

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ**

ДОПУСТИТИ ДО ЗАХИСТУ  
Завідувач кафедри

\_\_\_\_\_ Литвиненко О.Є.

«\_\_\_» \_\_\_\_\_ 2021 р.

**ДИПЛОМНИЙ ПРОЕКТ**  
(ПОЯСНЮВАЛЬНА ЗАПИСКА)

**ВИПУСКНИКА ОСВІТНЬОГО СТУПЕНЯ  
"БАКАЛАВР"**

Тема: \_\_\_\_\_ Програмна система для аналізу контенту *web*-сайтів

Виконавець: \_\_\_\_\_ Сакура В.І.

Керівник: \_\_\_\_\_ Артамонов Є.Б.

Нормоконтролер: \_\_\_\_\_ Тупота Є.В.

**Київ 2021**

# НАЦІОНАЛЬНИЙ АВІАЦІЙНИЙ УНІВЕРСИТЕТ

Факультет \_\_\_\_\_ кібербезпеки, комп'ютерної та програмної інженерії

Кафедра комп'ютеризованих систем управління

Освітнього ступеня \_\_\_\_\_ бакалавр

Напрямок (спеціальність) 123 "Комп'ютерна інженерія"  
(шифр, найменування)

ЗАТВЕРДЖУЮ

Завідувач кафедри

\_\_\_\_\_ Литвиненко О. Є.

« \_\_\_\_ » \_\_\_\_\_ 2021 р.

## ЗАВДАННЯ на виконання дипломної роботи (проекту)

Сакури Валентина Ігорівна

(прізвище, ім'я, по батькові)

1. Тема роботи: \_\_\_\_\_ Програмна система для аналізу контенту *web*-сайтів

затверджена наказом ректора від "04" \_\_\_\_\_ лютого 2021 року № 135 /ст.

2. Термін виконання роботи: з 17.05.2019 до 20.06.2019

3. Вихідні дані до роботи: 1) вимоги до змісту системи;

2) основні операції в системі

4. Зміст пояснювальної записки (перелік питань, що підлягають розробці):

1) аналіз сучасних методів аналізу контенту *web*-сайтів;

2) принципи роботи систем для аналізу контенту *web*-сайтів;

3) реалізація програмної системи аналізу контенту *web*-сайтів.

5. Перелік обов'язкового графічного матеріалу:

1) узагальнений принцип роботи системи;

2) структура системи аналізу контенту *web*-сайтів;

3) діаграма варіантів використання;

4) діаграма пакетів проекту;

5) приклади робочих вікон системи;

6) схема алгоритму рекурсивного пошуку словоформ;

7) схема алгоритму пошуку ключових слів.

## 6. Календарний план

№ п/п	Етапи виконання дипломного проекту	Термін виконання етапів	Примітка
1	Провести аналіз літератури за темою дипломного проекту та аналіз існуючих систем	178.05.21-19.05.21	
2	Провести аналіз принципів оптимізації текстового контенту	19.05.21-23.05.21	
3	Розробити модель роботи системи	24.05.21-26.05.21	
4	Розробити алгоритми обробки текстового контенту	27.05.21-29.05.21	
5	Розробити програмні засоби	30.05.21-02.06.21	
6	Написати пояснювальну записку	03.06.21-10.06.21	
7	Підготувати презентацію та графічні матеріали	11.06.21-14.06.21	

## 7. Дата видачі завдання « 17 » травня 2021 р.

Керівник дипломного проекту \_\_\_\_\_ Артамонов Є.Б.  
(підпис)

Завдання прийняв до виконання \_\_\_\_\_ Сакура В.І.  
(підпис студента)

## РЕФЕРАТ

Пояснювальна записка до дипломного проекту “Програмна система для аналізу контенту *web*-сайтів”: 60 с., 25 рис., 22 літературних джерела, 1 додаток.

*SEO*, АНАЛІЗ ТЕКСТУ, АВТОМАТИЧНА ОБРОБКА ТЕКСТУ.

Об’єкт дослідження – аналіз текстової інформації.

Предмет дослідження – програмна система для аналізу контенту *web*-сайтів.

Мета роботи полягає в розробці і реалізації на мові високого рівня модуля оцінки тексту щодо пошукових запитів, розбір тексту на природній мові передбачає розробку морфологічного словника сильнорозгалуженої структури, а також методів і алгоритмів роботи з цим словником. На основі отриманих результатів апробувати роботу алгоритмів у складі розробленого додатково програмного забезпечення для аналізу тексту за основними параметрами *SEO*.

## ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ, ТЕРМІНІВ .....	6
ВСТУП	7
РОЗДІЛ 1 АНАЛІЗ СУЧАСНИХ МЕТОДІВ АНАЛІЗУ КОНТЕНТУ <i>WEB-</i> САЙТІВ .....	10
1.1. Принципи оптимізації сайтів .....	10
1.2. Методи <i>SEO</i> .....	15
1.3. Висновки до розділу .....	18
РОЗДІЛ 2 ПРИНЦИПИ РОБОТИ СИСТЕМ АНАЛІЗУ КОНТЕНТУ <i>WEB-</i> САЙТІВ .....	20
2.1. Вимоги до систем аналізу текстів .....	20
2.2. Аналіз існуючих програмних рішень для аналізу текстів .....	29
2.3. Описання системи для аналізу текстів.....	31
2.4. Висновки до розділу .....	37
РОЗДІЛ 3 РЕАЛІЗАЦІЯ ПРОГРАМНОЇ СИСТЕМИ АНАЛІЗУ КОНТЕНТУ <i>WEB-САЙТІВ</i> .....	38
3.1. Робота з пошуковими фразами .....	38
3.2. Інтерфейс головного вікна системи .....	40
3.3. Висновки до розділу .....	54
ВИСНОВКИ .....	55
СПИСОК БІБЛІОГРАФІЧНИХ ПОСИЛАНЬ ВИКОРИСТАНИХ ДЖЕРЕЛ .....	59
ДОДАТОК А.....	61

## ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СКОРОЧЕНЬ, ТЕРМІНІВ

*UML – Unified Modeling Language*

АОТ – автоматична обробка тексту

ПМ – природня мова

## ВСТУП

Контент-аудит- це перевірка параметрів вмісту сайту, націлена на оцінку маркетингової ефективності контенту. Результат проведеного контент-аналізу - опис поточного стану контенту з достоїнствами і недоліками, а також рекомендації щодо виправлення. Залежно від специфіки веб-ресурсу оцінка може проводитися для вирішення наступних практичних завдань:

- визначення причин санкцій з боку пошукових систем.
- пошук сторінок, зміст яких має бути скоригована.
- визначення сторінок, які можна об'єднати або видалити.
- планування стратегії розміщення контенту.
- формування фактичної і ідеальної карти ключових слів для просування сайту.
- пошук можливостей для збільшення трафіку і розвитку.
- оцінка маркетингової ефективності контенту.

Частотний аналіз є одним з порівняно простих методів обробки тексту на природній мові (NLP). Його результатом є список слів, найбільш часто зустрічаються в тексті. Частотний аналіз також дозволяє отримати уявлення про тематику і основні поняття тексту. Візуалізувати його результати зручно у вигляді «хмари слів». Ця діаграма містить слова, розмір шрифту яких відображає їх популярність в тексті.

Проведення частотного аналізу можна умовно розділити на кілька етапів:

1. Завантаження та огляд даних
2. Очищення і попередня обробка тексту
3. Видалення стоп-слів
4. Переклад слів в основну форму
5. Підрахунок статистики зустрічальності слів в тексті

## Попередня обробка (препроцесінг) тексту

Для проведення частотного аналізу і визначення тематики тексту рекомендується виконати очистку тексту від знаків пунктуації, зайвих символів пробілів і цифр. Зробити це можна різними способами - за допомогою вбудованих функцій роботи з рядками, за допомогою регулярних виразів, за допомогою операцій обробки списків або іншим способом.

Набір спеціальних символів, які будуть видалені з тексту може бути розширено. Необхідно проаналізувати вихідний текст і виявити символи, які слід видалити. Додамо до знаків пунктуації символи розриву рядків, табуляції та інші символи, які зустрічаються в нашому вихідному тексті (наприклад, символ з кодом `\xa0`):

Для видалення символів використовуємо поелементну обробку рядка - розділимо вихідну рядок `text` на символи, залишимо тільки символи, що не входять в набір `spec_chars` і знову об'єднаємо список символів в рядок.

## Токенізація тексту

Для подальшої обробки очищений текст необхідно розбити на складові частини - токени. В аналізі тексту на природній мові застосовується розбивка на символи, слова і пропозиції. Процес розбиття називається токенізація. Для нашої задачі частотного аналізу необхідно розбити текст на слова. Для цього можна використовувати готовий метод бібліотеки NLTK:

## Підрахунок статистики зустрічальності слів в тексті

Для підрахунку статистики розподілу частот слів в тексті застосовується клас `FreqDist` (`frequency distributions`):

Спроба вивести змінну `fdist` відобразить словник, що містить маркери і їх частоти - кількість разів, які ці слова зустрічаються в тексті:

Частота розподілу слів тексті може бути візуалізувати за допомогою графіка. Клас `FreqDist` містить вбудований метод `plot` для побудови такого графіка. Необхідно вказати кількість токенів, частоти яких будуть показані на



графіку. З параметром `cumulative = False` графік ілюструє закон Ціпфа: Якщо все слова досить довгого тексту впорядкувати по спадаючій частоті їх використання, то частота  $n$ -го слова в такому списку опиниться приблизно обернено пропорційній його порядковому номеру  $n$ .

Можна помітити, що в даний момент найбільші частоти мають сполучники, прийменники та інші службові частини мови, що не несуть смислового навантаження, а тільки виражають семантико-синтаксичні відношення між словами. Для того, щоб результати частотного аналізу відбивали тематику тексту, необхідно видалити ці слова з тексту.

#### Видалення стоп-слів

До стоп-слів (або шумовим словами), як правило, відносять прийменники, сполучники, вигуки, частки та інші частини мови, які часто зустрічаються в тексті, є службовими і не несуть смислового навантаження - є надлишковими.

Об'єкт дослідження – аналіз текстової інформації.

Предмет дослідження – програмна система для аналізу контенту *web*-сайтів.

## РОЗДІЛ 1

### АНАЛІЗ СУЧАСНИХ МЕТОДІВ АНАЛІЗУ КОНТЕНТУ *WEB*-САЙТІВ

#### 1.1. Принципи оптимізації сайтів

Аналіз контенту сайту допоможе домогтися поліпшення конкретних метрик ефективності, що посприє розвитку не тільки інтернет-проекту, але і бізнесу в цілому.

Якщо ресурс невеликий за обсягом, то його перевірка буде максимально простий. Так, для блогу з десятками статей потрібно лише ознайомитися з керівництвом проведення аудиту і провести аналіз за допомогою Google Analytics. Для більших ресурсів, від 500 сторінок, дотримуйтеся наступного плану дій:

- Виписати в окрему таблицю всі посилання, які ведуть на сторінки з вмістом.
- Проаналізувати ефективність цих сторінок.
- Провести дослідження контенту. Використовуйте ручної аудит.
- Оцінити ситуацію по ключових запитах.
- Визначите, якого контенту бракує на сайті.
- Описати поточну ситуацію і сформулюйте рекомендації щодо її виправлення.

Цей метод контент-аналізу може бути доповнений окремими пунктами, в залежності від специфіки ресурсу.

Кафедра КСУ				НАУ 21 19 29 000 ПЗ			
<i>Виконав</i>	<i>Сакура В.І.</i>			Аналіз сучасних методів аналізу контенту <i>web</i> -сайтів	<i>Літера</i>	<i>Аркуш</i>	<i>Аркушів</i>
<i>Керівник</i>	<i>Артамонов Є.Б.</i>				<i>Д</i>	<i>10</i>	<i>60</i>
<i>Консульт.</i>					123 СП 436		
<i>Норм. контр.</i>	<i>Тупота Є.В.</i>						
<i>Зав. Каф.</i>	<i>Литвиненко О.Є.</i>						

У будь-якому випадку для зручності роботи використовуйте спеціальні таблиці, виписуйте підсумки оцінки і рекомендації.

Покрокове керівництво по проведенню контент-аудиту

Контент-аналіз - це довгий процес, який можна розбити на кілька етапів.

Розглянемо основні кроки і інструменти для кожного з етапів.

Підготовка сценарію оцінки

Кожен сайт має свою специфіку, а отже, вимагає індивідуального підходу.

Однак повинен бути визначений базовий алгоритм дій. На його вибір впливають два фактори:

- розмір ресурсу;
- наявність загрози накладення санкцій з боку пошукових систем.

Якщо явних проблем з контентом немає, сайт не потрапив під обмеження пошукачів і при цьому має невеликий об'єм, наприклад, до 100 сторінок, то цілком може бути використаний ручний аудит. Тут також допоможе налагодити роботу складання семантичного ядра і його дослідження в поточній ситуації. Для великих ресурсів, від 1 тис. сторінок, стратегія залишається тією ж, але проводите аудит для пріоритетних розділів.

Семантичне ядро- це таблиця, яка відображає всі ключові запити по яких просувається контент в мережі і ті, за якими просування не ведеться, але вони здатні принести хороший трафік на сайт.

У тому випадку, коли існує ризик накладення санкцій з боку пошукових систем, варто до описаної вище стратегії додати важливість видалення неактуальних сторінок. Це ж справедливо для великих сайтів вже потрапили під обмеження пошукових систем. Для малих ресурсів, які зіткнулися з санкціями через якість контенту можна провести аудит на кожній сторінці в ручному режимі.

сканування сайту

Цей крок необхідний для отримання списку активних URL, які ведуть на веб-сторінки вашого ресурсу. Проведіть сканування за допомогою онлайн-інструментів, наприклад, XML-sitemaps. Для сайтів з кількістю сторінок до 500 перевірка буде безкоштовною. зайдіть наXML-sitemaps. На головній сторінці в

рядок введення вставте адресу сайту і натисніть кнопку Start для початку перевірки.

Після закінчення сканування скачайте результати перевірки у вигляді окремої таблиці.

