

А. В. Смирнов, С. Ф. Маслович

(ГГУ им. Ф. Скорины, Гомель)

РАЗРАБОКА «WEB-ПАУКОВ» ДЛЯ СБОРА СТАТИСТИКИ НА РЫНКЕ ГРАЖДАНСКОЙ АВИАЦИИ

В моей диссертации, я должен проанализировать большое количество данных с рынка авиационных запасных частей и построить экономическую модель для покупки и продажи деталей. Для сбора статистических данных, я использую поисковых роботов.

Поисковые роботы (web-пауки) – программа, являющаяся составной частью поисковой системы и предназначенная для перебора страниц Интернета с целью занесения информации о них в базу данных поисковика. По принципу действия паук напоминает обычный браузер. Он анализирует содержимое страницы, сохраняет его в некотором специальном виде на сервере поисковой машины, которой принадлежит, и отправляется по ссылкам на следующие страницы.

Программная часть разработанных мною пауков реализована на PHP, и является интегрируемой частью CRM системы vTiger. В качестве базы данных используется реляционная система управления базами данных MySQL.

Системное и программное обеспечение информационных технологий
Телекоммуникационные системы и сети

Процесс работы пауков начинается после следующих действий: пользователь заходит на сайт <http://www.bai-inc.eu> и входит в личный кабинет. После этого, он может запросить квотацию необходимых деталей. Запрос будет отправлен в CRM vTiger. После этого запускается механизм пауков. Первым этапом идет проверка были ли подобные запросы в истории, если да то возвращается результат из истории. В противном случае паук подгружает сайт ILS и получает API, из которого собирает всю полученную информацию о запрошенных деталях. Если результатов не было обнаружено, то паук продолжает выполнять ту же операцию, только на сайте Stock Market. Если после окончания этих процедур, не было получено никаких сведений, то возвращается пустой результат, а затем операторы сообщают, что такие детали отсутствуют в системе. В случае положительного результата поиска, механизм пауков начинает обработку сайтов производителей, на которых были обнаружены запрошенные детали, что бы узнать окончательную информацию о количестве и цене.

Таким образом, механизм пауков имитирует действия пользователя на сайте: загрузка страницы, установка параметров для поиска, обработка полученной информации. Преимуществом «web-пауков» является автоматическая обработка информации с сайтов производителей, что значительно ускоряет процесс поиска. Основным недостатком является то, что если разработчики изменили структуру сайта производителя, необходимо изменить структуру паука для конкретного сайта.