***Зо Лин Кхаинг, асп.; рук. П.Р. Варшавский, к.т.н., доц. (МЭИ (ТУ))***

**Исследование и анализ подходов к построению информационно-Поисковых систем**

Информационно-поисковые системы (ИПС) появились на свет достаточно давно. Теории и практике построения таких систем посвящено множество статей [1]. ИПС в Интернете – это признание того, что ни иерархическая модель Gopher, ни гипертекстовая модель WorldWideWeb еще не решают проблему поиска информации в больших объемах разнородных документов. На сегодняшний день нет другого способа быстрого поиска данных кроме поиска по ключевым словам.

Основная задача любой поисковой системы – дать пользователю ответ на его запрос. Ответ поисковой системы в сети Интернет должен быть актуальным и информативным.

Первая задача поисковых систем (поисковых машин Интернет) – сбор информации или «индексация» сайтов. Для этого поисковые системы формируют свои базы знаний или так называемый «индекс», используя программных агентов (роботов, пауков и др.), перемещающихся по сети и собирающих необходимую информацию. Работа программных агентов строится на алгоритмах, созданных в результате совместной работы лингвистов, программистов и аналитиков. Из-за того, что объем информации в сети Интернет колоссален (по данным поисковой системы Яндекс, ими уже проиндексировано более 5 тысяч 610 Гб информации), сбор и хранение данных требует огромных мощностей – высокопроизводительных серверов с большим объемом памяти и мощными процессорами.

Следующая задача ИПС – определение тематических групп, рубрицирование сайтов по темам и т.д. Эта задача может выполняться автоматически или вручную (например, у многих поисковых систем существуют собственные каталоги сайтов, сформированные опытным редактором, который точно соотносит некий ресурс конкретной рубрике в огромном списке сайтов). Для определения важности ресурса в сети разработчиками ИПС используются различные коэффициенты и методы, основанные на учете внешних ссылок на сайты. Это тематический индекс цитирования у Яндекса, PageRank у Google, коэффициент популярности у Рамблера.

На этом подготовительная стадия работы ИПС заканчивается и выполняется основная задача – поиск по индексу ИПС с учетом заложенных алгоритмов релевантности и дальнейшего ранжирования результатов и вывод ответов на запросы пользователей в удобном виде (т.е. показываются сниппеты (выдержки из текста), в которых жирным выделяются слова из запроса).

Кроме того, поисковые системы анализируют реакцию пользователя на информацию, выданную ИПС. У поисковой системы Google есть панель-бар (надстройка к браузеру), с помощью которой пользователь может высказать свое мнение об отношении к тому сайту, на который его перенаправила поисковая система. Многие российские поисковые системы учитывают статистику переходов пользователя по результатам поиска, оценивая популярность соответствующих Интернет ресурсов.

На основе проведенного исследовании и анализа была реализована ИПС для библиотечного каталога [2, 3]. Особое внимание было уделено организации информационного поиска через Интернет и проектированию баз данных для ИПС. Программная реализация системы выполнена с использованием ASP.NET.

**Литература**

1. **Башмаков А.И., Башмаков И.А.** Интеллектуальные информационные технологии: Учеб. Пособие // –М.: Изд-во МГТУ им. Н.Э. Баумана, 2005. – 304 с.

2. **Дж. Солтон** Динамические библиотечно-информационные системы. Мир, Москва, 1979.

3. **G. Salton, C. Buckley** Term-Weighting Approaches in Automatic Text Retrieval. Information Processing & Management, 24(5), 1988, pp. 513-523.