

РАЗРАБОТКА WEB-ПРИЛОЖЕНИЯ ТЕМАТИЧЕСКОГО АНАЛИЗА ТЕКСТОВОГО
КОНТЕНТА ИНТЕРНЕТ

Е. А. Борвенко, студент,
Национальный аэрокосмический университет им. Н. Е. Жуковского «ХАИ»
katterina2009@yandex.ru

В настоящее время значительно увеличилось число прикладных задач, связанных с Интернет, что позволяет рассматривать Интернет в качестве источника знаний. Отсюда появляются различные идеи автоматизированной обработки и извлечения знаний из текстовой информации.

Разработка WEB-приложения заключается в использовании методов тематического анализа. Первым этапом лингвистического анализа является разбор текстовой информации на отдельные абзацы, предложения, слова. Далее происходит распознавание частей речи каждого слова текста. Затем выполняется синтаксический анализ, который состоит в выделении семантических элементов предложения. Следующим этапом является семантический анализ текста, который предполагает оценки смыслового содержания текста. А заключением становится статический анализ, который подсчитывает количества повторений слов в тексте и использовании результатов подсчета для вычисления весовых коэффициентов ключевых слов.

1. Сэлтон Г. Автоматическая обработка, хранение и поиск информации : пер. с англ. / под ред. А. И. Китова. – М. : Советское радио, 1973. – 560 с.