методи LSA та LDA, показав значне поліпшення в результативності рекламних оголошень. Застосування цих методів семантичного аналізу допомогло виявити ключові слова та теми, які найбільш резонують з цільовою аудиторією, що призвело до підвищення охоплення, кількості результатів та зниження вартості на досягнення цих результатів. Зростання результативності рекламних оголошень на 44% та зниження витрат на 31% демонструють, що використання даних методів може значно покращити маркетингові стратегії та оптимізувати бюджети на рекламу в соціальних мережах. Це підкреслює важливість інтелектуального підходу до аналізу даних у сучасному цифровому маркетингу, особливо у контексті залучення аудиторії у сфері освіти.

## ВИСНОВКИ

Рекламний контент є вирішальним елементом маркетингової стратегії, який впливає на залучення та взаємодію з цільовою аудиторією. Ефективний контент вимагає стратегічного підходу, креативності, естетичної привабливості та звернення до специфічних потреб аудиторії. Використання різних медіаформатів і адаптація контенту до особливостей платформ забезпечують високу ефективність комунікацій.

Обробка тексту та семантичний аналіз відіграють ключову роль у створенні рекламного контенту. Вони дозволяють глибше розуміти потреби аудиторії, створювати персоналізований та емоційно резонансний контент, прогнозувати тренди та точно вимірювати ефективність рекламних кампаній.

Аналіз сучасних підходів до інтернет-реклами в різних галузях підкреслює значення інтеграції обробки тексту та семантичного аналізу в маркетингові стратегії вищої освіти. Це відкриває нові можливості для створення більш ефективного та цілеспрямованого контенту.

Токенізація, лематизація, та ігнорування символів є критичними етапами передобробки тексту. Вони забезпечують покращення точності та ефективності подальшого аналізу тексту, що є важливим для розробки точних моделей машинного навчання та глибокого аналізу тексту.

Векторизація документу перетворює текст на числові дані, дозволяючи застосувати алгоритми машинного навчання для аналізу тексту. Вона враховує важливість слів у документі, сприяючи більш глибокому розумінню текстової інформації.

Метод декомпозиції єдиного значення (SVD) є ефективним інструментом для зменшення розмірності даних, дозволяючи зберегти ключову інформацію і відсіяти непотрібні деталі. Це сприяє ефективній обробці великих датасетів.

Методи LSA та LDA дозволяють виявити приховані тематичні структури в текстах, сприяючи глибокому аналізу контенту. Вони здійснюють зменшення розмірності даних та класифікацію текстів, відкриваючи нові можливості для аналізу та обробки тексту.

Розроблений інтелектуальний метод формування рекламного контенту на основі семантичного аналізу демонструє ефективність у створенні цілеспрямованого та резонансного контенту. Він включає ряд процесів, що дозволяють точно визначити ключові теми та слова для залучення цільової аудиторії.

Використання Python та його бібліотек для обробки даних з опитувань студентів забезпечує ефективний аналіз та інтерпретацію інформації, дозволяючи створити впливовий рекламний контент, що відповідає потребам та інтересам студентів.

Застосування методів LSA та LDA до аналізу текстових даних, зібраних з опитувань, показало їх ефективність у виявленні ключових слів та тем для створення резонансного рекламного контенту, що підвищує залученість та інтерес цільової аудиторії.

Порівняльний аналіз ефективності рекламних кампаній, створених за допомогою методів LSA та LDA, підтвердив значне поліпшення в результативності та оптимізацію витрат. Це підкреслює важливість інтелектуального підходу до формування рекламного контенту, особливо у контексті освітньої сфери.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. S. Hou, S. Zhou, W. Liu et al. "Classifying advertising video by topicalizing high-level semantic concepts." *Multimed Tools Appl*, Springer 77, 25475–25511, 2018. DOI: 10.1007/s11042-018-5801-3
2. F. García-Sánchez, R. Colomo-Palacios, R. Valencia-García. "A social-semantic recommender system for advertisements." *Information Processing & Management*, Elsevier, 57(2), 102153, 2020. DOI: 10.1016/j.ipm.2019.102153
3. J. Shokeen, C. Rana. "A trust and semantic based approach for social recommendation." *J Ambient Intell Human Comput*, Springer. 2021. DOI: 10.1007/s12652-020-02806-1
4. Y. Kalepalli, S. Tasneem, P. D. Pasupuleti, S. Manne. "IEEE 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS) - Effective Comparison of LDA with LSA for Topic Modelling." *IEEE*, 1245–1250, 2020. DOI: 10.1109/ICICCS48265.2020.9120888
5. N. Parveen, M. V. B. T. Santhi, L. R. Burra, V. Pellakuri, H. Pellakuri. "Women's e-commerce clothing sentiment analysis by probabilistic model LDA using R-SPARK." *Materials Today: Proceedings*, Elsevier, 2021. DOI: 10.1016/j.matpr.2020.10.064
6. L. Luo. "Network text sentiment analysis method combining LDA text representation and GRU-CNN." *Pers Ubiquit Comput*, Springer, 23, 405–412, 2019. DOI: 10.1007/s00779-018-1183-9
7. P. Kherwa, P. Bansal. "A Comparative Empirical Evaluation of Topic Modeling Techniques." *International Conference on Innovative Computing and Communications*, Springer, Singapore, 289-297, 2021. DOI: 10.1007/978-981-15-5148-2\_26
8. B. Aslam, H. Karjaluo. "Digital advertising around paid spaces, E-advertising industry's revenue engine: A review and research agenda."

