

ИНЖИНИРИНГ ОНТОЛОГИЙ

УДК 004.82:004.89:519.816

DOI: 10.18287/2223-9537-2020-10-4-489-502

Платформа для исследования аргументации в научно-популярном дискурсе

Е.А. Сидорова, И.Р. Ахмадеева, Ю.А. Загорулько, А.С. Серый, В.К. Шестаков

Институт систем информатики им. А.П. Ершова СО РАН, Новосибирск, Россия

Аннотация

Рассматривается программная система, предназначенная для поддержки исследования аргументации в русскоязычных научно-популярных текстах. Эта система базируется на онтологии, построенной на современных принципах моделирования аргументации. Данная онтология содержит формальные описания типовых схем рассуждений, на основе которых выполняется аннотирование текстов, анализ представленной в них аргументации и оценка её убедительности относительно заданной аудитории. Предлагается методика аргументативной разметки текста, которая обеспечивает выделение в нем утверждений и построение на их основе графа аргументации с использованием знаний о типовых схемах рассуждений. Описывается набор веб-инструментов, обеспечивающих создание тематических корпусов текстов, визуализацию используемой онтологии аргументации, построение графа аргументации, выделение в текстах индикаторов аргументации, а также поиск различных сущностей в корпусах текстов в терминах онтологии. Аналитические инструменты представлены средствами сбора статистической информации о встречаемости типовых элементов аргументации в корпусе текстов, средствами исследования индикаторов аргументации и средствами анализа убедительности аргументации. Представлена оригинальная методика исследования аргументации в научно-популярном дискурсе, базирующаяся на онтологии аргументации и поддерживаемая специализированной веб-платформой.

Ключевые слова: научно-популярный дискурс, онтология аргументации, аргументативная разметка текста, схема аргументации, индикатор аргументации, убедительность аргументации.

Цитирование: Сидорова, Е.А. Платформа для исследования аргументации в научно-популярном дискурсе / Е.А. Сидорова, И.Р. Ахмадеева, Ю.А. Загорулько, А.С. Серый, В.К. Шестаков // Онтология проектирования. – 2020. – Т. 10, №4(38). – С. 489-502. – DOI: 10.18287/2223-9537-2020-10-4-489-502.

Введение

В течение двух последних десятилетий роль Интернета как основной платформы для проведения споров и дебатов, распространения идей и обмена мнениями неуклонно возрастает. Этому способствуют онлайн-площадки и многочисленные Интернет-ресурсы, предоставляющие неограниченный доступ к различным источникам информации, в частности к научным и научно-популярным статьям. Потребителям этой информации хочется разобраться, насколько обоснованы высказываемые мнения и убедительны доводы авторов статей и участников споров и дебатов. Для решения этой задачи можно было бы привлечь методы и технологии обработки естественного языка, которые уже достигли впечатляющих результатов во многих областях. Однако, на данный момент эти технологии не предоставляют надёжных инструментов для выявления семантических связей между отдельными фрагментами текста. Например, они позволяют спрогнозировать, какого мнения будут придержи-

ваться люди через несколько лет, но не могут ответить на вопрос, почему люди придерживаются (или не придерживаются) такого мнения сейчас.

Обоснование мнений или тезисов составляет область теории аргументации, изучающей употребление аргументов в текстах и устной речи (дискурсе) с философской, лингвистической, когнитивной и вычислительной точек зрения. Анализ аргументации, в частности, включает преобразование неструктурированного текста в «цепочки» или графы связанных структурированных аргументов, что позволяет не только оценить отдельные высказывания, но и выявить отношения между ними, направленные на поддержку или нивелирование главного тезиса, предлагаемого автором публикации.

Автоматизация извлечения аргументации из текстов стала одним из приоритетных направлений лишь несколько лет назад [1]. Одним из основных условий развития данной области является создание корпусов текстов с аргументативной разметкой. На данный момент известны несколько аннотированных корпусов, включающих преимущественно англоязычные монологические тексты. Самым известным ресурсом с разметкой аргументации является *AlFdb*, бывший корпус *Araucaria* [2], который содержит новостные статьи, записи парламентских и политических дебатов. Созданы также ресурсы для немецкого языка: корпус Университета Дармштадта, который включает подкорпусы студенческих сочинений [3], новостных текстов и научных статей; Потсдамский корпус, содержащий небольшой набор микротекстов на заданную тему, позднее переведённых на английский язык [4]. Существуют проекты для некоторых других языков: итальянского, греческого, китайского. Для русского языка таких ресурсов, насколько нам известно, пока не создано.

