УДК 681.3

А.А. НИКИТИН, Л.А. НИКИТИНА, канд. техн. наук

АНАЛИТИЧЕСКОЕ ИССЛЕДОВАНИЕ СИСТЕМ ПОИСКА И ОБРАБОТКИ ИНФОРМАЦИИ В ИНТЕРНЕТ

В настоящее время, несмотря на насыщенность рынка поисковых систем, актуальными являются исследования, связанные повышением эффективности поиска информации в глобальных и локальных сетях. Проблемами современных поисковых систем в глобальной сети [2] являются возрастание общего объема материалов, расположенных в Интернете, переход к новым форматам представления информации (например, XML); увеличение количества нетекстовых материалов (MP3, RealVideo и т.д.). В проведенном исследовании охвачены поисковые системы AltaVista, Excite, FAST Search, Go (Infoseek), Google, Lycos и Northern Light.

В состав автоматической поисковой системы обычно входят следующие основные компоненты [1]: агент (паук или кроулер), который перемещается по сети и собирает информацию; база данных, содержащая всю информацию, собираемую пауками; поисковый механизм - интерфейс пользователя для взаимодействия с базой данных. Эффективность выполнения поиска во многом зависит от алгоритмов и методов, реализованных компонентами системы. Каждый поисковый механизм имеет собственный набор правил сбора и обработки документов.

В ходе исследования была выполнена сравнительная характеристика наиболее популярных поисковых систем на основе ключевых факторов сбора, анализа информации, ранжирования документов и борьбы с поисковым спамом. Для выполнения анализа и сравнения поисковых систем был выделен перечень характеристик: способы наполнения базы данных основных автоматических поисковых систем; методы выполнения индексации и ранжирования документов; способы обработки спама. Была предложена система оценок поисковых систем по выделенным характеристикам.

Список литературы: 1. П. Храмцов «Информационно-поисковые системы Internet», // Открытые Системы, № 3, Москва, 1996. 2. Robert Filman, Feniosky Pena-Mora «The compare of search systems of Internet», // ComputerWeekly, 40_98, p.15.