

ML АЛГОРИТМИ ДЛЯ SEO СТРАТЕГІЇ: КЛАСТЕРИЗАЦІЯ КЛЮЧОВИХ СЛІВ ТА СТВОРЕННЯ СЕМАНТИЧНОГО ЯДРА

Існує багато переваг SEO, які прямо чи опосередковано впливають на розвиток бізнесу або компанії. Якщо ви хочете залишатися конкурентоспроможними, важливо використовувати стратегічні переваги пошукової оптимізації. Це покращує зусилля в цифровому маркетингу. Щоб збільшити органічний трафік, контент повинен відображати реальність того, що насправді шукають користувачі. Ключові слова можуть надати цю інформацію і допомогти контенту відповідати ринковим запитам. Кластеризація ключових слів впорядковує контент-план SEO і підвищує продуктивність. Це один з найкращих способів збільшити органічний трафік і отримати у винагороду повноцінний, масштабований контент-план, який відповідає основним цілям. Крім того, що найважливіше, кожна частина контенту може бути легко оптимізована.

Кластеризація ключових слів - це просунута SEO-стратегія, яка може дати перевагу, необхідну для перемоги в конкурентній боротьбі. Чим більше досліджень і кластерів ключових слів можна створити на ранній стадії, тим більше це окупиться в рейтингах Google і бездоганній маркетинговій стратегії. Завдяки аналізу сторінок пошукової видачі можна легко зрозуміти свого клієнта на іншому рівні. SERP допомагає визначити, чи може компанія піднятися в рейтингу за ключовою фразою, і якщо так, то як, і чи варті зусилля винагороди. Таким чином, це створює контент, який буде націлений на важливі ключові слова. Крім того, він вказує, над чим потрібно працювати, щоб забезпечити собі перше місце на першій сторінці.

Раніше кластеризація ключових слів здійснювалася вручну і займала досить багато часу. Але зараз можна отримати переваги від впровадження ML у цій сфері та автоматизувати майже всі процеси.

Метод кластеризації не є одним конкретним алгоритмом. Це загальна задача, основна мета якої полягає в тому, щоб згрупувати набір об'єктів таким чином, щоб вони були більш схожі між собою в одній групі, ніж з об'єктами в інших групах (кластерах). Відповідний метод і налаштування параметрів (включаючи функцію відстані, поріг щільності або очікувану кількість кластерів) залежать від конкретного набору даних і передбачуваного використання результатів [2, с. 27]. Кожен метод кластеризації включає в себе тести і невдалі спроби. Часто доводиться змінювати параметри попередньої обробки даних і моделі, поки результат не досягне бажаних властивостей.

Існує кілька різних методів кластеризації ключових слів за допомогою алгоритмів ML та AI:

Тематичне моделювання - це метод обробки природної мови (Natural Language Processing, або NLP), який допомагає виявити кластери пов'язаних ключових слів. Вона передбачає аналіз великих обсягів тексту для виявлення основних тем і групування ключових слів на основі їхнього зв'язку з цими темами. Тематичне моделювання може сканувати набір даних, виявляти в них шаблони слів і словосполучень і автоматично кластеризувати групи слів і схожі вирази, які найкраще характеризують всі ці дані [4].

Під обробкою природної мови ми зазвичай розуміємо процес обчислення і розуміння людської мови за допомогою систем. Це різновид комп'ютерної лінгвістики, що дозволяє використовувати його в машинному перекладі, розпізнаванні мови, аналізі текстів та інших сферах. Таким чином, обробка природної мови (Natural Language Processing, або NLP) - це симбіоз штучного інтелекту та лінгвістики. В результаті його використання комп'ютери можуть розуміти, інтерпретувати та маніпулювати людською мовою не гірше, ніж ми з вами.

Один з найпоширеніших алгоритмів кластеризації - агломеративна кластеризація. Основна операція полягає в об'єднанні ключових слів у найбільш підходящі кластери. Модель USE (Universe Sentence Encoder) перетворює необхідні дані в цифрові значення - вектори (embeddings). Потім вони групуються за схожістю значення. Далі ключові слова, які знаходяться найближче одне до одного, додаються до найближчого кластеру. Якщо такого кластера ще не існує, створюється новий. Дані завжди кластеризуються відносно порогового значення [6].