- Telematics and Informatics, Elsevier, 34(8), 1650-1662, 2017. DOI: 10.1016/j.tele.2017.07.011
9. A. Peruta, A. B. Shields. "Social media in higher education: understanding how colleges and universities use Facebook." *Journal of Marketing for Higher Education*, 27:1, 131-143, 2017. DOI: 10.1080/08841241.2016.1212451
  10. J. Aguilar, G. Garcia. "An Adaptive Intelligent Management System of Advertising for Social Networks: A Case Study of Facebook." *IEEE Transactions on Computational Social Systems*, vol. 5, no. 1, pp. 20-32, 2018. DOI: 10.1109/TCSS.2017.2759188
  11. D. Lee, K. Hosanagar, H. S. Nair. "Advertising Content and Consumer Engagement on Social Media: Evidence from Facebook." *Management Science*, 64:11, 5105-5131, 2018. DOI: 10.1287/mnsc.2017.2902
  12. A. M. Jamison, D. A. Broniatowski, M. Dredze, Z. Wood-Doughty, D. Khan, S. C. Quinn. "Vaccine-related advertising in the Facebook Ad Archive." *Vaccine*, Elsevier, Volume 38, Issue 3, 512-520, 2019. DOI: 10.1016/j.vaccine.2019.10.066
  13. S. Youn, S. Kim. "Understanding ad avoidance on Facebook: Antecedents and outcomes of psychological reactance." *Computers in Human Behavior*, Elsevier, Volume 98, 232-244, 2019. DOI: 10.1016/j.chb.2019.04.025
  14. C. L. White, B. Boatwright. "Social media ethics in the data economy: Issues of social responsibility for using Facebook for public relations." *Public Relations Review*, Elsevier, Volume 46, Issue 5, 101980, 2020. DOI: 10.1016/j.pubrev.2020.101980
  15. A. Gitomer, P. V. Oleinikov, L. M. Baum et al. "Geographic impressions in Facebook political ads." *Appl Netw Sci*, Springer, 6, 18, 2021. DOI: 10.1007/s41109-020-00350-7

16. A. Galati, M. Crescimanno, S. Tinervia, F. Fagnani. "Social media as a strategic marketing tool in the Sicilian wine industry: Evidence from Facebook." *Wine Economics and Policy*, Elsevier, 6(1), 40-47, 2017. DOI: 10.1016/j.wep.2017.03.003
17. A. Sharma, S. Sharma, M. Chaudhary. "Are small travel agencies ready for digital marketing? Views of travel agency managers." *Tourism Management*, Elsevier, 79, 104078, 2020. DOI: 10.1016/j.tourman.2019.104078
18. G. Szymanski, P. Lininski. "Model of the effectiveness of Google Adwords advertising activities." In 2018 IEEE 13th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT), Vol. 2, pp. 98-101, IEEE, 2018.
19. J. Phua, J. S. E. Lin, D. J. Lim. "Understanding consumer engagement with celebrity-endorsed E-Cigarette advertising on Instagram." *Computers in Human Behavior*, Elsevier, 84, 93-102, 2018. DOI: 10.1016/j.chb.2019.04.025
20. H. Gupta, S. Singh, P. Sinha. "Multimedia tool as a predictor for social media advertising-a YouTube way." *Multimedia tools and applications*, Springer, 76(18), 18557-18568, 2017. DOI: 10.1007/s11042-018-5801-3
21. M. Xu, S. Liu. "Semantic-Enhanced and Context-Aware Hybrid Collaborative Filtering for Event Recommendation in Event-Based Social Networks." *IEEE Access*, 7, 17493–17502, 2019. DOI: 10.1109/ACCESS.2019.2895824
22. H. Lipyanina, S. Sachenko, T. Lendyuk, A. Sachenko. "Targeting model of HEI video marketing based on classification tree." 16th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer.

