

АСПЕКТНЫЙ АНАЛИЗ НАУЧНОЙ СТАТЬИ НА ОСНОВЕ ДИСКУРСИВНЫХ МАРКЕРОВ

Машигина Наталья Владимировна

Студент

Факультет ВМК МГУ имени М. В. Ломоносова, Москва, Россия

E-mail: mashigina.nata@mail.ru

Научный руководитель — Ефремова Наталья Эрнестовна

В связи с возрастающим количеством текстов особую актуальность приобретает задача их всестороннего анализа, в том числе и дискурсивного, при котором из текста извлекается главная мысль, отношение автора к предмету обсуждения, его культурное окружение и пр.

Фундаментальной теорией в области дискурсивного анализа является Теория риторических структур (TRC) [2], представляющая дискурс в виде множества дискурсивных единиц, связанных между собой дискурсивными отношениями. Слова-индикаторы дискурсивных отношений называются дискурсивными маркерами.

В данной работе рассматривается аспектный анализ, являющийся разновидностью дискурсивного и ставящий перед собой задачу выделения определенного аспекта научного текста, например, постановки проблемы или описания результатов.

Существует три основных подхода, которые можно применить для аспектного анализа:

- лингвистический;
- статистический;
- машинное обучение.

Лингвистический подход подразумевает поиск в тексте аспектов, опираясь на лингвистические правила. В статистическом подходе учитываются слова-индикаторы, наиболее часто указывающие на аспекты научного текста. При использовании машинного обучения задачу аспектного анализа рассматривают как задачу классификации фрагментов текста по искомым аспектам.

Поскольку задача аспектного анализа научных текстов не является широко исследованной, на данный момент нет размеченного корпуса, который можно было бы использовать для машинного

обучения. Поэтому наиболее перспективным выглядит комбинированный метод, сочетающий в себе элементы лингвистического и статистического подходов [1].

Для решения поставленной задачи был реализован метод выявления слов-индикаторов, помогающих определить аспекты текста. Полученный в результате список был разделен на группы, соответствующие каждому из аспектов. К примеру, аспекту *постановка проблемы* соответствуют индикаторы *проблема заключается, в работе рассматривается* и пр.

Заметим, что с помощью слов-индикаторов можно найти в тексте только место, где упоминается определенный аспект. Для выделения текстового фрагмента, именно описывающего аспект статьи, предложено использовать дискурсивные маркеры. Они отвечают за организацию рассуждений и могут помочь понять, где описание аспекта начинается и заканчивается. Для сбора маркеров были проанализированы существующие корпуса с дискурсивной разметкой на основе ТРС. В результате, к примеру, для отношения *противопоставление* были получены маркеры *несмотря на, однако, however, alternatively*.

Сейчас проводится оценка результатов работы предложенного метода. Одним из способов оценки является сравнение выделенных фрагментов с аннотациями статей, в которых обычно присутствует краткое описание основных аспектов.

Литература

1. Саломатина Н. В., Гусев В. Д. Автоматизация формирования индикаторных словарей и возможности их использования // Труды международной конференции «Диалог 2006». М., 2006. С. 459-463.
2. Mann W. C., Thompson S. A. Rhetorical structure theory: A theory of text organization. University of Southern California, 1987.