Далі розбиті на кластери дані аналізуються на предмет відповідності поставленим цілям. Якщо результат не задовольняє поставленому завданню, змінюються параметри порогу або інших компонентів до тих пір, поки не буде досягнуто необхідного результату. Існує також методика, яка вимірює згуртованість всередині кластерів - косинусоїдальна подібність. Його низька складність полягає у виявленні лише ненульових координат, які потрібно враховувати, і залежить лише від їхнього кута. Завдяки цьому ми можемо бути впевнені у схожості важливих для нас ключових слів [3, с.5].

Алгоритм К-середніх дозволяє вам самостійно задавати кількість кластерів. Це використовується, коли потрібно знайти суто найближчі значення відносно заданих кластерів [5]. Кожна точка даних займає свою позицію відносно семантичного змісту. Потім ми обчислюємо центроїд (функціональний центр) кожного кластера і перепризначаємо кожен точку даних до кластера з найближчим центроїдом. Ми повторюємо цей процес до тих пір, поки кластерні призначення для кожної точки даних не перестануть змінюватися [2, 28 с.].

Одним із важливих елементів SEO є також формування семантичного ядра. Семантичне ядро - один з головних елементів SEO-стратегії та основа будь-якого веб-ресурсу. Це набір слів та їхніх форм, які відповідають тематиці. Вони підбираються таким чином, щоб максимально точно характеризувати вид діяльності сайту. Правильно складене семантичне ядро допоможе створити логічну структуру сайту, організувати меню і внутрішній зв'язок між

сторінками. Завдяки цьому буде досягнутий високий рівень релевантності сторінок і, відповідно, зросте ранжування [1, с.12].

Багато компаній використовують машинне навчання (ML) для створення семантичного ядра, яке є основою їхніх стратегій оптимізації контенту. І, звичайно, найбільші з них - Google і Amazon. Google використовує машинне навчання для створення семантичного ядра, яке є основою пошукового алгоритму. Найважливішою функцією є розуміння намірів, що стоять за запитами користувачів, і підбір правильного контенту для релевантних користувачів. Аналізуючи слова і фрази, використані в запиті, а також контекст, в якому він був зроблений, Google може надавати більш релевантні результати пошуку. ML також допомагає розпізнавати закономірності в поведінці користувачів і пошукових запитах. Це впливає на персоналізовані результати пошуку.

З розвитком Штучного Інтелекту, SEO та контент-стратегії стають все більш витонченими, орієнтованими на споживача та зручними для користувача. Визначення ключових слів є важливою частиною SEO-стратегії і означає розуміння того, як Google ранжує нас за ними. Інформація про те, які ключові слова мають найбільше запитів, дозволяє нам визначити, де саме ми повинні докласти більше зусиль. Кластеризація ключових слів полегшує створення семантичного ядра, яке є фундаментом. На виході ми отримуємо основні ключові слова, на яких варто зосередитися. Це може допомогти покращити вашу загальну стратегію SEO для підвищення рейтингу та трафіку.

Список використаних джерел:

1. Павленко Ю.С. Пошукова оптимізація, технології та сервіси веб-аналітики : конспект лекцій [Електронний ресурс] / Ю.С. Павленко; ВНУ імені Лесі Українки. – Електронні текстові дані (1 файл: 968 КБ). – Луцьк: ВНУ імені Лесі Українки, 2022. – 51 с.
2. Bevza M. Analysis of clustering algorithms of scientific papers using summarisation via neural networks. Science and Education a New Dimension. 2021. Т. IX(255), № 32. С. 27–30. URL: <https://doi.org/10.31174/send-nt2021-255ix32-07> (дата звернення: 07.05.2024).
3. Giordano Bruno Beretta. Nominal Scaling of Print Substrates. 17th IS&T SID Color Imaging Conference. Albuquerque, NM. 2009. - 7 с.
4. Pykes K. What is Topic Modeling? An Introduction With Examples. Learn Data Science and AI Online | DataCamp. URL: <https://www.datacamp.com/tutorial/what-is-topic-modeling> (дата звернення: 07.05.2024).
5. What is K Means?. NVIDIA Data Science Glossary. URL: <https://www.nvidia.com/en-us/glossary/k-means/> (дата звернення: 07.05.2024).
6. Wijaya C. Y. Breaking down the agglomerative clustering process. Medium. URL: <https://towardsdatascience.com/breaking-down-the-agglomerative-clustering-process-1c367f74c7c2> (дата звернення: 07.05.2024).