Volume II: Workshops, ICTERI 2020, CEUR Workshop Proceedings, Vol.2732, P.487-498, 2020.

23. Н. Lipyanina, А. Sachenko, Т. Lendyuk, S. Nadvynychny, S. Grodskiy. "Decision tree based targeting model of customer interaction with business page." Proceedings of the third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020), CEUR Workshop Proceedings, vol. 2608, pp. 1001–1012, 2020
24. Загальні рекомендації з підготовки, оформлення, захисту та оцінювання випускних кваліфікаційних робіт здобувачів вищої освіти першого «бакалаврського» і другого «магістерського» рівнів / За ред. доц. М.І. Шинкарика. Тернопіль: ТНЕУ, 2018. 67 с.
25. Комар М.П., Саченко А.О., Васильків Н.М. Методичні рекомендації до виконання кваліфікаційної роботи з освітньо-професійної програми «Комп'ютерні науки» спеціальності 122 «Комп'ютерні науки» за другим (магістерським) рівнем вищої освіти. Тернопіль: ЗУНУ, 2021. 32 с.
26. Загальні рекомендації з підготовки, оформлення, захисту та оцінювання випускних кваліфікаційних робіт здобувачів вищої освіти першого «бакалаврського» і другого «магістерського» рівнів / За ред. доц. М.І. Шинкарика. Тернопіль: ТНЕУ, 2018. 67 с.
27. Комар М.П., Саченко А.О., Васильків Н.М. Методичні рекомендації до виконання кваліфікаційної роботи з освітньо-професійної програми «Комп'ютерні науки» спеціальності 122 «Комп'ютерні науки» за другим (магістерським) рівнем вищої освіти. Тернопіль: ЗУНУ, 2021. 32 с.
28. Загальні рекомендації з підготовки, оформлення, захисту та оцінювання випускних кваліфікаційних робіт здобувачів вищої



освіти першого «бакалаврського» і другого «магістерського» рівнів / За ред. доц. М.І. Шинкарика. Тернопіль: ТНЕУ, 2018. 67 с.

29. Комар М.П., Саченко А.О., Васильків Н.М. Методичні рекомендації до виконання кваліфікаційної роботи з освітньо-професійної програми «Комп'ютерні науки» спеціальності 122 «Комп'ютерні науки» за другим (магістерським) рівнем вищої освіти. Тернопіль: ЗУНУ, 2021. 32 с.

ДОДАТОК А  
АПРОБАЦІЯ ОТРИМАНИХ РЕЗУЛЬТАТІВ

**МАТЕРІАЛИ**  
IV ВСЕУКРАЇНСЬКОЇ СТУДЕНТСЬКОЇ НАУКОВОЇ  
**КОНФЕРЕНЦІЇ**

29 ВЕРЕСНЯ 2023 РІК • М. ЧЕРНІГІВ, УКРАЇНА

ЕКСПЕРИМЕНТАЛЬНІ ТА  
ТЕОРЕТИЧНІ ДОСЛІДЖЕННЯ В  
КОНТЕКСТІ СУЧАСНОЇ НАУКИ

ISBN 978-617-8126-63-6  
DOI 10.36074/liga-ukr-29.09.2023



**УДК 082:001**

**Е 45**

Голова оргкомітету: Коренюк І.О.

Верстка: Зрада С.І.

Дизайн: Бондаренко І.В.



*Конференцію зареєстровано Державною науковою установою «УкрІНТЕІ» в базі даних науково-технічних заходів України та інформаційному бюлетені «План проведення наукових, науково-технічних заходів в Україні» (Посвідчення №320 від 16.06.2023).*

*Матеріали конференції знаходяться у відкритому доступі на умовах ліцензії CC BY-SA 4.0 International.*



**Е 45** **Експериментальні та теоретичні дослідження в контексті сучасної науки:** матеріали IV Всеукраїнської студентської наукової конференції, м. Чернігів, 29 вересня, 2023 рік / ГО «Молодіжна наукова ліга». — Вінниця: ТОВ «УКРЛОГОС Груп», 2023. — 154 с.

ISBN 978-617-8126-63-6

DOI 10.36074/liga-ukr-29.09.2023

Викладено матеріали учасників IV Всеукраїнської мультидисциплінарної студентської наукової конференції «Експериментальні та теоретичні дослідження в контексті сучасної науки», яка відбулася 29 вересня 2023 року у місті Чернігів, Україна.

