

ОТСЛЕЖИВАНИЕ ПОИСКОВЫХ РОБОТОВ

Стрельцов В.В., к.ф.-м.н., доц. Черных Е.П.

Национальный технический университет «ХПИ», Харьков

Поисковые роботы («веб-пауки») – программы, являющиеся составной частью поисковой системы. Их принцип действия аналогичен обычным браузерам и предназначены они для перебора страниц Интернета с целью занесения информации об их посещениях в базу данных поисковика.

Владельцы поисковых машин нередко ограничивают глубину проникновения паука внутрь сайта и максимальный размер сканируемого текста, поэтому чересчур большие сайты могут оказаться не полностью проиндексированными поисковой машиной. Статистика посещений пользователей приводится обычно для распространенных сайтов. Для получения аналогичной информации с еще нераскрученных сайтов лучше написать свой скрипт. Также страницы могут посещать вредоносные боты. Установкой пароля на странице либо требованием заполнить регистрационную форму перед тем, как получить доступ к содержимому, не всегда можно обеспечить защиту сайта.

SEO-специалист может выполнить проверку на посещение сайтов поисковыми роботами, а также зафиксировать информацию об индексации сайта в «всемирной паутине». При анализе содержимого страницы, он сохраняет его в некотором специальном виде на сервере поисковой машины, которой принадлежит, и отправляется по ссылкам на следующие страницы.

Для решения данной проблемы был предложен подход – разработка программного модуля (скрипта), с помощью которого можно отследить визиты поисковых роботов. Скрипт будет содержать необходимую информацию: дату посещения, имя бота, IP-адреса бота и страницы, которые он посетил.

Данный подход позволит контролировать индексацию последних изменений страниц на определенном сайте и определить какие роботы посещали данный сайт. Это, в свою очередь, приведет к повышению эффективности индексации сайта.

Программная реализация может быть выполнена с помощью языков программирования PHP или JS.