В большинстве работ аргументативная разметка включает в себя сегментацию текста с выделением единиц аргументации, разметкой ролей (посылка, заключение) и отношений (поддержка / атака) без детализации структуры аргументов. Исключением являются корпуса, создаваемые с помощью системы *OVA* (*Online Visualisation of Argument* - наследник *Araucaria*) [5], где аннотация аргументативной структуры соотносится с конкретной схемой аргументации на основе теории Уолтона [6]. Исследование аргументации подразумевает её наглядное представление. Схемы аргументации формализуют определённые шаблонные конструкции, применяемые для убеждения целевой аудитории. При этом исследование статистики и контекстов использования той и иной схемы аргументации представляет больший интерес в рамках корпуса текстов, нежели в отдельно взятом тексте. Это подтверждается как быстрым развитием корпуса *AlFdb* [7], с которым уже объединены некоторые системы разметки аргументации [8-9], так и повышением интереса к задачам автоматического извлечения аргументации, где необходимы размеченные данные.

Для представления аргументации был разработан широкий спектр различных инструментов построения графов аргументации (*argumentation graph*). Они позволяют исследователям локализовать употребление аргументации в тексте и схематично представить аргументативную структуру текста в виде графа. Большинство этих инструментов, таких как *Araucaria* [2], *Rationale* [10], *OVA* [8], *Carneades* [11] и *DiGAT* [12] ориентированы на английский, немецкий и другие западноевропейские языки. Кроме того, они позволяют только представлять структуру аргументации в тексте, оставляя за рамками исследование её качественных характеристик. Качество же аргументации в первую очередь определяется степенью прозрачности и убедительности аргументов, то есть обоснованностью выводов.

В данной статье представлена программная система для поддержки исследования аргументации в русскоязычных научно-популярных текстах. Эта система предоставляет средства для работы с корпусами текстов, для моделирования и извлечения аргументации из текстов, для выполнения аргументативной разметки текстов с использованием онтологического представления схем аргументации, а также для анализа используемых в текстах аргументативных

стратегий и риторических приемов. Её особенностью является наличие средств моделирования и анализа убедительности аргументации. Такие средства необходимы для оценки качества аргументации, представленной в научно-популярных текстах, так как качество таких текстов, по нашему мнению, определяется не столько их литературными достоинствами, сколько качеством представленной в них аргументации.

1 Модель аргументации

Выявление в тексте типовых риторических приемов и рассуждений, убеждающих аудиторию в позиции автора, и их разметка является основой исследования аргументации, анализа её особенностей, проверки корректности и оценки убедительности относительно определённой аудитории.

Любая разметка текста опирается на модель, которая задаёт перечень сущностей, используемых при аннотировании, их типизацию, формат и интерпретацию. В работе в качестве такой модели используется расширенная версия онтологии аргументации (*AIF*-онтологии) [13], базирующейся на формате *AIF* (*Argument Interchange Format*) [14]. Отличительной особенностью данной онтологии [15] является её ориентированность на графовое представление аргументации, наличие разветвлённой системы классов для метаописания **типовых схем рассуждений** и средств для моделирования и анализа убедительности аргументации.

На рисунке 1 приведён пример схемы аргументации *Example_Inference* и её метаописание, построенное по её онтологическому представлению. Данная схема позволяет связывать найденные в тексте утверждения — две посылки и заключение, в единую структуру. Одно и то же утверждение может входить в разные структуры, тем самым осуществляя связывание выявленных в тексте «минимальных» единиц рассуждения - аргументов в единую цепочку, а в общем случае - в граф аргументации.

	Аргумент «Example_Inference»		
	Роль	Тип утверждения	Описание утверждения
CaseProperty_Premise	CaseProperty_Premise	CaseProperty_Statement	In this case, the individual a has property F and also property G
Conclusion	Conclusion	GeneralProperty_Statement	Generally, if x has property F then (usually, probably, typically) x also has property G
TypicalObject_Premise	TypicalObject_Premise	TypicalObject_Statement	a is typical of things that have F and may or may not have G
Найти примеры			Метаописание

Рисунок 1 – Пример схемы аргументации *Example_Inference* и её метаописание

Несмотря на указанный богатый набор свойств онтологии аргументации, её одной недостаточно для проведения полноценного исследования аргументации, используемой в текстах.

Во-первых, необходимо обеспечить не только представление и хранение аргументативной разметки текстов, но и информации об источнике аргументации. Для этого требуется создать хранилище корпусов текстов и их аннотаций, сопряжённых с онтологией.

Во-вторых, необходимы средства для поддержки жанровых и лингвистических исследований особенностей текста.

В-третьих, требуются инструменты для комплексного анализа созданных графов аргументации.

