

# АЛГОРИТМ ПОСТРОЕНИЯ СЕМАНТИЧЕСКОЙ СЕТИ С ОБРАБОТКОЙ МЕСТОИМЕНИЙ ТЕКСТА, ОСНОВАННЫЙ НА ДИАЛОГЕ С ПОЛЬЗОВАТЕЛЕМ

Потараев В. В.

Кафедра программного обеспечения информационных технологий, Белорусский государственный университет информатики и радиоэлектроники  
Минск, Республика Беларусь  
E-mail: vic229@rambler.ru

На протяжении последних десятилетий количество текстовой информации, хранимой в электронном виде, значительно выросло. Поэтому автоматизированная обработка текстов является актуальной. Многие тексты содержат местоимения, что приводит к многозначности. Рассмотрим алгоритм построения семантической сети для многозначного текста, основанный на диалоге с пользователем.

## ВВЕДЕНИЕ

К настоящему моменту количество информации, накопленной человечеством, довольно велико. Неструктурированная текстовая информация, или обычный текст, является одним из наиболее распространённых видов информации [1]. В связи с этим автоматизированная обработка текстов является актуальной. Семантический анализ является методом, позволяющим достаточно эффективно обрабатывать информацию за счёт выделения смысловых связей между различными понятиями.

Целью данной работы является построение методов автоматизированной обработки данных, основанных на семантическом анализе.

## I. ЗАДАЧА АВТОМАТИЗИРОВАННОГО ПОСТРОЕНИЯ СЕМАНТИЧЕСКОЙ СЕТИ

Информационные системы довольно часто используются для ответа на некоторый запрос [2]. Для решения данной задачи может быть использован алгоритм, основанный на представлении текста в виде семантической сети. В данной сети узлами являются слова текста, а связями – семантические отношения между словами [3].

## II. ОБРАБОТКА МЕСТОИМЕНИЙ ТЕКСТА

При обработке текста важно учитывать смысл всех его слов. Так, в естественном языке зачастую используются местоимения, и их корректный учёт важен для более полного извлечения информации из текста.

Местоимением называют самостоятельную часть речи, которая используется вместо существительных, прилагательных, числительных и наречий (или их характеристик), чтобы указать на эти существительные, прилагательные, числительные и наречия (а также их признаки и количество), не называя их. Личные местоимения указывают на объект речи – лицо, о котором идет речь. Местоимения 1 (я / мы) и 2 (ты / вы) лица обозначают участников речи. Местоимения

3 лица (он, она, оно / они) указывают на лиц, которые в речи не участвуют [4].

Одни и те же местоимения в каждом конкретном случае могут означать различные понятия. Поэтому при построении семантической сети имеет смысл заменять местоимения на те слова, которые они означают. Для обработки личных местоимений может быть использован специальный алгоритм [5]. Он основан на поиске предыдущего существительного, согласованного с местоимением по форме (род и число).

Пример построения семантической сети с обработкой местоимений представлен на рисунке 1. Сеть построена для текста, состоящего из двух предложений: «Лампа стоит на столе. Она освещает всю комнату». При её построении набор используемых связей ограничен связями «подлежащее-сказуемое» и «место», поэтому в сеть добавлены не все слова текста.

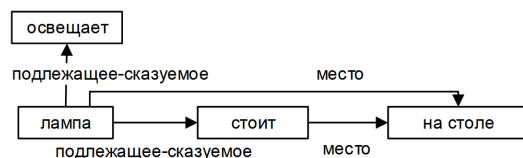


Рис. 1 – Пример семантической сети

## III. МНОГОЗНАЧНОСТЬ ТЕКСТА, СОДЕРЖАЩЕГО МЕСТОИМЕНИЯ

При автоматизированной обработке неструктурированной текстовой информации (т.е. обычного текста) могут возникнуть проблемы, связанные с неоднозначностью естественного языка. Рассмотрим пример текста: «Книга лежит в сумке. Она покупалась в подарок.»

После прочтения данного текста нельзя точно сказать, какой из предметов был куплен в подарок. Очевидно, в некоторых ситуациях можно сделать вывод о покупке конкретного предмета на основании некоторой заранее известной дополнительной информации (например, может быть известно, что сумка была куплена для себя); но такая информация доступна далеко не всегда.

Таким образом, при построении семантической сети в данном случае мы не можем точно сказать, о каком из предметов идёт речь.

