

Структурный метод формирования запросов к информационной системе

М.Г. Мальковский
МГУ им. М.В. Ломоносова
malk@cs.msu.su

С.Ю. Соловьев
Межведомственный суперкомпьютерный центр РАН
soloviev@jssc.ru

В статье рассматривается один из подходов к расширению поисковых возможностей пользователей сети Интернет. Основная идея подхода состоит в том, что пользователю предлагается удобный способ генерации "тонких" поисковых запросов к известным поисковым системам. В качестве модельного языка запросов выбран язык информационно-поисковой системы "Яндекс". Каждый запрос инициируется пользователем в связи с некоторым конкретным термином, занимающим определенное место в семантической сети понятий научной и деловой лексики. Совокупность семантических отношений, образующих контекст использования термина, позволяет автоматически генерировать конечное множество тонких запросов, связанных с этим термином. Фактически пользователю предлагается форма для уточнения тех характеристик запроса к информационно-поисковой системе "Яндекс", которые в дальнейшем транслируются в операторы языка запросов.

Все современные пользователи Интернет знают, что без поисковых систем обойтись нельзя, но работа с ними требует терпения и везения. В нашем случае под везением понимается быстрый выход на такое специфическое слово или словосочетание, которое выделит из океана документов ровно те, что хочется видеть. При этом наибольшую сложность представляет необходимость изложить поисковый запрос на "чистой" странице, любезно предоставленной Яндекс-ом или иной информационно-поисковой системой. Альтернативный, хотя и ограниченный по охвату проблемных областей подход к формированию запросов предложен в проекте www.glossary.ru, реализующим концепцию универсального терминологического пространства (УТП) [1,2]. УТП есть семантическая сеть, вершинами которой являются термины научной и деловой лексики, связанные бинарными отношениями (ребрами) типа "это-есть" и "относится-к". В общем случае вершина семантической сети задается термином, его синонимами, его определением и семантическими связями с вершинами более высокого уровня абстракции.

Семантическая сеть обеспечивает пользователям навигацию в терминологическом пространстве. Кроме того, семантическая сеть позволяет выполнять "интеллектуальные" преобразования терминологии, возникающие при формировании глоссариев.

Каждую вершину семантической сети можно рассматривать как потенциальный запрос, передающий информационно-поисковой системе термин-словосочетание. Такие потенциальные запросы активизируются пользователем по мере надобности. Это соображение было реализовано в проекте www.glossary.ru, причем для определенности информационные запросы передаются ИПС Яндекс. Статистика востребованности этой функциональной возможности показывает, что пользователи Интернет чрезвычайно охотно формируют запросы к ИПС в режиме меню.

Следующий этап в развитии функциональных возможностей интерфейса доступа к ИПС предусматривает формирование "тонких" запросов, использующих язык расширенных запросов. Фактически речь идет об отображении фрагмента топологии сети в запрос к ИПС.

Метод построения тонких запросов исходит из того, что пользователь, во-первых, инициирует обращение к ИПС в связи с некоторой вершиной семантической сети, и, во-вторых, утверждает окончательный вид запроса, используя предложенную заготовку. Алгоритм формирования заготовки выполняется на серверной стороне ресурса. В качестве исходных данных алгоритм получает адрес вершины семантической сети. Наименования понятий, с которыми связана вершина образуют потенциальный набор ограничений на класс документов, интересующих пользователя. В заготовке перечисляются все эти фразы-ограничения, однако окончательный выбор остается за пользователем. Термин, определяемый заданной вершиной, составляет обязательную основу запроса, которая может редактироваться и дополняться с учетом возможностей расширенного языка запросов. Обычно такие возможности позволяют зафиксировать порядок слов, расстояние между словами и т.п.

Особый случай обработки составляют многозначные термины, которым соответствуют несколько вершин семантической сети [3]. С формальной точки зрения контекст каждого обращения пользователя фиксирован, поэтому многозначностью можно пренебречь. Однако с практической точки зрения (массовому) пользователю совершенно безразлично из каких соображений термин, скажем, "Андеррайтер" порождает две различные заготовки формирования запросов. При обработке многозначных терминов предварительно выполняется виртуальная операция слияния соответствующих вершин, позволяющая определить единое множество семантических связей.

Заготовка запроса с отметками и уточнениями пользователя поступает на вход конструктора тонких запросов, который преобразует ее в собственно запрос на расширенном языке и передает его на исполнение ИПС. Конструктор тонких запросов также функционирует на серверной стороне ресурса.

Описанный метод формирования тонких запросов доступен в сети Интернет по адресу www.glossary.ru.

Литература

1. Мальковский М.Г., Соловьев С.Ю. Универсальное терминологическое пространство. Труды Международного семинара Диалог'2002 Компьютерная лингвистика и интеллектуальные технологии", т.1. М:Наука, 2002, с.266-277.
2. Мальковский М.Г., Соловьев С.Ю. Технология формирования универсального терминологического пространства. Сб. "Информационные компьютерные технологии и Интернет в образовании и науке". Москва: изд-во МИИ для инвалидов с нарушением ОДС, 2002, с.54-55.

3. Мальковский М.Г., Соловьев С.Ю. Методы формирования глоссариев в универсальном терминологическом пространстве. Труды Международного семинара Диалог'2003 "Компьютерная лингвистика и интеллектуальные технологии", М:Наука, 2003, с.438-440.