#### Аналіз контенту сторінки

На цьому етапі вам допоможе програма для контент-аналізу URL Profiler. Встановіть програму на комп'ютер. Перші 14 днів використання надаються в якості демо-версії. Відкрийте програму і в поле для введення URL імпортуйте скачаний раніше файл. Для цього викличте контекстне меню правою кнопкою миші і виберіть відповідний варіант.

Виберіть потрібні опції для аналізу ефективності контенту. Вони залежать від завдань аудиту і використовуваних інструментів веб-аналітики. Однак можна скористатися рекомендованими настройками.

За допомогою даних метрик ви отримаєте всю необхідну для аналізу контенту сайту інформацію:

Moz - це прогностична величина, що показує ефективність майданчика від 1 до 10.

HTTP Status відображає код стану сторінки. Чи є вона активної і працює без помилок.

Social Shares розповідає про кількість позначок «Поділитися».

Під параметром Readability розуміється читабельність контенту.

Значення uClassify говорить про те, що сайт буде проаналізовано на предмет його типу.

Copyscape відповідає за перебування плагіату і визначення загрози накладення санкцій на ресурс через неоригінальності контенту.

У розділі налаштувань Google вкажіть ті сервіси, до яких програма повинна підключитися для збору інформації:

Google Analytics. Підключення буде вироблено безпосередньо до API даного сервісу.

Google Indexation завантажує масові метрики посилань.

Desktop PageSpeed відповідає за швидкість завантаження сторінок сайту в звичайній десктопній версії. Дані беруться з Google PageSpeed.

Далі натисніть кнопку Run Profiler в нижньому правому куті вікна.

Результати будуть представлені у вигляді окремої таблиці.

Кількісні дані занесіть в таблицю результатів контент-аудиту та на їх підставі проведіть аналіз ефективності сторінок, наприклад, за кількістю репоста або на підставі інших метрик.

Дослідження ключових слів

Якщо у вашому сценарії контент-аудиту передбачено створення карти «ключових слів», використовуйте інструменти Google. Увійдіть в Search Console і відкрийте меню, присвячене пошукового трафіку і запитам.

Перейдіть на вкладку «Сторінки» і встановіть максимальний діапазон дат для аналізу.

Внизу вікна знайдіть кнопку «Показати рядків» і виберіть максимальне значення - 500.

Натисніть кнопку Enter. Завдяки цим кодом будуть відображені всі запити для кожного посилання.

Отриману таблицю збережіть у форматі CSV і імпортуйте дані в файл з результатами аудиту.

Потім вам буде потрібна допомога планувальника ключових слів. Ви можете придбати автоматичні програми або провести перевірку частоти ключових запитів в ручну за допомогою «Яндекс.Подбор Слов». Для цього введіть у пошуковий рядок сервісу цікавить запит і натисніть кнопку "Підібрати", щоб побачити статистику.

Об'єднання даних за ключовими словами

Планувальник Яндекс або Google, якщо ви використовуєте цей інструмент, надасть вам ключові запити з їх кількістю і рівнем конкуренції. Об'єднайте ці дані з тими, що були отримані за допомогою інструментів для web-майстрів: CTR, кількість переглядів, середньої позиції і так далі. Для цього створіть таблицю з наступними графами:

- ключовий запит;
- позиція;
- обсяг запитів;
- CPC (ціна за клік).

Отримані відомості використовуйте для створення семантичного ядра.

Визначення відсутнього контенту

Цей крок може входити не в усі сценарії. Однак при аналізі ключових запитів, ви можете побачити ті ключі, за якими жодна сторінка вашого сайту не ранжируется. Крім SEO оптимізованих текстів, потрібно буде створити окремі посадкові сторінки. У цьому допоможе розширення структури сайту на основі семантичного аналізу. Також вам стане ясно, який матеріал необхідно додати.

Наприклад, перевіряючи запит «що таке Лендінгем», ви можете з'ясувати, що є цілий ряд схожих фраз, наприклад, «лише однієї сторінки сайт» і «контекстна реклама». Кількість показів в місяць велика, але у вашому контенті вони не зустрічаються. Це означає, що якщо ви створите SEO-оптимізовані тексти під ці фрази, ваш трафік збільшиться, так як ресурс буде видаватися за ним в результатах пошукової видачі.

Аналіз додаткових характеристик

Для проведення останнього етапу SEO-аналізу контенту буде потрібно повернутися до даних, наданих програмою URL Profiler. Зверніть увагу на ті сторінки, які мають високим ризиком - Copyscape. Це означає, що в інтернеті вже є джерела з подібною інформацією і пошукові машини можуть накласти через це санкції.

Контент на таких сторінках відредагуйте для додання унікальності або ж видаліть. Не меншої уваги заслуговують сторінки з найменшою відвідуваністю.

Такий контент можна видалити, якщо у вас великий сайт. Якщо ж подібні сторінки знайдені на ресурсі з невеликою кількістю розділів, то варто окремо провести аналіз SEO показників, щоб виявити причину низької відвідуваності і виправити її.

## 1.2. Методи *SEO*

Вираз «якісний контент для сайту» перетворилося в розхожу фразу. Ним користуються всі кому не лінь, рекламуючи послуги посереднього рівня за високою вартістю. Що ж насправді ховається за шаблонним висловом? Відповідь на питання міститься в цій статті, де розглядаються критерії якісних текстів, зрозумілі простим обивателям і просунутим веб-майстрам.

1. Вирішує конкретні проблеми. Створення нового контенту підпорядковується головній меті - знайти відповідь на запитання потенційних читачів і надати практичне рішення: швидке, результативне і відчутне.

2. Володіє інноваційним підходом. Новий погляд на звичні речі, екстравагантні рішення складних питань і принципово інші підходи до вирішення вічних проблем - ще одна відмінність якісного контенту від низькопробної писанини.

3. Відрізняється глибиною опрацювання матеріалу. Надаючи великі знання і ділову компетенцію, контент для людей характеризується більшою глибиною опрацювання матеріалу і ґрунтовними пропозиціями для вирішення нагальних проблем.

4. Містить актуальну інформацію. Затребуваний контент надає актуальну інформацію, корисну для цільової аудиторії і сприяє вирішенню поточних завдань. На конкретних прикладах розкриваються потенційні вигоди і способи досягнення поставленого результату.

5. Має чітку логіку. Логічне оповідання в рамках обраної теми і тлумачні висновки завжди знаходять розуміння вдячної аудиторії, яка не тільки залишає коментарі та відгуки, але знаходить спільну точку зору на гострі проблеми і злободенні питання.

6. Володіє смисловий унікальністю. Безсумнівно, високий ступінь оригінальності текстів сприяє визнанню їх якісними. Адже подібний контент, який ще не розтиражований на просторах Інтернету, користуються особливим визнанням серед пошукових систем. Правильний контент містить нові ідеї,

надаючи можливість читачам обговорити написане і поділитися власними думками.

7. Містить тематичні підрозділи. Наявність логічних текстових блоків (не плутати з абзацами) полегшує пошук потрібної інформації, покращує сприйняття і запам'ятовування матеріалу.

8. Заслужує довіру читачів. Якісний контент подобається читачам. Адже в ньому містяться не тільки прописні істини і хрестоматійні приклади, але і емоційні розповіді, несподівані історії, переконлива аналітика, інтерв'ю з відомими людьми та багато іншого, що заслуговує не тільки увагу, але і довіру читачів.

9. Має лаконічну назву. Короткий, чіткий і нехитрий заголовок, що розкриває тему та одночасно інтригує читача, здатний забезпечити увагу потенційної аудиторії.

10. Володіє конкретною метою. Наявність конкретної мети відрізняє контент високої якості від пересічних текстів, позбавлених творчого змісту і смисловий унікальності.

11. цікавий читачам. Корисні і актуальні тексти на будь-яку тематику завжди цікаві читачам і вони з радістю діляться в соціальних мережах, залишаючи посилання і роблячи перепис хороших статей.

12. Релевантним тематиці сайта. Публікації в блогах і на спеціалізованих ресурсах повинні відповідати тематиці сайту. Не потрібно жертвувати релевантністю текстів, намагаючись бути «своїм серед чужих».

13. Має оптимальний обсяг. Для кожного випадку потрібен різний обсяг тексту. Починаючи від 500 знаків для коротких товарних описів і лаконічних продають текстів, до 3 000 знаків і більше для авторських і аналітичних статей, які читаються на одному диханні. Тут головне - розкрити тему і не гнатися за кількістю, насичуючи водою хороші тексти.

14. На нього посилаються інші сайти. Безсумнівний ознака якості контенту - це вхідні посилання з інших ресурсів. Цікаві та корисні матеріали не потребують додаткової реклами і розкрутки. Швидко обростаючи природної



посилальної масою вони отримують визнання інтернет-користувачів і пошукових систем.

15. Написаний простою мовою, зрозумілою для дилетантів. Контент для вузькоспеціалізованих сайтів на технічну або галузеву тематику, написаний для фахівців помітно відрізняється від текстів, призначених для звичайних користувачів, що не володіють належним досвідом і знаннями в предметній області. Щоб врахувати потреби широкої аудиторії, якісний контент пишеться в максимально спрощеному вигляді без використання специфічних термінів і глибокодумних міркувань.

16. Відрізняється бездоганним правописом. Відсутність граматичних помилок, прогалин в орфографії, помилок і стилістичних ляпів - ще один критерій текстового контенту в професійному виконанні.

17. містить списки. Наявність маркованих і нумерованих списків полегшує сприйняття структурованої інформації. Порядок дій або переваги товару, розкладені по полочках, наочно демонструють ознаки правильного контенту.

18. Чи не страждає зайвою оптимізацією. Якісний текст насичений ключовими словами, різноманітними словоформами і відповідними синонімами. Грамотні ключові фрази не кидаються в очі і не погіршують читабельність, не порушують правила орфографії і правопису.

19. Має природні анкор. Натуральні анкор без зайвого комерційного відтінку здатні надати виходить посиланнях природний вигляд і поліпшити пошукове ранжування сайту. Наприклад, можна використовувати безанкорніе посилання ([www.watermillsky.ru](http://www.watermillsky.ru)) або фрази типу «читати тут», «більше», «дізнатися більше» і ін.

20. Наділений візуальної привабливістю. Хороший контент зручний для читання і сприйняття. Він розбитий на логічні частини і окремі абзаци, має маркірування і нумеровані перерахування і в ньому є підзаголовки. Додаткову привабливість надають кольорові зображення (фотографії, малюнки, малюнки, схеми, графіки) і відеоматеріали. Графіка гарантує підвищений інтерес користувачів і помітно покращує поведінкові чинники.

21. Вихідні посилання ведуть на актуальні публікації. Якщо вихідні лінки ведуть на ресурси з оновлюваною інформацією - це ще одна ознака добротного контенту. Адже багато з того, що було опубліковано раніше, втрачає свою актуальність і перестає цікавити користувачів. Публікації на окремих інтернет-ресурсах в світлі нових даних надають недостовірні відомості або застарілу аналітику.

22. Автор не ховається від читачів. У хороших текстів і статей завжди є автор, який не приховує своє ім'я від зацікавлених читачів. Наявність фотографії та короткого життєпису, включаючи найбільш цікаві публікації, покращує якість текстового контенту і підвищує довіру з боку інтернет-користувачів.

23. Не містить суперечливих відомостей. Контент високої якості відрізняється стрункою логікою і не містить відомостей суперечливого характеру. Кожен абзац переконливо розкриває тему та ненав'язливо підводить читачів до певного висновку, мотивуючи на вчинення заданих дій.

24. Посилається на трастові першоджерела. Наявність посилань на оригінальні тексти (вітчизняні чи зарубіжні першоджерела) з авторитетних сайтів, що володіють високим трастом, сприяє більшій довірі користувачів і пошукових систем.

Містить конкретні приклади. Для текстів на складні тематики (фінанси, економіка, техніка, промислове обладнання, опис технологічних процесів, наукові публікації і ін.) Просто необхідні конкретні приклади, які допоможуть розібратися не тільки професіоналам, але звичайним користувачам без спеціальної освіти і належної компетенції.

### **1.3. Висновки до розділу**

Онлайн-аудит сайту дозволяє веб-майстру або власнику бізнесу знайти проблеми і недоліки свого сайту, які знижують продажі. Такий аудит обов'язково повинен включати в себе аналіз контенту сайту.

Якщо коротко позначити алгоритм проведення контент-аудиту, то він буде виглядати наступним чином:

– Виберіть стратегію для перевірки сайту виходячи з його розміру та наявності санкцій.

– список всіх активних посилань за допомогою XML-sitemaps.

– проведіть аналіз роботи сторінок сайту за допомогою URL Profiler.

– оцініть ситуацію по ключовим запитом за допомогою інструментів для веб-майстрів Google.

– на основі даних URL Profiler позбудьтеся від неунікального контенту і виявите сторінки.

– опишіть поточну ситуацію і сформууйте рекомендації щодо її виправлення.

На основі отриманих даних контент-аудиту ви сформуєте звіт із загальними враженнями від ефективності сайту. Залежно від специфіки і розмірів сайту може бути обрана й інша стратегія. Описані вище стратегія націлена на створення семантичного ядра і пошук неактуального контенту.

## РОЗДІЛ 2

### ПРИНЦИПИ РОБОТИ СИСТЕМ АНАЛІЗУ КОНТЕНТУ WEB-САЙТІВ

#### 2.1. Вимоги до систем аналізу текстів

Контент - це код, який визначає поведінку того, хто його читає. Крім тексту, до нього відносяться аудіо, відео, картинки, код розробників.

Цей матеріал буде стосуватися тільки контенту для сайтів і тільки для просування в Google. Тези статті не стосуються просування в YouTube, Instagram, Pinterest, Soundcloud і інших платформах, де текст не є ключовим змістом.

Моє завдання - допомогти вам провести якісний аудит вже опублікованого контенту і дати алгоритм підбору хороших авторів для сайту.

Ці рекомендації засновані на моєму досвіді розвитку блогу Sitechecker, журналу Brammelsiособистого блога.

Що відбувається в SEO в 2020 році

Читаючи телеграм канали SEO фахівців і спілкуючись з ними я бачу, що є два основних протилежних погляди на існуючий алгоритм ранжирування Google. Розберемо ключові ідеї кожного напрямку.