**УДК 082:001**

© Колектив учасників конференції, 2023

© ГО «Молодіжна наукова ліга», 2023

© ТОВ «УКРЛОГОС Груп», 2023

ISBN 978-617-8126-63-6

ВИКОРИСТАННЯ ТЕХНОЛОГІЇ БЛОКЧЕЙН ДЛЯ ЗАХИЩЕНОГО ВЕДЕННЯ ДОКУМЕНТООБИГУ Паплінський В.В., Науковий керівник: Устенко С.В. ....	60
ДОСЛІДЖЕННЯ МЕТОДІВ ПІДТРИМКИ КОНСИСТЕНЦІЇ У РОЗПОДІЛЕНИХ ТРАНЗАКЦІЯХ Савенков О.А., Науковий керівник: Білова Т.Г. ....	63
ІНТЕЛЕКТУАЛЬНИЙ МЕТОД ВИЗНАЧЕННЯ ШАХРАЙСЬКИХ ІНТЕРНЕТ-МАГАЗИНІВ Сукмановський О., Науковий керівник: Ліп'яніна-Гончаренко Х.В. ....	66
ІНТЕЛЕКТУАЛЬНИЙ МЕТОД ФОРМУВАННЯ КОНТЕКСТУ РЕКЛАМИ ТА ЦІЛЬОВОЇ АУДИТОРІЇ НА ОСНОВІ НАВЧАННЯ АСОЦІАТИВНИХ ПРАВИЛ Хам І., Науковий керівник: Ліп'яніна-Гончаренко Х.В. ....	68
ІНТЕЛЕКТУАЛЬНИЙ МЕТОД ФОРМУВАННЯ РЕКЛАМНОГО КОНТЕНТУ НА ОСНОВІ СЕМАНТИЧНОГО АНАЛІЗУ Курпела П., Науковий керівник: Ліп'яніна-Гончаренко Х.В. ....	70
КОМП'ЮТЕРИЗАЦІЯ ПРОЦЕСІВ ПРИЙНЯТТЯ РІШЕНЬ ПРИ УПРАВЛІННІ ЦІННИМИ ПАПЕРАМИ Ярцев С.Д., Науковий керівник: Устенко С.В. ....	72
МОДЕЛЬ НАУКОВОГО ДОСЛІДЖЕННЯ МЕТАВСЕСВІТУ Неделько В.Ю., Науковий керівник: Устенко С.В. ....	75
ОПТИМІЗАЦІЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ З ВИБОРУ ДИСЦИПЛІН ЗДОБУВАЧАМИ ЗАКЛАДІВ ВИЩОЇ ОСВІТИ Новак А.С., Науковий керівник: Онищенко В.В. ....	78
ПЕРЕХРЕСНА ОЦІНКА МОДЕЛЕЙ ІНВЕСТИЦІЙНИХ РИЗИКІВ ВІРТУАЛЬНОЇ ІТ-КОМПАНІЇ НА ОСНОВІ МАШИННОГО НАВЧАННЯ Дем'янова В., Науковий керівник: Ліп'яніна-Гончаренко Х.В. ....	82
ПІДХІД ПРОГНОЗУВАННЯ ПЛАНУ ПРОЕКТУ З ОБМЕЖЕНИМИ РЕСУРСАМИ Сікач Б., Науковий керівник: Саченко О.А. ....	84
ПРОЕКТУВАННЯ МОБІЛЬНО ОРІЄНТОВАНОЇ ІС «TASK-MENEDЖЕР» Щербак П.О., Науковий керівник: Устенко С.В. ....	86
РОЗРОБКА СИСТЕМИ КЕРУВАННЯ РОЗУМНИМ БУДИНКОМ НА ОСНОВІ ПРОТОКОЛУ MQTT Ніщенко Д.О., Науковий керівник: Баранюк О.Ф. ....	89

## СЕКЦІЯ 8.

### ФІЗИКО-МАТЕМАТИЧНІ НАУКИ

ПРИКЛАД ЗАСТОСУВАННЯ ДИНАМІЧНИХ СИСТЕМ РІВНЯНЬ ДО МАТЕМАТИЧНОГО МОДЕЛЮВАННЯ Бульботка Т.В., Науковий керівник: Тимчук М.В. ....	92
--	----

**Курпела Павло**, здобувач вищої освіти факультету комп'ютерних інформаційних технологій  
*Західноукраїнський національний університет, Україна*

**Науковий керівник: Ліп'яніна-Гончаренко Христина Володимирівна**, канд. техн. наук, доцент, доцент кафедри інформаційно-обчислювальних систем та управління  
*Західноукраїнський національний університет, Україна*

