

Разработать **класс** по анализу HTML-содержимого web-страниц заданного интернет ресурса, например, орг. структуры ЮУрГУ (<http://www.susu.ru/ru/university/old-departments>).

Анализ должен осуществляться по всем страницам, URI которых включает базовый URI ресурса (интернет-домен, например, [www.susu.ru](http://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, <http://www.susu.ru/ru/university/old-departments/branches>), *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.

Имя подразделения	Уровень
Южно-Уральский государственный университет	0
--Институты	1
-- --Высшая школа экономики и управления	2
-- --Институт дополнительного образования	2
-- -- --Центр дополнительного образования "Иностранный язык для специалистов"	3
-- -- --Центр дополнительного образования "Контроль и сварка"	3
-- -- --Центр дополнительного образования "Экономика и управление"	3
-- --Институт международного образования	2
-- -- --Научно-образовательный и координационный центр "Институт Пушкина"	3
-- -- --Профессиональные коммуникации	3
-- -- --Русский язык как иностранный	3

Рис. 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).

**Вариант 1**

Цель поиска – URI внешнего ресурса и название ссылки

**Вариант 2**

Цель поиска – E-Mail на странице

**Вариант 3**

Цель поиска – URI картинки и подпись к ней в ссылке

**Вариант 4**

Цель поиска – контактная информация (адрес, телефон)