

Алгоритм получения сети веб-страниц на основе поиска в ширину

ПАХОМОВ ИЛЬЯ АЛЕКСАНДРОВИЧ

Омский государственный технический университет (Омск), Россия
e-mail: utmo5t@yandex.ru

ЮДИН ЕВГЕНИЙ БОРИСОВИЧ

Институт математики им. С. Л. Соболева СО РАН (Омск), Россия
e-mail: udinev@asoiu.com

Рассматриваются вопросы обеспечения эффективности поиска информации в сети Интернет путем сбора и систематизации данных о структуре, содержании и взаимосвязи веб-ресурсов. Рассматриваются программные средства для осуществления сбора данных – веб-роботы. Приводится описание реализованного в ходе работы веб-робота: его компонентов, функций, принципа работы и получаемых результатов. Приводится алгоритм, используемый для сбора данных, основанный на реализации классического алгоритма поиска в ширину. Представляются результаты анализа данных, собранных в ходе работы веб-робота, а также характеристики веб-графа, сформированного роботом на их основе. Проводится сравнение результатов с данными о сети веб-ресурсов, полученными другими авторами. Приводятся характеристики скорости сбора данных программой и делаются выводы о возможностях ее расширения для получения более детальных результатов.