2 Инструменты создания и аннотирования корпусов текстов

Для поддержки исследования аргументации были разработаны инструменты для создания корпусов с аргументативной разметкой и работы с ними. Инструменты интегрированы в единую веб-платформу и предоставляют всю необходимую функциональность для проведения корпусных исследований (рисунок 2).

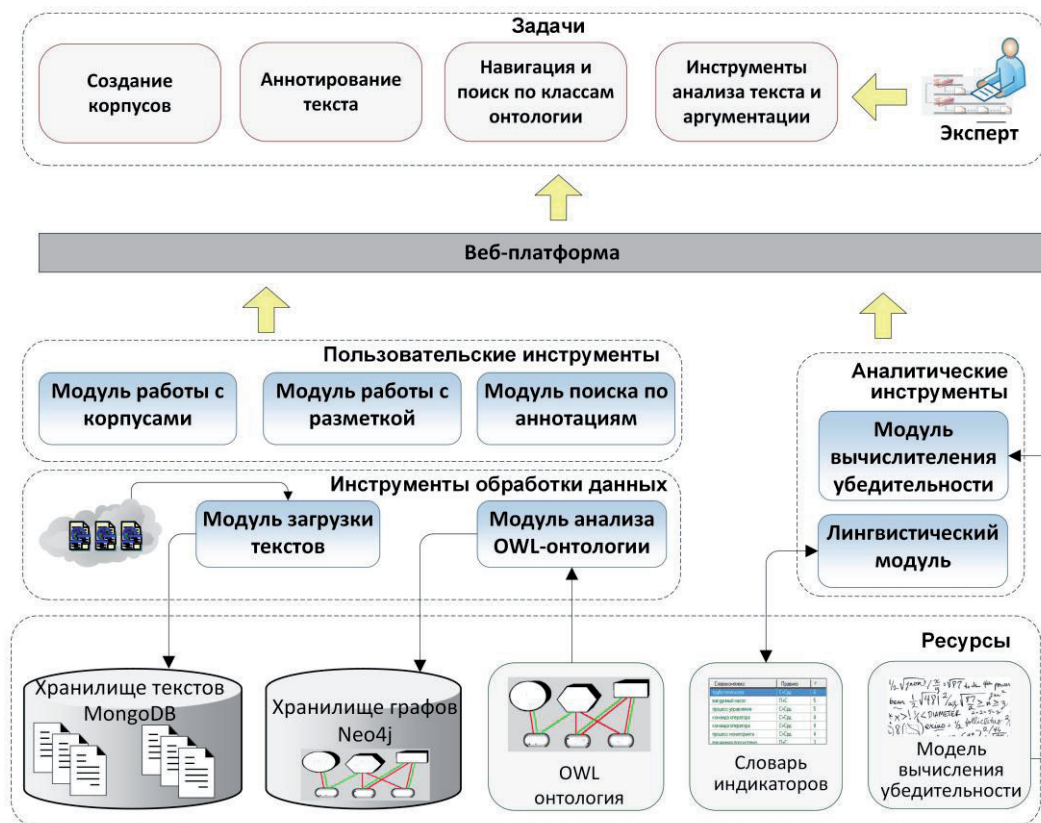


Рисунок 2 – Архитектура платформы для исследования аргументации

Платформа включает ресурсы, инструменты обработки данных, аналитические инструменты и пользовательские инструменты. Платформа поддерживает работу пользователей через веб-интерфейс, обеспечивая решение основных задач, возникающих в процессе построения и анализа аргументации.

Платформа позволяет загружать тексты на ресурс, снабжать их необходимыми метаданными и формировать тематические корпусы текстов (см. рисунок 3), корпусы организуются в иерархию и также снабжаются метаданными.

Основанием для объединения текстов в корпус может быть тематика, источник текста, автор текста, жанр и т.д. В рамках данной работы осуществляется сбор и аннотирование текстов научно-популярного жанра, а корпуса формируются на основе тематических журналов, использованных в качестве источников научно-популярных статей.

Научно-популярный жанр в основном предназначен для предъявления широкой публике результатов научной деятельности и обоснования их состоятельности. Как правило, тексты, написанные в этом жанре, содержат большое количество аргументации, т.е. типовых рассуждений, поддерживающих или опровергающих какой-либо тезис. При этом научно-популярный текст имеет небольшой объем (от 500 до 1000 словоупотреблений), в нём, обычно отсутствуют специальная лексика и явно выраженная структуризация (при помощи глав и подзаголовков).

