*Государственное образовательное учреждение высшего профессионального образования*

*«Московский государственный технический университет имени Н. Э. Баумана»*

ОТЧЕТ

По лабораторной работе №5

По курсу “Проектирование экспертных систем”

|  |  |
| --- | --- |
| Студент: | Бородин Д.С.  Федоров П.В, |
| Группа: | ИУ7-32М |
| Преподаватель: | Рогозин О.В. |

Москва, 2018г.

**Тема: «Проектирование экспертной системы»**

Описание системы:

Реализованная система представляет собой транслятор естественно-языковых запросов в запросы к реляционной базе данных. Естественный язык ограничен конкретной предметной областью (сфера кинематографа), а язык ограничен некоторым набором конструкций.

|  |
| --- |
| C:\Users\Дмитрий\Downloads\Untitled Diagram.png |
| Рисунок 1. Структура экспертной системы. |

В разработанной экспертной системе в качестве эксперта реализована предобученная система на основе некоторой базы знаний в соответствии с предметной областью. Представляет собой семантическую сеть, в качестве узлов которой однословные термины, в качестве связей – глаголы, связующие однословные термины. Алгоритм обучения опирается на заранее подготовленный корпус текстов, состоящий из словарных статей, обработка которого осуществляется в три этапа :

1. кластеризация текста;
2. выделение наиболее частотных классов для каждого кластера;
3. формирование списка терминов из слов, принадлежащих данным классам;

В качестве терминов выделялись однословные термины, а в качестве метода кластеризации используется метод Уорда, считающийся наиболее точным среди остальных методов.

Неоспоримым преимуществом данного метода является способность выделения даже тех терминов, которые редко встречаются в тексте (менее 3-х раз). Более того, метод не требует при себе обучающей выборки, так как реализует механизм кластеризации исходных текстов. Также достоинством алгоритма является способность выделять общие однословные термины с высокой степенью точности Однако, для работы этого алгоритма необходимо решить проблему подготовки специализированного текста для выделения терминов.

Выделение связей для алгоритма осуществлялось следующим образом: исходные словарные статьи очищались от стоп-слов, сохраняя порядок следования слов в статье. Далее использовалась маска «ОБЪЕКТ – СВЯЗЬ –ОБЪЕКТ» со всеми вариациями для формирования связи вида СВЯЗЬ(ОБЪЕКТ, ОБЪЕКТ).

Принцип преобразования естественно-языковых запросов описывается так: Обработка естественно-языковых запросов, как уже было описано ранее, осуществляется в три этапа: сначала исходный текст запроса проходит через лингвистический модуль, формируется запрос к базе знаний на Прологе и далее формируется запрос на выборку в зависимости от полученного запроса. Останавливаясь подробнее на описании первого этапа обработки естественно-языкового запроса, следует отметить, что написание собственного лингвистического модуля – очень сложный и трудоемкий процесс, который осуществляется командой разработчиков в течение нескольких лет, что явно выходит за рамки рассмотрения нашей работы. Поэтому принято решение об использовании в работе уже имеющихся технических средств, позволяющих осуществить требуемую работу на данном этапе. Условно, первый этап, получение запроса к базе знаний, может быть описан следующим алгоритмом (рисунок 2):

|  |
| --- |
|  |
| Рисунок 2. Схема алгоритма работы лингвистического модуля. |

В работе алгоритма делается копия исходного запроса с целью обеспечения проведения нескольких видов анализа за один прогон модуля независимо друг от друга. Поэтому для нас не имеет особого значения строгий порядок выполнения различных видов анализа. Следует отметить, что в данной версии алгоритма опускается графематический анализ текста, направленный на выделение словоформ в предложении. Это связано с ограничениями на естественно-языковые запросы, о которых будет рассказано далее.

