

## Реферат

### **Актуальность**

Рост количества информационных источников увеличивается с каждым днем, количество накопленных данных все возрастает. В лавине информации человеку работать не просто сложно, а практически невозможно (человек не успевает отреагировать на получаемую информацию-данные). Так читая новости человек, отсеивает интересные для себя, а остальные пропускает даже не взглянув на текст, а ориентируясь лишь на рейтинг новости или ее название. Преподаватель не может прочитывать полностью работу каждого своего студента. Становится невозможным ориентироваться в массе выпускаемых книг.

### **Цель и задания исследования**

Целью работы является рассмотрение такого аспекта работы информационно-поисковых систем на заданном множестве сайтов, как технологии сбора информации о содержимом Веб-ресурсов.

### **Решаемые в работе задачи**

Основными задачами разрабатываемого программного обеспечения, предназначенного для оценки параметров веб-страниц, являются:

1. рассмотреть организационную структуру сайтов локальной сети НТУУ «КГП»;
2. разработка автоматизированного модуля сбора информации;
3. рассмотреть существующие модели поиска информации;
4. рассмотреть организацию семантического веба;
5. рассмотреть существующие решения для полнотекстового поиска информации;
6. создание прототипа системы поиска на заданном множестве сайтов, включающего все этапы;
7. достижение высокой эффективности результатов

## **Достигнутые результаты**

Результатом работы является веб-ориентированное приложение, предназначенное для контроля информации, расположенной на страницах веб-сайтов порталов НТУУ «КПИ» и EGEE, с целью повышения рейтинга НТУУ «КПИ» в мировом рейтинге университетов Webometrics.

## **Научная новизна полученных результатов**

Научная новизна работы состоит в исследовании возможности применения веб-ориентированных приложений для осуществления поиска информации, отображенной на веб-сайтах.

## **Практическое значение полученных результатов**

Практическим применением разработанного приложения является возможность его использования для поиска информации, представленной на веб-сайтах НТУУ «КПИ», таким образом упрощая процесс поиска нужной информации.

## **Выводы**

В данной магистерской работе были исследованы различные системы поиска информации на множестве заданных сайтов. Для реализации поиска данных в веб-пространстве НТУУ «КПИ» был разработан и реализован модуль поиска информации для комплексной системы, созданной для контроля информации, представленной на веб-сайтах.

В ходе работы были решены следующие задачи:

1. Выбор механизма индексации и поиска, удовлетворяющего соотношению эффективности и скорости поиска и вычислительных мощностей, которые для этого необходимы.
2. Осуществление полнотекстового поиска по информации, собранной с заданных веб-сайтов.
3. Достижение высокой скорости обработки информации.
4. Построение пользовательского интерфейса для поиска.

В результате экспериментов было показано, что качество системы поиска является достаточным для использования результатов ее работы при решении дальнейших прикладных задач.

Работа на 81 страницах содержит 4 таблицы, 3 иллюстраций. При подготовке работы использовалась литература из 21 различного источника.

Ключевые слова: поиск, робот, паук, веб-сайт, веб-портал, индекс, модели.