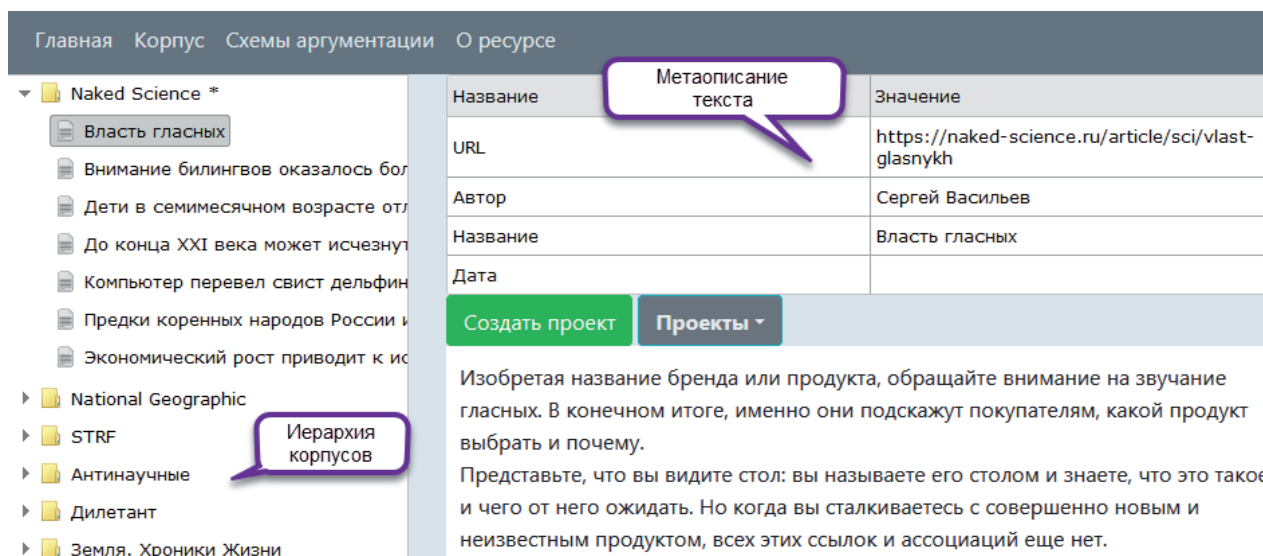


Рисунок 3 – Создание корпуса текстов

Для автоматизации процесса создания корпусов был разработан модуль загрузки текстов, который выполняет следующие задачи:

- загрузка интернет-страниц из открытых источников по ссылкам;
- выделение основного контента и очистка от ненужных элементов разметки;
- анализ контента интернет-страницы и выделение в нём текста публикации, её названия, автора (авторов), даты и тематики (если имеется);
- приведение текста к стандартным формату и кодировке, размещение его в БД системы (тексты снабжаются метаданными; при необходимости формируются новые корпуса).

С помощью автоматизированных средств на текущий момент собрана коллекция научно-популярных текстов объёмом около 3500 статей, сгруппированных по 11 корпусам, которые регулярно пополняются.

При размещении текста в БД системы из него удаляется всё форматирование, реализованное с помощью тэгов, за исключением перевода строки (символа абзаца). В дальнейшем текст редактироваться не может, что связано с особенностями представления аннотаций.

Обеспечение аргументативной разметки (аннотирования) текста – самая трудоёмкая и значимая задача, реализуемая системой. Такая разметка в соответствии с концепцией AIF помимо средств для непосредственной работы с текстом требует инструментов для построения графов и обеспечения методической поддержки исследователя (эксперта), осуществляющего разметку.

На рисунке 4 показаны три различных представления аннотированного текста: а) размеченный текст, б) список аргументов, с) аргументативный граф.

Представление аргументации в размеченном тексте (а) демонстрирует множество размеченных утверждений в связном тексте и дополнительно позволяет в пошаговом режиме просмотреть выделенные аргументы. В этом режиме аргумент показывается выделением всех утверждений, входящих в его состав (посылки и заключение).

Список аргументов (б) позволяет увидеть перечень всех размеченных в тексте аргументов и, в отличие от предыдущего представления, состав каждого аргумента, его схему и имплицитно представленные утверждения (данные утверждения отсутствуют в тексте, но восстанавливаются читателем из внешнего контекста). Совместно с текстовым данное представление даёт полный обзор размеченной аргументации и покрываемых ею фрагментов текста.

Для создания и редактирования аргументации преимущественно используется графовое представление (с). В соответствии с данным представлением множество взаимосвязанных аргументов, размеченных в тексте, называют сетью аргументов или графом аргументации. Онтология *AIF* уже содержит понятия для графового представления аргументов [2] — это класс *Node* и его потомки. На рисунке 4 утверждения (прямоугольные вершины) соответствуют экземплярам информационных вершин (*I-Node*), а аргументы (вершины со скруглёнными краями) [2] — экземплярам *S-Node*.

