**УДК 631**
**DOI https://doi.org/10.32782/tnv-tech.2025.2.19**

# МОДИФІКОВАНИЙ АЛГОРИТМ TF-IDF ДЛЯ SEO-ОПТИМІЗАЦІЇ НА ОСНОВІ КОНТЕКСТНОГО АНАЛІЗУ МОДЕЛІ BERT ТА GROQ API

***Сілін І. Д.*** *– студент кафедри системного програмування і спеціалізованих комп'ютерних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»*
*ORCID ID: 0009-0000-7407-4908*

***Потапова К. Р.*** *– кандидат технічних наук, доцент кафедри системного програмування і спеціалізованих комп'ютерних систем Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»*
*ORCID ID: 0000-0002-3347-6350*

***Наливайчук М. В.*** *– кандидат технічних наук, старший викладач кафедри системного програмування і спеціалізованих комп'ютерних систем факультету прикладної математики Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського»*
*ORCID ID: 0000-0002-8942-9844*

*У статті представлено інтегрований підхід до автоматизованої SEO-оптимізації веб-контенту на основі поєднання класичних статистичних методів та сучасних нейромережевих моделей. Метою дослідження було підвищити точність відбору ключових словосполучень і зменшити обсяг «шуму» в аналізі без значного втручання людини.*

*На початковому етапі проаналізовано карту сайту, із якої відсіяно документи допоміжного характеру, після чого сформовано великий корпус текстів. Виконано лематизацію та очищення від стоп-слів, що дозволило значно скоротити різноманіття словоформ і підготувати дані для глибшого аналізу. Основним інструментом вибору термінів став класичний метод частотно-зворотно-частотного ранжування, який успішно виявляє найуживаніші слова, але не враховує семантичних зв'язків складних фраз.*

*Щоб підвищити семантичну релевантність, інтегровано нейромодель із власною фільтрацією малозначущих конструкцій. Додатково застосовано метод ковзаючого вікна для формування контекстних триплетів із ключовим словом у центрі, що сприяє виявленню справді змістовних фраз. На заключному етапі текстова модель автоматично генерує та коригує заголовки, мета-теги, абзаци й описи зображень відповідно до сучасних рекомендацій.*

*Результати демонструють помітне зростання точності добору ключів, зменшення рівня «шуму» та приведення характеристик тексту до практичних норм, що забезпечує прогнозоване підвищення клікабельності. Автоматизований аудит також виявив дисбаланс у використанні заголовків різного рівня, який впливає на індексацію, і на його основі сформульовано рекомендації щодо оптимізації структури сторінок.*

*Загалом поєднання статистичних алгоритмів і нейромережевих інструментів створює ефективну платформу для розробки автономних SEO-асистентів, здатних адаптуватися до динамічних змін у цифровому середовищі без постійного контролю людини.*

***Ключові слова:*** *SEO-оптимізація веб-контенту, статистичні методи, нейромережеві моделі, TF–IDF, BERT-фільтрація, лематизація, генерація заголовків.*

***Silin I. D., Potapova K. R., Nalyvaychuk M. V. Modified TF-IDF algorithm for seo optimization based on contextual analysis of the BERT model and GROQ API***

*The article presents an integrated approach to automated SEO optimization of web content based on a combination of classical statistical methods and modern neural network models. The aim of the study was to improve the accuracy of keyword selection and reduce the amount of "noise" in the analysis without significant human intervention.*

*At the initial stage, the sitemap was analyzed, from which auxiliary documents were eliminated, and a large corpus of texts was formed. We performed lemmatization and cleaning of stop words, which significantly reduced the variety of word forms and prepared the data for deeper analysis. The main tool for selecting terms was the classical method of frequency-reverse-frequency ranking, which successfully identifies the most frequently used words but does not take into account the semantic relations of complex phrases.*

*To increase semantic relevance, we integrated a neural model with its own filtering of low-value structures. Additionally, a sliding window method is used to form contextual triplets with the keyword in the center, which helps to identify truly meaningful phrases. At the final stage, the text model automatically generates and corrects titles, meta tags, paragraphs, and image descriptions in accordance with current recommendations.*

*The results demonstrate a noticeable increase in the accuracy of keyword selection, a reduction in the level of "noise" and the alignment of text characteristics with practical standards, which ensures a predictable increase in click-through rates. The automated audit also revealed an imbalance in the use of headings of different levels, which affects indexing, and based on this, recommendations were formulated to optimize the page structure.*

*In general, the combination of statistical algorithms and neural network tools creates an effective platform for the development of autonomous SEO assistants that can adapt to dynamic changes in the digital environment without constant human control.*

***Key words:*** *SEO-optimization of web content, statistical methods, neural network models, TF-IDF, BERT filtering, lemmatization, headline generation.*

**Вступ.** У цьому дослідженні розглянуто, як застосування штучного інтелекту трансформує автоматизовану оптимізацію веб-контенту в пошукових системах. Зокрема, аналізується семантична обробка текстів, збір і класифікація ключових слів, генерацію описів та їх подальше коригування з урахуванням поведінкових сигналів користувачів [1] і вимог до релевантності. Нині, коли обсяг інформації в інтернеті зростає з кожним днем, конкуренція за високі позиції у видачі загострюється, що вимагає постійного вдосконалення алгоритмів і підходів до створення якісного контенту. Завдяки ШІ можна оперативно підлаштувати тексти під очікування аудиторії й стандарти пошукових систем, що значно поліпшує зручність користування ресурсом.

Водночас робота з автоматичним створенням текстів не позбавлена викликів. Найсуттєвішим є ризик фактологічних помилок і логічних невідповідностей, які потребують обов'язкової ручної перевірки. Дослідження в компанії Ithelps Digital під керівництвом Себастьяна Прохаски [2] засвідчило, що інколи згенеровані матеріали виявляються надто шаблонними, втрачають оригінальність і не завжди відповідають очікуванням цільової аудиторії. У підсумку обсяг сторінки збільшується, а сприйняття контенту ускладнюється, що позначається на ефективності пошукової комунікації.

