

Научная статья

УДК 81'322.2

DOI 10.25205/1818-7935-2022-20-1-21-36

Формальное выявление приемов аргументации в научных текстах

Иван Сергеевич Пименов¹
Наталья Васильевна Саломатина²
Мария Кирилловна Тимофеева³

^{1,3} Новосибирский государственный университет
Новосибирск, Россия

^{2,3} Институт математики им. С. Л. Соболева
Сибирского отделения Российской академии наук
Новосибирск, Россия

¹ pimenov.1330@yandex.ru, <https://orcid.org/0000-0001-5946-9469>

² salomatina_nv@live.ru, <https://orcid.org/0000-0001-2345-6789>

³ mtimof@inbox.ru, <https://orcid.org/0000-0001-8999-2330>

Аннотация

Представлены методы и полученные с их помощью результаты автоматизированного выявления приемов аргументации. Исследуемым материалом служили построенные путем экспертной разметки аргументационные аннотации 25-ти научных текстов двух тематик. Разметка проводилась с помощью веб-инструментов, позволяющих визуализировать аргументативные утверждения и схемы аргументации, строить аргументационную структуру текста в виде ориентированного графа. В графе содержатся два типа вершин: информационные вершины-утверждения и связывающие их модели (схемы) рассуждений из компендиума Уолтона. Под приемами аргументации в данной работе понимается применение повторяющихся отдельных моделей рассуждения и образуемых ими повторяющихся структур (подграфов) аргументации. Разметка приемов, включающих в свою структуру более одной схемы, в графе отсутствует. Использовано описание аргументационных аннотаций и аннотированной коллекции в форме совместного спектра характеристик, содержащих повторяющиеся приемы с их абсолютными и текстовыми частотами. Приемы аргументации, образуемые тремя и более схемами, были выявлены методами частотного анализа подграфов (FSM). Сравнение подграфов осуществлялось с помощью программы из библиотеки NetworkX, реализующей точный алгоритм установления изоморфизма подграфов VF2. Полученные частотные характеристики позволили обнаружить общие закономерности в использовании характерных для научных текстов приемов аргументации, а также специфические особенности их употребления в зависимости от темы (лингвистика и компьютерные технологии). Закономерности проявляются в применении как отдельных схем, так и структурно организованных. Конфигурация последних представляет собой либо цепочки аргументов (до 5 элементов), либо ветвящиеся деревья различной глубины (содержащие до 8 вершин). Данные результаты можно использовать для определения близости текстов по применяемым в них приемам аргументации, что, в свою очередь, может быть дополнительной полезной информацией при кластеризации и классификации текстов, оценке их убедительности, а также при формальном синтезе аргументации.

Ключевые слова

научные тексты, аргументационная структура текста, схемы рассуждения Уолтона, приемы аргументации, общие подграфы графов аргументации, корпусная лингвистика, лингвистика текста

Благодарности

Работа выполнена в рамках государственного задания ИМ СО РАН (проекты № FWNF-2022-0015 и FWNF-2022-0012).

© Пименов И. С., Саломатина Н. В., Тимофеева М. К., 2022

ISSN 1818-7935

Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2022. Т. 20, № 1. С. 21–36
Vestnik NSU. Series: Linguistics and Intercultural Communication, 2022, vol. 20, no. 1, pp. 21–36

Авторы выражают благодарность создателям инструмента аргументационной разметки текстов (лаборатория искусственного интеллекта ИСИ СО РАН) за предоставленную возможность его применения для аннотирования научных текстов

Для цитирования

Пименов И. С., Саломатина Н. В., Тимофеева М. К. Формальное выявление приемов аргументации в научных текстах // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2022. Т. 20, № 1. С. 21–36. DOI 10.25205/1818-7935-2022-20-1-21-36

Formal Identification of Argumentation Patterns in Scientific Texts

Ivan S. Pimenov¹, Natalia V. Salomatina²
Mariya K. Timofeeva³

^{1,3} Novosibirsk State University
Novosibirsk, Russian Federation

^{2,3} Sobolev Institute of Mathematics
of the Siberian Branch of the Russian Academy of Sciences
Novosibirsk, Russian Federation

¹ pimenov.1330@yandex.ru, <https://orcid.org/0000-0001-5946-9469>

² salomatina_nv@live.ru, <https://orcid.org/0000-0001-2345-6789>

³ mtimof@inbox.ru, <https://orcid.org/0000-0001-8999-2330>

Abstract

In this paper we present the methods for the automatic identification of argumentation patterns and the results of their application. These methods have been employed to analyze argumentation annotations of 25 scientific texts from two thematic areas. Under study were the expert annotations constructed manually with the help of web tools for visualizing argumentative statements and argumentation schemes, as well as for modelling the argumentation structure of a text as an oriented graph. Such graphs contain two node types: information nodes denoting statements and their connecting reasoning models (schemas) from Walton's compendium. The regularly employed reasoning models and their structural combinations (argumentation subgraphs) form argumentation patterns. Patterns containing more than one scheme are unmarked in initial graphs. As a result of processing argumentation annotations from the collection, we have constructed a joint spectrum of argumentation patterns with their absolute and text frequencies. The methods of frequent subgraph mining have been used to identify argumentation patterns containing three or more schemes. The subgraph matching has been performed through the use of the NetworkX package which implements the VF2 algorithm for subgraph isomorphism testing. We have analyzed the calculated frequencies to identify both the general principles behind the use of argumentation patterns typical of scientific texts, as well as the specific tendencies of their functioning within distinct thematic areas (linguistics and computer science). These general principles regulate the use of both separate schemes and their structural combinations. The latter appear in two configuration types: either as sequences of argumentation schemes (up to 5 elements) or as tree structures (which contain up to 8 nodes). Specifically, we demonstrate that branching within a tree-form pattern typically presupposes the parallel use of identical argumentation schemes. Additionally, branching of argumentation patterns occurs more often in proximity to the main thesis of a text, than near the initial premises. Finally, the thematic area can condition not only the use of separate schemes, but also their structural combination within complex patterns. The results obtained are applicable to the evaluation of text similarity based on argumentation patterns employed in them, which, in its turn, can improve the clustering and classification of texts, evaluation of their persuasiveness, as well as the formal synthesis of argumentation.

Keywords

popular science texts, scientific texts, argumentation graph, Walton's argumentation schemes, strategy-related patterns, common subgraph, corpus linguistics, text linguistics

Acknowledgements

The research was conducted within the framework of the state contract of Sobolev Institute of Mathematics (projects no. FWNF-2022-0015 and FWNF-2022-0012).

The authors are grateful to the Artificial Intelligence Laboratory of the IIS RAS SB for developing the argumentation annotation software and providing access to it

For citation

Pimenov, I. S., Salomatina, N. V., Timofeeva, M. K. Formal Identification of Argumentation Patterns in Scientific Texts. *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2022, vol. 20, no. 1, pp. 21–36. (in Russ.) DOI 10.25205/1818-7935-2022-20-1-21-36

Введение

Компьютерные исследования аргументации русскоязычных текстов начали проводиться только в последние два-три года, так как русскоязычных коллекций с аргументативной разметкой до последнего времени не существовало. Тем не менее, для многих приложений аргументационный анализ очень важен, в частности для оценки убедительности текстов разных жанров (коммерческих отзывов, научных докладов, студенческих эссе и т. д.), понимания и ведения дебатов, принятия решений в рекомендательных системах и пр. (см., например, [Xinyu Hua, Lu Wang, 2017; Madnani et al., 2012]).

