

## **ИССЛЕДОВАНИЕ МЕТОДОВ ПОИСКА И ОТБОРА ДАННЫХ В ИНТЕРНЕТ**

*аспирант М.В. Мясин, к.т.н, с.н.с. Н.С. Косило, НТУ "ХПИ", г. Харьков*

В условиях непрерывного развития информационного пространства и постоянного роста объемов информации возникает проблема поиска, отбора, анализа и качественной обработки (консолидации) необходимой информации. Решение этой проблемы с использованием сети Интернет предполагает оперирование огромными массивами разноплановой информации, содержащими миллиарды терабайт данных, циркулирующих в сети.

Для осуществления поиска необходимых данных используются специальные механизмы поиска и отбора данных на основе определенных алгоритмов по заданным условиям, реализованные в поисковых программах. Однако на практике, вследствие не идеальности поисковых программ, на конкретный поисковый запрос может быть получен довольно объемный набор разрозненных данных, процесс анализа которых и отбора необходимых из них становится достаточно трудоемким.

Проведенный сравнительный анализ наиболее популярных поисковых систем (таких, как Google, Яндекс, Mail.ru, Рамблер, Апорт, и др.) с точки зрения возможностей поиска и отбора необходимых данных позволил отметить такие недостатки, как спамдексинг (спам индексов поисковых систем), не совершенство системы поиска по meta keywords (ключевым словам), клоакинг и своп (техники замены содержимого проиндексированной страницы), незащищенность поисковых программ от некоторых "проблемных" запросов различных социальных групп, и др.

С учетом результатов анализа сформулированы практические рекомендации по использованию тех или иных методов поиска и отбора необходимых данных применительно к конкретным поисковым запросам.