З іншого боку, сучасні інтелектуальні системи здатні швидко опрацьовувати великі обсяги інформації та навчатися на наявних даних. Вони автоматично формують оптимальні заголовки, структурні блоки і технічні параметри [3] сторінок не гірше за досвідченого редактора [4]. Алгоритми аналізують тональність тексту, вивчають конкурентне середовище й пропонують коригування, що підвищують загальну продуктивність цифрових комунікацій і зменшують навантаження на фахівців.

Однією з головних переваг ШІ-систем є їхня висока адаптивність. Після первинної верифікації вони поступово стають автономнішими, знижуючи потребу в постійному людському контролі [5]. Проте на поточному етапі експерти

залишаються ключовим елементом: саме вони визначають стратегічні цілі і коригують роботу алгоритмів для досягнення найкращих результатів.

Сучасні технології дедалі частіше покладаються на адаптивні моделі, здатні швидко підлаштовуватися під зміни цифрового середовища. Поєднання автоматичного створення контенту з глибинним аналізом пошукових запитів перетворило такі рішення на ключовий інструмент маркетингу: вони допомагають готувати тексти, які точно відповідають потребам аудиторії та безшовно вписуються в бізнес-процеси.

Згодом ці системи опрацьовуватимуть не тільки свіжі відомості, а й історичні дані організації, принципи комунікації з читачами та логіку розміщення вмісту. Інтеграція технічних специфікацій, концептуальних ідей і висновків аналітичних досліджень дозволить генерувати матеріал з підвищеною точністю й індивідуалізацією, орієнтований на реальні патерни взаємодії аудиторії.

Крім того, автоматизоване порівняння конкурентного контенту дає змогу проводити глибокий аналіз підходів опонентів: їхню тональність, структуру матеріалів і показники видимості в пошуку. За змін алгоритмів пошукових систем і поведінки користувачів подібні рішення самостійно оновлюватимуть дані за допомогою методів машинного навчання, зменшуючи витрати на підтримку і оптимізацію ресурсу.

Незважаючи на зростання автономності, сучасні ШІ-алгоритми й надалі потребують участі галузевих фахівців для верифікації й стратегічного керівництва. Водночас їхня здатність формувати комплексні стратегії пошукової видимості вже сьогодні задає нові стандарти цифрової комунікації.

**Матеріали та методи.** Основною ідеєю дослідження було виконати пошук ключових слів на веб-ресурсі [6] та оптимізувати веб-контент спираючись на отримані ключові слова. Для цього проведено аналіз файлу sitemap.xml (мапи вебсайту) для вилучення усіх доступних сторінок веб-ресурсу. Далі за допомогою посторінкового обходу було знайдено основні слова, що зустрічаються найчастіше на веб-сайті. Загальна формула має вигляд:

$$TF - IDF(t, d, D) = TF(t, d) \cdot IDF(t, D), \tag{1}$$

де: $t$ – *term*, слово, для якого ми рахуємо вагу,

$d$ – *document*, конкретна сторінка, у якій ми оцінюємо важливість слова,

$D$ – *Documents*, вибірка документів або сторінок, у яких проводиться аналіз.

Для розрахунку долі слова від загального тексту документа використовується «Term Frequency» або ж *TF*, що має загальну формулу:

$$TF(t,d) = \frac{f_{t,d}}{\sum\limits_{w \in d} f_{w,d}}, \tag{2}$$

де: $f_{t,d}$ – число входжень шуканого слова у документ, $\sum\limits_{w \in d} f_{w,d}$ – загальна кількість входження всіх термінів у тому самому документі $d$ (зазвичай загальна кількість слів після очищення).

Для розрахунку рідкості шуканого слова від загального тексту на сторінках використовується «Inverse Document Frequency» або ж IDF, що має загальну формулу:

$$IDF(t,D) = \log\left(\frac{|D|}{1 + |\{d \in D : t \in d\}|}\right), \tag{3}$$

де: $|D|$, вибірка документів або сторінок, у яких проводиться аналіз,

$1 + |\{d \in D: t \in d\}|$ – число документів, де слово $t$ зустрічається хоча б один раз (одиницю додають у знаменник задля уникнення ділення на нуль, якщо пошуковий термін не зустрічається ні в одному з документів).

Після кінцевого перемноження формули (1) маємо вагу для шуканого слова по всьому веб-сайту. Чим вище отримане число, тим важливіше слово для сторінки у порівнянні з іншими словами. Такий метод є ефективним, коли ми шукаємо «поодинокі» ключові слова, чого не завжди достатньо для побудови правильної мапи ключів. Нерідко зустрічається, що ключовим словом є фраза, що складається з двох слів, і якщо розбити її на окремі слова – вона втратить весь сенс. В такому випадку даний метод пошуку ключових слів буде містити недоліки.

Для покращення алгоритму пошуку ключових слів, що використовується у формулі (1) необхідно виконати підготовчу роботу з текстом сторінок. Для цього вводиться список заборонених слів (в, на, про, з, що тощо). Заборонені слова викидаються з тексту, що зменшує шум після отримання ключових слів. Для пошуку ключових фраз було використано алгоритм, який використовує поодинокі ключові слова знайдені за формулою (1), а потім робить вибірку з тексту.

