

УДК 004.738

## **ПОИСК И УПРАВЛЕНИЕ ПОИСКОВЫМИ КРАУЛЕРАМИ НА САЙТЕ**

**В.В. СТРЕЛЬЦОВ<sup>1\*</sup>, Е.П. ЧЕРНЫХ<sup>2</sup>,**

<sup>1</sup> *магистр кафедры вычислительной техники и программирования, НТУ «ХПИ», Харьков, УКРАИНА*

<sup>2</sup> *доцент кафедры вычислительной техники и программирования, канд. физ.-мат. наук, НТУ «ХПИ», Харьков, УКРАИНА*

*\*email: strelcovv92@gmail.com*

Интернет прочно вошел в жизнь практически каждого современного человека. Поисковые системы стали не только местом, где можно найти ответы на все вопросы, но также способом общения, взаимодействия с другими людьми.

Поисковые системы выполняют сканирование и индексацию web-ресурсов. После чего в базы поисковых систем заносится информация о сайте и страницах сайта (ссылки, изображения, тексты, видео и прочее). Для того чтобы сайт или его страницы успешно продвигались, они обязательно должны быть проиндексированы поисковыми системами. Для этого используют поисковых роботов (краулеров), каждый из которых представляет автоматизированный скрипт, имеющий свой алгоритм работы, свое конкретное задание для определенного сайта. Системы обладают большим отрядом разных роботов, которые выполняют разные задачи: ищут новые страницы, отвечают за нахождение “мертвых” сайтов и чистку поисковых данных, индексируют картинки, находят видео. Не стоит забывать про безопасность данных на сайте при появлении сайта в сети Интернет. С этой целью для поисковых роботов, которые сканируют интернет, файл robots.txt предоставляет важную информацию. Он позволяет роботам с большей эффективностью сканировать сайт, приступить сразу к индексации действительно важной информации на сайте и обходить места с закрытым доступом для повышения безопасности сайта.

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-адресов бота и страницы, которые он посетил. Это позволит определить, какие роботы посещали данный сайт, или какую его страницу. Полученный отчет поможет определить страницы, на которые нужно запретить доступ для последующих посещений робота. Также возможно, внедрение данного скрипта в систему Telegram Bot API для его более удобного использования.

Предложенный подход позволит управлять поведением робота, правильно оформить файл robots.txt и повысить безопасность посещаемого сайта.