

ФОРМАЛЬНО-ЯЗЫКОВОЙ АППАРАТ ДЛЯ ФОРМИРОВАНИЯ ЗАПРОСОВ К СЕМАНТИЧЕСКОЙ МОДЕЛИ

Кушнеревич Павел Михайлович

*магистрант,
Белорусский государственный университет информатики и радиоэлектроники,
Беларусь, г. Минск*

Герман Юлия Олеговна

*канд. тех. наук, доц,
Белорусский государственный университет информатики и радиоэлектроники,
Беларусь, г. Минск*

FORMAL LANGUAGE APPARATUS FOR FORMING QUESTIONS TO THE SEMANTIC MODEL

Pavel Kushnerevich

*Magister, Belarusian state university of informatics and radioelectronics,
Belarus, Minsk*

Yuliya German

*PhD, associate Professor,
Belarusian state university of informatics and radioelectronics,
Belarus, Minsk*

АННОТАЦИЯ

Цель: Разработка и тестирование метода обработки семантических вопросов в контексте представления текста как базы знаний, позволяющего генерировать точные ответы на вопросы пользователя.

Метод: Текст представляется в виде формализованной семантической базы данных, где основные элементы - термины (существительные, глаголы, прилагательные и т.д.). Для обработки вопросов используется структурный анализ предложения, выделение субъекта, объекта и действия, с последующим сравнением выделенных частей с базой данных.

Результат: Был успешно реализован и протестирован метод обработки семантических вопросов. Это обеспечивает точную обработку вопросов, даже если текст вопроса содержит ошибки или опечатки.

Выводы: Разработанный метод позволяет эффективно работать с текстом как с базой знаний, обеспечивая качественную обработку вопросов пользователя и точный поиск ответов внутри текста. Он также позволяет существенно сократить время поиска ответа путем предварительного определения темы вопроса.

ABSTRACT

Goal: The development and testing of a method for processing semantic questions in the context of treating text as a knowledge base, aiming to generate accurate responses to user queries.

Method: Text is represented as a formalized semantic database where the main elements are terms (nouns, verbs, adjectives, etc.). For question processing, a structural analysis of the sentence is used, highlighting the subject, object, and action, followed by a comparison of the identified parts with the database.

Result: A method for processing semantic questions has been successfully implemented and tested. This provides accurate question processing even if the question text contains errors or typos.

Conclusions: The developed method allows effective work with text as a knowledge base, providing high-quality processing of user queries and accurate search for answers within the text. It also significantly reduces the response time by preliminary determination of the topic of the question.

Ключевые слова: работа с текстом, база знаний, семантический объект, интерфейс, семантическая обработка, поисковые вопросы, контекстная помощь, формализованное представление, семантическая база данных, грамматический разбор, субъектный блок, объектный блок, алгоритм обработки, хэш-таблица.

Keywords: text processing, knowledge base, semantic object, interface, semantic processing, search queries, contextual assistance, formalized representation, semantic database, grammatical parsing, subject block, object block, processing algorithm, hash table.

При работе с текстом как с базой знаний мы имеем дело с текстом как семантическим объектом. Это значит, что текст выступает в качестве базы зна-

ний. С такой базой знаний следует организовать интерфейс. Интерфейс предполагает обработку семантических и поисковых вопросов, организацию контекстной помощи и тестирование. Семантический

вопрос требует выдачи точечного ответа, в отличие от поискового вопроса, ответом на который является список ссылок на различные релевантные (подходящие по смыслу) места в тексте.

Чтобы обеспечить семантическую обработку вопросов, текст следует представить в формализованном виде. В принципе, имеется несколько способов формального представления текста для семантической обработки. Можно указать два основных. Во-первых, это семантическая сеть. Во-вторых, это семантическая база данных с интерфейсом. В наиболее простом случае в качестве модели знаний можно использовать обычный текст. В этой работе используем этот второй подход [1]. Согласно этому подходу, в базе данных будем хранить лексические еди-

ницы – термы (существительные, глаголы, прилагательные и др.), используемые при описании предметной области. С помощью термов формируем запросы (предложения). Структура предложений должна быть синтаксически корректной, во всяком случае, должно быть понятно, что интересует пользователя. При этом допускаем определенные отклонения как в правописании термов, так и в их расположении в запросах. Для корректной записи предложений языка следует определить грамматические правила (грамматику).

