

Дослідження методів релевантного пошуку та розробка пошукової системи комунальних підприємств

Приходько К.С., к.т.н.¹, Лівак М.О.²

¹²Національний університет кораблебудування ім. адм. Макарова

¹²Україна, Миколаїв

¹ORSID 0000-0003-0310-5724; ²gigofrimin@gmail.com

Анотація. Дана магістерська робота присвячена дослідженню методів релевантного пошуку та механізмів побудови пошукової системи. У роботі досліджуються сучасні рішення для побудови пошукової системи, їх сильні та слабкі сторони, технології та методи, що використовуються, а також рівень відповідності вимогам сучасного ринку інформаційних технологій.

Ключові слова: релевантний пошук. механізм пошуку. Solr.

Ми живемо в цифрову епоху. Люди генерують дані по-різному: вони фотографують, завантажують зображення, коментують чийсь фотографії, змінюють свій статус у соціальних мережах, роблять фінансові операції, пишуть електронні листи, зберігають дані у хмарних сховищах тощо. Розмір даних збільшився не лише в особистому просторі, але і в професійних службах, де людям доводиться мати справу з величезною кількістю даних. Для такого величезного потоку нам потрібні відповідні інструменти для організованого отримання даних, які можуть бути використані в різних галузях, таких як наукові дослідження, трафік у реальному часі, боротьба зі злочинністю, виявлення шахрайства, цифрова персоналізація тощо. Усі ці дані потрібно захоплювати, зберігати, шукати, ділити, передавати, аналізувати та візуалізувати.

Аналіз структурованих, неструктурованих і напівструктурованих даних допомагає виявити приховані закономірності, тенденції ринку, кореляції та особисті переваги. За допомогою правильних інструментів для обробки та аналізу даних організації можуть очікувати набагато кращих маркетингових планів, додаткових можливостей доходу, поліпшення обслуговування клієнтів, більш здорової операційної ефективності, конкурентних переваг та багато іншого. Важливо не тільки зберігати дані, але й обробляти їх для того, щоб генерувати необхідну інформацію. Кожна компанія збирає дані та використовує їх; однак, щоб потенційно процвітати ефективніше, компанії потрібно шукати відповідні дані. Кожна компанія повинна виводити дані, отримані за допомогою прямого пошуку, що може покращити їхній бізнес прямо чи опосередковано.

Розробка пошукової системи є нетривіальною задачею, оскільки кожна організація має свою структуру даних та вимоги до пошукової системи, що визначають релевантність даних. Відповідно до цих вимог використовуються різні методи та механізми для побудови пошукової системи.

Існує декілька найбільш відомих пошукових механізмів, які мають свої переваги та недоліки. Solr - один з найпопулярніших механізмів пошуку підприємств і широко використовується по всьому світу. Він написаний на базі Java і використовує пошукову

бібліотеку Apache Lucene. Solr - проект з відкритим кодом від Apache Software Foundation(ASF). Він надзвичайно швидкий, легко масштабується та ідеальний для пошуку релевантних даних. Основні переваги Solr:

- Розширений повнотекстовий пошук
- Оптимізація для роботи з великим обсягом трафіку
- Підтримка усіх сучасних форматів даних: XML, JSON, CVS тощо
- Комплексний інтерфейс адміністрування
- Зручна система моніторингу даних
- Можливість до масштабування
- Механізми для забезпечення відмовостійкості
- Гнучка система конфігурацій
- Велика кількість сторонніх плагінів

У даній системі пошуку комунальних платежів приділяється особлива увага до методів повнотекстового пошуку. У даній роботі досліджуються наступні методи повнотекстового пошуку:

- Пошук N-Gram
- Пошук за синонімами
- Пошук з приведенням слова до базової форми
- Пошук з розділенням
- Багатомовний пошук
- Фонетичний пошук

- Алгоритм Бейдер-Морза
- Алгоритм Дайч-Мокотоффа
- Алгоритм “Metaphone”
- Алгоритм “Soundex”
- Алгоритм “Caverphone”

На сьогоднішній день не існує пошукової системи, спеціалізується саме на пошуку комунальних підприємств. Сучасні пошукові системи такі як Google не спеціалізуються на конкретній предметній області, а отже не здатні у повному обсязі забезпечити релевантний пошук.

Тому є необхідність у пошуковій системі, яка буде працювати саме с комунальними підприємствами та на основі вхідних даних(пошуковий запит, місто користувача, тощо) з використанням унікального алгоритму, що складається з основних методів пошуку, згаданих вище. Користувач матиме змогу знаходити за ключовими словами саме ті комунальні підприємства, на рахунок яких йому необхідно провести платіж.

На рис.1 представлено концептуальну модель пошукової системи.

