Титульник

Реферат

**Содержание**

[Введение 5](#_Toc84061580)

[1 Анализ существующих аналогов 6](#_Toc84061581)

[2 Техническое задание 6](#_Toc84061582)

[3 Разработка парсера Excel файла 6](#_Toc84061583)

[4 Разработка исполнителя SQL-запросов 6](#_Toc84061584)

[5 Разработка пользовательского интерфейса 6](#_Toc84061585)

# Введение

# Анализ существующих аналогов

# Техническое задание

# Разработка парсера Excel-файла

Парсинг – это метод индексирования информации с последующей конвертацией ее в иной формат или даже иной тип данных. Парсинг позволяет взять файл в одном формате и преобразовать его данные в более удобоваримую форму, которую можно использовать в своих целях.

В нашем случае парсинг это процесс обработки данных из Excel-файла. Он подразумевает анализ текста, вычленение оттуда необходимых материалов и их преобразование в подходящий вид. Благодаря парсингу можно находить на страницах файла небольшие клочки полезной информации и в автоматическом режиме их оттуда извлекать, чтобы потом переиспользовать.

Вычленяемая информация в нашем случае это будет та информация, которая нужна для вставки в SQL-скрипт. Представлена она может быть как отдельными полями, так и табличными значениями.

## 3.1 Дескриптор Excel-файла

Под дескриптором Excel-файла будем понимать ту информацию, которая описывает извлекаемые объекты из файла. Дескриптор состоит из объектов и их атрибутов. На данный момент есть три вида объектов, которые заключаются в открывающий и закрывающий тег:

* <singleValue>…</singleValue> - информация об значении, хранившемся в одном поле;
* <table>…</table> - информация об таблице, в которой хранятся значения;
* <column>…</column> - является вложенным тегом в тег table и является информацией о столбце таблицы, в котором хранятся данные.

Атрибуты в свою очередь представляются в виде токенов. Под токеном подразумевается значение и короткого описания этого значения (имя токена). Токен представляется следующим образом:

Имя токена : значение токена;

Каждый токен объекта должен быть отделен точкой с запятой, при этом, если имя или значение имеют пробелы и табуляцию, то они должны быть заключены в кавычки. Таким образом дескриптор объекта, который имеет N атрибутов и один вложенный тег, выглядит следующим образом:

<tag>

nameToken\_1 : valueToken\_1;

nameToken\_2 : valueToken\_2;

…

nameToken\_N : valueToken\_N;

<nestedObject>

nameToken\_1 : valueToken\_1;

nameToken\_2 : valueToken\_2;

…

nameToken\_N : valueToken\_N;

</nestedObject>

</tag>

В таблицах 1-3 представлены возможные имена атрибутов тегов singleValue, table и column.

|  |  |  |  |
| --- | --- | --- | --- |
| Коды атрибутов | Допустимые значения | Описание | Если значение не указано |
| SHEET\_NUMBER | Целочисленные числа, больше 0 | Номер страницы, на которой производится поиск (отсчет страниц ведется с 1). | Поиск производится на всех страницах |
| SHEET\_NAME | Любая строка | Имя страницы, на которой производится поиск. | Поиск производится на всех страницах |
| SECTION\_NAME | Любая строка | Раздел на странице, относительно которого производится поиск (для полей с одинаковыми надписями). | Поиск производится на всей странице |
| SECTION\_BOTTOM\_LEFT | 1 или 0 | При условии, что атрибут SECTION\_NAME определен. Область снизу слева от раздела включать в поиск, рисунок 1. | Поиск не производится в левой нижней части от раздела (Аналогично значению 0). |
| SECTION\_BOTTOM\_RIGHT | 1 или 0 | При условии, что атрибут SECTION\_NAME определен. Область снизу справа от раздела включать в поиск, рисунок 2. | Поиск не производится в правой нижней части от раздела (Аналогично значению 0). |
| SECTION\_UP\_LEFT | 1 или 0 | При условии, что атрибут SECTION\_NAME определен. Область сверху слева от раздела включать в поиск, рисунок 3. | Поиск не производится в левой верхней части от раздела (Аналогично значению 0). |
| SECTION\_UP\_RIGHT | 1 или 0 | При условии, что атрибут SECTION\_NAME определен. Область сверху справа от раздела включать в поиск, рисунок 4. | Поиск не производится в правой верхней части от раздела (Аналогично значению 0). |
| FIELD | Любая строка | Текст ячейки, относительно которой ищется значение. | Аналогично пустой ячейке. |
| CODE | Любая строка | Код значения, по которому можно идентифицировать данное поле. | Значение данного объекта будет получено, но обратиться к нему будет нельзя в SQL-скрипте. |
| OFFEST\_ROW | Любое целочисленное число | Смещение по строке относительно FIELD. | Аналогично значению 0 |
| OFFEST\_COLUMN | Любое целочисленное число | Смещение по столбцу относительно FIELD. | Аналогично значению 0 |

* Одиночное значение (singleValue) – значение, которое представляет из себя значение, взятое из одного поля;
* Табличные значения (table) – значения, которые сформированы в таблицы;
* Столбец таблицы (column) -

# Разработка исполнителя SQL-запросов

# Разработка пользовательского интерфейса