

Лабораторна робота 16

Частина 1

Методичні вказівки щодо підбору семантичного ядра

(На прикладі тематики «дитячі коляски»)

1. Загальні положення

Семантичне ядро є структурованим набором пошукових запитів, що відображають інформаційні та комерційні потреби цільової аудиторії. У файлі ядро формується на основі:

- згрупованих словоформ,
- частотності вживання,
- прикладів пошукових фраз.

Це дозволяє перейти від сирого списку слів до осмисленої кластеризації.

2. Етапи формування семантичного ядра

2.1. Нормалізація та очищення даних

На основі файлу необхідно:

1. Видалити нерелевантні та шумові запити:

- "авіто", "б/в", якщо сайт не працює з цим сегментом
- інформаційні запити, що не відповідають цілям сайту (наприклад, ремонт)

2. Об'єднати словоформи:

- «коляска / коляски / коляску»
- «купити/куплю/купимо»

3. Виділити базові сутності:

- товар (коляска)
- дію (купити)
- атрибут (дитяча)

2.2. Визначення типів запитів

На основі прикладів із файлу можна виділити:

- Транзакційні («купити дитячу коляску»)
- Навігаційні («магазин дитячих колясок»)
- Інформаційні (наприклад: «яка коляска краще» — якщо є в даних)

2.3. Кластеризація (угруповання запитів)

Ключовий методичний етап – розбивка ядра на категорії (кластери), які відповідають структурі сайту.

3. Запропонована структура категорій (приклад)

3.1. Основна категорія (ядро)

дитячі коляски

коляска дитяча

купити дитячу коляску

👉 Це високочастотні запити (HF), що формують головну сторінку.

3.2. Типи колясок (ключовий комерційний кластер)

На основі логіки попиту:

- коляски 2 в 1
- коляски 3 в 1
- прогулянкові коляски
- коляски для новонароджених
- універсальні коляски
- коляски-трансформери

👉 Це основні категорії каталогу.

3.3. Комерційні інтенти (дії користувача)

- купити коляску
- магазин колясок
- ціна коляски

Підкатегорії:

- купити коляску недорого
- коляски ціна
- інтернет-магазин колясок

👉 Використовуються для посадкових сторінок та фільтрів.

3.4. Геозалежні запити

Якщо у даних присутні:

- купити коляску Харків
- коляски Київ

👉 Створюються регіональні сторінки.

3.5. Брендіві запити (якщо є у файлі)

- коляски Adamex
- коляски Apex
- коляски Chicco

👉 Окремі сторінки брендів.

3.6. Атрибутивні характеристики (фільтри)

На основі типових словоформ:

за віком:

- для новонароджених
- до 3 років

за характеристиками:

- легка коляска
- складаний візок
- зимовий візок

за ціною:

- недорогі
- преміум

👉 Використовуються у фільтрації каталогу.

3.7. Інформаційні запити (якщо присутні або будуть додані)

- як вибрати коляску
- який візок краще
- рейтинг колясок

👉 Формують блог/розділ статей.

3.8. Вторинні майданчики (маркетплейси)

- авіто коляска

👉 Зазвичай: або виключаються, або використовуються в конкурентному аналізі.

4. Методика кластеризації (алгоритм)

Рекомендується наступна послідовність:

1. Взяти стовпець «Приклад ключової фрази»
2. Розбити фрази за:
 - інтену (купити/вибрати/подивитися)

- сутності (тип коляски)
3. Згрупувати:
- вручну (для точності)
 - або через TF-IDF/clustering (якщо автоматизація)

5. Підсумкова структура ядра.

Категорія	Приклади
Основні	дитячі коляски
Типи	прогулянкові, 2 в 1
Комерційні	купити, ціна
Гео	коляски Харків
Бренди	Апех, Adamex
Атрибути	легка, зимова
Інформаційні	як вибрати

6. Практичні рекомендації

1. Не змішувати інтенти. Сторінка «купити» ≠ стаття «як вибрати»
2. Опиратися на частотність. Стовець «Зустрічається вкотре» — пріоритет для структури
3. Використовувати ієрархію: Головна → категорії → підкатегорії → фільтри
4. Мінімізувати канібалізацію. Один кластер = одна посадкова сторінка

Завдання

1. Скласти семантичне ядро на тематику свого сайту.
2. Визначити релевантні категорії.
3. Розподілити ключові фрази за категоріями.
4. Для кожної категорії написати статтю із включенням ключових фраз (ключові фрази виділити кольором).
5. Створити категорію "Цікаве". Наповнити її статтями, на які можна перейти за посиланнями з основних категорій.