## **ІНТЕЛЕКТУАЛЬНИЙ МЕТОД ФОРМУВАННЯ РЕКЛАМНОГО КОНТЕНТУ НА ОСНОВІ СЕМАНТИЧНОГО АНАЛІЗУ**

Для зменшення часових витрат на вибір популярного нового товару авторами розроблено інтелектуальний метод формування рекламного контенту закладів вищої освіти на основі семантичного аналізу. Запропонований метод ілюструється схематично (Fig.1) та представлений наступними кроками:

Крок 1. Проведення опитування студентів (Блок 1) та перетворення у формат .csv (Блок 2).

Крок 2: Токенізація документів (Блок 3). Токенізація - це процес розбиття великої кількості тексту на менші частини, відомі як лексеми. Ця процедура корисна для пошуку закономірностей і є основним кроком для лематизації і обгрунтування.

Крок 3: Лематизація документів (Блок 5). Лематизація - це процес перетворення слів у їх базову форму (Блок 4).

Крок 4: Видалення одиночних символів (".", ";", ":", "!", "", "?", ",", "''", "()", "[]"), які заважають роботі програми, на основі набору "стоп-слів" (Блок 7).

Крок 5: Векторизація документів (Блок 8) шляхом перетворення колекції необроблених документів на матрицю TF-IDF. TF (term frequency) - IDF (inverse document frequency) - це статистичний показник, який оцінює важливість слів у тексті.

Крок 6: Семантичний аналіз (Блок 9).

Крок 6.1: Зменшення розмірності даних (Блок 9.1) за допомогою методу декомпозиції єдиного значення (SVD) для латентного семантичного аналізу (LSA).

Крок 6.2: Моделювання тем у тексті на основі методу Latent Dirichlet Allocation (LDA).

Крок 7: Виведення ключових слів (Блок 10) для документів, використовуючи як LSA, так і LDA методи.

Крок 8: Формування рекламного контексту для вищих навчальних закладів (Блок 11).

Отже, розроблено інтелектуальний підхід до створення рекламного контенту для закладів вищої освіти на основі семантичного аналізу. Запропонований метод включає кроки, такі як опитування студентів, токенизація та лематизація текстів, векторизація документів, семантичний аналіз і вивід ключових слів. Цей підхід допомагає створювати ефективний рекламний контент, що може зменшити витрати на рекламу та полегшити вибір нових товарів у вищих навчальних закладах.



Fig. 1. Структура інтелектуального методу формування рекламного контенту закладів вищої освіти на основі семантичного аналізу

#### Список використаних джерел:

1. Lipyanina, H., Sachenko, A., Lendyuk, T., Nadvynychny, S., Grodskiy, S. (2020). Decision tree based targeting model of customer interaction with business page. Proceedings of the third International Workshop on Computer Modeling and Intelligent Systems (CMIS-2020), CEUR Workshop Proceedings, vol. 2608, pp. 1001–1012.
2. Lipianina-Honcharenko, K., Wolff, C., Sachenko, A., Desyatnyuk, O., Sachenko, S., & Kit, I. (2023). Intelligent Information System for Product Promotion in Internet Market. Applied Sciences, 13(17), 9585.
3. Gramyak, R., Lipyanina-Goncharenko, H., Sachenko, A., Lendyuk, T., & Zahorodnia, D. (2021). Intelligent Method of a Competitive Product Choosing based on the Emotional Feedbacks Coloring. In IntellTSIS (pp. 246-257).

