

Е. Е. Краснощек (Таганрог, ТТИ ЮФУ). **Применение парадигматических отношений при поиске релевантной информации в Интернете.**

Поиск документов можно отнести к наиболее важным задачам содержательной обработки текстовой информации. Применение нечеткого подхода повышает эффективность поиска, но недостаточно. На малых объемах текстов по-прежнему часто находится мало релевантных документов. При больших объемах текстовой базы релевантных документов более чем достаточно, однако качество ранжирования может оказаться неудовлетворительным. Одним из подходов к устранению существующих недостатков является добавление к запросу слов, связанных парадигматическими отношениями со словами запроса, и последующий поиск по расширенному запросу.

Существуют автоматические методы построения парадигматических отношений. При этом выделенные парадигматические отношения не разбиваются на типы, а сами отношения обычно называются ассоциативными, а не парадигматическими. Построение ассоциативных отношений основывается на данных о совместной встречаемости терминов в документах [1]. Автоматически построенные ассоциативные отношения могут применяться при поиске вместо вручную построенных парадигматических отношений либо в дополнение к ним. При этом эффективность поиска повышается только при использовании специальных алгоритмов ранжирования [2]. Максимальный эффект достигается в том случае, если к исходному запросу добавляются те ассоциативные термины, которые имеют высокое соответствие по отношению ко всему запросу, а не к его отдельным словам [3].

Расширение запроса ассоциативными словами поможет решить две задачи.

1. Выдачу дополнительных релевантных документов. Эта задача важна при поиске в небольших текстовых базах данных. Поскольку поиск нечеткий, то по расширенному запросу могут быть найдены документы, в которых интересующая тема выражена по-другому.

2. Построение тематического представления результатов поиска. Актуально при поиске в больших текстовых базах. Если по запросу найдено много документов, то, как правило, они разбиваются на несколько групп тематически однородных документов.

СПИСОК ЛИТЕРАТУРЫ

1. *Солтон Дж.* Динамические библиотечно-информационные системы. М.: Мир, 1979, 557 с.
2. *Харин Н. П.* Метод ранжирования выдачи, учитывающий автоматически построенные ассоциативные отношения между терминами. Научно-техническая информация. Сер. 2, 1989, № 9, с. 19–23.
3. *Qui Y., Frei H. P.* Concept based query expansion. — In: Proceedings of the SIGIR'93, Pittsburgh: 1993, p. 160–169.