На данный момент нам известны два русскоязычных корпуса, содержащие разметку аргументации, ссылки на них указаны в работах [Fishcheva, Kotelnikov, 2019; Сидорова и др., 2020]. Первый корпус состоит из размеченных и переведенных с английского и немецкого языков микротекстов (ArgMicro). Разметка включает теги, указывающие на тезис и аргументы «за» и «против». Второй корпус содержит научные и научно-популярные тексты, размеченные с помощью разработанного в ИСИ СО РАН набора веб-инструментов. Они позволяют создавать тематические корпуса, визуализировать аргументативные утверждения и схемы аргументации, строить аргументационную структуру текста в виде ориентированного графа. Граф содержит два типа вершин: информационные (вершины-утверждения) и аргументы (модели рассуждений). Инструмент также предоставляет возможность сохранения структуры графа в текстовом (в смысле линейности записи) представлении в формате .json, тем самым поддерживая компьютерную обработку аргументационных аннотаций. Приемы аргументации в аннотации в явном виде не представлены, но косвенно отражены в схемах (моделях рассуждений) и в содержании информационных узлов.

Настоящая работа проделана на коллекции научных текстов из второго корпуса. Жанровая специфика выбранного материала характеризуется значительным объемом аннотируемых текстов (несколько сотен предложений) в противовес англоязычным аргументационным корпусам, преимущественно состоящим из новостных сообщений, интернет-комментариев, эссе, микротекстов, имеющих на порядок меньший объем (несколько десятков предложений). Количественные характеристики исследованной коллекции текстов приведены в разделе 5, посвященном анализу результатов.

Цель работы – построение и апробация метода автоматического выявления приемов аргументации, а также анализ и интерпретация полученных результатов.

1. Исследовательская область

Одно из центральных направлений лингвистики текста – выявление и изучение средств, обеспечивающих связность текста. Такие средства многообразны и присутствуют на всех уровнях языка. Конкретные прикладные исследования, изучая те типы средств связности, которые важны для решаемого класса задач, могут в большей степени фокусироваться на языковой форме текста или на его содержании, в частности на аргументационной структуре представляемого посредством него рассуждения. Например, для решения многих задач автоматической обработки текста может быть полезен анализ лексико-синтаксических средств установления структурной связности (когезии, *cohesion*) [Лукашевич, 2011]. Тесная связь между формальными (прежде всего лексическими и синтаксическими) и логико-семантическими аспектами связности текста (*coherence*) отражена в теоретических работах данного направления лингвистики, например, в [Инькова, Манзотти, 2019]. Автоматизация аргументационного анализа требует рассмотрения как языковых индикаторов, сигнализирующих о границах между структурными единицами рассуждения или о типах связи между ними (моделях рассуждения), так и собственно логической структуры рассуждения, абстрагируемой от ее языковой формы и, возможно, от информационного содержания. Несмотря на важность обоих аспектов для разработки автоматических систем аргументационного анализа

текстов русского языка, одновременное их изучение на современном этапе представляется преждевременным из-за большой трудоемкости. Данная работа посвящена анализу закономерностей, проявляющихся в структурах рассуждения, и выявлению типовых структур (приемов). Языковые индикаторы, существенные для сегментации текста и идентификации моделей рассуждения, пока не рассматриваются.

Большинство компьютерных исследований аргументации посвящено автоматическому поиску и извлечению единиц аргументации разных типов, например предложений, клауз, элементов структуры аргумента – посылок и заключений, а также установлению отношений между ними: связности утверждений и типа связи, отношений «за» и «против» и др. (см. обзоры [Lippi, Torroni, 2016; Lawrence, Reed, 2019]). Приемы аргументации исследуются значительно реже, хотя статистика применения приемов (отдельных схем и образуемых ими структур) может быть полезна при распознавании аргументов с помощью поисковых шаблонов и в машинном обучении.

Авторы некоторых работ предлагают включать приемы аргументации в разметку, что упростит исследование ее эффективности. Например, в работе [Anand et al., 2011] показано, что информация о приемах аргументации оказывает существенную помощь в классификации блогов по степени убедительности.

Моделирование стратегий и приемов актуально для вычислительного синтеза аргументации, примерами такого рода исследований могут служить работы [Wachsmuth et al., 2018; Al-Khatib et al., 2017]. В работе [Al-Khatib et al., 2017] представлены результаты анализа аргументационных стратегий в редакционных новостных статьях. На размеченных текстах изучено применение трех приемов аргументации (статистика, анекдот¹, свидетельство (от эксперта, организации, наблюдателя)) в текстах разной тематики. Приведены примеры 15 встретившихся последовательно приемов (цепочек длиной от 1 до 7), использованных авторами. Выявлена корреляция между темами и используемыми приемами. Область применения результатов – синтез, идентификация аргументов, классификация текстов.

В работе [Wachsmuth et al., 2018] строится модель стратегии, которая опирается, по утверждению авторов, на каноны риторики, сформулированные Аристотелем. Модель реализуется в три этапа, которые выполняются «вручную», а ее оценка проводится экспертами. Для каждого заданного тезиса авторами из базы выбираются аргументативные дискурсивные единицы (ADU). Из них составляются структуры аргументации (цепочки из посылок, заключений). Выбранные утверждения оформляются в определенном стиле согласно принимаемой стратегии. Стратегия убеждения в этой работе определяется по пропорции средств соответствующей (контекстуальной) семантики, использованных в тексте, например, логос – 70 %, этос – 10 %, пафос – 20 %. Результаты синтеза коротких аргументативных текстов совпадают у разных экспертов примерно на 50 %. Изучение структурных особенностей в организации последовательности рассуждений, реализуемой в текстах, не входило в задачу авторов, и это их просчет, поскольку знания о такой организации могут быть полезны на этапе формирования структуры. Например, они могли бы отчасти нивелировать различия в оценке аргументации, синтезированной разными экспертами.

Авторы работы [El Baff et al., 2019] предложили алгоритм, согласно которому аргументационные цепочки строятся автоматически вне зависимости от темы ADU из базы, которая и содержит тезисы, аргументы «за» и «против» с указанием реализованной в них стратегии (логос, пафос). Выбранные ADU были упорядочены согласно критерию связности, определяемой путем вычисления семантического расстояния между ними. Совпадение с построенными экспертами структурами аргументации составило здесь порядка 50 % на уровне двух подряд следующих аргументов. Мы полагаем, что полезным дополнением к данному методу могли бы быть детализация связей между аргументами и учет организации схем в структуре.

¹ В английском понимании этого слова.