#### 1. Не потрібно переоцінювати можливості Google

Створюйте якомога більше сайтів. Пишіть контент, ставте посилання, через час відбирайте вижили і розвивайте їх. Попадання в топ - повний рандом. Сайт - видаткова одиниця. Краще багато маленьких сайтів, ніж один великий. Те, про що Google заявляє в своєму керівництві для ассесоров лише пропаганда, і ключове значення мають лише посилання.

Схожий погляд на алгоритм також проповідують все творці дорвеев і PBN. Ключова думка таких веб-майстрів - «скільки б Google не старалася поліпшити

Кафедра КСУ				НАУ 21 19 29 000 ПЗ			
Виконав	Сакура В.І.			Принципи роботи систем аналізу контенту web-сайтів	Літера	Аркуш	Аркушів
Керівник	Артамонов Є.Б.				Д	20	60
Консульт.					123 СП 436		
Норм. контр.	Тупота Є.В.						
Зав. Каф.	Литвиненко О.Є.						

свій алгоритм, в ньому завжди будуть уразливості або білі плями, до яких гугловці немає діла, а значить, наша технологія роботи і бізнес більш невразливий, ніж бізнеси, які просувають сайти білими методами».

2. Google навчилася ранжувати сайти по EAT, розпізнавати хороші і погані посилання, краще розуміти зміст тексту

Google вже складно обдурити, і тому не варто витратити на це зусилля. Потрібно просто виконувати вимоги гугловці. Робити якісний контент, залучати або вирощувати авторів-експертів і шукати креативні методи отримання посилань.

Але є проблема. Google еволюціонує, а власники сайтів і SEO фахівці не встигають за нею. Причина не в тому, що веб-майстри не знають, що потрібно робити. Успішні сеошники цілком регулярно оновлюють свою картину світу і швидко впроваджують необхідні зміни.

Веб-майстри шукають інші варіанти монетизації, щоб дозволити собі дорогих авторів, більше вкладають в посилання з авторитетних джерел, переходять на темну сторону просування.

Але для багатьох ніш, розвиток сайту за новими правилами все одно стає збитковим. Дмитро Лосьов останніх постах свого Telegram-каналу цікаво описує досвід перебудови бізнесу після зниження комісій Amazon і останнього апдейта Google.

Тільки зараз, під тиском Google, веб-майстри почали замислюватися, що ж таке якісний контент і як виявити людину, яка вміє його створювати.

Мені ближче другий погляд. Навіть якщо Google прикрашає свої можливості зараз, рано чи пізно вона навчиться робити те, про що заявляє. Тому питання, як відокремити якісний контент від неякісного, буде стає все популярнішим.

Якість контенту - це характеристика з багатьма змінними, частина з яких не можна вмістити в математичну формулу. Алгоритм Google може порахувати контрольний вагу сторінок, кількість входжень ключового слова, унікальність контенту.

Але без оцінок асесоров не можна навчити алгоритм оцінювати достовірність фактів наведених у статті, рівень експертності і авторитетності автора статті, якість редактури і оформлення.

Протягом останніх років веб-майстри при постановці ТЗ на текст приділяли більше уваги формі, а не змістом. Найпопулярніший метод постановки завдання на текст до сих пір - аналіз топа пошукової видачі. Детально розповідав про нього тут.

Цей метод корисний, але його вже недостатньо. Чим більше Google буде залучати асесоров і тренувати нейронні мережі пошукового алгоритму, тим важливіше буде проводити якісний аналіз текстів. Потрібно виявляти кращі практики створення контенту, його редактури, оформлення і застосовувати у себе на сайті.

Коли ви читаете якусь статтю в інтернеті, то можете помічати на рівні відчуттів, що вам матеріал подобається або не подобається. Але ви не завжди можете пояснити чому. Спробуємо розібратися, які параметри контенту впливають на його якість і задоволеність користувача.

Багато корисних параметрів перераховано вкерівництві Google для початківців веб-майстрів і керівництві для асесоров, Про який згадував вище. У своєму списку я спробую розібрати деякі параметри детальніше.

#### Параметри якісного контенту

##### 1. Автор

До параметрів автора відносяться: освіта, досвід, професійні досягнення, авторитет серед інших фахівців у конкретній сфері.

Насправді авторитетні і досвідчені автори не завжди роблять кращий контент, але це певний фільтр, який дає можливість Google відсікти спамерів за рахунок складності і дорожнечі накрутки такого фактора.

Краще втратити кілька цінних матеріалів від майбутніх, ще не визнаних зірок і сформуванати гарну видачу, ніж витратити ресурси на пошук самородків серед сотень тисяч сторінок непридатного контенту.

Розкажіть більше про ваших авторів, їхні досягнення та дайте посилання на їх соцмережі. Безликі сайти з нескінченим рерайтом відходять у минуле і це прекрасно.

### Профіль учасника з блогу Semrush

Крім того, коли людина знає, що під текстом або над текстом люди завжди будуть бачити його профіль, це серйозно впливає на його відповідальність і мотивацію.

## 2. Джерела

Згадайте, як ви писали курсову або дипломну роботу. Вступ до такої роботи має містити три важливих блоку: хто вже досліджував цю тему, чому вона актуальна і в чому новизна вашої роботи.

Більшість статей, які створюються в інтернеті, не відповідають всім параметрам наукової роботи. Але є просте правило, виконання якого завжди виділяє статтю серед всіх інших.

Це опис того, на чому ґрунтуються судження автора. Якщо це особистий досвід, то який саме. Якщо це дослідження інших матеріалів з цієї тематики, то яких саме.

### Приклад правильного вказівки джерела, блог Backlinko

Крім того, якщо людина приводить джерела, то відразу виростає шанс, що його матеріал не буде лише рерайтом чужих думок.

## 3. Структура

Структура виконує відразу кілька завдань:

- допомагає швидко оцінити, чи є в тексті відповіді на потрібні питання користувача;
- допомагає створити блок швидкого змісту і прискорити перехід до потрібної частини сторінки;
- спрощує розуміння всього змісту сторінки для користувача.

### Приклад оформлення швидкого змісту, блог Webtris

Саме через значущості структури, дослідження семантики і аналіз топа перед написанням тексту так важливі. Вони дозволяють спроектувати скелет сторінки, на який далі можна нанизувати контент.

Інші важливі характеристики якісного змісту:

класний заголовок- коротко і зрозуміло передає зміст сторінки. Також важливо щоб, матеріал всередині відповідав обіцянці в заголовку. Тема → швидке зміст → контент. Це три контрольних точки, на яких користувач приймає рішення чи дотримуються заглиблюватися далі в ваш матеріал або кинути його і піти.

різноманітність форматів- одна й та сама думка доноситься за допомогою різних форматів контенту. Крім тексту може бути інфографіка, таблиця, формули, картинки, відео, аудіо. Про різницю між типами і форматами контенту писав у статті проконтент-стратегію.

Інфографіка на основі дослідження технічних помилок з блогу Sitechecker, приклад різноманітності контенту

## 5. Оформлення

Гарне оформлення блогу і контенту на сайті починається з розробки редакційної політики. Щоб не витратити на це багато часу, можна переробити під себередполітику Т-Ж.

Головне, щоб така політика взагалі була присутня і була людина, яка контролює її виконання. Це не обов'язково повинен бути вже прокачаний редактор.

Тепер про характеристики якісного оформлення.

Найважливіші думки в тексті виділяються. Або жирним, або курсивом, або кольором, або дизайнерськими вставками. Важливо дати користувачеві за щось зачепитися, коли він вивчає контент по верхах. А як показують дослідження Nielsen Зараз таких користувачів більшість.

Текст розділений на невеликі абзаци. Так користувач може легко повернутися до потрібної ділянки, якщо він відволікся. Йому простіше скопіювати потрібну думку, щоб зберегти, прокоментувати або кому то відправити. До поділу на абзаци відносяться і використання Буллет і нумерованих списків.

Відсутні граматичні та пунктуаційні помилки. Тут без редактора складно обійтися. Але якщо у вас блог невеликий, то поки досить буде самому повторно



редагувати контент через час і використовувати допомогу онлайн-інструментів Copyscape, Grammarly.

Унікальний дизайн. Унікальний дизайн блогу і превью картинок відразу піднімає авторитет блогу на рівень вище в очах читача. Важливо щоб картинки схематично відображали текстовий зміст статті, категорії. Це все спрощує навігацію і допомагає недосвідченому користувачу швидше зрозуміти основи нової теми.

Серед сайтів по SEO одним з кращих прикладів цього прийому виступає Backlinko зі своїм хабом і блогом.

Ми на блозі Sitechecker теж заморочили над цим. Дизайнер вивчав зміст кожної статті і придумував як схематично передати сенс сторінки в одній картинці. В результаті зараз весь інтернет розтягує їх для своїх публікацій.

Єдині правила оформлення на всіх сторінках. Цього теж складно досягти без редактора. Різні відступи, лапки, різне застосування курсиву і жирного - все це виїдає очі. Якщо редактора найняти не можете, почніть приводити свої пости в порядок самі, проводячи аудит сторінок одна за однією, починаючи з самих трафікових.

## 6. Розмітка

Якщо читачі діляться вашим контентом, то відсутність Open Graph розмітки або помилки в її змісті можуть серйозно знизити конверсію в переходи за цим посиланням.

## 7. Оцінка читачів

Люди люблять читати відгуки і оцінювати популярність матеріалу, так як це знижує ризик витратити час, гроші даремно. Показуйте статистику переглядів, Шейру, коментарів, так, щоб її не треба було шукати.

Приклад хорошого відображення соціальних доказів

Як знайти хорошого автора для свого сайту

Тепер перейдемо до пошуку хороших авторів. Вам потрібно або знайти вже відбувся експерта, або розпізнати самородок, якого можна швидко виростити в профі. А можна робити і те й інше разом.

Найголовніше - вам потрібен той, хто добре розбирається або швидко розбереться у вашій ніші. Така людина не обов'язково повинен бути хорошим підприємцем, редактором, SEO-шником.

Найняти відбувся експерта

- Взяти у конкурентів. Відберіть найкращі блоги конкурентів. Складіть список авторів, які писали контент для блогу, але не працюють в штаті компанії. Запропонуйте їм написати матеріал для вашого блогу.

- Пошукати в Google. Використовуємо такі пошукові запити: кращі фахівці з ..., навчання ..., конференція по ....

- Пошукати в соцмережах і на форумах. У експерта може навіть не бути сайта або він буде слабо розвинений. Але в залежності від ніші і особистих переваг він може розвивати свій YouTube, Twitter, Facebook, LinkedIn, Pinterest, Quora, Instagram, профіль на форумах, подкаст. Не лінуйтеся подивитися всюди.

- Пошукати на біржах фріланса. У багатьох є уявлення, що на біржах фріланса працюють тільки новачки і середнячків. Це упередження може зіграти вам на руку. Там є профі.

- Пошукати на онлайн- і офлайн-конференціях. На хороших конференціях організатори приділяють велику увагу нетворкінгу. Вони пропонують заповнити свої дані і пояснюють як можна знайти цікавих для вас співрозмовників. Серед таких людей ви напевно зможете знайти і готового профі і початківця фахівця, який шукає шляхи самореалізації.

виростити експерта

Щоб виростити експерта в ніші необов'язково його брати штат, хоча так ефективність і буде вище.

рекрутинг

На цьому етапі важливо виявити здатного, мотивованого людини. Він може бути як студентом, так і фахівцем в іншій ніші, який вирішив змінити сферу діяльності. Щоб оцінити здатність кандидатів до створення контенту, спробуйте наступне.

- Дізнайтеся, чим він захоплюється, хто його улюблених авторів контенту, що йому подобається в творчості цих авторів. По тому куди людина

дивиться, що бачить і цінує, можна багато чого сказати про нього самого. Щоб розпізнати талант, потрібно самому володіти талантом.

- Дайте йому пару сторінок вашого сайту і запитайте у нього якась мета сторінок. Оцініть, наскільки швидко він зможе розібратися в цьому і наскільки потрапить в точку.

- На тих же кількох сторінках з контентом запропонуйте йому провести короткий аудит. Нехай дасть свої поради щодо поліпшення контенту на сторінці. Нехай виявить всі помилки: і в змісті, і в оформленні.

- Дізнайтеся на які етапи він буде ділити процес створення матеріалу, після отримання першого завдання. Важливо побачити наскільки він добре розуміє ключові етапи хорошого дослідження і підготовки тексту.

- Сплатіть виконання тестового завдання. Дайте написати статтю для вашого сайту на якусь просту тематику в якій можна досить швидко розібратися.

Ще в процесі підбору кандидатів потрібно пояснити, що ми чекаємо від кандидата, в який бік будемо його розвивати і як буде відбуватися навчання. Вам потрібно досягти одразу трьох цілей:

- створити такі умови, щоб автор навчився реальним навичкам вашої сфери;
- навчити його створювати контент високого рівня;
- допомогти йому виростити особистий бренд.

Весь процес розвитку автора можна поділити на такі етапи.

1. Показати кращі блоги у вашій ніші і донести, що саме в них вам подобається. Якщо таких блогів немає, можна показати просто хороші блоги, неважливо з якої сфери. Ще краще, якщо ви заздалегідь розробляєте редакційну політику і включіть в неї кращі практики і популярні помилки. Хороший приклад такої інструкції тут.

2. Скласти контент-план. Добре, якщо він у вас буде ще до найму автора в штат. Але крім складеного вами плану, у автора в процесі роботи будуть регулярно виникати нові ідеї того про що потрібно писати і що може бути цікаво аудиторії. Постійно оновлюйте контент-план його ідеями, навіть якщо по такій темі немає багато семантики.

3. Скласти план навчання навичкам вашої сфери. Ваш автор повинен стати профі, а не тільки зображувати його. Для цього йому потрібно відвідувати конференції, проходити курси і практично застосовувати всі отримані знання. Якщо сайт про подорожі - то він не повинен сидіти на місці, якщо сайт про кулінарію - то він повинен готувати.

