# Практическая работа №4

Делегаты и события. Механизм подписки на события. Класс web-клиент, http get-запрос.

Разработать класс по анализу HTML-содержимого web-страниц заданного интернет ресурса (выбирается студентом самостоятельно), например, орг. структуры ЮУрГУ (https://www.susu.ru/ru/structure).

Анализ должен осуществляться по всем страницам, URI которых включает базовый URI ресурса (интернет-домен, например, www.susu.ru). Предусмотреть настройку максимального уровня вложенности страниц в рекурсивном алгоритме анализа, а также максимального количества просматриваемых страниц.

В класс добавить событие (event) по определению цели поиска, с передачей в его обработчик информации о названии ссылки, ведущей на страницу (т.е. имя ссылки, которое видит пользователь на базовой странице в браузере, например, «Филиалы»), URI страницы (например, https://www.susu.ru/ru/university/structure/electronics-higher-school/kafedry), уровне вложенности (например, 1) и самой цели поиска (см. разбивку задания по вариантам). Если целей на странице несколько, то событие вызывается для каждой цели. Обработчик события должен выводить информацию на консоль (или в окно) и в CSV-файл в табличной форме. CSV-файл можно открыть в Excel.

Применение событийной модели позволит отделить друг от друга алгоритм поиска данных на страницах и алгоритм отображения/вывода информации на консоль/в файл.

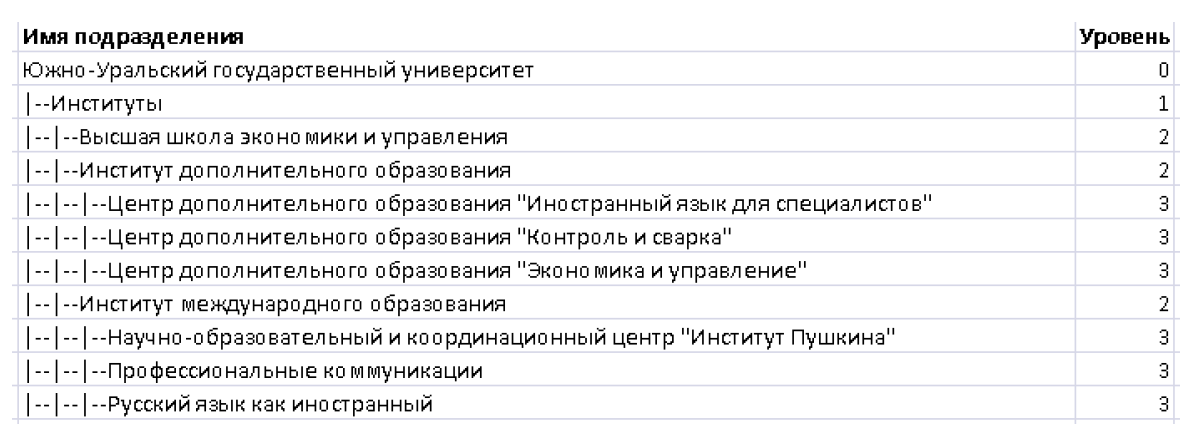


Рисунок 1. – Пример древовидной структуры в Excel

# Рекомендации

Использовать классы, реализующие HTTP-протокол, например, WebClient, HttpWebRequest, HttpWebResponse (using System.Net). Для анализа текста внутри HTML-страницы можно воспользоваться регулярными выражениями (класс System.Text.RegularExpressions.Regex).

Варианты 1 и 5

Цель поиска – URI внешнего ресурса и название ссылки

Варианты 2 и 6

Цель поиска – E-Mail на странице

Варианты 3 и 7

Цель поиска – URI картинки и подпись к ней в ссылке

Варианты 4 и 8

Цель поиска – контактная информация (адрес, телефон)