

РОЗРОБКА МЕТОДУ ТА ПРОГРАМИ З КОНТЕКСТНИМ ПОШУКОМ В ІНТЕРНЕТІ

Редька О.С.

Науковий керівник – проф. Кузьомін О.Я.

Харківський національний університет радіоелектроніки
(61166, Харків, просп. Науки, 14, каф. Інформатики, тел. +38 (057) 702 14
19) e-mail: oleksandr.redka@nure.ua

The work is devoted to solving the problem of contextual search. Methods for identifying keywords that allow searching information according to the query are developed. They consist in the analysis and processing of a search query, as well as information retrieval, based on this query. The methods are based on obtaining an existing query as a string, parsing it, and searching for information according to the query. The developed methods allow finding appropriate matches for the further search of the information on the Internet.

У наш час проблема пошуку інформації в Інтернеті є дуже актуальною. Інформаційний пошук - це те рішення, завдяки якому кожен день мільйони людей прискорюють свою роботу. Кожен з нас має своє бачення щодо пошукової системи, але вона завжди повинна виконувати одні і ті ж завдання – точний та швидкий пошук інформації згідно пошукового запиту. Одним з кращих рішень такого пошуку є реалізація контекстного пошуку. Пошукові машини, що підтримують цей метод, видають посилання на інформацію, яка точно відповідає ключовим словам у пошуковому вікні.

Актуальність завдання реалізації контекстного пошуку в інтернеті полягає в тому, що з кожним днем збільшується кількість користувачів та запитів, тому пошукові системи повинні ставати «розумнішими», задля більш точної видачі результату. Такі системи дуже зручно використовувати у будь-якій сфері: медицина, навчання, спорт тощо. Тому реалізація методу контекстного пошуку є дуже актуальною на сьогодні.

Основною формою контекстного пошуку є процес сканування повнотекстового запиту, щоб зрозуміти, що потрібно користувачеві. Всі технології повнотекстового запиту працюють за одним принципом. На основі текстових даних будується індекс, який здатний дуже швидко шукати відповідності ключового слова. Така система, зазвичай, складається з двох компонент: пристрій та індексатор.

При контекстному пошуку система надає наступні можливості:

1. Завдання в запиті логічних формул, в тому числі з операторами відстані. Як «слів» можуть виступати цифри, букви і цифри послідовності, слова які повинні бути знайдені в заданому вигляді, а також шаблони –

буквено-цифрові послідовності з символами "*" (будь-який підрядок, в тому числі порожня) і "?" (Будь-яка буква або цифра).

2. Вбудований морфологічний аналізатор, що дозволяє автоматично знайти всі існуючі словоформи для більшості слів російської та англійської мов.

Для прискорення контекстного пошуку система створює індекси за текстом. З метою зменшення розміру, ці індекси не містять докладної інформації про становище слова всередині документа, тому в разі використання операторів відстані підсистема пошуку працює в два проходи: на першому відбирає документи, що містять задані слова, на другому зчитує їх тексти в пам'ять і перевіряє виконання умови по відстані.

Основні пошукові служби, такі як Google, Bing і Yahoo, мають систему автоматичного визначення контексту конкретних запитів користувачів. Google відстежує попередні запити користувача і вибрані результати для подальшої персоналізації результатів для цих осіб. Пошукові служби прагнуть вчитися у користувачів на основі попереднього досвіду, щоб автоматично надавати контекст по поточним запитам. Bing також забезпечує автоматичний контекст для конкретних запитів на основі вмісту самого запиту.

Контекстний пошук є одним з найскладніших методів пошуку інформації в Інтернеті. Зазвичай цей метод краще реалізовувати за допомогою нейронних мереж. Пошуковий запит такого методу потрібно обробляти таким чином, щоб машина за сенсом могла зрозуміти, що саме має на увазі користувач. Тому реалізація такого методу дає змогу вирішувати багато проблем у пошуку інформації користувачами.

Перелік посилань:

1. Семакин, И.Г., Хеннер, Е.К. Информационные системы и модели [Текст]//И.Г. Семакин, Е.К. Хеннер – Москва : БИНОМ – 2005.

2. Батура Т.В., Еркаева О.Н., Мурзин Ф.А. К вопросу об анализе текстов на естественном языке. // Новые информационные технологии в науке и образовании[Текст]//Т.В. Батура, О.Н. Еркаева, Ф.А. Мурзин – Новосибирск : Под ред. проф. В.Н Касьянова – 2003.

3. Краюшкин, Д.В. Анализ технологий предварительной обработки документальной информации [Текст]//Д.В. Краюшкин – Москва : РАН И.А. – 2005.