

4. Вычисление значения критерия качества

Значения компонент критерия качества (1) определяются следующим образом: $L = (\bar{\Pi}_1(1) + \bar{\Pi}_2(1))\bar{e}$, $P_2 = \Lambda \bar{P}_2(1) \bar{b}_2^{(1)} \bar{e}$, $P_1 = 1 - P_2$, $R = (\lambda^{(1)} P_1 + \lambda^{(2)} P_2)(1 - P_+)$.

Заключение

Рассмотрена управляемая система BMAP/SM/1 с MAP-потокм сбоев. Найдены стационарное распределение вероятностей состояний системы, вероятность успешного обслуживания произвольной заявки, среднее время между уходами заявок и значение критерия качества при заданном значении порога.

Литература

1. *Artalejo J.* G-networks: a versatile approach for work removal in queueing networks // *Eur. J. Oper. Res.* 2000. P. 233–249.
2. *Dudin A. N., Nishimura S.* A BMAP/SM/1 queueing system with Markovian arrival of disasters // *J. Appl. Prob.* 1999. V. 36. P. 868–881.
3. *Dudin A. N., Karolik A. V.* BMAP/SM/1 queue with Markovian input of disasters and non-instantaneous recovery // *Performance Evaluation*. 2001. V. 45. № 1. P. 19–32.
4. *Dudin A. N., Nishimura S.* Optimal hysteresis control for a BMAP/SM/1/N queue with two operation modes // *Mathematical Problems in Engineering*. 2000. V. 5. P. 394–419.
5. *Dudin A. N., Nishimura S.* Optimal control for a BMAP/G/1 queue with two service modes // *Mathematical Problems in Engineering*. 1999. V. 5. P. 255–273.
6. *Dudin A. N.* Optimal multithreshold control for a BMAP/G/1 queue with N service modes // *Queueing Systems*. 1998. V. 30. P. 22–25.
7. *Choi B. D., Chung Y., Dudin A. N.* The BMAP/SM/1 retrial queue with controllable operation modes // *European Journal of Operational Research*. 2001. V. 131. P. 16–30.
8. *Dudin A. N., Klimenok V. I.* Alternative algorithm for characteristics calculation of a BMAP/SM/1 system with the multi-threshold control // *Queues: flows, systems, networks*. 2001. V. 16. P. 81–86.
9. *Скоруход А. В.* Теория вероятностей и случайных процессов. Киев: Высш. шк., 1980.

ОТКРЫТАЯ АРХИТЕКТУРА ЕЯ-ИНТЕРФЕЙСА ПОЛЬЗОВАТЕЛЯ В ЗАДАЧЕ ИНФОРМАЦИОННОГО ПОИСКА

О. В. Сергей

Наш век – век информации. Но развитию информационных технологий в современном мире сопутствует рост объема информации, что в свою очередь порождает проблему быстрого и удобного доступа к ней. Усиливающаяся тенденция к хранению информации в компью-

терных структурированных источниках данных и широкое распространение глобальной компьютерной сети Интернет делают проблему поиска данных одной из первостепенных в мире информационных технологий. [1] Особенно остро она стоит для пользователей, не являющихся специалистами в конкретной области знаний. Точность, быстрота и полнота поиска во многом зависят от того, какие возможности дает система пользователю при создании запроса и насколько хорошо информационно-поисковая система смогла распознать ту информацию, которая была заложена в запросе.

Функции приема, анализа и предварительной обработки информации с целью создания ПОЗа выполняет интерфейс пользователя. [2] Среди всего многообразия интерфейсов пользователя особо выделяется естественно-языковой интерфейс (ЕЯ-интерфейс) – разновидность пользовательского интерфейса, который принимает запросы на естественном языке. ЕЯ-интерфейс очевидно выигрывает перед всеми другими видами интерфейсов в гибкости, простоте и высокой скорости создания запросов, к тому же запрос на естественном языке наиболее полно отражает предметную область. [3]

Нами была разработана система семантического поиска информации, ориентированная на то, чтобы производить поиск не только по словам или ключевым словам, но и по семантическим конструкциям, выделяемым из запроса. Блок ЕЯ-интерфейса, реализованный в нашей системе, совмещает в себе несколько стадий автоматического анализа и обработки текста.

