

Булавко А. Г.

Белорусский национальный технический университет

Наибольшей проблемой при написании универсальных web-парсеров, является разработка архитектуры, позволяющей оптимально хранить и обрабатывать полученную информацию. Для этого может применяться подход, при котором используются цепочки действий с использованием промежуточных блоков, т.е. после того как парсер получил данные страницы сайта, он начинает к ним последовательно применять действия обработки, в результате выполнения которых из исходного текста страницы выделяются подблоки текста, и последующие действия уже могут быть применены к ним.

Рассмотрим подробнее функционирование данного алгоритма. На входе универсального парсера дан следующий текст

"произвольный_текст

<блок1><поле1>Текст1</поле1><поле2>Текст2</поле2></блок1>".

Цепочка действий состоит из восьми пунктов:

- 1) создать поле B1;
- 2) копировать в поле B1 текст тега "блок1";
- 3) создать поле T1;
- 4) копировать в поле T1 текст тега "поле1" из поля B1;
- 5) создать поле T2;
- 6) копировать в поле T2 текст тега "поле2" из поля B1;
- 7) сохранить поле T1;
- 8) сохранить поле T2;

Важным нюансом, является создание промежуточного поля B1, что позволяет вести поиск не во всем тексте, а только в тексте поля B1, т.е. фактически действия 4 и 6 применяются не к исходному тексту а к полю B1. В случае необходимости к полям T1 и T2 могут быть так же разбиты на более мелкие части. Благодаря своей гибкости данный алгоритм может быть выбран для обработки сложных текстов. Для тестирования быстройдействия алгоритма было разработано тестовое приложение.

На рисунке 1 изображена зависимость времени обработки от объема входящего текста для метода с выделением подблоков и стандартного подхода. Как видно из приведенного выше графика метод с применением промежуточных блоков проигрывает по скорости обработки на текстах маленького объема до 1Кб, однако с возрастанием объема исходного текста эффективность данного метода значительно увеличивается.

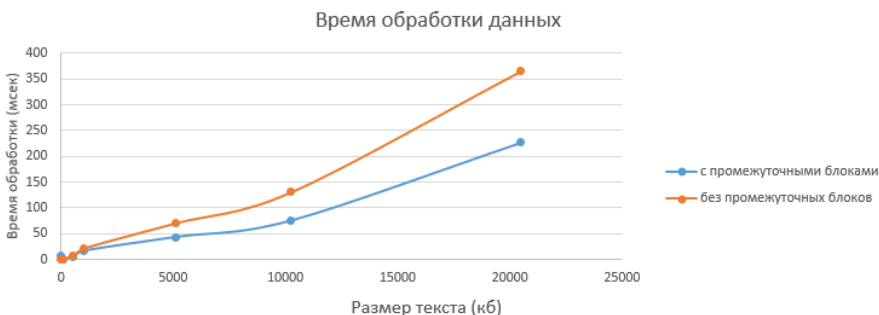


Рисунок 1. Снижение суммарных потерь активной мощности в сети в режиме наибольших нагрузок.

Литература:

1. A language independent web data extraction using vision based page segmentation algorithm [Электронный ресурс] / arxiv — Электронные данные. — Режим доступа: <http://arxiv.org/abs/1310.6637> — Дата доступа: 10.05.2015.
2. Extraction of Flat and Nested Data Records from Web Pages [Электронный ресурс] / arxiv - Электронные данные. Режим доступа: <http://arxiv.org/abs/1002.0139> — Дата доступа: 10.05.2015.

УДК 629.11

Моделирование управления колебаний подвески мобильной машины нейроконтроллером с эталонной моделью

Иванченко В.В.

Белорусский национальный технический университет

Снижение динамической нагруженности мобильных машин при движениях их в изменяющихся дорожных условиях требует использования активных элементов поддресоривания узлов и агрегатов таких объектов. Для изучения и разработки эффективных схем активного поддресоривания транспортных систем необходимы математические и программные модели, поддерживающие различные этапы имитационного моделирования колебательных процессов мобильных машин.

В данной работе представлены математические модели, описывающие динамическое состояние твердотельных систем многоопорных мобильных машин с элементами активного поддресоривания их несущих конструкций, функционирующих в условиях случайной внешней среды.