

П.В.Ильюшко (асп., каф. САиУ), Б.И.Морозов, проф.

ИНТЕЛЛЕКТУАЛЬНАЯ ИНФОРМАЦИОННО-ПОИСКОВАЯ СИСТЕМА ДЛЯ ДИСТАНЦИОННОГО ОБРАЗОВАНИЯ

В настоящее время трудно переоценить влияние сети Интернет на образовательную и научно-исследовательскую деятельность. Однако по мере развития глобальной сети обостряется парадокс: вероятность существования нужной информации возрастает, а возможность ее нахождения уменьшается. Теоретически гипертекстовая природа WWW обеспечивает нахождение любой информации в процессе целенаправленного продвижения по ссылкам. Однако, согласно оценкам, в Интернет сегодня существует более 60 млн. документов и найти нужную информацию в этом пространстве, продвигаясь от ссылки к ссылке, практически невозможно. Для решения этой проблемы в последнее время интенсивно развиваются информационно-поисковые системы (ИПС), предназначенные для свободного поиска информации в информационных массивах по совокупности признаков (обычно ключевых слов).

Индексирование документов, т.е. фиксацию положения ключевых слов ИПС обычно производят одним из следующих методов: векторная модель информационного потока, нечеткие множества, вероятностная модель. Для поиска документов пользователь может ввести булевский запрос с логическими выражениями (AND, OR, NOT и т.д.), использовать информационно-поисковые языки типа "Like this" или общаться с ИПС на естественном языке. Основной недостаток запросов с логическими выражениями - плохая масштабируемость выдачи: OR приводит к слишком большому расширению списка релевантных документов, а AND резко сужает отклик.

Дальнейшее содержательное развитие ИПС идет в сторону упрощения работы с ними и их интеллектуализации. Идеальный язык запросов ИПС - естественный. Для этого широко используют словари, ряды синонимов, тезаурусы слов. В данной работе разрабатывается система, позволяющая преобразовать запрос на естественном языке в форму, улучшающую релевантность поиска, т.е. меру соответствия получаемого результата желаемому или, в терминах поисковых систем, меру соответствия ответа запросу.

Первый этап обработки запроса - морфологическая обработка - выделение основы слова, т.е. ядерной части слова, с которой связано его вещественное лексическое значение и которая остается за вычетом из него словоизменительных морфем (окончаний). Основа отличается от частей слова своим неслужебным характером; от корня - тем, что выражает категориальную отнесенность к одной из частей речи; от слова - тем, что не обладает ни завершенностью, ни оформленностью последнего и существует всегда только как часть производного слова или словоформы.

Далее модуль обработки запроса выделяет ключевые и неключевые слова (или «стоп-слова», т.е. предлоги, союзы, частицы, местоимения и т.д.). Для ключевых слов на основе известных законов Джорджа Зипфа (George K. Zipf) можно присвоить каждому термину весовой коэффициент, отражающий его значимость.

Так как большинство пользователей механизмов поиска просто выбирают одно или несколько ключевых слов для поиска и не применяют сложных функций булевой логики, то поисковый механизм сам должен быть достаточно интеллектуальным. Один из способов использования контекстно-зависимой информации для повышения точности результатов - нечеткий поиск с дополнением запроса синонимами ключевых слов. В разрабатываемой

системе каждое ключевое слово запроса может дополняться синонимами, соответствующими одному или нескольким значениям слова, в зависимости от контекста.

Авторам представляется перспективным дальнейшее развитие ИПС, дополнение ее возможностями поиска с учетом устойчивых словосочетаний, нечеткого поиска с учетом опечаток, как в индексированных текстах, так и в запросе, а также расширение поддерживаемых форматов файлов.

ЛИТЕРАТУРА:

1. Попов А. Эффективная методика поиска информации в сети Интернет: Журнал «Internet», N 2 (7), 1998.
2. Частиков А.П., Дедкова Т.Г., Алешин А.В. Системы искусственного интеллекта: Учебное пособие. - Краснодар: Изд - во КубГТУ, 1998. 166 с.
3. R.Murphy. English grammar in use. UK, Oxford university press, 1997. P.350.
4. B. Katz. From Sentence Processing to Information Access on the World Wide Web. Cambridge, MA: MIT Artificial Intelligence Laboratory, 1997.
5. Evan L. Antworth. Morphological Parsing with a Unification - based Word Grammar - Arlington, TX: University of Texas, 1994.