В нашей работе основное внимание уделяется исследованию организации схем рассуждений в структурах (графах) аргументации. Под приемами нами понимается применение отдельных повторяющихся моделей рассуждений и образуемых ими повторяющихся структур в аргументации, реализуемой в авторских текстах. Задача поиска приемов аргументации в нашем случае – это извлечение всех повторяющихся подграфов в заданной коллекции. Задачи такого сорта могут решаться методами частотного анализа подграфов *FSM* (Frequent subgraph mining) [Jiang et al., 2004], который включает генерацию подграфов, представляющих интерес, а также подсчет частоты встречаемости этих подграфов в заданном наборе данных. В нашем случае генерацию кандидатов заменяет разбиение графа аргументации текста на подграфы с фиксированным числом вершин. Для подсчета частоты в этом случае требуется провести процедуру сравнения полученных подграфов, которая известна как проверка подграфов на изоморфизм. Два самых известных на данный момент алгоритма, позволяющих получить точное решение, предложены в работах [Ullmann, 1976; Cordella et al., 2001]. Мы применяем последний, использующий стратегию поиска в глубину и набор правил для сокращения времени поиска (и потребляемой памяти компьютера).

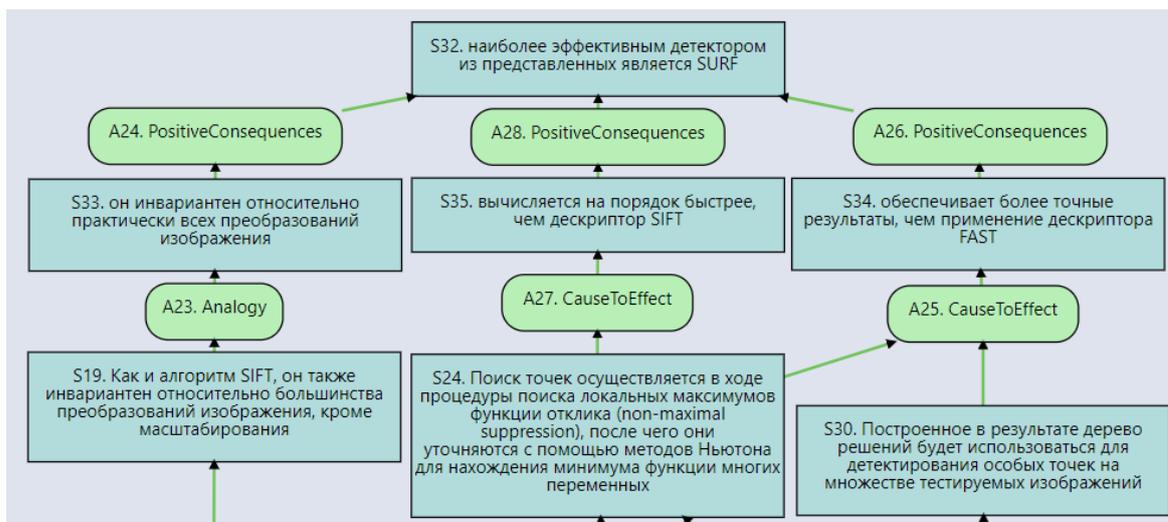
2. Моделирование аргументации посредством схем рассуждения

Под аргументом здесь понимается совокупность связанных утверждений, применяемых для доказательства заключительного утверждения. В составе утверждений аргумента выделяют утверждения-посылки и утверждение-заключение. Они объединяются определяемыми моделью (схемой) рассуждения отношениями. Заключение одного утверждения может служить посылкой для одного или нескольких других утверждений или использоваться в качестве заключения у других аргументов. Утверждение представляет собой сформулированное на естественном языке высказывание. Связи, устанавливаемые между аргументами, позволяют объединить их в единую аргументативную структуру.

Для описания аргументов и аргументативных структур приняты стандарты, фиксированные форматом AIF (Argument Interchange Format) [Rahwan, Reed, 2009]. Согласно формату, аргументы представляются ориентированными графами, в которых выделяют два типа вершин: информационные вершины (вершины-утверждения) и вершины-схемы. С вершинами-утверждениями сопоставляются посылки, заключения, а с вершинами-схемами – схемы (модели) рассуждений. В используемом нами для разметки текстов инструменте доступно порядка сорока схем аргументации из компендиума Уолтона [Walton, 2008]. В графе выделяется корневая вершина, соответствующая главному тезису. Тем самым предполагается, что в любом тексте можно выделить ровно один главный тезис (в научных статьях он часто отражается в названии, но также в явно сформулированной или неявно выраженной цели работы, может дублироваться в заключении).

Пример фрагмента аргументационной разметки текста приведен на рисунке. Вершины-утверждения изображены прямоугольниками, вершины-схемы – эллипсами. Утверждения и схемы обозначены символами S и A соответственно и снабжены номерами согласно линейному порядку утверждений в тексте. Стоит, однако, отметить, что линейный порядок отличается от порядка в аргументационной структуре: посылки могут выражаться в тексте как до, так и после заключения, а между тезисом и доводами в его поддержку может приводиться обоснование этих доводов.

На рисунке представлены основные типы связей между утверждениями [Lawrence, Reed, 2019], к примеру: Convergent Arguments (S33, S35 → S32), Linked Arguments (S24 & S30 → S34), Divergent Arguments (S24 → S34, S35), Sequential Arguments (S24 → S35 → S32).



Фрагмент аргументационной разметки текста
A fragment of the argumentation annotation of a text

3. Методы выявления приемов аргументации

Как уже было определено в п. 1, под приемами аргументации в данном исследовании понимается применение как отдельных моделей рассуждений, так и подграфов (т. е. графов, образованных некоторым подмножеством вершин-схем графа аргументации и некоторым подмножеством смежных с ними рёбер), повторяющихся в аргументации в разных текстах.

Пусть $A = \{a_i\}$ – множество аргументационных аннотаций коллекции текстов. Каждая аннотация a_i представляет собой связный граф G_i , $G_i = \langle V_i^{inf} \cup V_i^{ch}, E_i \rangle$, где V_i^{inf} – множество информационных вершин, V_i^{ch} – множество вершин-схем, E_i – множество рёбер, соединяющих вершины. На данном этапе нами проводится анализ структур (графов) \hat{G}_i , образуемых из графов G_i с сохранением только вершин-схем: $\hat{G}_i = \langle V_i^{ch}, \hat{E}_i \rangle$. Рёбра \hat{E}_i строятся следующим образом: 1) в случае удаления информационной вершины при цепочечной их организации – путем замены входящего и исходящего ребра на ребро, соединяющее две схемы и сохраняющее направление; 2) в случае конвергенции информационных вершин аналогичным образом строятся рёбра при удалении каждой из них; 3) в случае дивергентного связывания информационных вершин строятся рёбра, объединяющие каждую схему с входящим в информационную вершину ребром с каждой схемой, соединенной исходящими из информационной вершины рёбрами. Аннотацию a_i удобно представлять совместной частотной характеристикой, а именно совокупностью характеристик $\Phi(a_i) = \{\Phi_1(a_i), \dots, \Phi_n(a_i), \dots, \Phi_{N_{max}}(a_i)\}$, где $\Phi_1(a_i)$ – статистика отдельных схем (разных вершин из V_i^{ch} и частот их встречаемости $F_{abs}(V_i^{ch})$). Характеристика $\Phi_n(a_i)$ ($n > 1$) аккумулирует – множество подграфов $\{sg_{ik}^n\}$ графа \hat{G}_i с фиксированным числом вершин, равным n ($n = 2, \dots, N_{max}$, N_{max} – задаваемая константа, ограниченная сверху значением $|V_i^{ch}|$): $\Phi_n(a_i) = \{sg_{ik}^n\}$, $k = 1, \dots, K_n$ – число *разных* подграфов с n вершинами. Для каждого подграфа sg_{ik}^n вычисляется число его вхождений $F_{abs}(sg_{ik}^n)$ в граф \hat{G}_i .