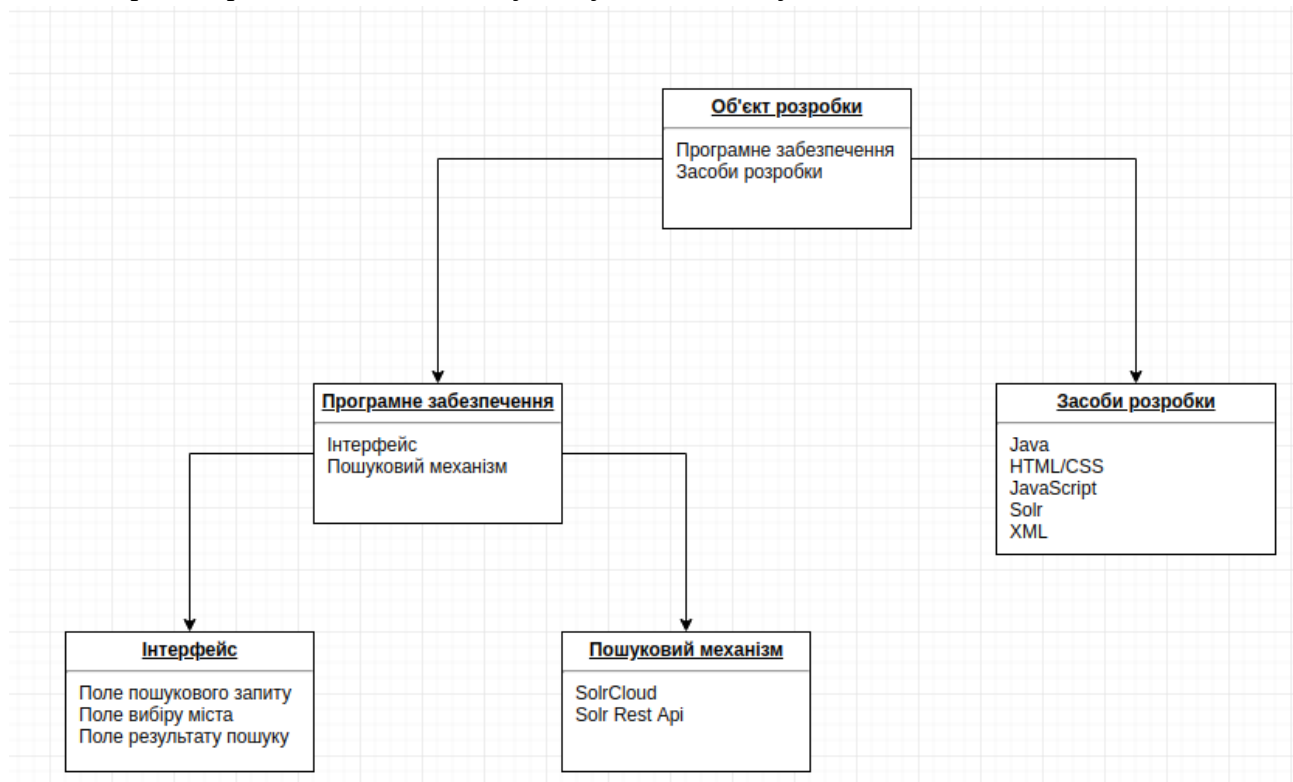


Рисунок - 1. Концептуальна модель пошукової системи

Пошукова система на даний час знаходиться в активній розробці, має усі переваги пошукового механізму Solr, зручний інтерфейс, високу швидкість обробки даних. А отже відповідає всім вимогам до сучасного програмного забезпечення. Оскільки система має комерційну цінність для комунальних підприємств(збільшення кількості платежів при меншому кількості відділень) та банків, що проводять платежі(комісія), планується подальше впровадження в такі системи як Приват24, Монобанк, тощо.

REFERENCES

- [1] Sandeep N., Chintan M., Dharmesh V. (2018). Mastering Apache Solr 7.x. Birmingham. Packt Publishing Ltd
- [2] Тарнбулл Д., Берримен Дж., (2018). Релевантний поиск с использование Elasticsearch и Solr. Москва. ДМК Пресс
- [3] Kumar J, (2015). Apache Solr Search Patterns. Birmingham. Packt Publishing Ltd

Prykhodko Kateryna, Livak Maksim

Research on methods of relevant search and development of search system of utility companies

This master's thesis contains research of methods of relevant search and mechanisms of search engine construction. The work examines the modern search engine solutions, their strengths and weaknesses, the technologies and methods used, as well as the level of compliance with the requirements of the modern information technology market.

Keywords: relevant search. search engine. Solr.

Приходько Е.С., Ливак М.О.

Исследования методов релевантного поиска и разработка поисковой системы коммунальных платежей

Данная магистерская работа посвящена исследованию методов релевантного поиска и механизмов построения поисковой системы. В работе исследуются современные решения для построения поисковой системы, их сильные и слабые стороны, технологии и методы, используемые, а также уровень соответствия требованиям современного рынка информационных технологий.

Ключевые слова: релевантный поиск. механизм поиска. Solr.