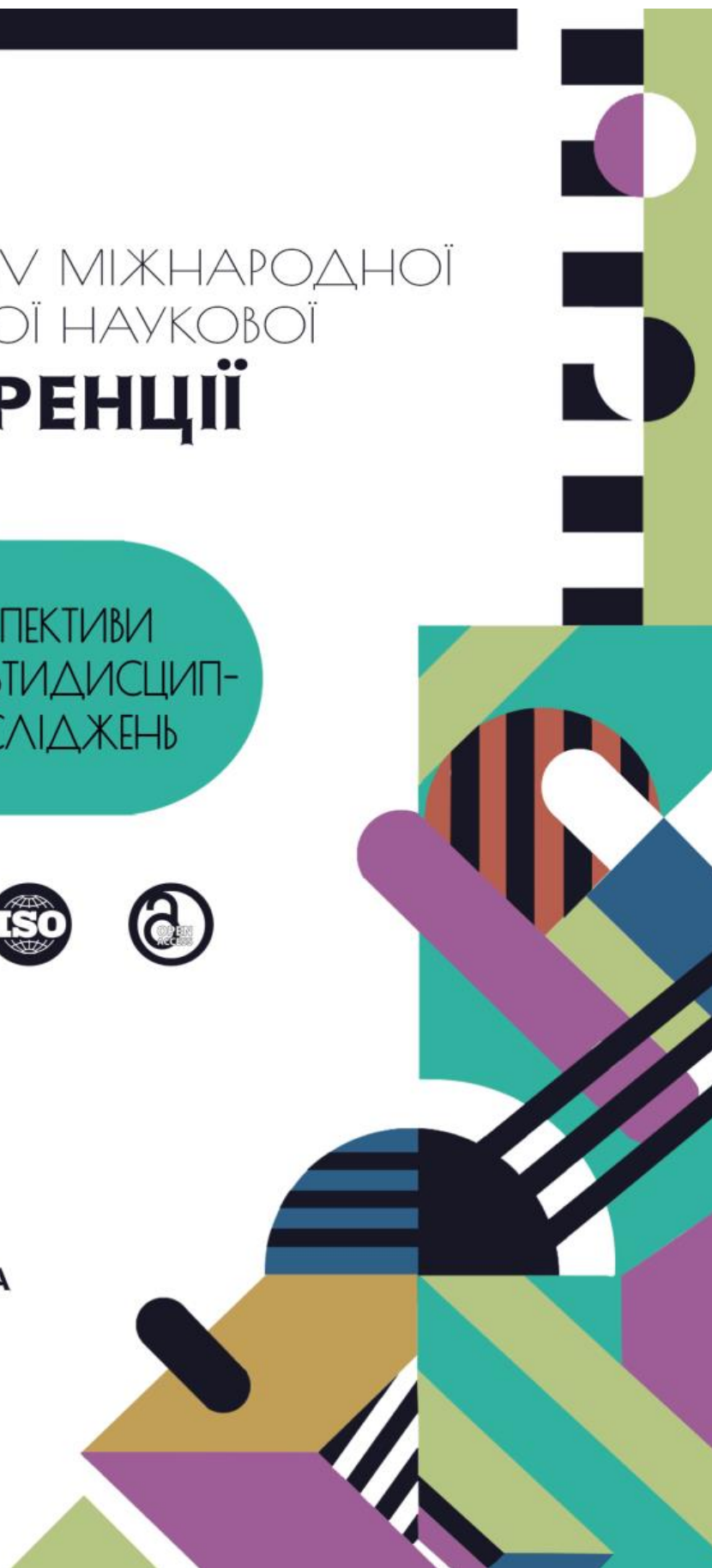
МАТЕРІАЛИ ІV МІЖНАРОДНОЇ  
СТУДЕНТСЬКОЇ НАУКОВОЇ  
**КОНФЕРЕНЦІЇ**

ТРЕНДИ ТА ПЕРСПЕКТИВИ  
РОЗВИТКУ МУЛЬТИДИСЦИП-  
ЛІНАРНИХ ДОСЛІДЖЕНЬ



М. ЛУЦЬК, УКРАЇНА

**1 ГРУДНЯ  
2023 РІК**



**УДК 082:001**  
**Т 66**



Голова оргкомітету: Коренюк І.О.

Верстка: Зрада С.І.

Дизайн: Бондаренко І.В.



*Конференцію зареєстровано Державною науковою установою «УкрІНТЕІ» в базі даних науково-технічних заходів України та бюлетені «План проведення наукових, науково-технічних заходів в Україні» (Посвідчення №329 від 16.06.2023).*

*Матеріали конференції знаходяться у відкритому доступі на умовах ліцензії Creative Commons Attribution-ShareAlike 4.0 International License (CC BY-SA 4.0).*

Т 66

**Тренди та перспективи розвитку мультидисциплінарних досліджень:** матеріали IV Міжнародної студентської наукової конференції, м. Луцьк, 1 грудня, 2023 рік / ГО «Молодіжна наукова ліга». — Вінниця: ТОВ «УКРЛОГОС Груп», 2023. — 500 с.

ISBN 978-617-8126-79-7

DOI 10.36074/liga-inter-01.12.2023

Викладено матеріали учасників IV Міжнародної мультидисциплінарної студентської наукової конференції «Тренди та перспективи розвитку мультидисциплінарних досліджень», яка відбулася 1 грудня 2023 року у місті Луцьк, Україна.