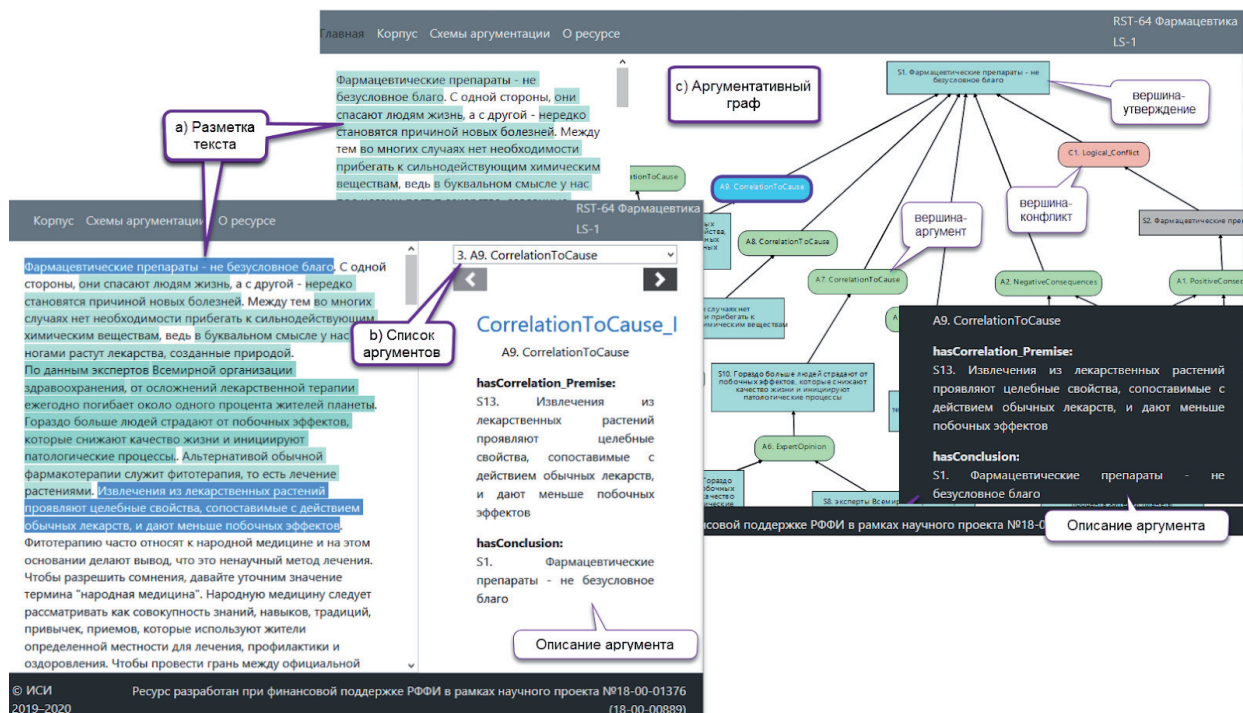


Рисунок 4 – Аргументативная разметка текста

Сценарий работы пользователя включает основной этап, на котором осуществляется построение графа аргументации, и этап, на котором выполняется анализ аргументации.

Основной этап включает следующие шаги:

- сегментация текста с выделением в нём аргументативных дискурсивных единиц (АДЕ) и формирование на их основе утверждений (при формулировании утверждений должны сниматься неоднозначность, разрешаться анафора, восстанавливаться эллипсис и т.п.);
- определение роли для каждой АДЕ (заключение или посылка) и построение направленных отношений [2] аргументов, связывающих аргументативные единицы;
- детализация структуры аргументов на основе соответствующих схем аргументации;
- выявление имплицитных и эквивалентных утверждений и обеспечение максимальной связности графа.

Для перехода к следующему этапу — анализу и оценке аргументации — полученный граф полезно верифицировать, поскольку даже небольшое изменение в его структуре может повлечь существенные расхождения в итоговой оценке. Автоматическая верификация графа может включать такие процедуры как поиск циклов, анализ связности, учёт текстовых индикаторов аргументации, сравнительный анализ с другими разметками. На данный момент системой поддерживаются первые три возможности. Верификация, осуществляемая пользователями, может заключаться в согласовании мнений нескольких экспертов, а также в проведении сравнительного исследования аргументативных связей (например, анализ корреляции с риторической разметкой [16-17]).

3 Инструменты поддержки исследования аргументации

Поддержка исследования аргументации на предлагаемой платформе обеспечивается следующим набором специализированных средств: поисковым сервисом, позволяющим ознакомиться с уже представленными в системе результатами работы экспертов; лингвистическим модулем, выполняющим предварительную обработку текста, в ходе которой фиксируются специальные языковые конструкции — индикаторы; вычислительным модулем, поддерживающим оценку убедительности представленных в тексте аргументов.

