УДК 004.01(082)

К.А. Мацуева

*Национальный авиационный университет*

технологии информационного поиска
(Тематический поиск)

Развитие индустрии систем электронного документооборота, сопровождающееся ростом массивов обрабатываемых полнотекстовых документов, требует новых средств организации доступа к информации, многие из которых следует отнести к разряду систем искусственного интеллекта.

Основной задачей, возникающей при работе с полнотекстовыми базами данных, является задача поиска документов по их содержанию. Однако, ставшие традиционными средства контекстного поиска по вхождению слов в документ зачастую не обеспечивают адекватного выбора информации по запросу пользователя.

Основная проблема заключается в сложности точной формулировки запроса - подбора ключевых слов, которые предстоит искать в телах документов. Это может быть связано с рядом причин, как то: недостаточным знанием пользователем терминологии предметной области, трудностью определения границ своих интересов, наличием в языке многозначных и синонимичных слов, и даже орфографическими ошибками в написании искомых слов, которые могут встречаться как в текстах, так и в самом запросе.

Отдельную сферу приложения технологий обработки текста занимает широкий класс задач информационного поиска, относимых к разряду аналитических. Это задачи, требующие "раскопки" скрытых взаимосвязей между целевыми объектами, выявления ключевых факторов, влияющих на развитие ситуации, прогнозирование тенденций ее развития и т.п.

Над решением указанных проблем работает наше подразделение RCO Research Group. Результат многолетней работы воплотился в ряду технологий, основанных на автоматическом анализе содержания текстовых документов, применение которых призвано облегчить процессы поиска и эксплицирования скрытой информации. Часть технологий реализована в наших "коробочных" продуктах, другая часть находит внедрение в заказных решениях.