ЕЯ-интерфейс, опирающийся на лингвистическую базу знаний (ЛБЗ) и обеспечивающий окончательное выделение вычислительных шагов из общей сферы управления, является интеллектуальной информационной системой. Принцип открытой архитектуры, в нашем случае ЕЯ-интерфейса, т. е. вынесение правил обработки запроса из кода программы с тем, чтобы они были доступны экспертам для исправления и дополнения, – это необходимое условие качества любой системы искусственного интеллекта, а в частности ИПС. Лингвистические алгоритмы являются отдельным инструментом анализа и разбора запроса на естественном языке, в них заложены знания о языке и его особенностях. Поэтому так важно обеспечить экспертам доступ к ЛБЗ, чтобы улучшать показатели полноты и точности, не завися от программной среды.

Лингвистические алгоритмы – это морфологические, синтаксические и семантические правила построения слов, фраз, предложений,

это знания о значениях слов, различного рода классификаторы свойств ЕЯ на различных уровнях его глубины. [3] Вместе со словарями они образуют лингвистические базы знаний (ЛБЗ). Лингвистический алгоритм оперирует текстом на естественном языке и производит действия над ним. В правиле можно выделить правую и левую части: схему и действия.

В нашей системе семантического поиска запрос пользователя проходит через многие этапы автоматической обработки. С помощью лингвистической базы знаний и алгоритмов автоматического анализа текст тегируется (раскрашивается) синтаксически и семантически, выделяются отношения между лингвистическими единицами. После обработки пользовательского запроса на выходе получается поисковый образ запроса в виде графа лингвистических отношений, узлами которого являются слова или другие лингвистические единицы, а дугами – лингвистические отношения. В качестве лингвистических единиц выступают: именная группа, предложная группа, глагольная группа, SAO, предложение, абзац.

Главная цель при обработке – привести запрос к виду Subject-Action-Object (SAO). Subject – это сущность, производящая действие. Обычно это именная группа или существительное, являющееся подлежащим в предложении. Action – собственно действие. Обычно это глагол, который играет роль сказуемого. Object – это сущность, над которой производится действие. Например, в предложении «Fire heats water» Subject: «Fire», Action «heats», Object «water» присутствует полная тройка SAO. Если любое из полей отсутствует (невозможно выделить), тройка SAO – неполная. При трансформации запроса может быть выделено более одного поискового образа.

Тройка SAO – это важный узел графа лингвистических отношений. Специальная база правил QBank, разработанная в нашей системе, предназначена для обработки коротких запросов пользователя в виде вопросов. Она содержит шаблоны вопросов и действия, которые надо совершить над частью вопроса, – например выделить группу слов в именную группу или представить глагол в каноническом виде и выделить его как Action в тройке SAO.

При компиляции правила QBank обрабатываются отдельно от кода программы и, следовательно, проходят все этапы анализа: этап разбивки текста правил на слова, лексический анализ, синтаксический анализ и компиляцию в бинарный формат (на стороне сервера), интерпретацию (на клиентской стороне). [4]

Рассмотрим каждый из этапов подробнее. На стадии разработки нотации главной задачей является достижение максимальной простоты и наглядности правил. В нотации записи правил QBank каждое правило разделено на левую и правую части. В левой части – цепочка синтаксических категорий (тегов), объединяющая множество предложений одной синтаксической структуры. В правой части записываются лингвистические конструкции, которые строятся на основе слов предложения. В правилах используются ссылки на слова и ссылки на конструкции. Ссылки на слова (имеют префикс \$) непосредственно связаны с левой частью правила и указывают на последовательность слов. Ссылкам на слова соответствуют идентификаторы треугольных скобок в левой части правила. Ссылки на конструкции (имеют префикс &) – это ссылки на структуры, которые были описаны выше. Ссылки на конструкции позволяют построить древовидную структуру, или граф. В правой части правила для указания действий используются функции. Функции оперируют ссылками на слова.

