

**В. А. Башаев** (Ульяновск, УлГТУ). **Извлечение терминов как прикладная задача в САПР.**

На современных предприятиях большая часть данных хранится и обрабатывается в электронном виде, и очевидно, что с течением времени доля данных в цифровом виде будет только увеличиваться. В то же время, существующие ГОСТы и терминология ориентированы, прежде всего, на подготовку бумажной документации. Они не учитывают специфику проектирования с помощью САПР. Это приводит к двум проблемам: недостатку терминов и их нечеткости при описании работы САПР [1]. Для содействия в решении этих проблем нами была поставлена задача разработки программного обеспечения по автоматическому извлечению терминологии применительно к задачам САПР. Начальным этапом работы является определение наилучшего алгоритма для нашей системы на основе комбинации или модификации уже известных.

Интерес представляют работы двух направлений: выделение ключевых слов и терминов; нечеткий информационный поиск с индексацией. В рамках нашего междисциплинарного исследования интеллектуальный лингвистический анализ будет выполняться на основе нечеткой логики.

Отечественные и зарубежные исследования в области извлечения терминологии и ключевых слов сводятся к сочетанию различных вариантов трех анализов: синтаксического, морфологического и частотного, с возможными добавлениями. Существующие методы ориентированы на выделение одно-, двух- и многословных терминов и коллокаций. На данный момент задача автоматизации отбора терминов в словарь далека от полного решения. Стандартные методы приводят, по большому счету, к одному и тому же результату. Обычно в результате программы отбора словосочетаний порождается список, упорядоченный по весу в соответствии с заложенной моделью [2].

Отечественные и зарубежные исследования в области нечеткого поиска находят свое главное отражение в поисковых системах, принцип работы которых будет реализован в нашей системе. В этой связи представляют интерес алгоритмы Bitap, расширения выборки, метод N-грамм, хеширование по сигнатуре, ВК-деревья, расстояние Левенштейна и расстояние Дамерау–Левенштейна.

#### СПИСОК ЛИТЕРАТУРЫ

1. <http://www.electricspro.ru/?q=node/13> (2011 г.).
2. Лукашевич Н. В., Добров Б. В., Чуйко Д. С. Отбор словосочетаний для словаря системы автоматической обработки текстов. — Компьютерная лингвистика и интеллектуальные технологии. — По материалам ежегодной Международной конференции «Диалог» (Бекасово, 4–8 июня 2008 г.). В. 7 (14). М.: РГГУ, 2008, с. 339–344.