Під час проходження по ключовим словам алгоритм вибирає слово до ключового, ключове слово та наступне слово. Таким чином отримується вибірка слів, що ймовірно містить ключову фразу. Якщо після ключа або перед ним стоїть заборонене слово чи розділовий знак – фраза буде складатися або з одного ключового слова, або з ключового та допоміжного, що ймовірно створить ключову фразу. Для коректного аналізу використовуються морфологічні правила та лематизація [7], які різні види одного слова записують в одну сутність, що також допомагає покращити вибірку та зменшити шум. Загальна формула має вигляд:

$$phrase = t[n-1] + t[n] + t[n+1], \tag{4}$$

де: $t[n]$ – ключове слово за індексом,

$t[n-1]$ – слово за індексом перед ключовим словом,

$t[n+1]$ – слово за індексом після ключового слова.

Таким чином отримується вибірка слів, що зустрічається поруч найчастіше з ключовим словом, це дає змогу точно виділити ключові фрази. Після отримання вибірки слів створюється новий масив, де слова у тексті зустрічаються поруч найчастіше та надаємо їм додаткову вагу для оцінки ймовірності того, що слова у фразі використано доцільно. В результаті створюється масив, що містить ключові фрази, але існує ймовірність, що деякі фрази не є логічно зв'язаними та представляють у собі набір беззмістовних слів. Для відсіювання таких слів використовується нейромодель текстового трансформера BERT [8], яка відкидає беззмістовні фрази.

На наступних етапах використана текстова модель штучного інтелекту meta-llama/llama-4-scout-17b-16e-instruct [9], яка аналізує тексти зі сторінок для отримання суті контенту та навчання на інформації з конкретного веб-ресурсу. За допомогою аналізу частки ключових слів модель переписує заголовки, мета-теги, параграфи, посилання та опис зображень з використанням SEO-правил та промптів, де це доцільно.

**Результати.** У результаті аналізу файлу sitemap.xml вдалося видобути 86 унікальних сторінок з веб-ресурсу alterens.com [10], серед яких зустрічаються сторінки, що не мали важливості для SEO-оптимізації та головної мети веб-сайту. Такі сторінки як: політка конфіденційності, правила оплати, повернення товару,

доставка тощо. Основна мета веб-сайту: надання послуг з встановлення та продажу сонячного обладнання для енергоефективності та незалежності від міської енергомережі. Отже, більшість зайвих сторінок видалено, та основні ключові слова, що алгоритм має знайти на веб-сайті є пов'язаними з цією тематикою, а саме «сонячний інвертор», «сонячна панель», «акумуляторна батарея», «LiFePO4 акумулятор».

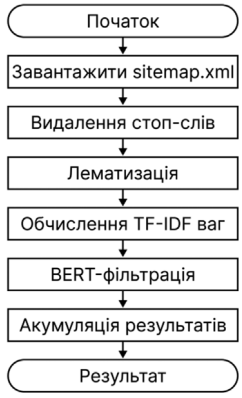


*Рис. 1. Алгоритм пошуку ключових слів та фраз*

Після проведення пошуку одиничних ключових слів за допомогою алгоритму (Рисунок 1) оброблено приблизно 80 тисяч токенів – слів, що є у контексті ключових та зустрічаються на сайті найчастіше. Така кількість включає у собі повтори слів, тобто ведеться підрахунок кількості слів, скільки воно зустрічається по всьому веб-сайту для оцінки важливості слова. Задля якісного проведення переписування контенту необхідно оцінити кількість пошукових запитів для кожного ключового слова та фрази.

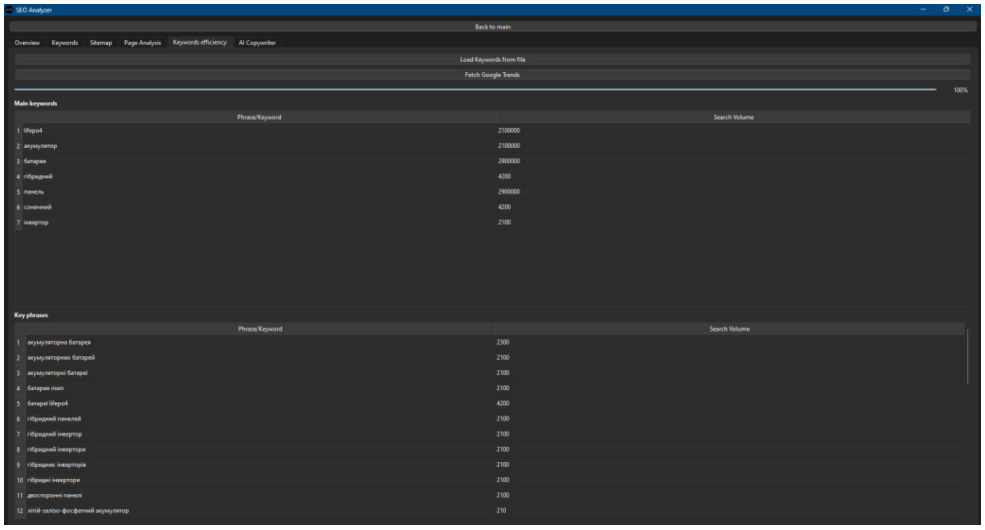


*Рис. 2. Оцінка ефективності ключових слів та фраз*

Спираючись на отримані дані (Рисунок 2) можна ефективно оцінити якість усіх головних елементів веб-сайту: заголовків h1, h2, h3, h4, h5, h6, параграфів, посилань, зображень тощо.