**УДК 082:001**

ISBN 978-617-8126-79-7

© Колектив учасників конференції, 2023

© ГО «Молодіжна наукова ліга», 2023

© ТОВ «УКРЛОГОС Груп», 2023



**Тренди та перспективи розвитку мультидисциплінарних досліджень**

ПЕРЕВАГИ ТА ВИКЛИКИ ПРОМИСЛОВОГО ІНТЕРНЕТУ РЕЧЕЙ Гарасимчук А.О., Хомік М.А., <i>Науковий керівник: Гарасимчук О.І.</i> .....	327
ПЕРИФЕРІЙНІ ОБЧИСЛЕННЯ В АНАЛІЗІ ВЕЛИКИХ ВІДЕОДАНИХ Каліта В.І., <i>Науковий керівник: Устенко С.В.</i> .....	331
ПОКРАЩЕННЯ ЕФЕКТИВНОСТІ РЕКЛАМНОГО КОНТЕНТУ ДЛЯ ВИЩИХ НАВЧАЛЬНИХ ЗАКЛАДІВ ЧЕРЕЗ ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ТЕКСТУ Курпела П., <i>Науковий керівник: Лип'яніна-Гончаренко Х.В.</i> .....	335
РЕАЛІЗАЦІЯ WEB-ЗАСТОСУНКУ ДЛЯ МУЗИЧНОГО СТРІМІНГУ Білоконь О.Ю., <i>Науковий керівник: Вельмагіна Н.О.</i> .....	337
РОЗРОБКА АВТОМАТИЗОВАНОЇ СИСТЕМИ МАГАЗИНУ МОБІЛЬНИХ ТЕЛЕФОНІВ Горбатко А.В., <i>Науковий керівник: Шибко О.М.</i> .....	339
СИСТЕМИ ЗАХИСТУ В ОПЕРАЦІЙНИХ СИСТЕМАХ MICROSOFT WINDOWS Ляшук Є.Б., <i>Науковий керівник: Деркач Т.М.</i> .....	342
ШТУЧНИЙ ІНТЕЛЕКТ: КОРИСТЬ ЧИ ШКОДА Довженко В.О., <i>Науковий керівник: Крисак Л.В.</i> .....	344
ШТУЧНИЙ ІНТЕЛЕКТ ТА АВТОНОМНІ ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ: СУЧАСНИЙ СТАН ТА МАЙБУТНІ ПЕРСПЕКТИВИ Лагода Ю.О. ....	346

**СЕКЦІЯ 19.****ФІЗИКО-МАТЕМАТИЧНІ НАУКИ**

ВИКОРИСТАННЯ ЕКОНОМІКО-МАТЕМАТИЧНИХ МЕТОДІВ У ПРОГНОЗУВАННІ ОБСЯГІВ ПРОДАЖ ПРОДУКЦІЇ ПІДПРИСМСТВА Ксьондз А.Ю., <i>Науковий керівник: Кулинич Р.О.</i> .....	348
СТІЙКІСТЬ РОЗВ'ЯЗКІВ ЛІНІЙНИХ СИСТЕМ ЗІ СТАЛИМИ КОЕФІЦІЄНТАМИ Моторко В.В., <i>Науковий керівник: Бичков О.С.</i> .....	351

**СЕКЦІЯ 20.****ФІЛОЛОГІЯ ТА ЖУРНАЛІСТИКА**

HISTORICAL PREREQUISITES FOR COMBATING OFFENSES COMMITTED BY MINORS Вівденко К., <i>Науковий керівник: Декусар Г.Г.</i> .....	354
БУДОВА НОМІНАЦІЙ ГАЛУЗІ ЦИВІЛЬНОГО ЗАХИСТУ В АНГЛІЙСЬКІЙ МОВІ Білан Д.І., <i>Науковий керівник: Цимбал Н.А.</i> .....	357
ВПЛИВ ХРОНОТОПУ НА ФОРМУВАННЯ СЮЖЕТУ ТРИЛОГІЇ УЛАСА САМЧУКА «ВОЛИНЬ» Хавалюк М.М., <i>Науковий керівник: Погребенник В.Ф.</i> .....	359
ДІСЛОВО УКРАЇНСЬКОЇ І АНГЛІЙСЬКОЇ МОВ У ПЛОЩИНІ ГРАМАТИЧНОЇ ПРАГМАТИКИ (НА МАТЕРІАЛІ ПУБЛІЦИСТИЧНОГО ДИСКУРСУ) Цінська Д.Р.....	362

**Курпела Павло**, здобувач вищої освіти факультету комп'ютерних інформаційних технологій  
Західноукраїнський національний університет, Україна

**Науковий керівник: Ліп'яніна-Гончаренко Христина Володимирівна**, канд. техн. наук, доцент, доцент кафедри інформаційно-обчислювальних систем та управління  
Західноукраїнський національний університет, Україна

## ПОКРАЩЕННЯ ЕФЕКТИВНОСТІ РЕКЛАМНОГО КОНТЕНТУ ДЛЯ ВИЩИХ НАВЧАЛЬНИХ ЗАКЛАДІВ ЧЕРЕЗ ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ТЕКСТУ

Західноукраїнський національний університет провів дві рекламні кампанії в соціальних мережах. У першій кампанії використовувався звичайний текстовий контент, у другій - контент, сформований на основі семантичного аналізу.