3.1 Поисковый сервис

Поисковый сервис позволяет ориентироваться в представленных корпусах и обеспечивает в них поиск в терминах онтологии аргументации. С его помощью исследователь может найти в аннотированных текстах примеры использования заданных схем аргументации или примеры утверждений, участвующих в аргументах в определённой роли.

На рисунке 5 представлены результаты поиска всех утверждений, участвующих в аргументах, представляющих причинно-следственные связи в текстах (*Causal_Statement*). Результаты поиска отображаются в виде таблицы, в которой для каждого найденного утверждения даются ссылки на текстовый источник (Первоисточник) и аннотацию (Проект), в которых оно встречается, название аргумента (Аргумент), в котором оно участвует, и само утверждение с ближайшим окружением (Контекст).

Таким образом, поисковый сервис позволяет пользователю ознакомиться с примерами употребления тех или иных схем аргументации или утверждений и статистикой их встречаемости в текстах корпуса.

► BadConsequence_Statement

► BadOtherConsequence_Statement

BiasConclusion_Statement

BiasConditional_Statement

BiasedClassificationDefinition_Statement

BiasedPerson_Statement

CaseOutcome_Statement

CaseProperty_Statement

Causal_Statement

ChainRequired_Statement

ChainedCases_Statement

ChainedImplications_Statement

ChainedIndistinguishable_Statement

► CharacterAttack_Statement

CharacterRelevance_Statement

ClassificationProperty_Statement

CommitmentEvidence_Statement

CommitmentException_Statement

CommitmentInconsistency_Statement

CommitmentLinkage_Statement

Commitment_Statement

CommonKnowledgeBackup_Statement

ConflictingGoals_Statement

ConsequenceBackupEvidence_Statement

ContainsProperty_Statement

ContinueAction_Statement

Correlation_Statement

CredibilityOfSource_Statement

DifferencesUndermineSimilarity_Statement

Disjunction_Statement

Утверждение «Causal_Statement»

Описание утверждения

A causes B

Роли утверждения в аргументации

Схема аргументации

Роль утверждения в схеме

CauseToEffect_Inference

Causal_Premise

CorrelationToCause_Inference

Conclusion

Найти примеры

Первоисточник	Проект	Аргумент	Контекст
*Самые необычные звуки в языках мира	ИльинаД-Л	от причины к с...	одну из песен ее народа — свадебную «Qongqothwane» — в Америке и Европе именуют «The Click Song», поскольку не могут произнести ее оригинальное название
*Самые необычные звуки в языках	ИльинаД-Л	от причины к с...	в Америке и Европе именуют «The Click Song»
*RST-Ling-43	IP-1	CorrelationToCa...	мотивированность неофициальных именований лица может рассматриваться под разным углом зрения и с учетом разных факторов
*RST-Ling-43	IP-1	CauseToEffect_I...	Экстралингвистическая мотивированность

Previous

Page 1 of 43

10 rows

Next

Рисунок 5 – Поиск утверждений по заданной роли в структуре аргументов

3.2 Анализ индикаторов аргументации

Для привлечения внимания пользователя к аргументам, представленным в тексте явно, и оказания помощи в выделении в тексте границ АДЕ и в выборе схемы аргументации выполняется предварительная лингвистическая обработка текстов, которая позволяет обнаружить в тексте специфические подсказки в виде различного рода словесных клише. Эти клише являются индикаторами, указывающими на факт присутствия аргумента в тексте.

Автоматический поиск индикаторов осуществляется с помощью шаблонных конструкций, описывающих классы языковых выражений, и учитывает возможные грамматические формы, их сочетаемость в многословных цепочках и пунктуацию [18]. Шаблоны формулируются экспертами на основе анализа текстов, содержащих аргументацию, после чего расширяются путём учёта вариантов методами построения образцов с переменными и итерационного поиска [19].

Можно выделить следующие типы соотношений между индикатором и сигнализируемыми им аспектами аргументации:

- 1) сила или убедительность аргумента;
- 2) степень уверенности автора в утверждении;
- 3) отношение вывода между двумя утверждениями (наличие аргументации);
- 4) тип аргументативного отношения (поддержка vs. конфликт);
- 5) роль утверждения в отношении вывода (посылка vs. заключение);
- 6) семантико-онтологическое отношение, на котором основана применяемая в данном случае типовая схема рассуждения;
- 7) структура аргументации (множественная vs. последовательная аргументация).

Аспекты 2–3 и 5–6 соотносят индикатор не столько с аргументом, сколько с конкретным утверждением (или несколькими утверждениями) в структуре аргумента. Аспекты 1, 3 и 7 сигнализируют только о наличии аргументации, что в случае ручной разметки фокусирует внимание эксперта на соответствующих фрагментах текста, а при автоматической — требует наличия других показателей для более точного выявления класса аргумента. Аспекты 1–2 влияют на оценку убедительности аргументации.

