

«Высокие интеллектуальные технологии образования и науки».

Материалы X Международной научно-методической конференции. С.121-122, 2003. © Санкт-Петербургский государственный технический университет, 2003

АВТОМАТИЧЕСКАЯ ОБРАБОТКА ТЕКСТА В ИНТЕЛЛЕКТУАЛЬНОЙ ИНФОРМАЦИОННО-ПОИСКОВОЙ СИСТЕМЕ

Ильюшко П.А., Морозов Б.И.

Санкт-Петербургский государственный политехнический университет

Одним из способов использования контекстно-зависимой информации для повышения точности результатов запроса – нечеткий поиск с дополнением искомого запроса синонимами ключевых слов. В предлагаемой интеллектуальной системе каждое ключевое слово запроса может дополняться синонимами, соответствующими одному или нескольким значениям слова, с учетом морфологии русского языка.

Задачей автоматической обработки текста является получение семантического представления входного языка. Этот процесс включает три этапа: морфологический, синтаксический и семантический.

На морфологическом этапе к каждому слову предложения добавляется его полная грамматическая характеристика, взятая из словаря. Для этого каждое слово расчленяется на составляющие такие как: основа, префикс, корень, суффикс, окончание, постфикс. Возможность такого разбиения слов обеспечивается наличием в словаре системы основ всех лексических единиц входного языка и списков допустимых элементов для всех остальных позиций разбираемого предложения.

На синтаксическом этапе строится дерево грамматического разбора фразы, которое устанавливает синтаксические зависимости между словами. Это производится на основе специальных правил вывода следующего вида:

$$H; S_1(s_1)m_1; S_2(s_2)m_2; D \Rightarrow S_1 R^{12} S_2,$$

где s_1 и s_2 – две словоформы, между которыми мы хотим установить связь R^{12} ; S_i и m_i – соответственно синтаксическая и морфологическая информация, приписанная этим словам; D – информация о расположении этих словоформ в анализируемой фразе; H – некоторое условие применимости этого правила.

На семантическом этапе осуществляется преобразование синтаксического дерева в модельное представление, отражающее смысл фразы. Указанный этап является наиболее сложным, так как основными задачами его являются определение смыслового значения каждого из словосочетаний исходного текста и отношений, в которых находятся эти словосочетания.

Таким образом, в данной системе проводится анализ и преобразование запроса к поисковой системе на естественном языке к форме, позволяющей достичь более эффективных результатов поиска.