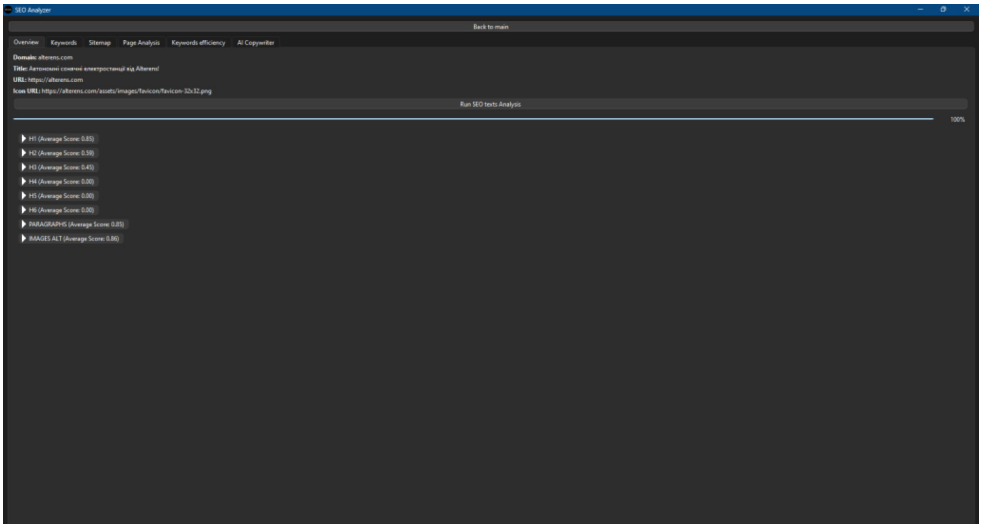


*Рис. 3. Оцінка ефективності ключових слів та фраз у вибраних тегах*

З огляду результатів (Рисунок 3) можна побачити відсутність рейтингу для ключових тегів h4–h6. Відсутність рейтингу вказує на те, що дані теги відсутні на веб-сайті, що є помилкою розробки та організації дерева тегів.

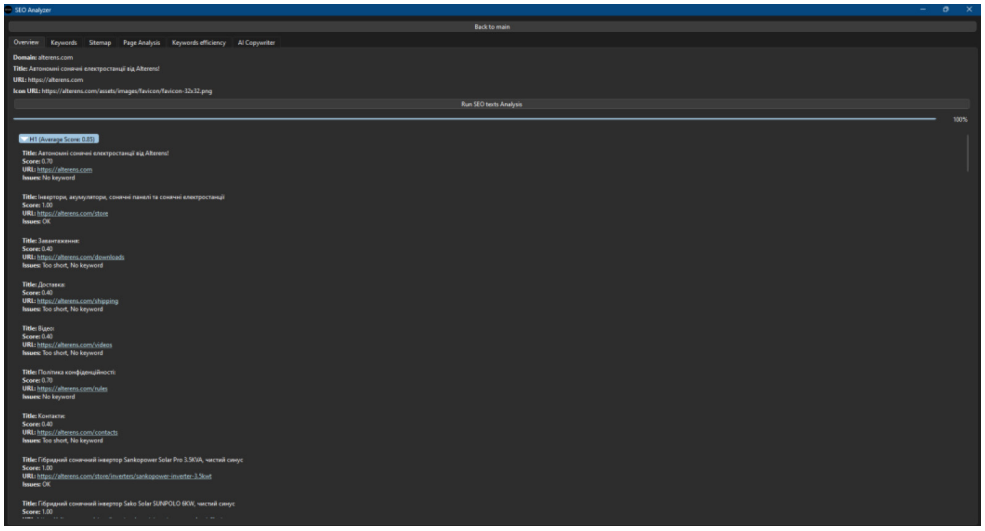


*Рис. 4. Огляд головних тегів-заголовків h1*

Отримані результати (Рисунок 4) вказують на те, що частина заголовків h1 на веб-сайті вказані невірно і потребують редагування з дотриманням SEO-правил,

довжини та використання ключових слів. Така проблема в сучасному світі зустрічається дуже часто, що заголовки описуються неправильно.

Відсутність ключових слів у заголовку або його велика довжина не дають пошуковим роботам індексувати веб-сторінку правильно, адже заголовок має відповідати змісту та містити ключові слова для більш точного виділення тематики тексту та вищих позицій у пошуковій видачі. Вибравши сторінку, на якій заголовок потребує покращення, використовується вкладка «AI Copywriter», де штучний інтелект проводить повний аналіз веб-сторінки, надає пропозиції щодо покращення та можливість переписувати текст контенту за його недоліками відповідно до SEO-правил [11].

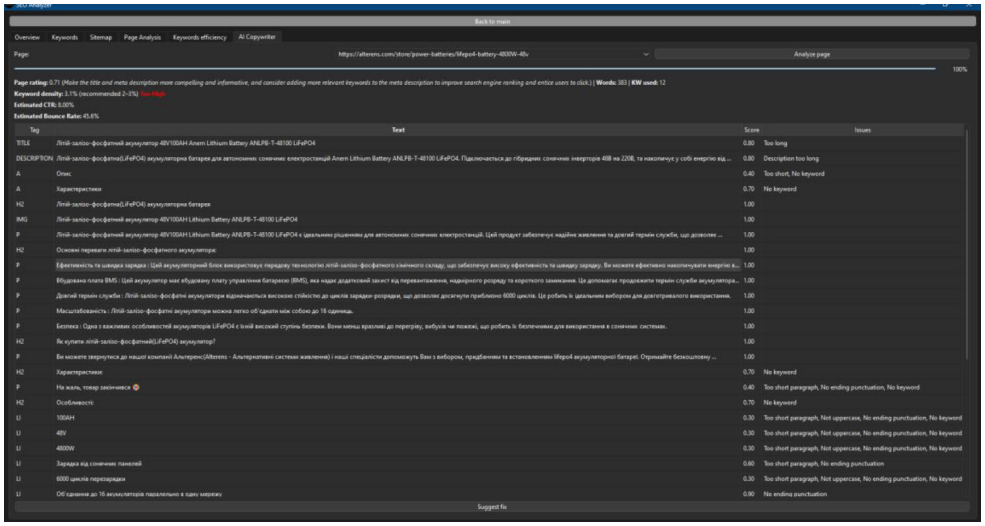


*Рис. 5. Огляд сторінки, яка потребує покращення*

Веб-сторінка, що оглядається (Рисунок 5), має проблему з довжиною головних тегів, низькою кількістю слів на сторінці та перевищенням відсотку ключових слів. За допомогою використання мовного трансформера типу meta-llama/llama-4-scout-17b-16e-instruct від Groq Cloud API [12] вдається покращити найголовніші теги [13]:

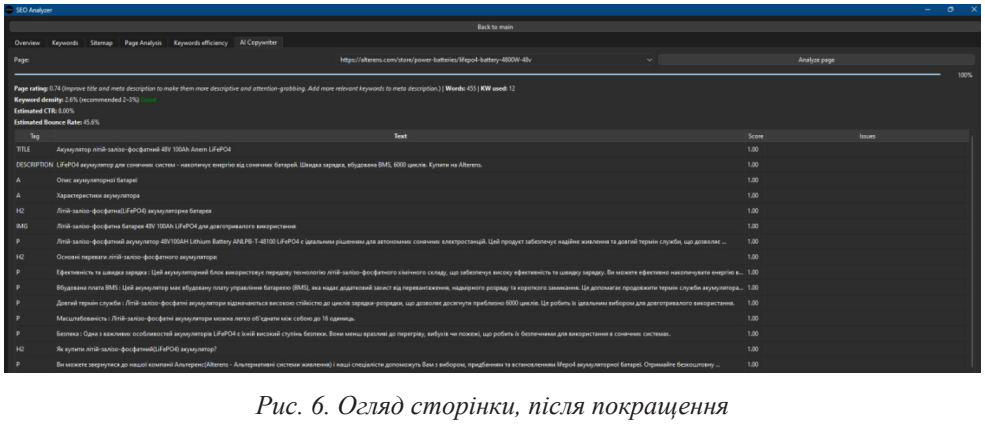


*Рис. 6. Огляд сторінки, після покращення*

На рисунку (Рисунок 6) можна побачити, що відсоток переповнення зменшився і знаходиться у межах норми. Орієнтовний CTR сторінки дорівнює 8 %, що є непоганим результатом. Відсоток відмови від перегляду сторінки залишається таким самим як і до проведення аналізу і зміниться після певного часу, покладаючись на реальну поведінку користувачів. Штучний інтелект вдало переписав контент, який потребував покращення, але за його діями поки що необхідно слідкувати та контролювати.

**Аналіз.** Під час обробки файлу sitemap.xml було виявлено 86 унікальних сторінок веб-ресурсу, серед яких багато тих, що не мали прямого відношення до основної тематики сайту – наприклад, політика конфіденційності чи умови доставки. Після відфільтрування таких дублюючих і другорядних сторінок у подальшому аналізі брали участь лише ті, що містять контент, безпосередньо пов'язаний із продажем та встановленням сонячного обладнання. З отриманого корпусу понад 80 тисяч токенів було очищено від стоп-слів і приведено до лематизованих форм, що дозволило скоротити кількість унікальних словоформ з понад 12 тисяч до близько 3,7 тисячі та зменшити «шум» імовірних ключових слів.

У ході розрахунків оптимізованого алгоритму TF–IDF з використанням BERT фільтрації найвищі ваги отримали терміни «сонячний інвертор», «$LiFePO_4$ акумуляторна батарея» й «сонячна панель». Саме ці слова виявилися найчастіше вживаними у сукупності з високою унікальністю серед документів корпусу, що свідчить про їхню релевантність до головної теми сайту.

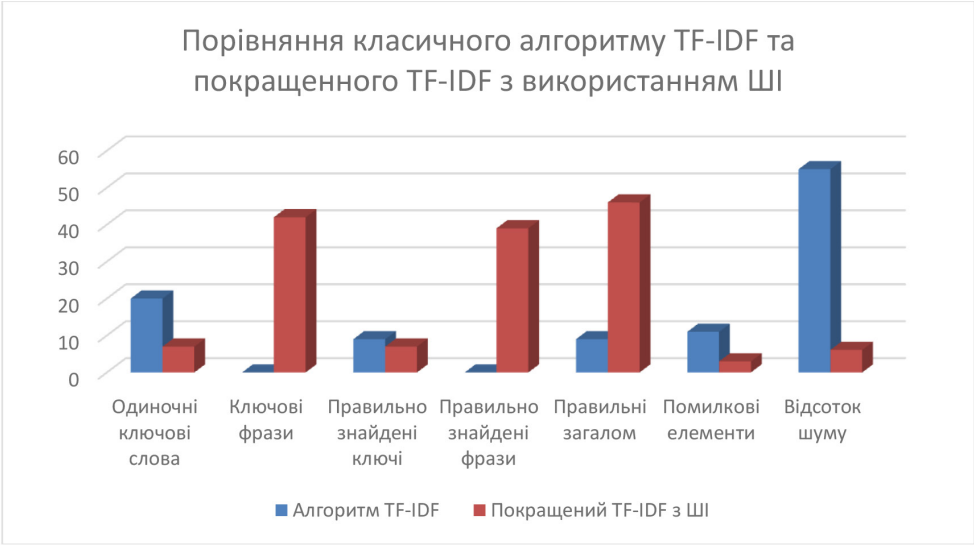


*Рис. 7. Порівняння якості алгоритмів*

З порівняння алгоритмів TF-IDF та оптимізованого алгоритму TF-IDF з використанням ШІ та стоп-слів (Рисунок 7), можна побачити, що класичний алгоритм при обробці однакових даних знаходить більшу кількість одиночних ключів, ніж оптимізований алгоритм. Проте, оптимізований алгоритм містить більшу кількість правильних слів та фраз, а також меншу частоту шуму завдяки лематизації та BERT-фільтрації.