Представление множества A включает аналогичные характеристики и является совместным частотным спектром коллекции: $\Phi(A) = \{\Phi_1(A), \Phi_2(A), \dots, \Phi_n(A), \dots, \Phi_{N_{max}}(A)\}$, где N_{max} – максимальное число вершин в подграфах, общих хотя бы для двух графов аргументации разных текстов: $\Phi_n(A) = \{sg_{ik}^n\}_{i=1}^{|A|}$. Для каждого подграфа sg_{ik}^n вычисляются $F_{abs}(sg_{ik}^n)$ и $F_{txt}(sg_{ik}^n)$ – абсолютная и текстовая частота в коллекции A . Под текстовой частотой понимается число текстов из коллекции A , в графах которых встречается данный подграф.

Задача построения $\Phi(A)$ эквивалентна отысканию общих совпадающих по структуре подграфов sg_{ik}^n и sg_{jk}^n в графах аргументации \hat{G}_i и \hat{G}_j для всех $i, j = 1, \dots, |A|$; $i \neq j$; $k = 1, \dots, K_n$. Как сказано во введении, для определения идентичности структур двух графов используется понятие «изоморфизм графов». По определению два графа изоморфны, если у них одинаковое число вершин (n) и вершины каждого из них можно переименовать так, что в первом графе две вершины соединены ребром тогда и только тогда, когда вершины с такими же именами соединены во втором графе.

При поиске изоморфного подграфа для каждой комбинации значений индексов i, j, k, n ($i \neq j$) в каждом отдельном случае нужно определять, не содержит ли граф \hat{G}_j подграф, изоморфный подграфу $sg_{ik}^n \subset \hat{G}_i$. Для установления изоморфизма используется реализация алгоритма *VF2* из библиотеки *NetworkX*, предназначенной для обработки графов на Python [Hagberg et al., 2008].

4. Поиск подграфов в графах аргументации

Построение совместного частотного спектра $\Phi(A)$ производится в три этапа.

I. *Этап предобработки*. Основные функции этого этапа – извлечение из каждого аргументационного графа, считываемого из описания в формате *.json*, множества применяемых схем и множества его подграфов согласно заданному числу n вершин-схем $\{sg_{ik}^n\}$ ($n > 1$). Для этого выполняются следующие шаги.

1. Преобразование текстового представления графа из *.json*-формата в матрицу смежности.

2. Построение по матрице смежности компактного представления графа \hat{G}_i , состоящего только из вершин-схем (без информационных вершин с текстовым содержанием отдельных утверждений, но с сохранением всех связей, проходящих через вершину-схему). В условиях отсутствия информационных вершин для обеспечения связности графа \hat{G}_i добавляется техническая вершина *MainThesis*, обозначающая главный тезис исходного графа (с указанием всех переходов от ведущих к нему вершин-схем).

Удаление из рассмотрения информационных вершин обусловлено абстрагированием на данном этапе исследования от содержания тезисов при анализе приемов аргументации (приемы определяются на уровне типовых моделей рассуждения). В этом случае число вершин графа уменьшается почти в два раза, что заметно сокращает трудоемкость поиска изоморфных подграфов.

3. Извлечение пар связанных вершин.

4. Извлечение подграфов $sg_{ik}^n \subset \hat{G}_i$ для всех $i = 1, \dots, |A|$; $k = 1, \dots, K_n$. Нижняя граница диапазона n соответствует наименьшему числу вершин, при котором проявляется структурная вариативность между подграфами (все подграфы с $n = 2$ образуют линейную цепочку, три же вершины могут соединяться последовательно, либо через конвергентный переход от двух к третьей, либо через дивергенцию от одной к двум другим). Верхняя граница изменения n определяется в ходе работы алгоритма на этапе исследования подграфов с $n + 1$ вершиной.

Извлечение всех подграфов sg_{ik}^n осуществляется полным перебором: из множества вершин графа поочередно выбирается n вершин ($n > 2$), для которых по матрице смежности графа \hat{G}_i проводится проверка связности вершин.

II. *Построение совместного частотного спектра* $\Phi(A)$.

1. Вычисление характеристик $\Phi_1(A)$ на базе V_i^{ch} для всех $i = 1, \dots, |A|$.

2. Вычисление характеристик $\Phi_2(A)$ по парам связанных вершин из (V_i^{ch}, \hat{E}_i) для всех $i = 1, \dots, |A|$.

3. Вычисление характеристик $\Phi_n(A)$, $n > 2$.

Поиск изоморфных подграфов. На этом этапе производится попарный анализ преобразованных графов \hat{G}_i всех текстов коллекции с помощью алгоритма *VF2*. Выявляемые изоморфизмы sg_{ik}^n фиксируются в $\Phi(A)$ с указанием вычисленных $F_{abs}(sg_{ik}^n)$ и $F_{txt}(sg_{ik}^n) > 1$.

Итеративный поиск изоморфных подграфов с числом вершин $n + 1$ ($n > 2$) реализуется только при обнаружении для пары текстов хотя бы одного изоморфизма на n вершинах. Из анализа исключаются вершины, не вошедшие в изоморфные подграфы с n вершинами; они вообще не могут входить ни в какой изоморфный подграф размером $n + 1$: в противном случае они бы содержались в его вложенном подграфе размером n , также изоморфном для двух текстов.

III. *Постобработка результатов*. Как было отмечено выше, алгоритм *VF2* выявляет изоморфизмы, т. е. графы одинаковой структуры (без учета имен вершин). На данном этапе исследования при анализе аргументационных графов мы ограничиваемся рассмотрением случая точного совпадения имен вершин-схем. Исследование одинаковых конфигураций, образуемых изоморфными подграфами с разными именами вершин, несомненно, также представляет интерес и будет проведено в следующих работах. Постобработка изоморфных подграфов осуществляется с целью фильтрации подграфов с несовпадающими именами вершин.

5. Результаты эксперимента

В данном эксперименте использована коллекция из 25 научных текстов двух предметных направлений: 14 – лингвистика (*Ling*) и 11 – информационные технологии (*Comp*). Аргументационная разметка каждого текста проводилась одним и тем же участвующим в эксперименте экспертом. Объем аннотируемых текстов варьируется от 800 до 1 500 слов; они отличаются выделением большого числа тезисов и подробной детализацией при фиксировании аргументативных связей. Все тексты скачаны из онлайн библиотеки (<https://cyberleninka.ru/>). Всего в получившихся аннотациях насчитывается 1 066 аргументативных утверждений и 925 схем.

