



СМЫСЛОВОЙ АНАЛИЗ ТЕКСТА В СИСТЕМЕ АВТОМАТИЧЕСКОГО РЕФЕРИРОВАНИЯ

Лазаренко О. В.

*Харьковский гуманитарный университет «Народная украинская академия»
г. Харьков, ул. Лермонтовская, 27, тел. 050-300-82-83
e-mail: lazolvlad@yandex.ru*

Дыбина А. В.

*Харьковский гуманитарный университет «Народная украинская академия»
г. Харьков, ул. Лермонтовская, 27, тел. 066-613-52-33
e-mail: generosite@mail.ru*

Сабалаев Д. Н.

*Харьковский гуманитарный университет «Народная украинская академия»
г. Харьков, ул. Лермонтовская, 27, тел. 096-276-93-53
e-mail: sab_den@mail.ru*

Современное состояние исследований в области прикладных информационных систем, связанных с обработкой текстов, характеризуется повышением уровня их интеллектуальности. Эти исследования направлены, в частности, на разработку методов смыслового (семантического) анализа текстов, извлечения и использования знаний в компьютерных системах. Смысловой анализ текста – «одна из ключевых проблем как теории создания систем искусственного интеллекта, относящаяся к обработке естественного языка (Natural Language Processing, NLP), так и компьютерной лингвистики» [1].

Результаты семантического анализа могут применяться для решения многих прикладных задач в самых разных областях деятельности человека, связанных с обработкой информации, представленной в текстовом виде. К числу таких систем относятся и системы автоматического реферирования. Несмотря на большую востребованность смыслового анализа текста в работе различных компьютерных систем, эта задача далека от своего решения в силу ее чрезвычайной сложности. Это объясняется, в первую очередь, тем, что смысловой анализ предполагает изучение процессов понимания текста.

Существуют различные подходы к автоматизированному семантическому анализу, «некоторые ученые предлагают рассматривать не просто текст на основе его составляющих слов, предложений, абзацев, но и попытаться выявить тот смысловой образ, который автор хотел создать в сознании своих читателей, посредством этого текста» [1]. Однако многие при этом считают, что на современном этапе развития искусственного интеллекта такая задача является практически неразрешимой.

Мы не разделяем эту точку зрения и считаем, что работа системы автоматического реферирования должна основываться на использовании алгоритма семантического анализа текста. При этом мы исходим из гипотезы,



предложенной голландским ученым А. ван Дейком о том, что при чтении текста люди часто обрабатывают информацию не полностью или неточно и, тем не менее, понимают текст. «Языковому пользователю нет необходимости дожидаться конца абзаца, главы или целого текста, чтобы понять, о чем идет речь в тексте или в его фрагменте, ... пользователь языка может догадаться о теме текста уже после минимума текстовой информации из первых пропозиций. Догадку может подтвердить самая различная информация: заглавие, тематические слова, тематические первые предложения...» [2]. Все это позволяет рассматривать задачу смыслового анализа текста как практически разрешимую даже на нынешнем уровне знаний в области когнитивной лингвистики.

Наш подход к разработке процедуры смыслового анализа предполагает изучение структурной и смысловой схем текста и построение текстовой базы, представляющей собой его «информационное ядро». Текстовая база является семантическим представлением воспринимаемого текста в эпизодической памяти. Однако ван Дейк отмечает, что «понимание текста подразумевает не только представление текстовой базы в эпизодической памяти, но также активацию, обновление и другие способы функционирования так называемой ситуационной модели в эпизодической памяти: это когнитивное представление событий, действий, лиц и вообще ситуаций, о которых говорится в тексте». Из этого следует, что ситуационную модель можно наполнить необходимой семантической информацией в процессе формирования текстовых баз при анализе различных текстов, связанных с похожими ситуациями. Включение эпизодического и семантического знания в ситуационную модель позволяет в процессе смыслового анализа текста постоянно сравнивать текстовую базу с этой моделью «без интерполяции большого объема активированного знания» [2].

Эти обстоятельства легли в основу разрабатываемой нами модели смыслового анализа текста с использованием текстовых баз. Процедура построения текстовой базы в системе АР состоит в выделении с помощью слов-указателей на *объект, результат, цель и метод* предложений, раскрывающих основной смысл анализируемого текста [3]. Эти предложения формируют текстовую базу, из которой выбирается информация для синтеза реферативного текста в соответствии с разработанной ранее моделью индикативного реферата [4].

Примеры слов-указателей на смысловые аспекты в тексте:

Объект: анализируется, изучается, рассматривается, описывается, описываем, рассмотрим, исследуется, реализуется, представляет собой.

Результат: был получен, предложен, дано определение, мы пришли к выводу, в заключение следует отметить, можно сделать взвод.

Цель: нашей целью было, целью данной работы является, данная статья представляет собой попытку.

Метод: при моделировании используются, математическим аппаратом для построения модели является, в качестве средств моделирования используется.



Для выбора предложений в текстовую базу кроме выделения в тексте слов-указателей на соответствующие смысловые аспекты мы используем и результаты анализа заголовка, который рассматриваем как концептуальный инвариант (по определению В. А. Широкова) текста и реферата. Включение в процедуру формирования текстовой базы анализа заголовка позволило более точно определять предложения для текстовой базы, образующие макроструктуру текста. Понятие макроструктуры применяется «для обозначения того, что обычно называется сутью, общим содержанием, темой или топиком текста» [2].

Выводы. Предложенная в работе процедура смыслового анализа текста, основанная на построении текстовой базы, обеспечивает:

1. эффективный поиск и быстрый доступ в тексте к необходимой для реферата информации;
2. возможность извлечения знания, имеющегося в тексте и представленного в виде макроструктуры – семантического плана текста, состоящего из элементов общего знания и, в особенности, из элементов ситуационной модели;
3. возможность накопления текстовых баз для формирования ситуационных моделей, позволяющих обеспечить стратегическое понимание связного текста.

Список литературы

1. Чапайкина Н. Е. Семантический анализ текстов. Основные положения [Текст] / Н. Е. Чапайкина // Молодой ученый. – 2012. – № 5. – С. 112–115.
2. Дейк ван Т. А. Стратегии понимания связного текста / Т. А. ван Дейк, В. Кинч // Новое в зарубежной лингвистике. – Вып. 23: Когнитивные аспекты языка. – М., 1988. – С. 153–211.
3. Дыбина А. В. Побудова текстової бази в системі автоматичного реферування на основі структурно-семантичного аналізу тексту / А. В. Дыбина, О. В. Лазаренко. – Киев, 2010. – 136 с.
4. Лазаренко О. В. Моделювання процесу узагальнення в системі автоматичного реферування / О. В. Лазаренко, А. А. Яковенко. – Х. : Изд-во НУА, 2007. – 136 с.