4. Скласти план розвитку особистого бренду автора. Реєстрація та кросспостінг в соцмережах, участь в оффлайн і онлайн дискусіях, розміщення гостьових постів на авторитетних майданчиках, проведення вебінарів. Чим раніше ви почнете розвивати публічні профілі тим краще.

5. Влаштувати змагання. Це найкраще. Немає нічого продуктивніше для творчості, ніж бажання перемогти. Якщо у вас буде кілька авторів, то влаштуйте змагання між ними. Якщо автор тільки один, знайдіть кого на схожій позиції в конкуруючій компанії і перетворите його в віртуального суперника.

#### Технологія розповсюдження контенту

В процесі розвитку автора буде також розвиватися і ваша технологія поширення контенту. У цьому розвитку є кілька ключових правил, які завжди потрібно пам'ятати.

Основна помилка всіх творців хорошого контенту - вони не доносять його до всієї можливої аудиторії.

Щоразу після створення одиниці контенту (статті, відео, аудіозаписи, картинки) і первинного її посіву в доступних каналах перед вами стоїть вибір - перейти до створення нової одиниці контенту або продовжити просувати існуючу.

Вибір цей нелегкий. Бо якщо у вас ще немає налагодженого процесу донесення контенту до аудиторії, то дії ваші будуть хаотичні, є ризик витратити час даремно і втратити мотивацію вашого зростаючого автора.

Значить, потрібна технологія, яка позбавить вас від цього гнітючого вибору. І потрібна вона якомога раніше, з трьох причин.

1. Кількість тем для висвітлення і творчий потенціал творця в конкретній ніші обмежений.

Так, це так, щоб вам не говорили тренера по успішному успіху. І це завжди потрібно пам'ятати. Якщо ви досягнете піку зібраної аудиторії тоді, коли у вас 80-90% матеріалів буде вже створено, то незабаром вам нічого не залишатиметься, крім як постити старі Використання матеріалів [povynar.com.ua](http://povynar.com.ua) дозволяється індустрії.

Чим раніше ви досягнете піку передплатників, яких можете залучити доступними вам способами, тим краще у вас буде коефіцієнт утримання цієї аудиторії, коефіцієнт вірального зростання і кількість органічних посилань, побудованих на нові матеріали. Потрібно давати людям скуштувати хороші речі поступово.

2. Найбільшою технології поширення контенту потрібен час на еволюцію.

Ви не знаєте який канал поширення контенту буде більш ефективним, який контент для якого каналу краще підійде, скільки буде коштувати передплатник, клієнт з кожного каналу, де краще працюватимуть механізми вірального зростання. Хіба що хтось із конкурентів вже повністю розкрив свою технологію і ви плануєте її повторити. А так, чим раніше ви зробите технологію, тим раніше вона вийде на максимальну ефективність.

3. Чим більше даних, тим краще наступна одиниця контенту.

Ваш контент - продукт. І правила розвитку контенту такі ж як правила розвитку інших продуктів. Вам потрібен зворотній зв'язок від аудиторії, аналітика її поведінки, тестування різних типів змісту і форм контенту. Вам потрібен холодний душ, який зможе вилікувати вас або автора, який для вас пише, від зіркової хвороби.

## **2.2. Аналіз існуючих програмних рішень для аналізу текстів**

Які параметри важливі в контент-аналізі для оцінки з точки зору користувачів

- **Читабельність.** Її краще оцінювати вручну або за допомогою сервісу [ru.readability.io](http://ru.readability.io). На параметр впливають неочевидні аббревіатури, довгі речення, дієприслівникові звороти, повтори, складні формули, великі абзаци. Сервіс

аналізує текст і визначає рівень читабельності в балах і вік аудиторії - чим він нижчий, тим зрозуміліше текст.

- Орфографія і пунктуація. Будь-який текст повинен бути грамотним. Для перевірки контенту можна використовувати сервіс "Орфограммка" або "Текст.ру", але вони не можуть замінити перевірки людиною.
- Оформлення. Хороший текст - структурований. У нього є заголовок і підзаголовок, блоки прості, немає великих абзаців, частина інформації - у вигляді списків або таблиць.
- Ілюстрації. Досить побіжно переглянути сторінку, щоб оцінити, чи несуть додаткову цінність контенту зображення. Якщо в тексті є доречні скріншоти і фотографії, він буде привабливий для користувачів.
- Посилання на пов'язані з темою сторінки. Вони потрібні для внутрішньої перелінковки і допомагають користувачам розібратися в темі.
- Зміст і відповідність назві. На жаль, жоден сервіс не зможе оцінити зміст написаного. Текст буде корисний, якщо написаний як відповідь на запит користувача, а зміст пов'язано з заголовком.

#### 2.2.1. Параметри, що важливі при контент-аналізі для оцінки з точки зору машин

Релевантність. Оцінити можна за допомогою семантичного аналізу. Так, на першому місці повинні надаватися слова і словосполучення, що характеризують основну тему. Для семантичного аналізу можна використовувати сервіси "Адвего", "Текст.ру".

Унікальність. Висока унікальність впливає на те, як текст буде прийнятий пошуковими системами. Оптимальний рівень - 85-100%. Для визначення використовуйте "Текст.ру", "Адвего", Etxt.ru, content-watch.ru. Але важливо розуміти, що вже опубліковані тексти за умови індексації мають нульову унікальність, тому в цьому випадку орієнтуйтеся на посилання після перевірки. Якщо посилання тільки на ваш сайт, все в порядку. Якщо немає - текст хтось скопіював, краще унікалізувати. До речі, деякі сервіси, наприклад, Content-Watch, дозволяють ігнорувати URL.

Входження ключових слів. Оптимізований текст - той, де є ключові словосполучення. Добре, якщо вони будуть знаходитися на перших позиціях семантичного ядра - так пошуковик зрозуміє, про що текст, і буде показувати його за запитами. Важливо стежити, щоб не виникло "переспама", зайвої оптимізації тексту - занадто частого вживання одного слова або фрази на сторінці. Для перевірки тексту можна використовувати сервіс "Тургенєв", який оцінює сторінку і радить, що виправити. Оцінка до 8 балів вважається середньою і низькою. При показниках вище є ризик потрапити під фільтр пошукових систем.

Об'єм. Тут немає якихось універсальних рекомендацій. Користувачі можуть читати і короткі тексти і лонгріди. Але важливо відстежити, що подобається вашій аудиторії більше і доопрацювати при необхідності.

Метатеги. Потрібно оцінити Title і Description, а також заголовки від h1 і нижче. Робити це доведеться вручну. Добре, якщо в метатегах будуть входження основних ключових слів - це теж допоможе пошуковій системі зрозуміти, про що текст. Унікальність метатегів можна перевірити в Serpstat.

Поведінковий фактор. Оцініть число відвідувань, час сеансу, глибину перегляду сайту. Чим вище число унікальних відвідувачів, відсоток конверсії, час сеансу - тим краще.

Видимість в пошукових системах. Весь контент має бути видно в пошукових системах. В іншому випадку користувачі не побачать ваш сайт, а значить, доведеться витратити гроші на додаткову рекламу.

Чим більше репоста в соцмережах, тим більше довіри. Досить подивитися статистику сторінки, щоб оцінити цей параметр.

### **2.3. Описання системи для аналізу текстів**

Існує кілька основних критеріїв, згідно з якими можна оцінити якість матеріалів на вашому сайті (блозі). В рамках даної статті ми розберемося, як проводити всебічну перевірку існуючих матеріалів на відповідність критеріям якісного контенту.

#### **1. Унікальність**

Перша вимога, якій повинен відповідати ваш контент всередині сайту, - унікальність. І це зовсім не означає, що вам потрібно замінити всі звичні слова на їх дореволюційні аналоги, або ж намагатися описати мовою Достоевського процедуру вибору, скажімо, зимової гуми для автомобіля. Подібного роду унікальність точно не зіграє вам на руку

Можна умовно поділити унікальність на два типи:

- Абсолютна (унікальність на просторах інтернету) .Для визначення унікальності тексту можна використовувати такі онлайн-сервіси, як Content WatchабоAdvego.

Бажаний показник унікальності - від 80 до 100 відсотків.

- Відносна (унікальність всередині сайту) Унікальність контенту, а також вмісту метатегів всередині сайту є одним з важливих чинників пошукової оптимізації. Від неї значною мірою залежатиме ранжування сторінок сайту в пошуковій видачі.

Для аналізу сайту на предмет дублювання контенту усередині сайту вам знадобиться Google Search Console і краулер, в якості якого я буду використовуватиNetpeak Spider. В GSC повідомлення про дублях тих чи інших сторінок усередині сайту будуть з'являтися у міру їх появи. Netpeak Spider ж допоможе вам контролювати дублювання контенту в режимі реального часу.

Програма дозволить вам знайти:

- повні дублікати сторінок,
- дублікати текстового вмісту тега <body>,
- дублікати заголовків H1 і метатегів Title і Description.

## 2. Орфографічна і лексична грамотність

Думка користувача про ваш професіоналізм буде формуватися під впливом відразу декількох факторів, і одним з них може стати ваша грамотність (або безграмотність) - орфографічна, синтаксична і лексична. Якщо у вас немає можливості звертатися за допомогою до професійного редактору, який прибере з тексту все зайве і виправить всі помилки, скористайтесяГлавредомабоТЕХТ.RU.

До речі, якщо ви працюєте також з англійськими матеріалами, рекомендую скористатися сервісомGrammarly.



### 3. Структурування тексту

Структурування інформації за допомогою заголовків другого-шостого порядку, нумерованих і маркованих списків не тільки робить текст візуально більш привабливим, але і полегшує сприйняття його користувачем. До того ж, виходячи з численних досліджень, заголовки H1 і H2 грають не останню роль в SEO і враховуються пошуковими системами.

Якщо у вас на сайті всього кілька десятків сторінок з текстовим вмістом, перевірка не загрожує обернутися справжнім нічним кошмаром довжиною в кілька тижнів. Однак, якщо сайт налічує тисячі тематичних сторінок, то краще все ж скористатися автоматизованим рішенням. У ролі такого рішення може виступити згаданий вище краулер.

Для перевірки наявності заголовків того чи іншого порядку, пошуку дубльованих заголовків, а також занадто довгих H1 (більше 65 символів за замовчуванням) вам слід діяти відповідно до наступного алгоритму:

1. Запустіть Netpeak Spider.
2. Відкрийте вкладку «Параметри» на бічній панелі.
3. Перейдіть до пункту «Контент» → «Заголовки H1-H6». Відзначте пункти «Вміст H1», «Заголовки H1» і «Заголовки H2». Якщо основний контент сайту - складні і об'ємні матеріали, які повинні бути структуровані за допомогою заголовків нижчого порядку (H3-H6), відзначте також відповідні пункти для аналізу.

4. У полі «Початковий URL» введіть адресу сайту. Якщо вас цікавить аналіз всіх сторінок сайту, слід просто прописати домен. Якщо потрібно просканувати, наприклад, тільки сторінки блогу (і адреси розділу і сторінок всередині нього починаються однаково - скажімо, website.com/blog), введіть у рядок пошуку адресу розділу, а потім в розділі «Налаштування» → «Основні» включите опцію «Перевіряти тільки всередині розділу».

Після закінчення сканування ви побачите таблицю з результатами такого вигляду:

У стовпчиках, що відповідають обраним заголовкам (тих, що ви відзначили в «Параметрах»), ви побачите їх кількість на кожній сторінці сайту (або на всіх сторінках розділу), а також дані про довжину H1.

В ідеалі на кожній сторінці має бути один заголовок H1 (довжиною не більше 65 символів в середньому) і кілька заголовків H2. Загальна кількість заголовків того чи іншого порядку не має значення з точки зору SEO, але відіграє значну роль в структуруванні контенту, як ми вже писали вище.

Розібравшись з даними про заголовках в основній таблиці, переходимо на бічну панель. Там на вкладці «Звіти»→ «Помилки» ви побачите зведення по помилках, зокрема помилок, пов'язаних з заголовками H1.

#### 4. Обсяг тексту і вмісту метатегів

Як би не були хороші ті три пропозиції, якими ви заповнили порожнечу посеред сторінки, вони залишаються всього-на-всього трьома реченнями. Незважаючи на те, що якість цінується на порядок вище кількості, останнім все ж має велике значення для оптимізації контенту всередині сайту. Це стосується як основного вмісту всередині тега <body>, так і метатегів Title і Description.

Для перевірки обсягу контенту (і його наявності як такого) я скористаюся вищезазначеним Netpeak Spider. Однак, перш ніж описати послідовність перевірки, розберемося з рекомендованими показниками обсягу контенту і довжини вмісту метатегів:

- Основне вміст: не менше 500 символів.

Якщо у вас контентний проект, і всі матеріали повинні бути більш-менш об'ємними, ви можете задати значення мінімально допустимого обсягу в розділі «Налаштування» → «Обмеження».

- Title: не менше 10 і не більше 50-70 символів в середньому.
- Description: для головної сторінки - від 50 до 230 в середньому, для всіх інших сторінок - від 50 до 320 символів.

Перевірка проводиться таким чином:

1. Запускаємо Netpeak Spider.
2. Вводимо адресу сайту в полі «Початковий URL».

3. На бічній панелі відкриваємо вкладку «Параметри». У пункті «Head Теги» вибираємо перші чотири параметри, в пункті «Показники» відзначаємо «Розмір контенту» (кількість тексту в блоці <body> з пробілами) та «Кількість символів» (кількість тексту в блоці <body> без пробілів).

5. За підсумком сканування в таблиці будуть зібрані дані по кожному з аналізованих показників, а на вкладці «Звіти» → «Помилки» - зведення з проблем, пов'язаних із занадто маленьким або занадто великим обсягом.

#### 5. Ключові слова

Часом навіть сама класна експертна стаття, що відповідає якійсь дуже злободенною потреби аудиторії, може виявитися десь в хвості пошукової видачі. Звідси впливає необхідність як слід пробігтися по семантиці і скористатися сервісами текстової аналітики, щоб поліпшити матеріал для ранжирування в органічній видачі за ключовими словами. Я присвятив цьому окремий матеріал, так що прооптимізації контенту під цільові запити можна детально прочитати тут.