Анализ встречаемости схем аргументации (из характеристики $\Phi_1(A)$) показал, что десяти самых частотных из них достаточно, чтобы хорошо описать употребляемые авторами отдельные модели (приемы) рассуждений: они покрывают 90 % всех схем, примененных для разметки научных текстов. В табл. 1 приведены их названия, обозначения в полной и краткой (используемой далее) формах, абсолютные и текстовые частоты встречаемости в коллекции A (F_{abs} , F_{txt}), а также относительные частоты встречаемости разных схем для отдельных тем, вычисленные по значениям F_{abs} .

Самой частотной в текстах коллекции является схема с поддержкой тезиса примером (*Example*). Эта модель применяется почти во всех работах (24 из 25), но значимо то, что в лингвистических статьях она встречается значительно чаще, чем в статьях по информационным технологиям (71 против 29 %) даже при неравенстве числа тематических текстов в нашей коллекции. Прагматическое обоснование состоит в том, что в работах по лингвистике автору удобно иллюстрировать свои тезисы типичными речевыми примерами. Кроме того, статьям этой тематики свойственен анализ глубинных языковых структур через их наглядные внешние проявления (по схемам *CorrelationToCause* для общих языковых тенденций и *Sign* на уровне частных семиотических переходов; две эти указанные модели встречаются в *Ling* текстах в 69 и 70 % случаев). Наоборот, прикладная направленность работ по ИТ влечет за собой предпочтение практических доказательств (*PracticalReasoning*). Наконец, обоим предметам присущи прямые причинно-следственные связи (от структуры программы к особенностям функционирования, по схеме *CauseToEffect*) и основательная систематизация описываемых явлений по модели *VerbalClassification*.

Соотношение текстовой и абсолютной частот характеризует избирательность применения отдельных схем авторами. Например, рассуждения от экспертного мнения или конфликта применяются в большем количестве текстов (15), чем модели *PartToWhole* и *Sign* (11 и 12), но при этом уступают им по абсолютным частотам (47 и 34 против 68 и 56). Это объясняется особенностями их употребления в научных текстах.

Таблица 1
 Частоты встречаемости отдельных схем аргументации
 Table 1
 Frequencies of individual argumentation schemes usage.

Схема	Обозначение	Аббр.	F_{abs}	F_{txt}	$F_{Comp}, \%$	$F_{Ling}, \%$
От примера	<i>Example</i>	<i>E</i>	183	24	29	71
От взаимосвязи к причине	<i>Correlation to Cause</i>	<i>CtoC</i>	136	22	31	69
Через классификацию	<i>Verbal Classification</i>	<i>VC</i>	133	21	49	51
От причины к следствию	<i>Cause to Effect</i>	<i>CtoE</i>	114	23	43	57
От практической цели	<i>Practical Reasoning</i>	<i>PR</i>	69	20	58	42
От части к целому	<i>PartToWhole</i>	<i>PtoW</i>	68	11	34	66
От знака к означаемому	<i>Sign</i>	<i>S</i>	56	12	30	70
От экспертного мнения	<i>Expert Opinion</i>	<i>EO</i>	47	15	34	66
Через конфликт	<i>LogicalConflict</i>	<i>LC</i>	34	15	44	56
От применяемого метода	<i>AppliedMethod</i>	<i>AM</i>	23	12	43	57

Так, при доказательстве главного тезиса достаточно привести ограниченное число чужих авторитетных суждений, на основе которых затем развивается подробный самостоятельный анализ, в котором логические переходы между частью и целым, знаком и означающим способны выстраиваться неоднократно. Конфликты же могут применяться в обосновании главного тезиса через указание возможных возражений и их встречное опровержение, при этом рассмотрение большого числа таких возражений нетипично ввиду отвлечения внимания от основного доказательства. Схожим образом при рассуждении от практической цели (на уровне всего исследования или его отдельного этапа) могут подробно анализироваться пути ее достижения, что отражается в соотношении текстовой и абсолютной частоты схемы: *PracticalReasoning* применяется 69 раз в 20 текстах, в то время как для *PartToWhole* отмечаются схожие 68 реализаций лишь в 11 работах.

В работе [Пименов, 2021] было, однако, показано, что приемы аргументации на уровне применения авторами отдельных схем отличаются и для номинально близких жанров, а именно научного и научно-популярного.

Более сложным приемам аргументации соответствуют подграфы двух типов. Совокупность связанных вершин в них образует либо дерево (далее – «цепочка»), в котором каждая вершина имеет только одно входящее и исходящее ребро, кроме корневой (у нее нет исходящего ребра) и листовой (у нее нет входящего ребра), либо дерево, не обладающее таким свойством (далее – дерево). Следует отметить, что согласно сложившейся в области аргументационных исследований традиции, дуги в дереве ориентированы нетрадиционным образом: от потомков к родительским узлам, что, впрочем, характерно и для некоторых других предметных областей, например теории риторических структур [Taboada, Mann, 2006]. Переориентация дуг может быть проведена однозначно заменой каждой дуги в дереве на направленную противоположным образом: если дуга вела из вершины a в вершину b , то она заменяется на дугу из b в a .

Количество всех встретившихся в 2-х и более текстах коллекции разных цепочек (из характеристики $\Phi_n(A)$, $n = 2, \dots, 5$) и разных деревьев (из характеристики $\Phi_n(A)$, $n = 3, \dots, 8$) – 231 и 221 соответственно. Самые часто встречающиеся в разных текстах коллекции цепочки приведены в табл. 2, а деревья – в табл. 3.

Таблица 2

Частоты встречаемости цепочек аргументов

Table 2

Frequencies of argument sequences usage

<i>N</i>	<i>F_{txt}</i>	<i>F_{Comp}</i>	<i>F_{Ling}</i>	Цепочки
<i>n</i> = 2				
1	14	7	7	<i>CtoE</i> → <i>PR</i>
2	13	4	9	<i>CtoC</i> → <i>CtoE</i>
3	12	3	9	<i>E</i> → <i>CtoC</i>
4	12	2	10	<i>CtoE</i> → <i>CtoC</i>
5	11	5	6	<i>E</i> → <i>CtE</i>
<i>n</i> = 3				
1	8	1	7	<i>E</i> → <i>CtoC</i> → <i>CtoE</i>
2	7	4	3	<i>CtoE</i> → <i>PR</i> → <i>CtoE</i>
3	6	0	6	<i>CtoC</i> → <i>CtoE</i> → <i>CtoC</i>
4	5	1	4	<i>CtoC</i> → <i>CtoE</i> → <i>MainThesis</i>
5	5	4	1	<i>PR</i> → <i>CtE</i> → <i>PR</i>
<i>n</i> = 4				
1	5	4	1	<i>CtoE</i> → <i>PR</i> → <i>CtoE</i> → <i>PR</i>
2	4	0	4	<i>E</i> → <i>CtoC</i> → <i>CtoE</i> → <i>CtoC</i>
3	3	0	3	<i>CtoE</i> → <i>CtoC</i> → <i>CtoE</i> → <i>MainThesis</i>
4	3	0	3	<i>EO</i> → <i>CtoE</i> → <i>CtoC</i> → <i>CtoE</i>
<i>n</i> = 5				
1	2	2	0	<i>E</i> → <i>CtoE</i> → <i>PR</i> → <i>CtoE</i> → <i>PR</i>
2	2	0	2	<i>CtoC</i> → <i>CtoE</i> → <i>CtoC</i> → <i>CtoE</i> → <i>MainThesis</i>
3	2	0	2	<i>CtoE</i> → <i>E</i> → <i>CtoC</i> → <i>CtoE</i> → <i>MainThesis</i>
4	2	1	1	<i>CtoE</i> → <i>PR</i> → <i>CtoE</i> → <i>PR</i> → <i>MainThesis</i>

