

Исследование применения тезаурусов для полнотекстового поиска

Ф.М. Бикмуратов

Московский государственный университет им М.В. Ломоносова

Область современного информационного поиска чрезвычайно разнообразна. Она включает такие задачи, как собственно поиск информации, фильтрация, рубрикация и кластеризация документов, поиск ответов на вопросы, автоматическое аннотирование документа и группы документов, поиск похожих документов и дубликатов, сегментирование документов и многое другое.

подавляющее большинство современных методов обработки неструктурированной информации решают эти задачи на основе минимальных дополнительных предварительных знаний и базируются на моделях текста как набора слов (bag of words), предлагая различные методы учета частотностей встречаемости слов в предложении, тексте, наборе документов, совместной встречаемости слов и т.п. Пословные модели не учитывают такие языковые явления как синонимия, гипонимия, существование лексических отношений между словами. Такой подход нередко приводит к нерелевантному поиску особенно в тех случаях, когда один и тот же запрос может быть сформулирован различным набором слов.

Одним из решений такого рода проблем является расширение поискового запроса за счет имеющегося тезауруса – словарь, в котором указаны семантические отношения (синонимы, антонимы, паронимы, гипонимы, гиперонимы и т. п.) между лексическими единицами.

В данной работе рассматриваются способы расширения поисковых запросов с помощью тезаурусов. Рассматриваются такие связи между понятиями как гипонимия, гиперонимия и ассоциативные связи, приводятся примеры, когда следует использовать те или иные отношения.

Литература

- *Jean Aitchison, David Bawden, Alan Gilchrist, Thesaurus Construction and Use: A Practical Manual*, London, Aslib IMI, The Association for Information Management, Staple Hall, Stone House Court, 2005, 187 p.
- *Лукашевич Н.В.*, Тезаурусы в задачах информационного поиска, МГУ, Москва, 2011, 387с.
- *Добров Б.В., Лукашевич Н.В.*, (), Тезаурус РуТез как ресурс для решения задач информационного поиска, Труды Всероссийской Конференции Знания-Онтологии-Теории (ЗОНТ-09), Новосибирск, 2009, 10 с.