#### 6. Вхідні і витікаючі посилання

Додаючи черговий неймовірно корисний і цікавий матеріал на свій сайт, не забудьте подбати про те, щоб цей матеріал був знайдений. Якщо ж контент не буде виявлений користувачами і роботами пошукових систем, його цінність буде прагнути до нуля. Що можна зробити, щоб виправити ситуацію? Відповідь проста: залишити в потрібних місцях заклики на ваш контент, а саме:

- в блоці з останніми записами на головній;
- в блоках «схожі записи» під іншими публікаціями;
- в блоці з закріпленими постами (візуально виділений блок на головній, слайдер);
- в блоці з «останніми постами»;
- в тексті тематично пов'язаних матеріалів.

Тепер ви напевно знаєте, як будете робити з новими публікаціями в майбутньому, але не варто забувати про матеріали, які ви встигли наплудити в минулому. Щоб з'ясувати, на яких сторінках у вас немає входять (з інших

сторінок) і вихідних (на інші сторінки) посилань, знову відкрийте вікно краулер і виконайте ряд нехитрих махінацій:

1. Зайдіть в розділ з настройками програми і відзначте параметри, як показано на зображенні нижче.

2. В «Параметрах» на бічній панелі виберіть шаблон з мінімальним набором параметрів. Активуйте параметри, які стосуються пункту «Посилання».

3. Введіть адресу сайту в полі «Початковий URL» і запустіть сканування.

Якщо за підсумком сканування на сайті будуть знайдені сторінки, на яких немає внутрішніх посилань, або ж на які не веде жодна посилання всередині сайту, Netpeak Spider відзначить їх як сторінки з помилками. У першому випадку висвітиться помилка «PageRank: відсутні вихідні посилання», у другому - «PageRank: відсутні зв'язку».

7. Зображення та відео

Пам'ятайте нарцісичного красеня Гастона з мультфільму «Красуня і Чудовисько»? В одній зі сцен він здивовано вигукує: «Книга? Але в ній же зовсім немає картинок »! До чого ми ведемо? Багато з користувачів, які втомилися від нескінченного потоку інформації, яка щомиті ллється буквально звідусіль, чимось схожі на Гастона.

Для тих, хто працює з контентом, це означає тільки одне: у всіх ваших постах повинні бути картинки. Фото, скріншоти, інфографіка, гифки і відео - все це робить контент більш привабливим і «дружнім» для читачів. Якщо ж у вас контент освітнього характеру, картинки і відео роблять матеріал більш наочним і інформативним, розкриваючи все те, що складно відобразити в тексті.

Щоб перевірити, чи на всіх тематичних сторінках сайту є зображення, проскануйте сайт за допомогою краулер. Процедура виконується за аналогією з попередніми, з тією лише різницею, що єдиний параметр, який вам в даному випадку потрібен, - це «Зображення» (пункт «Контент»).

Коротко про головне

Якість контенту і його оптимізації грають важливу роль як для його потенційних читачів, так і пошукових систем, які прагнуть показувати ті матеріали, які будуть найкраще відповідати запитам користувачів. Всього можна

виділити кілька основних аспектів оптимізації контенту з точки зору його якості, простоти сприйняття, а також відповідності вимогам пошукових систем:

- унікальність,
- орфографічна і лексична грамотність,
- структурування тексту,
- обсяг тексту і вмісту метатегів,
- ключові слова,
- вхідні і витікаючі посилання,
- зображення і відео.

Перевіряти і оптимізувати контент по кожному з пунктів можна за допомогою різних онлайн-сервісів (Главред, Advego, Content Watch, TEXT.RU, Grammarly), платформи Serpstat, Google Search Console і краулер Netpeak Spider.

#### **2.4. Висновки до розділу**

На оцінку сайту може піти багато часу, тому вибирайте для аналізу найважливіші категорії. Контент повинен бути корисний користувачам, але в той же час високо оцінюватися пошуковими системами, тому краще використовувати обидва види контент-аналізу.

РОЗДІЛ 3  
РЕАЛІЗАЦІЯ ПРОГРАМНОЇ СИСТЕМИ АНАЛІЗУ КОНТЕНТУ  
WEB-САЙТІВ

**3.1. Робота з пошуковими фразами**

Для роботи з модулем необхідно додати в проект пошукові фрази. Кожна пошукова фраза являє собою конкретний пошуковий запит, який користувач міг би ввести в пошукову систему. Наприклад, «кімнатні рослини» або «купити стару іномарку», або «створення сайтів» і т.д. Залежно від тематики вашого сайту, ті чи інші фрази будуть для вас цільовими, тобто користувачі, які шукають саме ці фрази складають вашу аудиторію. Якщо на вашому сайті встановлена система статистики (наприклад, *Google Analytics*) То пошукові фрази краще взяти з неї, це будуть ті фрази за якими користувачі вже знаходили ваш сайт в пошукових системах.

Можна ввести цільові фрази в відповідне поле по одній на рядок або вставити їх з буфера обміну і натисніть кнопку «Далі».

На наступному кроці майстра створення проекту ви повинні задати цільові веб-сайти, тобто ті сайти, позиції яких за раніше заданим фразам ви б хотіли відстежити. Це можуть бути ваш сайт або сайти, або сайти ваших конкурентів. Однак як мінімум одна з цілей повинна бути позначена як «Своя мета».

Необхідно натиснути кнопку «Додати мету» і введіть *URL* сайту, опустивши «*http: //*» частина. Потім виберіть його тип: свій це ресурс або конкуруючий. Натисніть ОК, щоб додати мета в список.

Кафедра КСУ				НАУ 21 19 29 000 ПЗ			
Виконав	Сакура В.І.			Реалізація програмної системи аналізу контенту web-сайтів	Літера	Аркуш	Архів
Керівник	Артамонов Є.Б.				Д	38	60
Консульт.					123 СІ 436		
Норм. контр.	Тупота Є.В.						
Зав. Каф.	Литвиненко О.Є.						

Після додавання всіх необхідних цілей, натисніть «Далі».

Тепер потрібно вибрати джерела даних про позиції, тобто самі пошукові системи, положення сайту в яких ви хочете контролювати.

У випадяючому списку вибрати пошукову систему, наприклад, *Google*. Потім в таблиці параметрів даної пошукової системи вкажіть домен (*Domain*). Інші параметри потрібні для більш тонкої настройки пошуку, і їх, як правило, можна не чіпати.

Так, наприклад, якщо ви хочете здійснити пошук в Яндекс-Україна, то вибрати слід домен *.UA* (Україна).

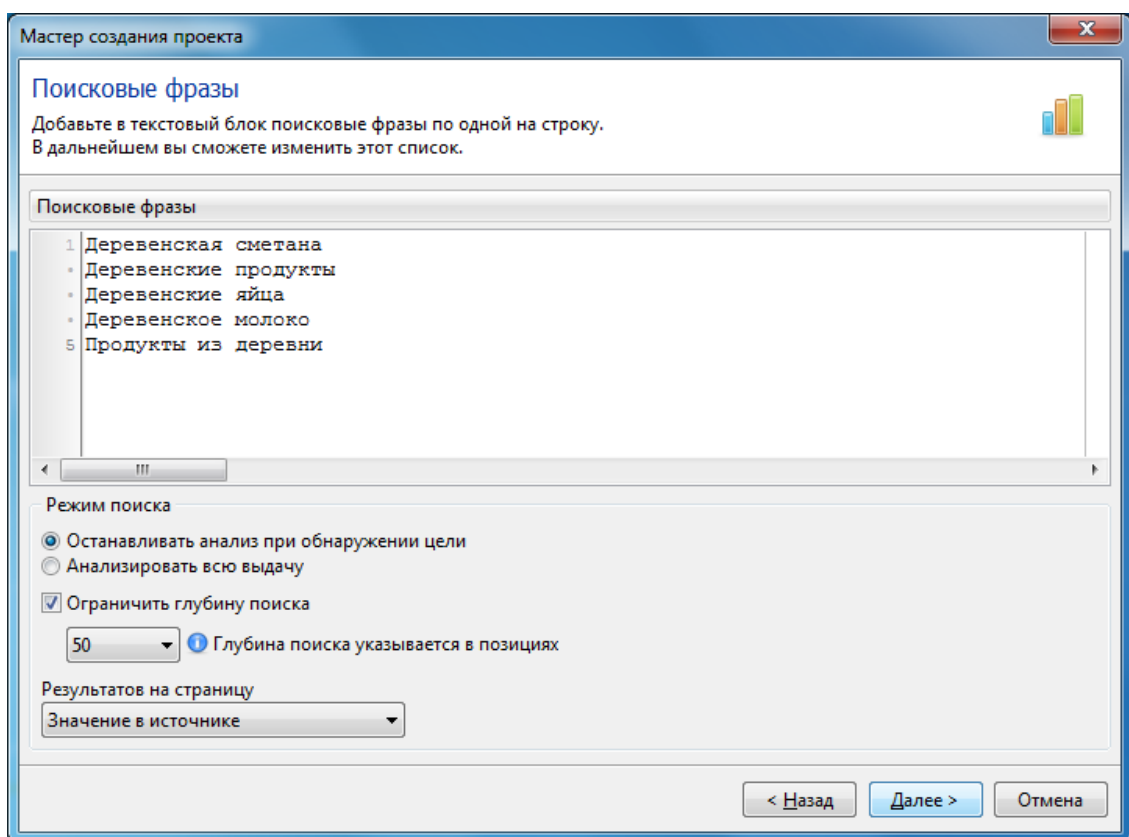


Рис. 3.1. Работа мастера створення проекту аналізу контенту сайту

При бажанні, заданий набір налаштувань можна зберегти у вигляді профілю.

Натисніть «Далі» і перевірте правильність всіх даних, після чого завершіть майстер. Програма запропонує вибрати файл для збереження проекту і запише всі внесені зміни.

Для виконання перевірки необхідно обрати команду меню «Проект | Оновити» або просто натиснути відповідну кнопку на панелі інструментів (рис. 3.2).

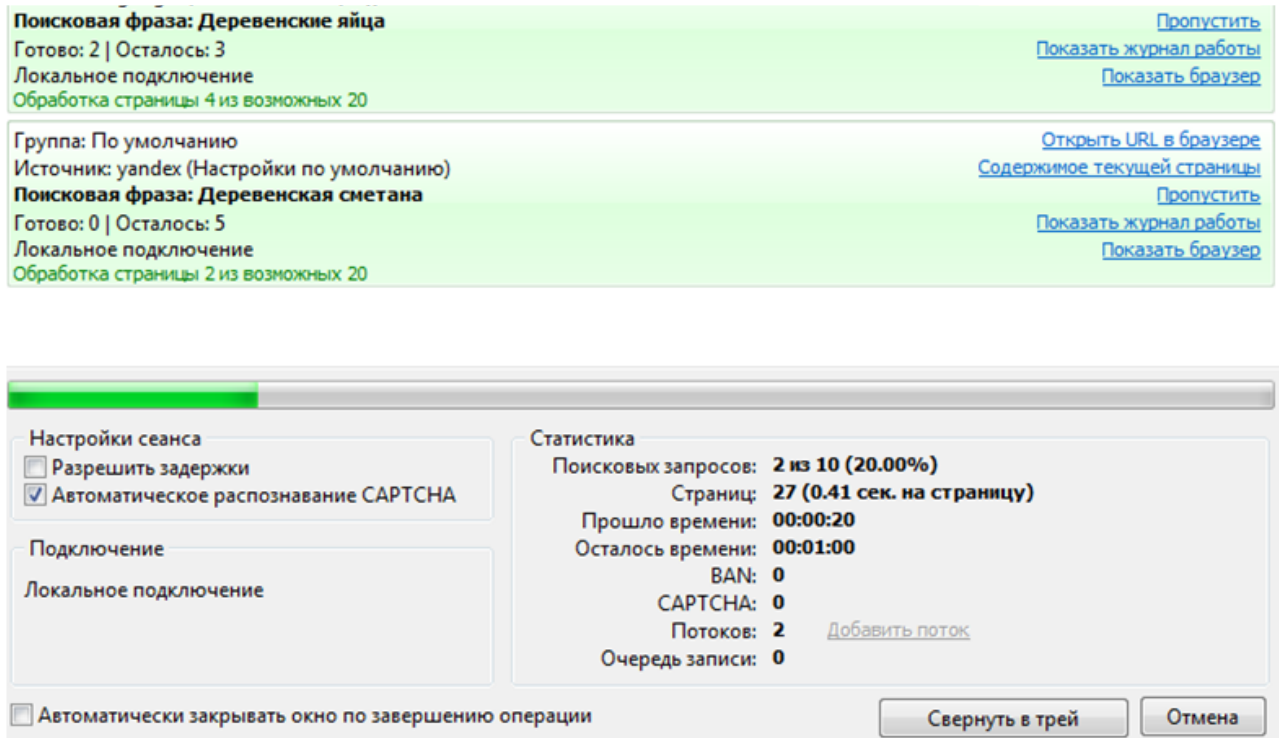


Рис. 3.2. Роботи системи аналізу сайту

### 3.2. Інтерфейс головного вікна системи

Інтерфейс головного вікна програми розділений на три частини. Зверху знаходиться рядок меню і панель інструментів. Зліва знаходиться дерево проекту. У центральній частині знаходиться область звітів.

Більшість операцій доступні як з головного меню програми так і з контекстного меню дерева проекту і таблиці звітів (рис. 3.3).

Для роботи з системою необхідно призначити мету (рис. 3.4).

*URL* цільової сторінки – адреса цілі, наприклад *example.com*. Важливо відзначити, що без особливої необхідності не потрібно вводити нічого крім імені домена, наприклад введення *www.example.com* може привести до небажаного ефекту, коли пошукова система індексує сайт без вказівки "www". У цьому випадку мета визначена не буде.