В настоящий момент мы используем только одну функцию `Noun2Verb`, преобразующую существительное в глагол.

При построении грамматики выделяются группы лексических единиц (строится база лексем), указываются правила продукции. На следующем этапе происходит разбивка текста правил на токены (токен – последовательность символов) для дальнейшего анализа.

На этапе лексического анализа каждому токену ставится в соответствие лексема из базы лексем QBank. Лексема – это стандартная лексическая единица, возможная для языка, порождаемого грамматикой правил QBank. Таким образом, все токены разбиваются на классы лексем. В базе лексем QBank используются следующие классы лексических единиц: данные, служебные слова, названия синтаксических и семантических конструкций.

На этапе синтаксического анализа (парсинга) происходит построение дерева синтаксического разбора. Если во время парсинга произошла ошибка, значит, дерево синтаксического разбора построить не удалось из-за ошибок в записи правила. К синтаксическим ошибкам, выявляемым на этом этапе, можно отнести пропуск специальных символов, неправильное использование разделителей.

Этап компиляции совмещает семантический анализ и запись внутреннего представления (дерева синтаксического разбора) в бинарный формат. В результате семантического анализа выявляются ошибки: использование имен функций, не известных системе, использование

названий синтаксических и семантических конструкций, не известных системе, превышение пределов допустимых констант и др. Далее, если ошибок не было обнаружено, левые и правые части правил отдельно компилируются в бинарный формат и затем склеиваются в один файл. Все эти действия происходят на стороне сервера и являются частью процесса компиляции ЛБЗ.

При интерпретации правила распаковываются из бинарного формата в память машины и используются блоком автоматического анализа при условии, если входной запрос пользователя совпал с шаблоном в левой части правила.

База правил QBank – удачное решение для обработки коротких запросов пользователя; с ее помощью можно охватить большой класс вопросов и тем самым гарантировать для них правильный разбор.

Литература

1. Жигалов В. А. Технология к построению естественно-языковых интерфейсов к структурированным источникам данных: Дис...канд. тех. наук. М., 2000.
2. Храмцов П. Информационно-поисковые системы Internet / Интернет адрес: <http://www.osp.ru/os/1996/03/46.html>. 1996.
3. Совпель И.В. Инженерно-лингвистические принципы, методы и алгоритмы автоматической переработки текста Мн.: Выш. шк., 1991.
4. Ахо А., Сети Р., Ульман Дж. Компиляторы. Принципы, технологии, инструменты / Пер. с англ. М.: Издательский дом «Вильямс», 2001.

ИССЛЕДОВАНИЕ ПОВЕДЕНИЯ ОПТИМАЛЬНОГО СПРОСА И ПРЕДЛОЖЕНИЯ ДЛЯ ЗАДАЧИ АНАЛИЗА СПОСОБОВ ПРОИЗВОДСТВЕННОЙ ДЕЯТЕЛЬНОСТИ

С. В. Скаскевич

Одним из участников экономической системы является фирма, под которой понимают участника экономической системы, занимающегося выпуском некоторого продукта, потребляя при этом определенный набор факторов производства (которые, в свою очередь, могут быть продуктами других фирм). Будем рассматривать фирму, которая занимается выпуском одного вида продукции и при этом использует m производственных факторов. Количество i -го фактора, закупаемого фирмой, обозначим через x_i , $i \in I = \{1, \dots, m\}$. Пусть $x = (x_1, \dots, x_m)'$. В простейшем случае выполняется условие неотрицательности

$$x \geq 0. \quad (1)$$