Индикаторы, автоматически найденные в тексте, выделяются полужирным шрифтом (см. рисунок 6), привлекая внимание пользователя к фрагментам текста, потенциально содержащим аргументацию. При наведении курсора на индикатор отображается подсказка, где указано имя соответствующего шаблона. В некоторых случаях имя шаблона косвенно указывает на присутствие в тексте аргументов, соответствующих определённой схеме.

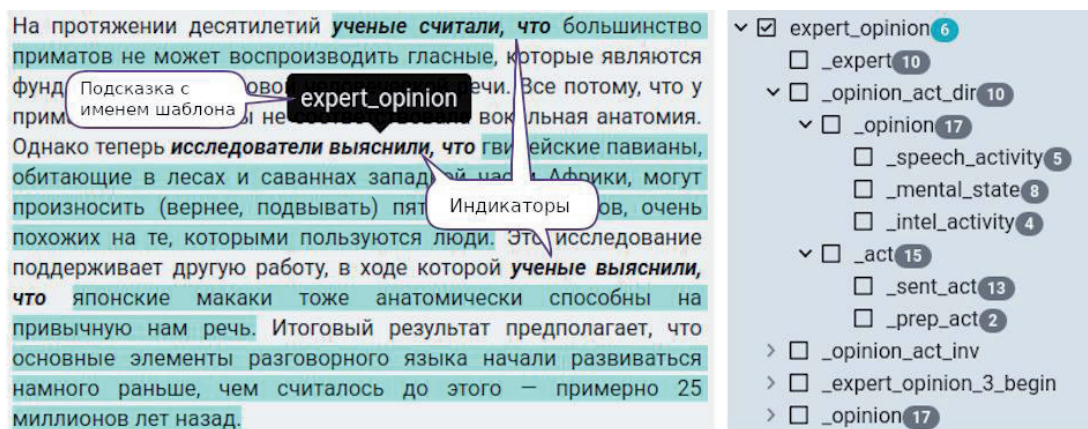


Рисунок 6 – Поиск индикаторов аргументации

В левой части рисунка 6 представлен фрагмент размеченного текста, в котором полужирным шрифтом выделены индикаторы, сигнализирующие о наличии аргументации «От эксперта», а в правой части - фрагмент иерархии шаблонов, в которой пользователь может выбрать те шаблоны, которые он хочет видеть в тексте.

Таким образом, аннотация состоит из множества фрагментов текста, каждому из которых поставлено в соответствие утверждение или индикатор. Аргументы определяются как местные отношения над размеченными текстовыми фрагментами. Роль индикаторов заключается в фиксации свойств и границ аргументов и их структурных элементов.

3.3 Оценка убедительности аргументации

Чтобы оценить, насколько рассматриваемый текст успешен в донесении своих тезисов до читателя, требуется установить, какие в нём используются риторические приёмы и схемы рассуждений (т.е. аргументативные структуры), сравнить их с приемами и схемами, использованными в качественных, с точки зрения убедительности, научно-популярных текстах, а также определить убедительность аргументов исследуемого текста. Для этого необходимы методика оценивания убедительности аргументации и достаточно репрезентативный корпус текстов, содержащий максимально разнообразную аргументацию.

Существуют различные модели численного представления и вычисления убедительности аргументов. Как правило, убедительность выводов, отстаиваемых при помощи аргументов, зависит от убедительности посылок и самих аргументов. Популярным подходом при измерении и вычислении убедительности является рассмотрение степени убедительности как вероятности утверждения оказаться истинным. Это значит, что степень убедительности (вес) выводов, посылок и аргументов представляется числом из интервала от 0 до 1.

При создании веб-платформы применялся специально разработанный алгоритм, основанный на операциях нечёткой логики, в которой определена алгебра истинностных значений. Если рассматривать аргумент как имеющее определённую силу доказательство некоторого утверждения, то чем более обосновано утверждение, тем большей полагается степень его истинности.

Упомянутый алгоритм по заданным экспертом весам посылок и аргументов вычисляет веса выводов, в том числе проводит вычисления по цепочке, когда вывод одного аргумента одновременно является посылкой для другого, как это показано на рисунке 7, или когда в графе помимо выводов содержатся и конфликтующие с ними тезисы. При этом очевидно, что цепочка рассуждений в графе не должна заикливаться, так как в противном случае убедительность вывода будет зависеть в том числе от себя самой, и вычисление не будет корректным. Каждой вершине графа соответствуют три веса: начальный вес; вес, вычисленный без учёта конфликтных ситуаций (только выводы); вес, вычисленный с учётом атак со стороны конфликтующих утверждений (если таковые имеются).