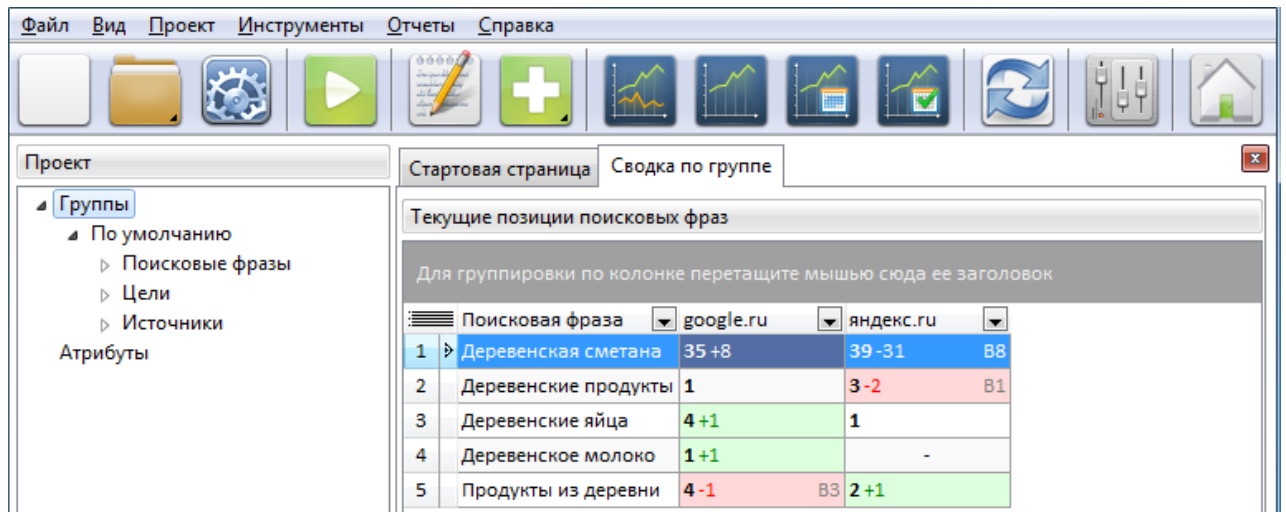


Рис. 3.3. Інтерфейс головного вікна системи

редактор мети

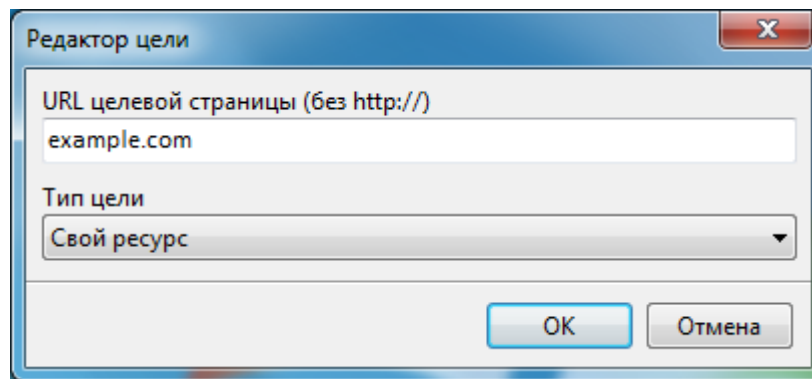


Рис. 3.4. Призначення мети

Префікси цілей – до цілей при необхідності можна додати префікси:

– префікс «!» – це точна відповідність введеної цілі, наприклад якщо вказати! *Example.com*, то буде визначена тільки мета *example.com*, *example.com/page.html* визначена не буде;

– префікс «#» – точна відповідність домену, наприклад якщо вказати # *example.com*, то будуть впізнані мети *example.com*, *example.com/page.html* і *example.com/path/page.html*, але не буде визначена мета *www.example.com*.

Тип цілі- в програмі є два типи цілей: Свій ресурс і конкурує ресурс. Необхідно відкрити вікно «Управління цілями», для цього знайти команду з меню Проект -> Управління цілями, або з контекстного меню дерева проекту виділивши мета або список цілей (рис. 3.5).

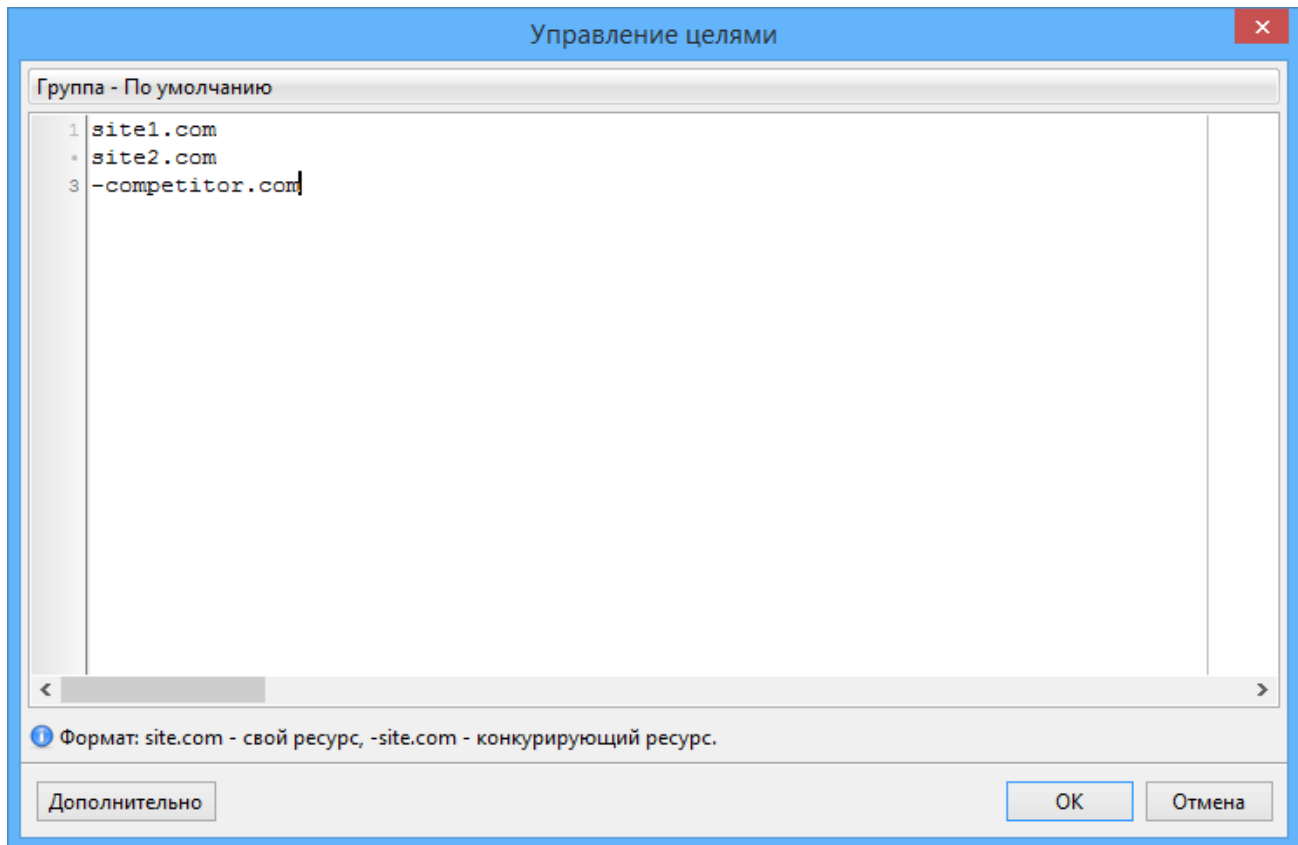


Рис. 3.6. Вікно призначення цілей

У текстовому полі можна додавати, видаляти або переміщати цілі.

Додати мету можна через меню Проект -> Додати мета. Або через контекстне меню дерева проекту, перебуваючи на списку цілей. Редактор Мети викликається подвійним клацанням по цілі в дереві проекту. Видалити мету можна через контекстне меню дерева проекту, виділивши конкретну мету.

У списках параметрів є параметри з позначкою "змінювати не рекомендується", ці параметри відносяться до фільтрів, відтинає видачу за мовою \ регіону, і користувачі при звичайному пошуку їх практично не використовують. У більшості пошукових систем присутній настройка "Результатів на сторінці", її зміна так само може вплинути на якість видачі, наприклад якщо при 10 результати на сторінку сайт знаходиться на 5-й позиції то при 100 результати на сторінку він може бути на 7-му або 3-м місці. Враховуйте це при налаштуванні пошукових систем.

Приклад налаштування пошукової системи *Google* на рис. 3.7.

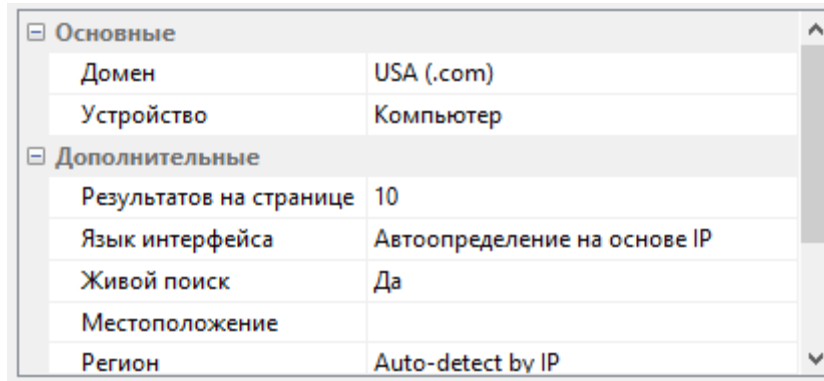


Рис. 3.7. Приклад налаштування пошукової системи *Google*

Додаткові налаштування доступні в дереві проекту з контекстного меню статистики (рис. 3.8).

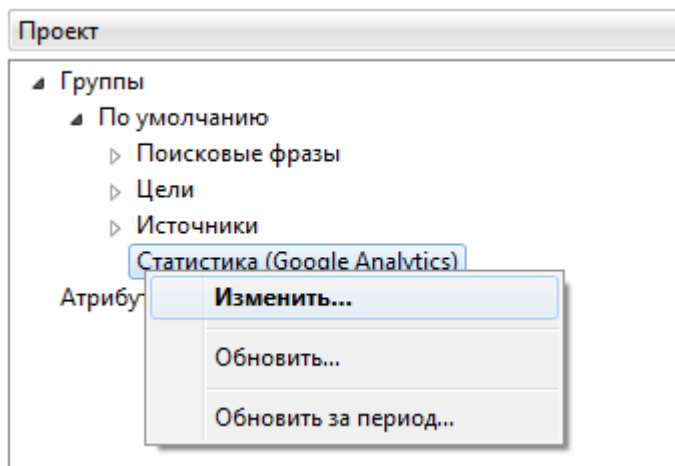


Рис. 3.8. Додаткові налаштування

Оновлення статистики відбувається автоматично разом з оновленням позицій. Статистика оновлює дані з затримкою в 2 дня. Наприклад: дані за 1-е число будуть отримані 3-го числа. Це зроблено для того щоб виключити завантаження в повному обсязі оброблених даних.

Якщо проект оновлюється не щоденною, то програма автоматично завантажить інформацію за пропущені дні. Таким чином виключаються пропуски в статистичних даних.

Після включення статистики в звітах по позиціях з'являються додаткові колонки (Відвідини, Глибина, Час на сайті, Відмови).

Дані статистики можна отримати за будь-який період. Для цього з контекстного меню статистики потрібно вибрати пункт Оновити за період.

У діалозі потрібно вказати часовий період для отримання статистичних даних. Менеджер облікових записів викликається кнопкою знаходиться поруч з полем введення логіна. У вікні можна вибрати наявну або додати новий обліковий запис (рис. 3.9).

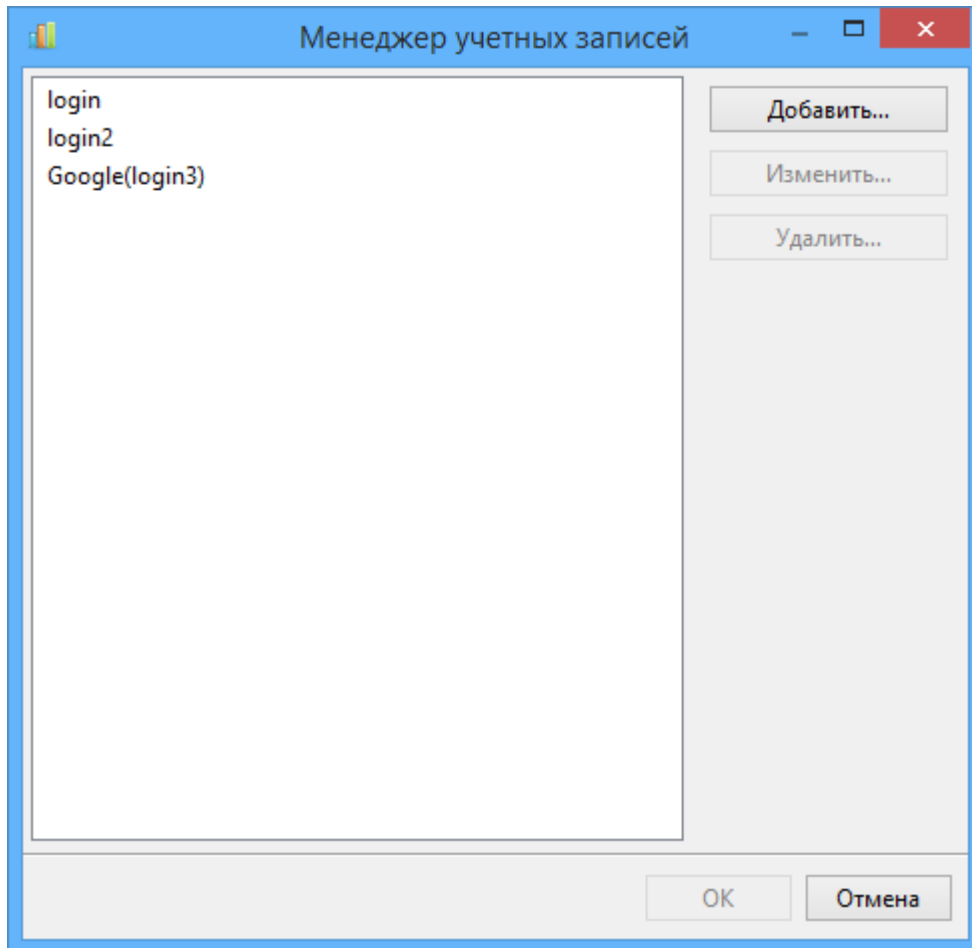


Рис. 3.9. Вікно менеджера облікових записів

### 3.2.1. Робота з атрибутами

Атрибути потрібні для того, щоб персоніфікувати проект під потреби користувача. В першу чергу атрибути – це призначені для користувача колонки в звітах, але за допомогою них можна групувати і \ або стилізувати (колір, шрифт і т.д.) пошукові фрази в звітах. Значення атрибутів можна редагувати як в редакторі пошукової фрази, так і безпосередньо в самій таблиці звіту.

