

Модели обработки естественного языка в задачах информационного поиска

- Ахмедов Гаджи Омар оглы
- Научный руководитель: к.ф.-м.н., доцент кафедры СП Д.В.Луцев

Введение

Одной из основных функций современных информационных систем (ИС) является поиск элементов данных, удовлетворяющих некоторым признакам (информационный поиск). Специфика методических и технологических проблем, возникающих при организации такого поиска, обуславливается несколькими факторами. Прежде всего, это - характер контента, содержащегося в информационных ресурсах, входящих в систему.

В современных ИС попрежнему преобладает текстовый контент, однако все большее распространение приобретают мультимедийные ресурсы, содержащие мультимедиа контент (графика, аудио и видео информация), а также использующие для повышения эффективности функционирования различные формы структуризации контента.

Перечислим основные трудности,
возникающие при обработке текстов на NLP:

- Проблема синонимии;
- Проблема омонимии;
- Устойчивые сочетания слов;
- Морфологические вариации.

Рассмотрим два основных подхода к обработке и анализу текстов NLP – статистический и лингвистический (рис.1).



Рис.1. Методы обработки естественного языка

Лингвистический анализ



Рис.2. Уровни лингвистического анализа

Модели информационного поиска



Рис.3. Модели традиционного информационного поиска

Заключение