В качестве пояснения рассмотрим предложение: the average price of chocolate bar is about 5 dollars

Это предложение соответствует приведенной ниже грамматике, что поясняется следующей схемой (рисунок 1).

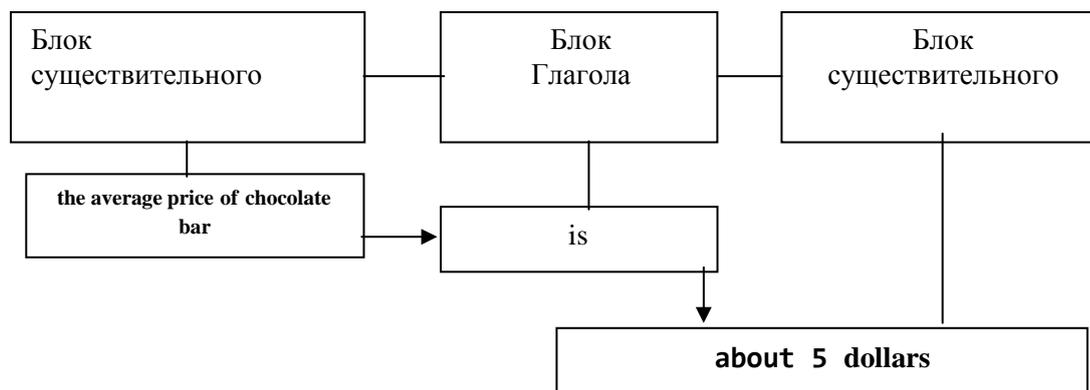


Рисунок 1. Разбор предложения

Первый блок существительного можно определить как субъектный блок (блок субъекта – того, кто или что производит действие). Второй блок существительного можно определить как объектный блок, т.е. того, на кого или на что направлено действие или посредством кого или чего оно реализуется. В дальнейшем будем иметь в виду это определение блока существительного. Оно соответствует стандартной схеме: субъект – действие – объект.

Очевидным образом к рассматриваемому нами предложению можно задать несколько вопросов:

What is average price of chocolate bar?

Is average price of chocolate bar 5 dollars?

How much dollars are average price of chocolate bar?

Для правильной обработки вопросов к семантической базе знаний следует представить текстовую базу знаний в определенном виде. Каждое предложение текстовой базы знаний должно иметь структуру, определяемую правилами вида

```

ПРЕДЛОЖЕНИЕ :- <БЛОК СУЩЕСТВИТЕЛЬНОГО>
<БЛОК ГЛАГОЛА>
<БЛОК СУЩЕСТВИТЕЛЬНОГО>
БЛОК СУЩЕСТВИТЕЛЬНОГО :- <ПРИЛАГАТЕЛЬНОЕ>
<СУЩЕСТВИТЕЛЬНОЕ>
БЛОК СУЩЕСТВИТЕЛЬНОГО: - <СУЩЕСТВИТЕЛЬНОЕ>
<СУЩЕСТВИТЕЛЬНОЕ> (1)
БЛОК СУЩЕСТВИТЕЛЬНОГО: - <СУЩЕСТВИТЕЛЬНОЕ>
БЛОК ГЛАГОЛА :- <НАРЕЧИЕ> <ГЛАГОЛ>
БЛОК ГЛАГОЛА :- <ГЛАГОЛ>
СУЩЕСТВИТЕЛЬНОЕ :- <любое слово Z – аргумент функтора noun(Z)>
ПРИЛАГАТЕЛЬНОЕ :- <любое слово Z – аргумент функтора adj(Z)>
ГЛАГОЛ :- <любое слово Z – аргумент функтора verb(Z)>
    
```