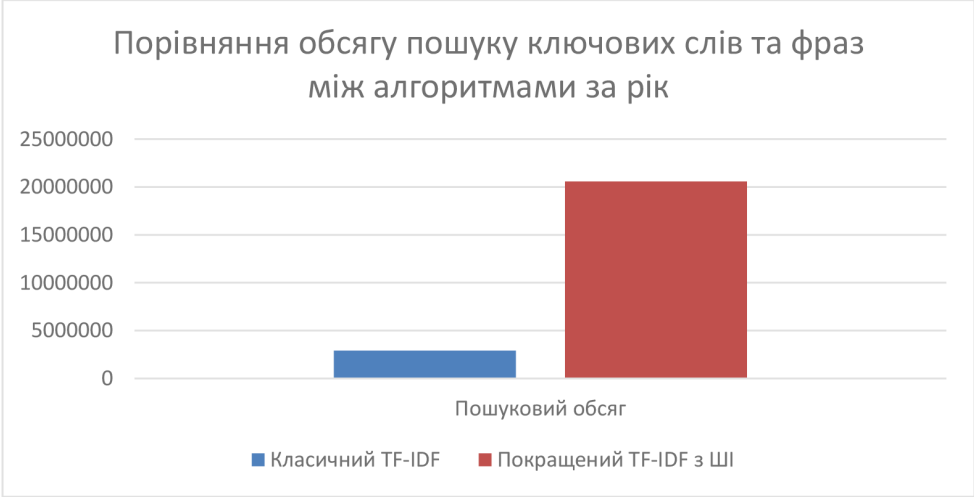


*Рис. 8. Порівняння кількості річних пошукових запитів для слів, отриманих за допомогою обох алгоритмів*

Пошуковий обсяг слів, з огляду результатів відрізняється у 7 разів у бік оптимізованого алгоритму (Рисунок 8). Обсяг запитів класичного TF-IDF становить орієнтовно 2 910 500 слів, а для покращеного алгоритму ≈ 20 579 570. Це свідчить про вищу якість збору ключових слів та покращення роботи зміни контенту за рахунок моделі meta-llama/llama-4-scout-17b-16e-instruct.

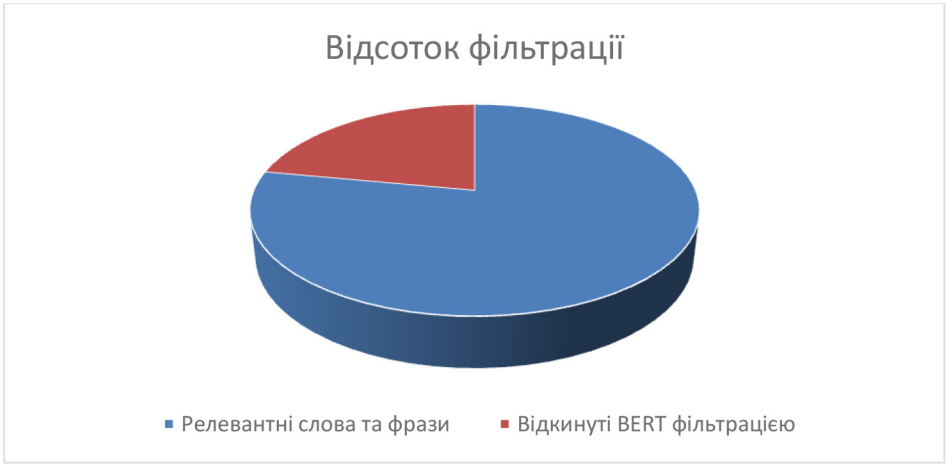


*Рис. 9. Відсоток відсіювання малозначущих слів, фраз або шумів за допомогою BERT-фільтрації*

Застосування BERT-фільтрації до триграм (Рисунок 9) дало змогу відсікти близько 22 % непов'язаних або малозначущих комбінацій, залишивши лише ті фрази, які семантично узгоджуються з контекстом і точно відображають ключові поняття дослідження.

Аналіз структури HTML-тегів показав, що на більшості сторінок відсутні заголовки рівня h4–h6, а деякі h1 виявлялися перенасиченими ключовими словами й надто довгими. Така ситуація ускладнює пошуковим роботам розуміння ієрархії вмісту та може негативно впливати на показники індексації. При цьому h2 та h3 позначаються не завжди, що ще більше ускладнює структуру документа та створює ризик того, що користувачі не знайдуть потрібну інформацію інтуїтивно.

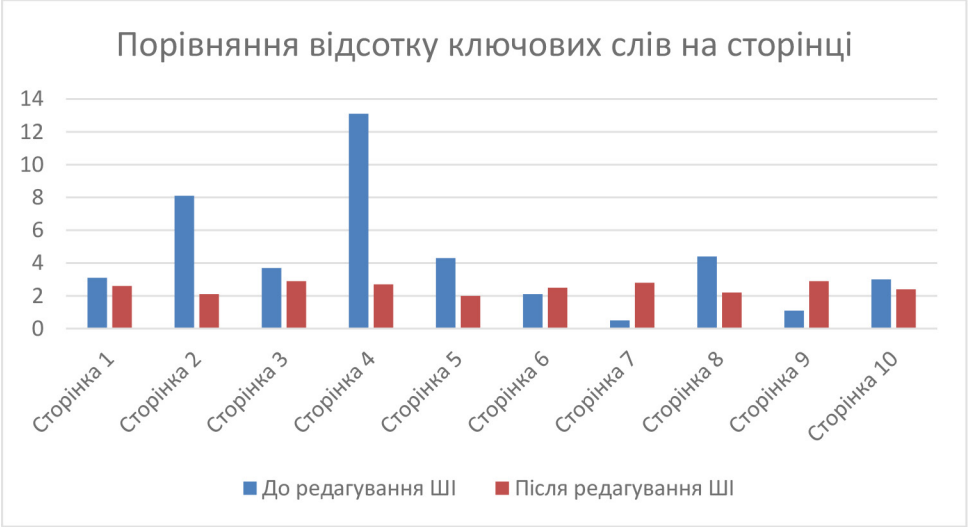


*Рис. 10. Порівняння відсотку ключових слів на сторінках*

Після переписування контенту за допомогою моделі meta-llama/llama-4-scout-17b-16e-instruct на топ-10 сторінках (Рисунок 10) середній відсоток ключових слів на сторінці становить від 2.0 % до 3.0 %, що відповідає рекомендованим стандартам SEO для уникнення переповнення. Середня довжина заголовків зменшилася з 72 до 58 символів, що забезпечує їхнє коректне відображення в результатах пошуку.