Также стоит отдельно упомянуть принципы работы каждого из оставшихся видов анализа применительно к нашей задаче. Синтаксический анализатор, принимая текст исходного запроса, строит синтаксическое дерево зависимостей, в котором определены отношения между частями предложения. Так как русский язык обладает большой степенью омонимии, результатом синтаксического разбора в зависимости от контекста могут быть несколько деревьев.

На этапе морфологического анализа проверяется валидность входящих в предложение словоформ посредством поиска в имеющемся словаре, формируется список словарных статей с информацией о каждой словоформе, затем этот список возвращается на верхний уровень работы программы. Данный список словоформ нужен для корректного формирования запроса к базе знаний.

Этап формирования запроса к базе знаний может осуществляться несколькими способами, о которых будет рассказано далее. В качестве результата работы этой части алгоритма планируется получать текст запроса к базе знаний и список инициализированных параметров.

После того, как произведен этап преобразования запроса на естественном языке, возникает проблема приведения полученных данных к конечному результату – запросу на формальном языке. С учетом модели реализуемой базы данных, имеется два пути обработки полученного синтаксического дерева:

1) с помощью ***семантических сетей***, отражающих связи между объектами, а также связи между связями;

2) с помощью ***средств логического программирования***;

Первый способ очень удобен при реализации запросов к графовым базам данных, так как под семантической сетью понимается граф информационной модели, который может быть переложен на базу данных. Второй способ пригоден для работы как с графовыми базами данных, так и с реляционными базами данных, так как средства логического программирования позволяют представить синтаксическое дерево как в виде семантической сети, так и в виде различных таблиц, свойств и атрибутов, а также связей между таблицами. Более того, анализ доступных информационных источников показал, что на данный момент имеются весьма эффективные трансляторы из языка логического программирования в формальный язык работы с базами данных, что существенно упрощает процесс составления запросов [39]. Однако, представленный в данной работе транслятор имеет ряд ограничений, которые впоследствии могут быть отброшены при соответствующей модификации.

Стоит также отметить, что использование языка логического программирования в качестве промежуточного средства преобразования запросов необходимо, так как преобразование с его помощью дает простор для последующей модификации.

С учетом описанных выше фактов и достоинств, можно сказать, что использование языка логического программирования применительно к данной задаче весьма оправдано.

Таким образом, можно сделать вывод о том, что мы описали два равновозможных в использовании метода преобразования получаемой после анализа предложения на естественном языке информации, которые условно могут быть представлены следующей схемой (рисунок 3):

|  |
| --- |
|  |
| Рисунок 3. Принцип работы экспертной системы. |

Формирование запросов к базе данных можно разделить на три этапа в соответствии с выбранной структурой системы:

1. обработка исходного естественно-языкового запроса;
2. преобразование в запрос на языке логического программирования;
3. транслирование запроса на языке логического программирования в запрос к реляционной базе данных;

При формировании запросов к нашей базе данных, мы сталкиваемся с проблемой произвольности порядка слов в предложении на выбранном естественном языке. Так, в английском языке мы можем наблюдать очень строгую последовательность слов (например, вопрос на английском языке всегда начинается со вспомогательного слова – «Who are you?», «How old are you?» и другие). Такой строгий порядок позволяет в достаточно четкой и корректной форме приводить имеющиеся естественно-языковые запросы к удобному для работы программы виду. Однако, выбранный нами естественный язык проявляет гибкость в отношении порядка слов в предложении без нарушения смысловой нагрузки («Я люблю тебя.», «Тебя я люблю.» и «Люблю я тебя.» - конструкции, которые могут в равной степени существовать в тексте). Поэтому, так как в данной работе опускается семантический анализ предложений на естественном языке, было принято решение ограничить рамки вводимых запросов различными «схемами» предложений, которые позволят добавить предсказуемости в трансляции запросов к конечной базе данных и сделать корректное преобразование. Под «схемой» предложения понимается определенный порядок слов в предложении, наиболее часто употребляемый в повседневной речи.

Пример работы системы представлен на рисунке 4.

|  |
| --- |
|  |
| Рисунок 4. Пример работы экспертной системы |