В примере, который мы рассматриваем, соответствующее представление текстовой базы знаний реализуется на основе hashtable:

```

sentences.Add(1, "chocolate and sweets are of high demand today");
    
```

```

sentences.Add(2, "the cheapest chocolate is in
China");
sentences.Add(3, "the average price of chocolate
bar is about 5 dollars");
sentences.Add(4, "chocolate is a top sold product
of all sweets");
sentences.Add(5, "the biggest chocolate
manufacturer is USA");
sentences.Add(6, "the most popular sweets are
chocolate, cookies and ice cream");
sentences.Add(7, "chocolate stimulates mental
activity");
sentences.Add(8, "the largest consumer of
chocolate is Germany");
sentences.Add(9, "there is a lot of competition on te
sweets market");
sentences.Add(10, "sweets prices are quite stable");

```

Очевидно, что обработка вопроса прямым обра-
зом связана с вопросным словом и структурной ор-
ганизацией предложения. Повторим сказанное

выше: мы должны выделять субъект, объект и дей-
ствие. Субъект – это то лицо или предмет, которое
производит действие. Объект – это то лицо или
предмет, на которое направлено действие или по-
средством чего оно выполняется. Таким образом, в
общей структуре (рисунок 2): Субъект – Действия –
Объект.

Первый блок существительного характеризует
субъект, а второй – объект. Далее мы будем назы-
вать их соответственно субъектный блок (существи-
тельного) и объектный блок существительного. Во-
просы к субъектному блоку начинаются со слов

кто (производит действие)?
что (производит действие)?
какой (-ая, ое) производит действие?
что такое ...?
про что (говорится) ...?
о чем (говорится) ...?

Пример вопроса

Which activity stimuli chocolate

Результат обработки таков (рисунок 2).

```

Input a phrase:
which activity stimuli cocolate
QUEST is: why
VERB BLOCK IS: stimulate
Second NB IS: chocolate

THE ANSWER IS: chocolate stimulates mental activity
correctquestion IS: which activity stimulates chocolate

```

Рисунок 2. Пример разбора

Наконец, последнюю группу вопросов образуют
вопросы к объектной части. Это вопросы следую-
щего вида

что делает (субъект)
чем (характеризуется, является) субъект
от чего (зависит) ...
чему равен ...
какой вид имеет ...

Итак, мы рассматриваем общую структуру пред-
ложения в форме триады

Субъект (SB) – действие (PR)– объект (OB).

Ограничиваясь этой общей схемой, можно вы-
делить следующие типы вопросов к семантической
модели

T1. QW PR OB?

T2. QW OB PR?

Здесь QW – это вопросительное слово.

Таким образом, алгоритм обработки семантиче-
ских вопросов сводится к следующему.

1. Произвести грамматический разбор вопроса
и выделить в нем субъектную, объектную части,
действие и вопросное слово. Если соответствующая
часть отсутствует в тексте вопроса, то она остается

пустой. Итак, первым этапом разбора является раз-
бор текста вопроса с формированием соответствую-
щих частей. При этом важной задачей является об-
работка слов с ошибками и их последующее исправ-
ление.

2. На следующей фазе из текстовой базы знаний
последовательно считываются предложения (clause)
и производится их разбор на предмет выделения
субъектной, объектной и глагольной частей. Вып-
няется сравнение выделенных частей текста вопроса
и выделенных частей предложения (clause) на пред-
мет их идентичности. Здесь проблема может состо-
ять в том, что текст вопроса может содержать
ошибки, описки, пропуски и т.п.

Значительно более удобно выполнять обработку
вопросов, если предварительно определить, к ка-
кому разделу (теме) знаний относится вопрос. Это
позволит существенно сократить время поиска от-
вета. Таким образом, фаза разбора делится на две:
определение темы, по которой требуется выполнить
консультацию, и выполнение поиска ответа в рам-
ках выбранной темы.

Список литературы:

1. Гурин Н.И. Компьютерные обучающие системы в издательском деле : учеб. пособие : в 2 ч. / Н.И. Гурин, О.В. Герман. – Минск : БГТУ, 2015. – Ч. 2. – 192 с

