

## Сети ЭВМ и телекоммуникации (весенний семестр)

### Лабораторная работа №1. «Разработка программы АНАЛИЗАТОР САЙТОВ»

#### Задание:

1. Разработать клиентское Desktop-приложение, которое позволяет по протоколу HTTP заходить на любую WWW-страницу по начальному заданному адресу и далее осуществляет на этой странице действия согласно варианту задания. Разрешается использовать готовые библиотеки, например, QT для организации работы с сокетами.
2. Результаты работы программы сохраняются в БД, структура которой создается и разрабатывается самостоятельно в зависимости от варианта.

#### Полный список операций.

**ВНИМАНИЕ!!!** Все их не нужно реализовывать. Из набора этих операций строится конкретный вариант задания.

Номер операции	Название операции	Описание операции
O1	Поиск внутренних ссылок на странице	Осуществляется поиск всех ссылок на странице, которые принадлежат тому же домену, что и адрес страницы анализа. Сохраняется адрес ссылки и ее содержимое в текстовом виде, которое находится между открывающим и закрывающим тегом А
O2	Поиск всех внешних ссылок на странице	Осуществляется поиск всех ссылок на странице, которые не принадлежат тому домену, которому принадлежит текущая страница анализа. Сохраняется адрес ссылки и ее содержимое в текстовом виде, которое находится между открывающим и закрывающим тегом А
O3	Поиск списков (UL,OL) и их элементов (LI)	В документе ищутся все списки, заданные с помощью тегов UL или OL. При нахождении списка также осуществляется поиск информации об элементах списка. Сохраняются элементы списка и информация о их принадлежности с самому списку
O4	Поиск изображений и их сохранение (IMG)	В документе осуществляется поиск изображений, которые находятся в теге IMG. Сохраняется само изображение, а также параметры его отображения (атрибуты тега: размер, рамка и т.п.)
O5	Поиск ссылок на документы (архивы, текстовые файлы, PDF файлы и т.п.) и их	В документе осуществляется поиск ссылок на объекты, которые не являются HTML страницами. Сохраняется сам документ и его MIME-

	сохранение.	тип
O6	Извлечение списка ключевых слов, которые содержатся в разделе HEAD	В документе осуществляется поиск ключевых слов, которые указаны в разделе HEAD. Список ключевых слов сохраняется таким образом чтобы отдельные слова сохранялись в БД отдельно
O7	Поиск и извлечение текстовой информации с разделением на блоки	В документе осуществляется поиск текстовой информации, которая находится внутри тегов P. Сохраняется сама текстовая информация, а также атрибуты
O8	Создание скриншота страницы	Создается скриншот страницы, который сохраняется в виде изображения. При создании изображения указывается его размер. Сохраняется само изображение, его размер и дата формирования.
O9	Поиск управляющих элементов на странице (кнопки, поля ввода и т.п.)	В документе осуществляется поиск управляющих элементов (INPUT, TEXTAREA). Сохраняются все элементы и их атрибуты
O10	Поиск заголовков (H1,...,H6)	В документе осуществляется поиск заголовков, отмеченных тегами H1,H2,H3,H4,H5. Сохраняется информация, которая содержится внутри указанных тегов, а также их атрибуты

**Примечание:**

- На оценку «**Удовлетворительно**» достаточно реализовать перечисленный в задании функционал с минимальным интерфейсом
- На оценку «**Хорошо**» необходимо реализовать графический интерфейс управления работы приложением, который позволяет:
  - Работать с несколькими потоками параллельно
  - Останавливать/запускать все/отдельные потоки анализатора
- На оценку «**Отлично**» необходимо реализовать все вышеперечисленные требования, а также добавить интерфейс управления БД с результатами работы анализатора, который позволяет:
  - Выбирать данные по названию сайта и типам элементов
  - Удалять данные по названию сайта и типам элементов

**Требования для сдачи лабораторной работы:**

1. Демонстрация работы реализованной вами системы.
2. АВТОРСТВО
3. Оформление и представление письменного отчета по лабораторной работе, который содержит:
  - Титульный лист
  - Задание на лабораторную работу
  - Структура БД
  - Листинг программы

**Варианты заданий:**

№	Список реализуемых операций
	04, 03, 01
	05, 09, 02
	07, 010, 06
	08, 03, 01
	04, 09, 02
	05, 010, 06
	07, 03, 01
	08, 09, 02
	04, 010, 06
	05, 03, 01
	07, 09, 02
	08, 010, 06
	04, 03, 06
	05, 09, 01
	07, 010, 02
	08, 03, 06
	04, 09, 01
	05, 010, 02
	07, 03, 06
	08, 09, 01