Таблица 3

Частоты встречаемости структурных приемов аргументации

Table 3

Frequencies of tree-structured argumentation patterns usage

<i>N</i>	<i>F_{txt}</i>	<i>F_{Comp}</i>	<i>F_{Ling}</i>	Деревья
<i>n</i> = 3				
1	6	0	6	<i>CtoC</i> ↘ <i>CtoC</i> → <i>CtoE</i>
2	5	3	2	<i>CtoC</i> ↘ <i>CtoC</i> → <i>MainThesis</i>
3	5	0	6	<i>E</i> ↘ <i>E</i> → <i>CtoC</i>
4	4	0	4	<i>CtoE</i> ↘ <i>CtoE</i> → <i>PR</i>
<i>n</i> = 4				
1	3	0	3	<i>PtoW</i> ↘ <i>E</i> → <i>PtoW</i> → <i>PtoW</i>
2	3	3	0	<i>E</i> ↘ <i>E</i> → <i>CtoE</i> → <i>PR</i>

Окончание табл. 3

<i>N</i>	<i>F_{txt}</i>	<i>F_{Comp}</i>	<i>F_{Ling}</i>	Деревья
3	3	1	2	$E \rightarrow$ $CtoE \rightarrow PR \rightarrow CtoE$
<i>n = 5</i>				
1	3	0	3	$CtoC \rightarrow CtoC \rightarrow$ $CtoC \rightarrow CtoC \rightarrow MainThesis$
2	2	2	0	$VC \rightarrow$ $VC \rightarrow CtoC$ $PtoW \rightarrow VC \nearrow$
3	2	1	1	$VC \rightarrow$ $VC \rightarrow PtoW \rightarrow CtoE$ $VC \nearrow$
<i>n = 6</i>				
1	2	2	0	$E \rightarrow$ $E \rightarrow$ $E \rightarrow CtoE \rightarrow PR$ $CtoC \nearrow$
2	2	0	2	$E \rightarrow CtoC \rightarrow$ $CtoC \rightarrow CtoE \rightarrow CtoC \rightarrow MainThesis$
3	2	1	1	$CtoC \rightarrow$ $CtoC \rightarrow$ $CtoC \rightarrow MainThesis$ $CtoE \rightarrow CtoC \nearrow$
<i>n = 7</i>				
1	2	2	0	$CtoC \rightarrow$ $CtoC \rightarrow$ $CtoC \rightarrow MainThesis$ $CtoE \rightarrow PR \rightarrow CtoE \nearrow$
2	2	0	2	$CtoC \rightarrow CtoE \rightarrow$ $CtoE \rightarrow E \rightarrow CtoC \rightarrow CtoE \rightarrow MainThesis$
3	2	1	1	$VC \rightarrow$ $VC \rightarrow$ $VC \rightarrow PR$ $CtoE \rightarrow PR \rightarrow CtoE \nearrow$
<i>n = 8</i>				
1	2	0	2	$E \rightarrow CtoC \rightarrow$ $CtoE \rightarrow CtoE \rightarrow CtoC \rightarrow$ $CtoC \rightarrow$ $CtoC \rightarrow MainThesis \nearrow$

Исследование цепочек, проведенное на коллекции научно-популярных и научных текстов и представленное в работе [Саломатина, Пименов, 2021], выявило сильную вариативность в организации схем – самые длинные общие цепочки характеризовались длиной 4 (без учета узла MainThesis).

Анализ приемов аргументации (цепочек и деревьев с тремя и более вершинами) позволил выявить нижеперечисленные зависящие и не зависящие от предмета особенности организации рассуждений в научных текстах коллекции.

Во-первых, если в узлах дерева возникает ветвление, то вне зависимости от уровня узла в большинстве случаев схемы в узлах следующего уровня на разных ветвях будут одинаковы. Это значит, что построение параллельных доводов чаще происходит по одной и той же модели вне зависимости от предметной области текста (см. примеры в табл. 3: схема *MainThesis* ($n = 3$, 5–8) поддерживается несколькими *CorrelationToCause*; *CorrelationToCause* ($n = 5$), *PracticalReasoning* ($n = 7$) и *PartToWhole* ($n = 5$) – несколькими *VerbalClassification*; *CauseToEffect* поддерживается двумя *CorrelationToCause* ($n = 3$) и несколькими *Example* ($n = 4$, 6) и др.

Такое построение свидетельствует о частом использовании так называемой «интенсивной аргументации»: через организацию рассуждения от нескольких аргументов одного типа. Доказательство тезиса через несколько аргументов разных типов является менее регулярным.

Иными словами, если авторы или составители научных текстов выстраивают рассуждение от нескольких доводов, то они регулярно организуют эти аргументы по одинаковой модели рассуждения (при возможном применении нескольких: *CorrelationToCause*, *Example*, *VerbalClassification*). Сочетание двух доводов через разные модели является менее типичным. Отмеченная тенденция может объясняться удобством восприятия: читателю проще осмыслить доказательство от нескольких доводов тогда, когда они все соотносятся с тезисом одинаково. И наоборот, использование нескольких доводов разного типа способно оказаться избыточным, и тогда более убедительным окажется применение лишь одного довода из нескольких доступных.

Во-вторых, корневая (техническая) вершина в графе *MainThesis* (см. п. 2.1) чаще является корнем в дереве, чем конечным звеном в цепочке: в 63-х подграфах из 221 (28,5 %) аргументы направлены непосредственно на доказательство главного тезиса, тогда как среди цепочек примыкают к главному тезису только в 25-ти из 122 (с тремя или более вершинами), т. е. в 20 % случаев. Важно подчеркнуть, что цепочки встречаются и внутри подграфов (при рассмотрении их отдельных путей), в то время как обратное структурно невозможно. Таким образом, для научных текстов в окрестности главного тезиса чаще отмечаются сочетания нескольких аргументов, чем одного завершающего в цепочке.