В таблиці 1 представлено порівняння ефективності сформованого рекламного контенту на основі LSA та LDA-методу. В період з Jan 4, 2021 - Jan 31, 2021 була проведена рекламна компанія з великим текстовим контентом. В період з Feb 1, 2021 - Feb 28, 2021 проведено рекламну кампанію з текстовим контентом, що розроблений на основі LSA та LDA-методу.

Таблиця 1.

### Порівняння ефективності сформованого рекламного контенту

Показник	Jan 4, 2021 – Jan 31, 2021	Feb 1, 2021 – Feb 28, 2021	Change
Coverage	6792	7350	558
Frequency	1,18	1,19	0,02
Results	9,00	13,00	4,00
Price for the result	1,42	0,98	-0,44
Amount of expenses (USD)	12,77	12,75	-0,02
CPM (Cost per 1000 views)	1,60	1,46	-0,14
Link clicks	46,00	43,00	-3,00
CPC (cost-per-click)	0,28	0,30	0,02
CTR (clickability)	0,58	0,49	-0,09

Для порівняння ефективності кампаній були використані такі показники:

- Охоплення - кількість осіб, які побачили рекламу.
- Частота показів - середня кількість разів, коли кожен користувач бачив рекламу.
- Результати - кількість переходів в месенджер.
- Ціна за результат - витрати на один результат.
- Середня ціна за 1000 показів (CPM) - витрати на 1000 показів.
- Кількість кліків по посиланням - кількість переходів на сайт компанії або в месенджер по посиланню в рекламі.
- Ціна за клік по посиланням (CPC) - витрати на один клік по посиланню.
- Частка кліків по посиланням (CTR) - частка людей, які побачили рекламу і клікнули по посиланню.

## Тренди та перспективи розвитку мультидисциплінарних досліджень

Результати порівняння ефективності кампаній представлені в таблиці 1.

Як видно з таблиці, в результаті другої кампанії:

- Охоплення збільшилося на 558 осіб.
- Середня частота показів на особу збільшилася на 0,02.
- Кількість результатів збільшилася на 4.
- Ціна за результат зменшилася на 0,44 цента.
- Середня ціна за 1000 показів зменшилася на 0,14 цента.
- Кількість кліків по посиланням зменшилася на 3.
- Ціна за клік по посиланням збільшилася на 0,02 цента.
- Частка кліків по посиланням зменшилася на 0,09.

З цих результатів можна зробити важливий висновок, що результативність оголошення збільшилася на 44%, а ціна за результат зменшилася на 31%.

На відміну від аналогів [1, 2] розроблений інтелектуальний метод формування рекламного контенту закладів вищої освіти на основі семантичного аналізу, дозволить виділити ключові слова на основі, яких можна сформувавши рекламний текст оголошення та результативність оголошення збільшилась на 44% та при цьому ціна за результат зменшилась на 31%.

### Список використаних джерел:

1. García-Sánchez, Francisco; Colomo-Palacios, Ricardo; Valencia-García, Rafael (2020). A social-semantic recommender system for advertisements. *Information Processing & Management*, Elsevier, 57(2), 102153. doi:10.1016/j.ipm.2019.102153.
2. Xu, Minghua; Liu, Shenghao (2019). Semantic-Enhanced and Context-Aware Hybrid Collaborative Filtering for Event Recommendation in Event-Based Social Networks. *IEEE Access*, 7, 17493–17502. doi:10.1109/ACCESS.2019.2895824.
3. Gramyak, R., Lipyana-Goncharenko, H., Sachenko, A., Lendyuk, T., & Zahorodnia, D. (2021). Intelligent Method of a Competitive Product Choosing based on the Emotional Feedbacks Coloring. In *IntellITSIS* (pp. 246-257).
4. Dulishkovych, O., Lipyana, H., Sachenko, O., Lendyuk, T., Osolinskiy, O., & Kit, I. (2021). Business Process Management of Virtual Enterprise based on Augmented Reality. In *ITPM* (pp. 422-431).