

Главный файл - Краулер.xmlz, именно его нужно запускать. В файл Domens.txt нужно добавить брошенные домены (сейчас для теста там домены по ключу *zdor*. В файле Dorvey.txt сохраняются подозрительные домены, требующие ручной проверки (слишком много файлов, как правило, в 99% это дорвей).

Файл scrapeg.txt отвечает за путь подключения скрипта full-text-rss. В папку Готовые сайты складываются скаченные сайты. Остальные файлы промежуточные.

Добавлен файл конфигурации Config.cfg. Это главная настройка отвечает за минимальную границу текста в символах. По умолчанию стоит 300 символов, программа проверяет страницу и берет количество символов, если их меньше 300, то берется следующая страница. 300 символов выглядят вот так:

Невероятный семейный проект – грузовик, трансформируемый в замок

Что такое «дом мечты»? Для кого-то это простенький, но уютный загородный домик, для других – роскошная квартира в самом сердце большого города. Человек с нестандартным мышлением мгновенно нарисует в своем воображении бескрайние пески пустыни, высокое раскидистое дерево и причудливой формы бунгало среди его ветвей. А вот супруги Джола и Джастина решили свить себе гнездышко в самом обычном грузовике. Жилые автомобили, тем более для Новой Зеландии, откуда родом пара – уже давно не новость. Однако молодоженам все же удалось удивить мир своим уникальным архитектурным изобретением, возведенным ими также собственноручно. Благодаря своей гениальной конструкции грузовик Джола и Джастина способен в считанные минуты превратиться в самый настоящий замок. Здесь есть в буквальном смысле все: оборудованные гостиная, кухня, спальня, туалет, душевая и даже терраса на крыше автомобиля. Вот оно – слияние уюта и комфорта, творчества и фантазии!

Чем ниже граница парсинга, тем больше парсится данных и захватывается мусора. Чем выше число, тем меньше данных, но и количество случайного мусора стремится к нулю. Здесь нужно подбирать число опытным путем.