#### IV. РАЗРЕШЕНИЕ МНОГОЗНАЧНОСТИ МЕСТОИМЕНИЙ

Модифицируем модель семантической сети, чтобы учитывать многозначность. Для этого при построении сети будем использовать особый вид связей - назовём их «вероятными».

Для приведённого примера текста с многозначностью получится семантическая сеть, представленная на рисунке 2.



Рис. 2 – Пример семантической сети, построенной для текста с многозначностью

Две связи, образованные в результате обработки местоимения, для удобства последующей обработки помечены как относящиеся к одной группе (группа 1).

Для уточнения, какая из вероятных связей действительно имеет место, можно задать вопрос пользователю. Очевидно, что используя союз «или» и информацию, содержащуюся в сети, можно сформулировать вопрос: «Книга или сумка покупалась в подарок?». Ответ пользователя позволяет вместо группы вероятных связей оставить одну «достоверную» связь.

Алгоритм построения семантической сети с обработкой личных местоимений 3 лица может быть представлен следующим образом:

1. Перейти к очередному предложению текста (начиная с первого).
2. Найти в предложении существительные. Для каждого существительного определить его тип (род и число).
3. Найти в предложении личное местоимение 3 лица (он, она, оно, они).
4. Если личное местоимение было найдено, то загрузить существительные из предыдущего предложения, соответствующие ему по типу (род и число). Каждое из них является вероятным подлежащим.
5. Найти сказуемое и добавить его в сеть, а также добавить в сеть соответствующие связи «подлежащее-сказуемое» для каждого вероятного подлежащего. Если таких подлежащих несколько, то связи помечаются как вероятные и относящиеся к одной группе.
6. Добавить остальные виды связей, на основе которых строится сеть.
7. Повторять шаги 1-6 для каждого предложения в тексте.

8. Перейти к очередной группе вероятных связей.
9. Составить вопрос из подлежащих, входящих в группу, а также соответствующего сказуемого и обстоятельств.
10. Отобразить вопрос пользователю и получить ответ.
11. В зависимости от полученного ответа оставить только одну связь «подлежащее-сказуемое» и пометить её как достоверную. Остальные связи данной группы необходимо удалить из сети.
12. Повторять шаги 8-11 для каждой группы вероятных связей в семантической сети.

Таким образом, постановка вопросов перед пользователем позволяет построить корректную семантическую сеть для многозначного текста. Пользователю стоит один раз ответить на вопрос, и тогда последующие пользователи смогут получить ответ много раз. Если ответ не известен, в программном средстве можно предусмотреть возможность пропустить вопрос.

#### V. ВЫВОД

Семантическая сеть является моделью, которая довольно часто используется при обработке информации. Автоматизированное построение семантической сети позволяет значительно ускорить обработку информации, однако для текста с многозначностью может быть необходимо вмешательство человека.

Предложен алгоритм, который позволяет разрешить неоднозначность местоимений при помощи формулировки вопроса пользователю. Аналогичный подход может быть использован и при разрешении других неоднозначностей, которые могут встречаться в обрабатываемом тексте.

#### СПИСОК ЛИТЕРАТУРЫ

1. Усталов, Д. А. Модели, методы и алгоритмы построения семантической сети слов для задач обработки естественного языка: дис. канд. физ.-мат. наук. Институт математики и механики им. Н. Н. Красовского Уральского отделения Российской академии наук, Екатеринбург, 2017.
2. Информационные системы и сети. [Электронный ресурс]. – Электронные данные. – Режим доступа: [https://tsput.ru/res/informat/sist\\_seti\\_fmolekicii/lekciy-1.html](https://tsput.ru/res/informat/sist_seti_fmolekicii/lekciy-1.html). Дата доступа : 08.10.2019.
3. Потараев, В. В. Алгоритм применения семантической сети для поиска ответа на вопрос / В. В. Потараев // Компьютерные системы и сети: Материалы 54-ой научной конференции аспирантов, магистрантов и студентов. – Минск : БГУИР, 2018. – С. 103–105.
4. Все, что вы хотели узнать о местоимениях, но не знали, как спросить [Электронный ресурс]. – Электронные данные. – Режим доступа: <https://blog.tutoronline.ru/vse-cto-vy-hoteli-uznat-o-mestoimenijah-no-ne-znali-kak-sprosit>. Дата доступа : 08.10.2019.
5. Потараев, В. В. Алгоритм построения семантической сети с обработкой местоимений текста / В. В. Потараев // Компьютерные системы и сети: Материалы 55-ой юбилейной научной конференции аспирантов, магистрантов и студентов. – Минск : БГУИР, 2019. – С. 147–148.