Причиной этому служит предпочтение в научных текстах многоуровневых доказательств главного тезиса: ключевые выводы анализируются в различных аспектах (возможно, разрозненных, если пути рассуждения не пересекаются вне главного тезиса), а их обоснование обеспечивает аргументационную связность и, следовательно, логико-семантическую целостность всего текста. Поскольку же главный тезис служит связующим элементом для всех путей рассуждения, типичным является выделение смысловых подблоков текста в непосредственной с ним связи.

Наглядный пример для двух указанных наблюдений приведен на рисунке. На нем представлен главный тезис текста (без исходящих связей к каким-либо иным утверждениям), в котором утверждается преимущество одного алгоритма над двумя другими. Три довода в поддержку главного тезиса приведены по одной и той же модели (от положительных результатов), а их обоснование состоит в подробном представлении каждого из алгоритмов по отдельности.

Наконец, в употреблении приемов аргументации отмечается заметное влияние предметно-тематической области. Так, среди 15 наиболее частотных приемов (реализуемых в 5 или более текстах) треть используется исключительно в *Ling*. К этим приемам относятся, например, $[E, E \rightarrow CtC]$, $[CtE \rightarrow CtC \rightarrow CtE]$, $[CtC, CtC \rightarrow CtE]$, $[CtC \rightarrow CtE \rightarrow CtC]$ и $[E \rightarrow CtC \rightarrow CtE]$. Несмотря на высокую частоту каждой из этих схем по отдельности (в том числе внутри текстов по *Comp*: по абсолютным значениям частоты указанные модели уступают в них только классификационной схеме VC), их указанные комбинации применяются только в текстах одной группы.

Такая особенность обуславливается предметно-тематическими ограничениями на сочетаемость схем: модели *Example*, *CauseToEffect* и *CorrelationToCause* применяются в работах

по *Comp* не совместно друг с другом, а с добавлением иных схем. К примеру, цепочка [CtE → PR → CtE], отличающаяся на одну замену от [CtE → CtC → CtE], встречается в 3-х статьях по *Ling* и в 4-х по *Comp*.

Предметно-тематические особенности аргументации проявляются и в конфигурации деревьев, что особенно заметно при $n > 4$. Деревья, характеризующие приемы аргументации, применяемые в *Ling*, значительно реже демонстрируют интенсивное ветвление в вершинах (что часто проявляется в корневой вершине) в отличие от текстов *Comp*. Исключение составляют случаи применения моделей *Example* и *VerbalClassification*, которые встречаются при интенсивном ветвлении в вершинах деревьев, характеризующих аргументационные приемы в текстах обеих предметных тематик.

Заключение

На коллекции русскоязычных научных текстов с «ручной» разметкой аргументации, построенной в виде графа с вершинами-утверждениями и вершинами-схемами, связывающими эти утверждения согласно реализуемой модели рассуждения, проведено количественное исследование по применению приемов аргументации разными авторами. Под приемами в данной работе понимается использование повторяющихся отдельных моделей рассуждения и образуемых ими повторяющихся структур (подграфов). Использовано описание аргументационных аннотаций и всей коллекции в форме совместного спектра характеристик, содержащих повторяющиеся приемы с их абсолютными и текстовыми частотами. Получены частотные характеристики употребления отдельных схем аргументации и подграфов двух видов: цепочек схем и деревьев разной глубины. Для выявления структурных приемов, содержащих три и более схем аргументации, использован трехэтапный алгоритм поиска общих подграфов в графах аргументации, включающий 1) предобработку графов аргументации, главным в которой является разбиение графов на подграфы с фиксированным числом вершин; 2) поиск изоморфных подграфов с помощью алгоритма *VF2* и 3) постобработку для выявления рассматриваемых на данном этапе подграфов с точным совпадением имен вершин-схем. На основе частотных характеристик приемов аргументации продемонстрированы их общие свойства для научных текстов и специфические особенности в зависимости от темы текста (предпочтительное употребление отдельных схем, ограничения на сочетаемость схем внутри цепочек, различия в конфигурации деревьев).

Полученные результаты могут быть использованы для определения близости текстов по применяемым в них приемам (что может быть полезно при классификации и кластеризации текстов), для оценки убедительности текстов, для формального синтеза аргументации.

Список литературы

- Инькова О., Манзотти Э.** Связность текста: меререологические логико-семантические отношения. М.: ИД ЯСК, 2019. 376 с.
- Лукашевич Н. В.** Тезаурусы в задачах информационного поиска. М.: Изд-во Моск. ун-та, 2011. 512 с.
- Пименов И. С.** Специфика аргументационного аннотирования научных и научно-популярных текстов // Корпусная лингвистика – 2021: Тр. Междунар. конф. СПб.: Скифия-принт, 2021. ISBN 978-5-98620-557-1
- Саломатина Н. В., Пименов И. С.** N-граммный анализ аргументационных аннотаций текстов // Знания – Онтологии – Теории (ЗОТ): Тр. Междунар. конф., 2021. URL: https://drive.google.com/file/d/1vLIHJk_cjYaSK9lwpw1i2HPNr8Qtmy1t/view
- Сидорова Е. А., Ахмадеева И. Р., Загорюлько Ю. А., Серый А. С., Шестаков В. К.** Платформа для исследования аргументации в научно-популярном дискурсе // Онтология проектирования. 2020. Т. 10, № 4 (38). С. 489–502.

- Al-Khatib, K., Wachsmuth, H., Hagen, M., Stein, B.** Patterns of Argumentation Strategies across Topics. In: Proc. of the 2017 Conference on Empirical Methods in Natural Language. Copenhagen, Denmark, 2017, pp. 1351–1357.
- Anand, P. et al.** Believe Me – We Can Do This! Annotating Persuasive Acts in Blog Text. In: Computational Models of Natural Argument. San Francisco, CA, USA, 2011, August 7.
- Cordella, L. P., Foggia, P., Sansone, C., Vento, M.** A (Sub)Graph Isomorphism Algorithm for Matching Large Graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, vol. 26, no. 10, pp. 1367–1372.
- El Baff, R., Wachsmuth, H., Al Khatib, Kh., Stede, M., Stein, B.** Computational Argumentation Synthesis as a Language Modeling Task. In: Proc. of the 12th International Conference on Natural Language Generation. Tokyo, Japan, 2019, pp. 54–64.
- Hagberg, A. A., Schult, D. A., Swart, P. J.** Exploring network structure, dynamics, and function using NetworkX. In: Proc. of the 7th Python in Science Conference (SciPy2008). Pasadena, CA, USA, 2008, pp. 11–15.
- Fishcheva, I., Kotelnikov, E.** Cross-Lingual Argumentation Mining for Russian Texts. In: Proc. of the 8th International Conference “Analysis of Images, Social Networks and Texts”. Kazan, 2019, pp. 134–144.
- Jiang, C., Coenen, F., Zito, M.** A Survey of Frequent Subgraph Mining Algorithms. *The Knowledge Engineering Review*, 2004, no. 000(1), pp. 1–31.
- Lawrence, J., Reed, C.** Argument Mining: A Survey. *Computational Linguistics*, 2019, vol. 45, no. 4, pp. 765–818.
- Lippi, M., Torrony, P.** Argumentation Mining: State of the Art and Emerging Trends. *ACM Transactions on Internet Technology*, 2016, vol. 16, article 10.
- Madnani, N., Heilman, M., Tetreault, J., Chodorow, M.** Identifying high-level organizational elements in argumentative discourse. In: Proc. of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Montréal, Canada, NAACL, Association for Computational Linguistics, 2012, pp. 20–28.
- Rahwan, I., Reed, C.** The argument interchange format. In: Rahwan, I., Simari, G. (eds.). *Argumentation in artificial intelligence*. Springer, 2009, pp. 383–402.
- Taboada, M., Mann, W.** Rhetorical Structure Theory: Looking back and moving ahead. *Discourse Stud.*, 2006, no. 8, pp. 423–459.
- Ullmann, J. R.** An Algorithm for Subgraph Isomorphism. *Journal of the ACM*, 1976, vol. 23 (1), pp. 31–42.
- Wachsmuth, H., Stede, M., El Baff, R., Al-Khatib, K., Skeppstedt, M., Stein, B.** Argumentation Synthesis following Rhetorical Strategies. In: Proc. of the 27th International Conference on Computational Linguistics. Santa Fe, New Mexico, USA, 2018, pp. 3753–3765.
- Walton, D., Reed, C., Macagno, F.** *Argumentation schemes Fundamentals of critical argumentation*. New York, Cambridge Uni. Press, 2008, 443 p.
- Xinyu Hua, Lu Wang.** Understanding and Detecting Supporting Arguments of Diverse Types. In: Proc. of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada, ACL Association for Computational Linguistics, 2017, vol. 2, pp. 203–208.