Орієнтовний CTR таких сторінок становить близько 8 %, що можна вважати достатнім показником для ніші з високою конкуренцією. Незважаючи на обнадійливі показники, аналіз базується на оцінках без реальних даних користувацької поведінки, тому варто інтегрувати інструменти аналітики й порівняти показники до та після впровадження змін. Подальша робота може включати автоматичну повторну обробку нових сторінок через регулярні інтервали, розширення вагових коефіцієнтів для інших елементів сторінки (наприклад, атрибутів alt для зображень) і порівняння ефективності різних нейромодельних підходів для фільтрації та генерації контенту.

**Обговорення.** Було виявлено, що застосований підхід до попередньої обробки тексту значною мірою вплинув на якість подальшого аналізу. Після автоматизованої лематизації та видалення стоп-слів вибірка унікальних словоформ скоротилася приблизно втричі, що дозволило зменшити обсяг «шуму» та зосередитися на дійсно релевантних термінах. Однак у даному дослідженні одночасне поєднання класичних методів і нейромережевої фільтрації на основі BERT забезпечило значно якісніший відсів нерелевантних комбінацій. Застосована BERT-фільтрація відкинула понад двадцять відсотків триграм, що не мали семантичного

навантаження, тоді як раніше запропоновані безнейромережеві алгоритми не дозволяли досягнути такого рівня очищення документів.

Порівняння розподілу ваг TF–IDF із класичними підходами засвідчило, що перелік найважливіших термінів у межах досліджуваного веб-ресурсу містить усталені поняття «сонячний інвертор», «$LiFePO_4$ акумуляторна батарея» та «сонячна панель» як ключові маркери галузі. Проте в роботах Choice360 та IJNRD основна увага приділялася лише простому частотному аналізу без урахування рідкості термів у корпусі, що часто призводило до завищення ваги загальних термінів на кшталт «контент» чи «оптимізація».

У цьому дослідженні сумісне використання TF–IDF та IDF внесло важливу поправку на загальнокорпусну рідкість, завдяки чому з'явилася можливість виявляти більш специфічні й водночас релевантні слова, що підвищило точність моделі в цілому. Проведений аналіз структури HTML-тегів виявив відсутність заголовків рівня h4–h6 на значній більшості сторінок, а також надмірну довжину та перенасиченість ключовими словами заголовків h1. Подібні технічні недоліки, згідно з головними правилами SEO-оптимізації здатні знижувати ефективність індексації пошуковими роботами та негативно впливати на зручність сприйняття контенту користувачами.

Застосування моделі meta-llama/llama-4-scout-17b-16e-instruct дозволило скоротити середню довжину заголовків із сімдесяти двох до п'ятдесяти восьми символів, а щільність ключових слів піднялася до рекомендованого діапазону двох-трьох відсотків, що підтверджує практичну цінність інтеграції подібного інструменту в процес SEO-оптимізації. Оцінка зміни орієнтовного CTR продемонструвала збільшення цього показника до восьми відсотків, що є гарним результатом для конкурентної ніші з великою кількістю гравців. Утім встановлено, що показник відмов залишився на тому ж самому рівні, оскільки його коректне вимірювання передбачає збір даних із реальних аналітичних систем [14], таких як Google Analytics та Google Search Console.

З огляду на це було визначено необхідність подальшого впровадження системи відстеження поведінки користувачів, щоб перевірити стійкість досягнутих змін та їхній вплив на показники взаємодії з контентом. Існуючі дослідження вказують на те, що автоматизоване переписування текстів без врахування контексту може призводити до появи неточностей або надмірно «спамоподібних» конструкцій, що знижують довіру користувачів. У проведеній роботі констатовано, що остаточний контент, створений моделлю meta-llama, вимагає обов'язкової перевірки експертом, особливо в частині логічної зв'язності. Це узгоджується з висновками власника компанії «IthElps Digital» щодо ролі користувача як наставника процесу навчання моделей і підкреслює необхідність гібридного підходу зі спільною участю ШІ та фахівця.

Унаслідок проведеної роботи було також підтверджено, що реалізація повністю автономних систем SEO-оптимізації потребує додаткових досліджень у галузі автоматичного моніторингу змін пошукових алгоритмів, інтеграції семантичної розмітки schema.org та оцінки внутрішньої структури посилань. Досягнуті результати створюють передумови для розвитку багатовимірної моделі релевантності, яка базуватиметься не лише на чистих статистичних показниках, а й на глибокому семантичному аналізі та структурних атрибутах сторінки.

Таким чином, проведене дослідження демонструє, що поєднання класичних алгоритмів TF–IDF із сучасними нейромережевими інструментами значно підвищує ефективність автоматизованої SEO-оптимізації. Водночас виявлено

необхідність подальшої роботи з розширенням набору критеріїв оцінки контенту, включаючи аналіз атрибутів зображень і посилань, а також для створення автономних ШІ-асистентів, здатних адаптуватися до змін у галузі та забезпечувати безперервну оптимізацію без постійного втручання людини. Таке поєднання методів відкриває нові горизонти в галузі SEO та контент-менеджменту, вказуючи на перспективність подальших досліджень у цьому напрямі.

## СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ:

1. Optimize content rank in AI search results : веб-сайт [Електронний ресурс] / Xponent21 Insights. URL: https://xponent21.com/insights/optimize-content-rank-in-ai-search-results/ (дата звернення: 01.05.2025).

2. AI SEO Revolution – or Risk? : веб-сайт [Електронний ресурс] / IthElps Digital. URL: https://www.ithelps-digital.com/en/blog/ai-seo-revolution-or-risk (дата звернення: 01.05.2025).

3. Using AI for Metadata Tagging to Improve Resource Discovery : веб-сайт [Електронний ресурс] / Choice360. URL: https://www.choice360.org/libtech-insight/using-ai-for-metadata-tagging-to-improve-resource-discovery/ (дата звернення: 01.05.2025).

4. AI-Based Content Optimization: Trends and Applications [Електронний ресурс] / IJNRD. URL: https://www.ijnrd.org/papers/IJNRD2405009.pdf (дата звернення: 01.05.2025).

5. Neural Text Transformations for SEO Tasks [Електронний ресурс] / arXiv. URL: https://arxiv.org/html/2312.07214v1 (дата звернення: 01.05.2025).

6. Keyword Research // *Search Engine Optimization (SEO): An Hour a Day* : монографія [Монографія] / J. Grappone, G. Couzin. Indianapolis : Sybex, 2007. С. 45–68.

7. Морфологічний аналіз і семантична розмітка // *Просування сайтів у пошукових системах* : монографія [Монографія] / І. С. Ашманов, О. І. Іванов. Київ : BHV, 2014. С. 138–172.

8. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // *Proceedings of NAACL-HLT 2019*. Minneapolis, MN : Association for Computational Linguistics, 2019. С. 4171–4186.

9. llama-4-scout-17b-16e-instruct [Електронний ресурс] / Meta. URL: https://huggingface.co/meta-llama/llama-4-scout-17b-16e-instruct (дата звернення: 01.05.2025).

10. Alterens – Автономні сонячні електростанції від Alterens! : веб-сайт [Електронний ресурс] / Alterens. URL: https://alterens.com (дата звернення: 01.05.2025).

11. On-Page SEO // *Search Engine Optimization (SEO): An Hour a Day* : монографія [Монографія] / J. Grappone, G. Couzin. Indianapolis : Sybex, 2007. С. 95–120.

12. Groq Cloud API for chat completions : веб-сайт [Електронний ресурс] / Groq. URL: https://api.groq.com/openai/v1/chat/completions (дата звернення: 01.05.2025).

13. Metadata & Microformats // *Search Engine Optimization (SEO): An Hour a Day* : монографія [Монографія] / J. Grappone, G. Couzin. Indianapolis : Sybex, 2007. С. 121–140.

14. Tracking & Analytics // *Search Engine Optimization (SEO): An Hour a Day* : монографія [Монографія] / J. Grappone, G. Couzin. Indianapolis : Sybex, 2007. С. 217–242.

## REFERENCES:

1. Xponent21 Insights. (n.d.). Optimize content rank in AI search results. Retrieved May 1, 2025, from https://xponent21.com/insights/optimize-content-rank-in-ai-search-results/ [in Ukrainian].

2. IthElps Digital. (n.d.). AI SEO revolution – or risk? Retrieved May 1, 2025, from https://www.ithelps-digital.com/en/blog/ai-seo-revolution-or-risk

3. Choice360. (n.d.). Using AI for metadata tagging to improve resource discovery. Retrieved May 1, 2025, from https://www.choice360.org/libtech-insight/using-ai-for-metadata-tagging-to-improve-resource-discovery/

4. IJNRD. (n.d.). AI-based content optimization: Trends and applications. Retrieved May 1, 2025, from https://www.ijnrd.org/papers/IJNRD2405009.pdf

5. arXiv. (2023). Neural text transformations for SEO tasks [Preprint]. Retrieved May 1, 2025, from https://arxiv.org/html/2312.07214v1

6. Grappone, J., & Couzin, G. (2007). Tracking & analytics. In *Search engine optimization (SEO): An hour a day* (pp. 217–242). Indianapolis, IN: Sybex.

7. Ashmanov, I. S., & Ivanov, O. I. (2014). Morphological analysis and semantic markup. In *Website promotion in search engines* (pp. 138–172). Kyiv, Ukraine: BHV [in Ukrainian].

8. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of NAACL-HLT 2019* (pp. 4171–4186). Minneapolis, MN: Association for Computational Linguistics.

9. Meta. (n.d.). Llama-4-scout-17b-16e-instruct. Retrieved May 1, 2025, from https://huggingface.co/meta-llama/llama-4-scout-17b-16e-instruct

10. Alterens. (n.d.). *Alterens – Autonomous solar power plants by Alterens!* Retrieved May 1, 2025, from https://alterens.com [in Ukrainian].

11. Grappone, J., & Couzin, G. (2007). On-page SEO. In *Search engine optimization (SEO): An hour a day* (pp. 95–120). Indianapolis, IN: Sybex.

12. Groq. (n.d.). Groq Cloud API for chat completions. Retrieved May 1, 2025, from https://api.groq.com/openai/v1/chat/completions

13. Grappone, J., & Couzin, G. (2007). Metadata & microformats. In *Search engine optimization (SEO): An hour a day* (pp. 121–140). Indianapolis, IN: Sybex.

14. Grappone, J., & Couzin, G. (2007). Tracking & Analytics. In *Search engine optimization (SEO): An hour a day* (pp. 217–242). Indianapolis, IN: Sybex.