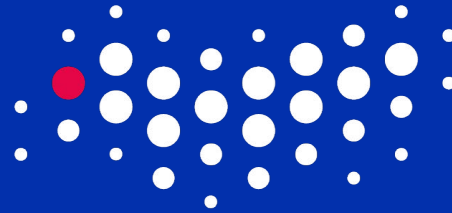


УНИВЕРСИТЕТ ИТМО



УНИВЕРСИТЕТ ИТМО

Разработка виртуального тренажера процесса пастеризации в среде Labview

Студент: Тишков И.В., Менщиков А.А.

Научный руководитель – д.т.н., профессор Гатчин Ю.А.

Санкт-Петербург 2017



- **Цель работы.** Проектирование и разработка системы противодействия парсингу веб-ресурса на основе комбинации технических подходов.
- **Актуальность** выбранной темы обусловлена увеличением активности веб-роботов, при этом методы и механизмы защиты веб-роботов так же стремительно развиваются.



Классификация веб-роботов

- **1. Любительский.** Средства автоматизации не обладают достаточными механизмами противодействия обнаружению.
- **2. Профессиональный.** Средства автоматизации содержат механизмы противодействия обнаружению.
- **3. Передовой.** Данные средства автоматизации содержат самые передовые методы противодействия обнаружению.



Методы обнаружения веб-роботов

- Бан по IP адресу
- Использование учетных записей
- Использование CAPTCHA
- Использование сложной JavaScript логики
- Динамическое изменение структуры страницы
- Ограничение частоты запросов и объемов загружаемых данных
- Отображение важных данных в виде изображений



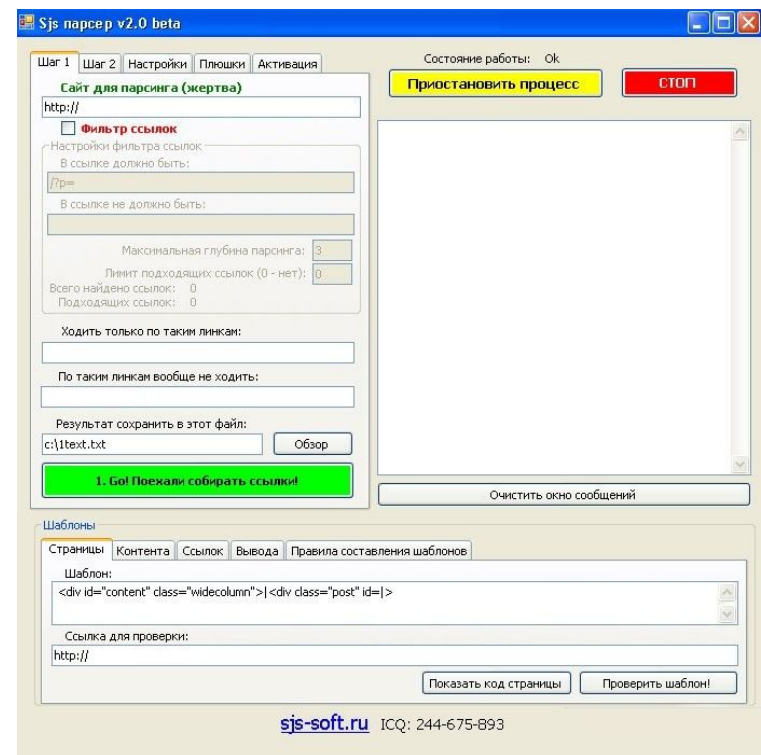
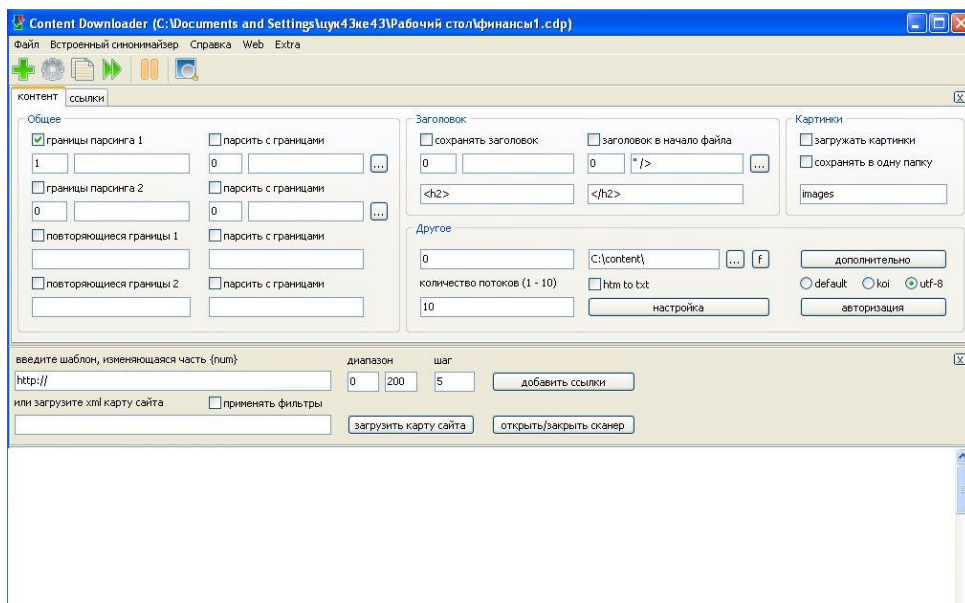
Использование JavaScript логики

The screenshot shows a web browser window with a single tab titled "Тест". The address bar displays "file:///C:/123.html". The browser's developer tools are open to the "Elements" panel, showing the DOM tree. The selected element is the <body> tag, which contains the following HTML and JavaScript code:

```
<!--?
header('Content-type: text/html; charset=utf-8');
mb_internal_encoding('utf-8');

if (isset($_POST['char'])) {
    if (mb_strlen($_POST['char']) == 1) {
        echo 'Человек';
    } else {
        echo 'Робот';
    }
}
?-->
<form method="post" action=".">...</form>
<!--?highlight_file(__FILE__);?-->
<script src="http://www.google-analytics.com/urchin.js" type="text/
javascript">
</script>
<script type="text/javascript">
_uacct = "UA-2822435-2";
urchinTracker();
</script>
```

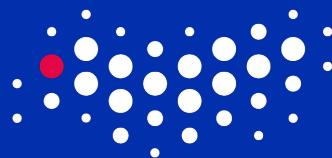






Результаты

Создан прототип веб-ресурса, включающий в себя систему противодействия парсингу, содержащую совокупность методов от учетных записей до включения ловушек в содержимое веб-ресурса. Разработанный прототип системы противодействия автоматизированному сбору информации с веб-ресурсов показал отличные результаты, сделан вывод о применимости использования включенных в него методов, которые позволили полностью исключить деятельность веб-роботов на ресурсе. Далее планируется исследование работы и производительности данной системы на крупных веб-ресурсах.



УНИВЕРСИТЕТ ИТМО

Спасибо за внимание