References

- Al-Khatib, K., Wachsmuth, H., Hagen, M., Stein, B.** Patterns of Argumentation Strategies across Topics. In: Proc. of the 2017 Conference on Empirical Methods in Natural Language. Copenhagen, Denmark, 2017, pp. 1351–1357.
- Anand, P. et al.** Believe Me – We Can Do This! Annotating Persuasive Acts in Blog Text. In: Computational Models of Natural Argument. San Francisco, CA, USA, 2011, August 7.

- Cordella, L. P., Foggia, P., Sansone, C., Vento, M.** A (Sub)Graph Isomorphism Algorithm for Matching Large Graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, vol. 26, no. 10, pp. 1367–1372.
- El Baff, R., Wachsmuth, H., Al Khatib, Kh., Stede, M., Stein, B.** Computational Argumentation Synthesis as a Language Modeling Task. In: Proc. of the 12th International Conference on Natural Language Generation. Tokyo, Japan, 2019, pp. 54–64.
- Hagberg, A. A., Schult, D. A., Swart, P. J.** Exploring network structure, dynamics, and function using NetworkX. In: Proc. of the 7th Python in Science Conference (SciPy2008). Pasadena, CA, USA, 2008, pp. 11–15.
- Fishcheva, I., Kotelnikov, E.** Cross-Lingual Argumentation Mining for Russian Texts. In: Proc. of the 8th International Conference “Analysis of Images, Social Networks and Texts”. Kazan, 2019, pp. 134–144.
- Inkova, O., Manzotti, E.** Text coherence: mereological logico-semantic relations. Moscow, LRC Publ., 2019, 376 p. (in Russ.)
- Jiang, C., Coenen, F., Zito, M.** A Survey of Frequent Subgraph Mining Algorithms. *The Knowledge Engineering Review*, 2004, no. 000(1), pp. 1–31.
- Lawrence, J., Reed, C.** Argument Mining: A Survey. *Computational Linguistics*, 2019, vol. 45, no. 4, pp. 765–818.
- Lippi, M., Torrony, P.** Argumentation Mining: State of the Art and Emerging Trends. *ACM Transactions on Internet Technology*, 2016, vol. 16, article 10.
- Lukashevich, N. V.** Thesauri in Information Retrieval Problems. Moscow, Moscow State Uni. Press, 2011, 512 p. (in Russ.)
- Madnani, N., Heilman, M., Tetreault, J., Chodorow, M.** Identifying high-level organizational elements in argumentative discourse. In: Proc. of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. Montréal, Canada, NAACL, Association for Computational Linguistics, 2012, pp. 20–28.
- Pimenov, I. S.** On argumentation annotation of scientific and popular science texts. In: Corpus Linguistics – 2021. Proc. of the International Conference. St. Petersburg, Skifia Print, 2021. (in Russ.) ISBN 978-5-98620-557-1
- Rahwan, I., Reed, C.** The argument interchange format. In: Rahwan, I., Simari, G. (eds.). *Argumentation in artificial intelligence*. Springer, 2009, pp. 383–402.
- Salomatina, N. V., Pimenov, I. S.** The N-gram analysis of argumentation annotations of texts. In: Knowledge – Ontology – Theory. Proc. of the VIII International Conference, 2021. (in Russ.) URL: https://drive.google.com/file/d/1vLIHJk_cjYaSK9lwpw1i2HPNr8Qmy1t/view
- Sidorova, E. A., Akhmadeeva, I. R., Zagorulko, Yu. A., Sery, A. S., Shestakov, V. K.** Research platform for the study of argumentation in popular science discourse. *Ontology Design*, 2020, vol. 10, no. 4 (38), pp. 489–502. (in Russ.)
- Taboada, M., Mann, W.** Rhetorical Structure Theory: Looking back and moving ahead. *Discourse Stud.*, 2006, no. 8, pp. 423–459.
- Ullmann, J. R.** An Algorithm for Subgraph Isomorphism. *Journal of the ACM*, 1976, vol. 23 (1), pp. 31–42.
- Wachsmuth, H., Stede, M., El Baff, R., Al-Khatib, K., Skeppstedt, M., Stein, B.** Argumentation Synthesis following Rhetorical Strategies. In: Proc. of the 27th International Conference on Computational Linguistics. Santa Fe, New Mexico, USA, 2018, pp. 3753–3765.
- Walton, D., Reed, C., Macagno, F.** Argumentation schemes Fundamentals of critical argumentation. New York, Cambridge Uni. Press, 2008, 443 p.
- Xinyu Hua, Lu Wang.** Understanding and Detecting Supporting Arguments of Diverse Types. In: Proc. of the 55th Annual Meeting of the Association for Computational Linguistics. Vancouver, Canada, ACL Association for Computational Linguistics, 2017, vol. 2, pp. 203–208.

Информация об авторах

Иван Сергеевич Пименов, аспирант

Наталья Васильевна Саломатина, кандидат физико-математических наук

Мария Кирилловна Тимофеева, доктор филологических наук

Information about the Authors

Ivan S. Pimenov, Postgraduate Student

Natalia V. Salomatina, Candidate of Sciences (Physics and Mathematics)

Mariya K. Timofeeva, Doctor of Sciences (Philology)

*Статья поступила в редакцию 25.11.2021;
одобрена после рецензирования 01.02.2022; принята к публикации 01.02.2022
The article was submitted 25.05.2021;
approved after reviewing 01.02.2022; accepted for publication 01.02.2022*