Типи атрибутів:

– строковий - звичайний рядок (наприклад, Коментар);

– цілочисельний - число без дробової частини (наприклад, кількість відвідувань);

– речовий- число з дробовою частиною (наприклад – Гроші);

– логічний- галочка Так \ Ні (наприклад – Бере участь в контекстній рекламі чи ні);

Перераховуються- набір іменованих елементів (наприклад – Тематика (Телевізори \ Пилососи \ Телефони))

Створити атрибут можна з контекстного меню дерева проекту (рис. 3.15).

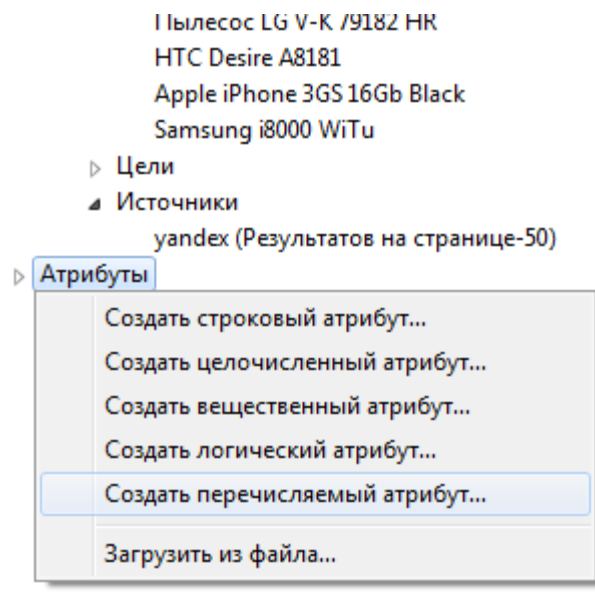


Рис. 3.10. Меню створення атрибуту

Вікна роботи редактору атрибутів для різних типів атрибутів представлено на рис. 3.10. Кожному атрибуту можна задати значення за замовчуванням, а так само участь у звітах (бачимо \ невидимий \ відсутня). Невидимий і Відсутня відрізняються тим що при перегляді звіту невидимі атрибути можна відобразити. Після створення атрибута він стає доступним для редагування в редакторі пошукової фрази,

Редатор строкового атрибута

Название  
Атрибут1

Видимость в отчетах  
Видим по умолчанию

Значение по умолчанию

OK Отмена

А)

Редатор целочисленного атрибута

Название  
Атрибут1

Видимость в отчетах  
Видим по умолчанию

Значение по умолчанию

OK Отмена

Б)

Редатор логического атрибута

Название  
Атрибут1

Видимость в отчетах  
Видим по умолчанию

Значение по умолчанию

Не определено  
 Истина  
 Ложь

OK Отмена

В)

Рис. 3.10. Вікна редагування атрибутів для: А) строкового атрибута, Б) цілочисельного атрибута, В) логічного атрибута

По одному з атрибутів можна зробити угруповання пошукових фраз. Так само один з перераховуються атрибутів може бути заданий як стилю пошукових фраз в звітах, тоді стиль атрибута передається пошуковій фразі, а сам атрибут відображається без стилю. Атрибути групування і стилю називаються Керуючими атрибутами. Встановити керуючі атрибути можна внастройках проекту.

### 3.2.2. Робота зі звітами у системі аналізу контенту сайтів

У програмі передбачено чотири режими звітів: зведення по групі, зведення по джерелу, зведення по днях і зведення по датам. Кожен режим зручний для аналізу отриманих даних в тому чи іншому розрізі. Створити звіт можна через меню «Звіти».

Даний режим показує зведену інформацію по групі пошукових фраз. При виклику команди «Звіти | Зведення по групі »вам буде запропоновано вибрати групу, по якій ви хочете побудувати звіт. У нашому випадку, ми маємо тільки одну групу «За замовчуванням», а значить і звіт автоматично буде побудований для неї.

Всі звіти в програмі мають схожу структуру, тут для прикладу наведено звіт Зведення по джерелу.

У центрі звіту розташована таблиця з даними, знизу – графік. Вміст таблиці розбите на три логічні частини: пошукові фрази, атрибути, якщо вони присутні в проекті, і дані звіту.

Сортування доступне з будь-якої колонці звіту. Для сортування по колонці клацніть по її заголовку, для зворотного сортування по колонці клацніть по її заголовку повторно, для скасування сортування клацніть по заголовку колонки, натиснувши клавішу *CTRL*. Для сортування за кількома колонкам клацніть по заголовку колонки, натиснувши клавішу *SHIFT*.

Для групування рядків за даними колонки досить перетягнути заголовок колонки в сіру область над таблицею. На малюнку показаний приклад перетягування колонки ТОП.

Після того як колонка буде перенесена, рядки в звіті згрупуються за даними колонки.

Для скасування угруповання необхідно перетягнути заголовок колонки назад в таблицю.

Щоб встановити фільтр на дані колонки, потрібно натиснути на кнопку в заголовку колонки і в випадаючому списку вибрати пункт Свій фільтр.

У діалозі потрібно вказати умову і значення фільтра (рис. 3.17).

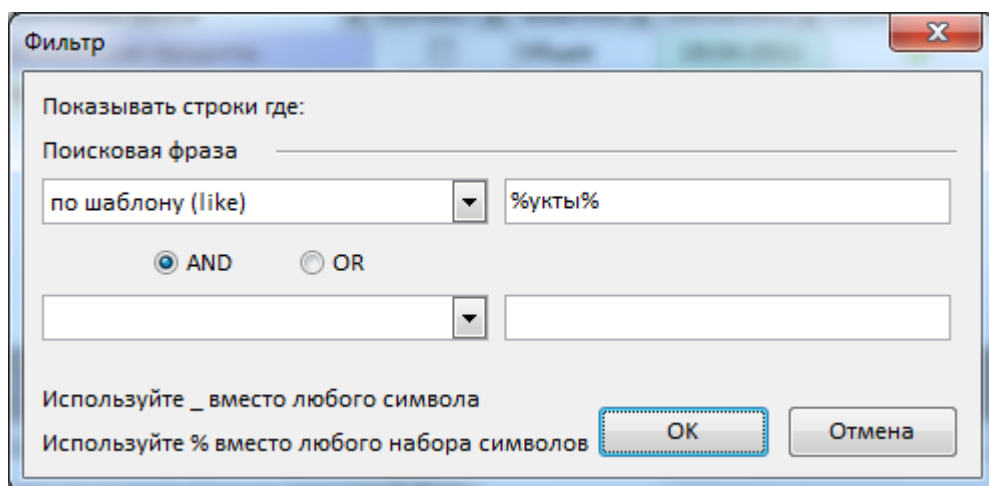


Рис. 3.11. Вікно налаштування фільтра для звіту

У прикладі умовою є відповідність шаблону, а значенням -% укты%.

В результаті в звіті залишилися тільки ті пошукові фрази в частині яких присутній підрядок "укты". Так само в таблиці звіту знизу з'явилася панель фільтра. Щоб прибрати фільтр, потрібно натиснути на хрестик в панелі фільтра.

Після введення фільтра, в таблиці залишаються пошукові фрази в частині яких присутній введений підрядок. Фільтр працює в кожному звіті окремо. Значення фільтра зберігається між сеансами роботи.

Діалог пошуку викликається з меню пошук-> Знайти (*CTRL + F*). Пошук ведеться по виділеній в звіті колонці. Для продовження пошуку використовується команда пошук-> Знайти далі (*F3*).



Щоб підвищити інформативність звітів, непотрібні колонки можна приховати, а потрібні показати. Для цього потрібно клацнути в верхній лівий кут таблиці, далі галочками встановити видимість колонок.

Видимість колонок зберігається між сеансами роботи для кожного проекту індивідуально.

Щоб експортувати дані звіту потрібно з контекстного меню таблиці вибрати команду Експорт звіту (рис. 3.11).

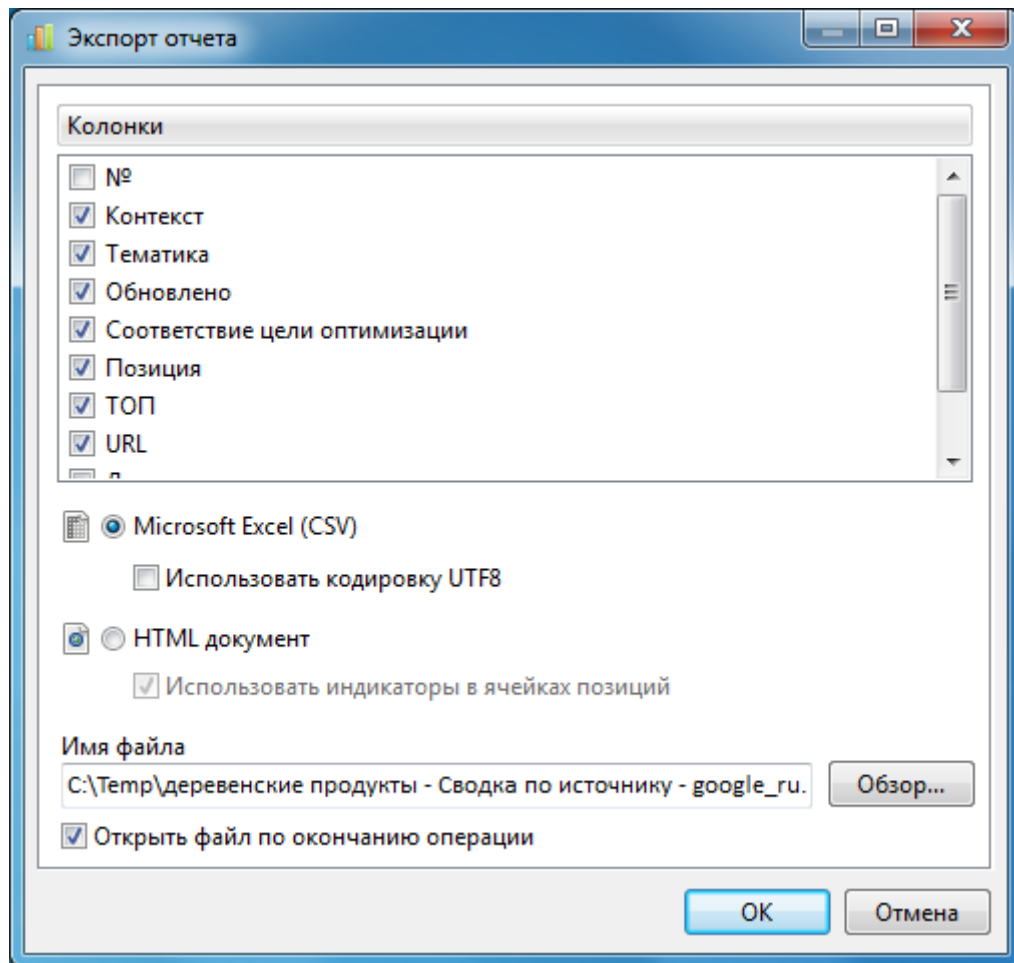


Рис. 3.11. Вікно налаштування експорту звіту

У діалозі експорту потрібно вибрати формат *CSV* або *HTML*. Галочками вибираються колонки які повинні бути присутніми в звіті.

Щоб виділити або зняти виділення з рядка клацніть на ній, утримуючи *CTRL*. Для виділення діапазону рядків клацніть на першій та утримуючи клавішу *SHIFT* клацніть на останньому рядку.

Комплексний звіт, містить як загальну інформацію по всьому проекту так і основні звіти по всіх групах і всіма джерелами.

На сусідніх вкладках знаходяться зведення по групах, всередині яких містяться зведення за джерелами.

ТОП 3 – ТОП 100- процентне співвідношення фраз входять до відповідного ТОП, якщо це значення змінилося з попереднього поновлення то різниця відображається в дужках праворуч від значення. Процентне значення в топах накопичувальне, в ТОП10 входять фрази з першої позиції по десяту, в ТОП20 з першої по двадцятку і т.д.

Звіт викликається з головного меню Звіти -> Зведення за проектом.

Якщо в налаштуваннях програми встановлена опція Побудувати звіт Зведення за проектом після поновлення, то після поновлення проекту звіт буде викликаний автоматично.

Звіт з налаштуванням періоду викликається з головного меню Звіти -> Зведення за проектом (з налаштуванням періоду). Або кнопкою на панелі інструментів Зведення за проектом, натиснувши клавішу *SHIFT*.

Бувають випадки, коли за однією і тією ж пошуковою фразою позиція в браузері відрізняється від позиції, яку визначила програма. Причин може бути декілька, розглянемо кожну:

1. Порівнюються дві різні видачі. Найпростіший спосіб перевірки програми виглядає так: відкрити в браузері сторінку ПС, Ввести пошукову фразу, далі знайти свій сайт на одній зі сторінок видачі, чекаючи що в програмі позиція повинна бути та ж, а якщо не так, то вважати це помилкою програми, але помилкою це вважати передчасно. Справа в тому, що пошукові системи враховують безліч параметрів, і від кожного з них залежать результати видачі. Налаштування видачі в браузері повинні бути в точності такими як вони задані під час налаштування джерела пошукової системи в програмі. Більш того, результати видачі залежать від попередніх пошуків в цій ПС.

Як відкрити в браузері ту ж видачу яку аналізує програма без врахування попередніх пошуків? Для цього в програмі є команда Відкрити пошукову видачу

Команда Відкрити пошукову видачу доступна з контекстного меню клітини позиції. В результаті виконання команди буде відкритий браузер з чистою історією пошуку і тієї видачею, яку аналізує програма. Але навіть цих умов може не вистачити.

2. Видача не статична. Складні алгоритми ранжирування пошукових систем, плюс не припиняється ні на секунду завантаження нових даних пошуковими системами, призводять до того, що одна і та ж видача, відкрита в двох різних вкладках браузера з інтервалом в декілька секунд, може істотно відрізнятись.

Беручи до уваги все перераховане вище, для перевірки програми залишається тільки один варіант – потрібно отримати доступ до точних копій сторінок видачі, які пошукова система видала на запит програми, що і робить команда Контрольне визначення позиції.

Команда Контрольне визначення позиції доступна з контекстного меню клітини позиції.

Після запуску команди відкриється вікно поновлення з одним завданням – визначення позиції за поточною фразі з поточного джерела. Після того як завдання буде виконано (при цьому в проект результати не заносяться, позиція в проекті залишиться колишня) в браузері відкриється звіт з результатами перевірки.

Після виконання команди, в правій частині вікна ви можете бачити зведену таблицю позицій по кожній пошуковій фразі обраної групи. У рядках таблиці розташовуються фрази, а в стовпчиках – джерела. На перетині відповідного рядка і стовпця знаходиться поточна позиція по цій фразі в даній пошуковій системі (рис. 3.12). Графіки різних кольорів відповідають різним пошуковим системам. Даний режим звітності зручний для швидкої оцінки ваших позицій відразу у всіх обраних джерелах.

Файл Вид Проект Инструменты Отчеты Справка

Проект

Группы

- По умолчанию
  - Поисковые фразы
    - Деревенская сметана
    - Деревенские продукты
    - Деревенские яйца
    - Деревенское молоко
    - Продукты из деревни
  - Цели
  - Источники
- Атрибуты

Стартовая страница Сводка по группе

Текущие позиции поисковых фраз

Для группировки по колонке перетащите мышью сюда ее заголовок

	Поисковая фраза	google.ru	яндекс.ru
1	Деревенская сметана	43	8
2	Деревенские продукты	1	1
3	Деревенские яйца	5	1
4	Деревенское молоко	2	-
5	Продукты из деревни	3	3

Рис. 3.12. Звіти системи аналізу контенту сайту

Для зручності подання інформації, її можна відфільтрувати, а також згрупувати, перетягнувши заголовки стовпців в відповідну область. Крім того, дані будь-якого звіту можна експортувати в *Excel*, клацнувши правою кнопкою миші і вибравши в контекстному меню пункт «Експорт звіту».

Режим «Зведення за джерелом» відображає детальну статистику по одному конкретному джерелу. При виборі даної команди з меню «Звіти | Зведення по джерелу », ви повинні вказати групу (в нашому прикладі вона одна) і бажаний джерело, тобто пошукову систему.

У правій частині вікна ви побачите список перевічених пошукових фраз, а також наступну інформацію по кожній з них:

- дата останньої перевірки;
- поточна позиція цієї фрази в обраному джерелі;
- входження в ТОП, тобто в перші  $N$  результатів видачі, наприклад, ТОП 3, ТОП 10 і т.д.;
- *URL* сторінки сайту, що займає цю позицію.

У нижній частині вікна також відображається графік зміни позиції обраної фрази в даній пошуковій системі.

Режим «Зведення по джерелу» зручний для перегляду конкретних результатів в тій чи іншій пошуковій системі, а також для відстеження досягнення поставлених цілей, наприклад, входження в перші три результату на сторінці, або в першу двадцятку і т.д.

Щоб відкрити звіт «Зведення по днях», скористайтеся однойменною командою в меню «Звіти».

При виклику команди відображається діалог, в якому слід задати групу, джерело, а також бажаний діапазон відображення історичних даних. Зазначена інформація буде показана внизу вікна у вигляді відповідного графіка, а також в текстовому вигляді в основному вікні.

Даний звіт дозволяє зробити довільний зріз історичний даних по позиціях в декількох точках. Іншими словами, він дозволяє скласти часткове уявлення про динаміку зміни позицій з тих чи інших пошукових фраз. При виклику команди програма запропонує вибрати групу і джерело даних, а також задати список дат, за якими повинен бути виконаний зріз. Внесіть кілька дат в список за допомогою кнопки «Додати» поруч з полем для завдання дати. Після закінчення натисніть кнопку «ОК»

### 3.2.3. Імпорт даних в системи аналізу контенту сайтів

Імпорт даних відбувається з *CSV* файлів, або з буфера обміну, якщо в ньому містяться дані в форматі *CSV* (наприклад – скопійовані з *Excel* або з таблиці в веб браузері і т.д.). Однією з колонок вихідних даних обов'язково повинна бути колонка зв'язки – пошукова фраза. Першим рядком в даних можуть бути присутніми імена колонок.

Підготовка даних з *Google Analytics*

1. Зайдіть в акаунт *Google Analytics*.
2. Перейдіть до звіту Джерела трафіку -> Ключові слова.
3. Експортуйте звіт вибравши Експортувати -> *CSV* для *Excel*.
4. Відкрийте файл в *Excel*
5. Скопіюйте необхідне.

Дані для імпорту готові!

### 3.3. Висновки до розділу

В даному розділі було розглянуто систему аналізу контенту сайтів

Бібліотека NLTK містить готові списки стоп-слів для різних мов.

Отримаємо список сто-слів для російської мови:

Слід зазначити, що стоп-слова є контекстно залежними - для текстів різної тематики стоп-слова можуть відрізнятися. Як і у випадку зі спец.символов, необхідно проаналізувати вихідний текст і виявити стоп-слова, які не увійшли в типовий набір.

## ВИСНОВКИ

В межах дипломного проекту було реалізовано систему аналізу контентів сайтів. Базові параметри, аналіз яких був реалізований в програмі аналізу текстового контенту наступний:

1. Об'єм. Наявність оригінального заповнення сайту дозволяє індексувати сторінки, це особливо актуально, якщо мова заходить про Яндексі. Щоб сайт потрапив в індекс, на ньому повинен бути присутнім текст, чий обсяг становить не менше 500 знаків. Оптимальний варіант - текст обсягом понад 2000 символів. При розміщенні тексту потрібно обов'язково брати до уваги тип виділеного Інтернет-ресурсу. Якщо мова йде про віртуальному магазині, тоді текст товарної картки повинен складатися з 500 символів, для блогів же зазначеного обсягу недостатньо.

2. Регулярність оновлення. Від того, як часто проводиться оновлення контенту, знаходиться в залежності ступінь довіри пошукових систем до ресурсу. Наприклад, якщо в віртуальному магазині оновлення товарного асортименту та збільшення контенту відбувається приблизно раз на рік, варто очікувати, що з плином часу сайт втратить колишні позиції і трафік.

3. Вхідження ключових фраз. При проведенні аналізу, варто приділити увагу цьому фактору, що має значення. Він впливає на ранжування, виступає в якості основи для визначення ступеня відповідності ресурсу запитам користувачів в пошуковій системі. Експерти рекомендують не обмежуватися законом Ціпфа, не довіряти даним, які втратили актуальність. На одну тисячу символів краще використовувати не більше 1-2 ключових слів, розміщуючи, в середньому по 3-4 штуки на одній сторінці. Це актуально, якщо обсяги статей перевищують 5 тисяч символів. Перед тим, як приступити до роботи над інформацією, варто виконати аналіз віртуальних ресурсів, які видаються в ТОПі пошукової системи за конкретним запитом.

4. Переспа ключових слів. Тут трохи важко говорити про якийсь конкретний показник щільності, так як він знаходиться в залежності від теми

сайту. Проте, експерти кажуть про те, що щільність ключових слів не повинна бути більше 5%. Якщо одне ключове слово або фраза були впроваджені по 10 і більше разів, а положення сторінки в пошуковому сервісі так і не змінилося, ймовірно, це пов'язано з надмірною оптимізацією сторінки, і потрапляння під фільтр. Іноді фахівці стикаються зі сторінками, де застосовується більше двох десятків ключових слів. Якщо взяти до уваги мета-теги і alt / title, то число включень досягало 30 штук.

5. Застосування форматування не за призначенням (це актуально для таких тегів, як: b, strong, i). Були часи, коли веб-майстри виділяли ключові слова або фрази курсивом, або напівжирним шрифтом. Цей метод надавав позитивний вплив на ранжирування сторінки. Однак, після того, як було проведено вдосконалення пошукових алгоритмів, даний спосіб втратив колишню актуальність. Тут є ще важливий момент: прогресивні фахівці відмовляються від будь-яких виділень, однак, це теж не зовсім правильно. Зазвичай виділення використовується тоді, коли потрібно привернути увагу користувачів, які відвідали ресурс, до будь-якої фрази або ж підкреслити найбільш важливу думку.

6. Унікальність текстової інформації. При виконанні аналізу контенту, слід приділити увагу унікальності. При самостійному створенні тесту не можна на сто відсотків бути впевненим в його оригінальності. Не можна забувати і про те, що існує злодійство контенту, яке зустрічається не так рідко. Якщо конкуруюча фірма прочитає статтю, і вона їй сподобається, то вони можуть абсолютно спокійно копіювати її, щоб використовувати для заповнення своєї сторінки. Якщо їх ресурс функціонує довше, велика ймовірність того, що пошукові системи віддадуть перевагу саме йому. При цьому текст на сторінці, де він був розміщений спочатку, порохують плагіатом і не будуть індексувати.

7. Унікальність графічного матеріалу. Велике значення для трафіку з пошукового сервісу має інформація в графічному вигляді. Найбільш популярні сервіси, призначені для пошуку в Мережі, індексують і видають в ТОП ті Інтернет-ресурси, які містять оригінальні матеріали. Це дає можливість користувачам, хто зацікавився вашими матеріалами, переходити на необхідну сторінку прямо з результатів пошуку по картинках.



8. Структура контенту. Основні вимоги, які пред'являються до структури: чіткість, легкість сприйняття. Неприпустимо, щоб текст був представлений єдиним полотном. Важливо включати списки і виноски, ділити текстову інформацію на абзаци і розділи, забезпечені заголовками.

9. Помилки, допущені в текстових фрагментах. Досить часто недостатньо досвідчені веб-майстри самостійно роблять тексти для заповнення Інтернет-ресурсів, проте, вони можуть допускати серйозні недоліки. З іншого боку, професійні копірайтер також не завжди здатний надати матеріал, що відрізняється високою якістю. Звичайно, скоєні помилки не можуть не вплинути на стан ресурсу. виправлення текстових помилок - це робота для коректорів. Вони здатні виконати аналіз контенту, виявити існуючі орфографічні, а також синтаксичні помилки.

10. Надлишок мета-тегів - це одна з найбільш поширених помилок. Потрібно поставитися до проблеми з особливою увагою. Досить часто зустрічаються тексти, де є безліч мета-тегів, що включають ключі, вставлені в якості перерахування, розділені тільки комами. Це абсолютно неприпустимо. Суть таких тегів, як title і description полягає в смисловому розкриттям Інтернет-сторінки в якості узагальнених відомостей, представлених там. Важливо, щоб Мета-теги могли бути прочитані.

11. Дубляж мета-тегів. Проблема актуальна для багатьох порталів, незалежно від ступеня їх популярності. Зазвичай з нею стикаються на Інтернет-сторінках з параметрами GET і пагінацію. Така проблема може бути вирішена за рахунок класичних сервісів URL і robots.txt. Якщо при аналізуванні контенту були виявлені певні електронні сторінки з дублями, то їх буде досить замінити на унікальні, щоб вирішити питання.

12. Заголовки H1. Це найбільш груба помилка, яка включається в категорію «технічних». На жаль, на сьогоднішній день вона дуже часто зустрічається. Суть полягає в тому, що на одній сторінці розташовується кілька заголовків першого рівня. Це може спричинити за собою санкції з боку пошукового сервісу.

13. Заголовки H2-H3. Такі заголовки не мають особливого значення для SEO порталу, однак, вони беруться до уваги при деякій релевантності

віртуального ресурсу. Іноді можна зустріти сайти, які містять недозволітельно багато заголовків. Це дуже великий недолік, який потрібно виправляти. При формуванні контенту, не рекомендується впадати в фанатизм: цілком достатньо 5 заголовків H2, H3. Звичайно, можна додавати більше, головне, не перестаратися. Також важливо простежити за тим, щоб вони не були включені в заголовок сайту і в навігаційну систему.

14. Довжина заголовків H1-H6. Немає певних вказівок на цей рахунок. Однак, експерти припускають, що будь-які заголовки здатні зробити негативний вплив на ранжування, якщо включають кілька пропозицій. У деяких випадках за розмірами їх можна порівняти з цілим абзацом. Краще використовувати лаконічні заголовки, що містять основну інформацію. Небажано, щоб їх довжина була більш пари рядків. Не варто забувати про те, що заголовки потрібні для того, щоб привертати увагу користувачів, які відвідали сайт. Важливо, щоб при першому ж огляді він зрозумів: це те, що йому буде корисно.

15. Застосування в тезі title ключових слів / фраз. Такий заголовок значиться при визначенні релевантності веб-сторінки запиту, заданого в пошуковій системі. Саме з цієї причини при виконанні аналізу контенту сайту, цим пунктом варто приділити пильну увагу. Багато, щоб заголовок title включав ключове слово або фразу.

16. Наявність атрибутів «alt» і «title» для графічного матеріалу. Перший з перерахованих атрибутів може бути корисний, якщо користувач не в змозі розглянути зображення. При цьому title виникає за рахунок наведення курсору на картинку, по суті, це і є її найменування. У перерахованих атрибутах слід прописати релевантний (тобто, основний) ключ сторінки, що дозволить істотно підняти релевантність за запитом в пошуковій системі.

17. Релевантність тега «title» до контенту. Не можна забувати про те, що цей тег впливає на релевантність сайту, отже, важливо, щоб він відображав суть вмісту Інтернет-порталу. К, наприклад, title містить дані щодо цін, знижок, безкоштовної доставки, при цьому сам контент не включає перераховані обставини.

## СПИСОК БІБЛІОГРАФІЧНИХ ПОСИЛАНЬ ВИКОРИСТАНИХ ДЖЕРЕЛ



## Додаток А

### Лістинг коду аналізатора тексту