Следует заметить, что эксперты могут заранее задать каждой схеме аргументации так называемый априорный вес, учитывающий образ мышления и систему ценностей одной из трёх аудиторий: «широкая публика», «научная аудитория», «школьники». Априорный вес схемы аргументации по умолчанию используется в качестве начального веса всех построенных на её основе аргументов графа, если исследователь сам их не задаст.

Отсюда следует, что графы аргументации, построенные для разных аудиторий, могут не совпадать. В целом, каждый текст в корпусе может иметь произвольное количество независимых аннотаций.

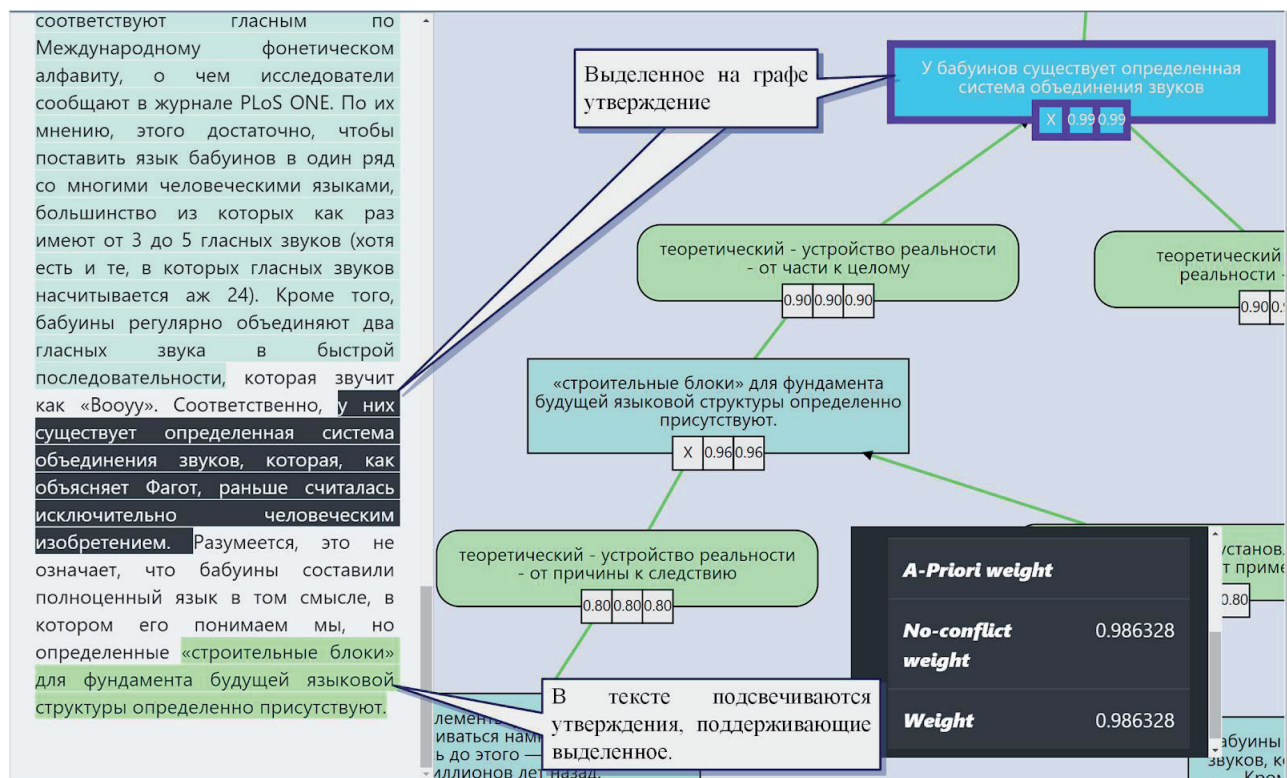


Рисунок 7 – Пример убедительности аргументации

Заключение

Предложенный в работе подход к исследованию аргументации в научно-популярных текстах включает два этапа:

- этап аннотирования текстов и построения графов аргументации на основе знаний о типовых схемах рассуждений и примеров их употребления в корпусе,
- анализ статистики употребления типовых схем рассуждений в исследуемом корпусе и оценку убедительности аргументации относительно различных аудиторий.

Разработан набор инструментов для поддержки исследования аргументации, интегрированных в единую веб-платформу, которая позволяет создавать корпуса текстов, извлекать индикаторы аргументации, снабжать тексты аргументативной разметкой, осуществлять поиск в терминах онтологии и анализировать качество аргументации на основе реализованных в системе моделей вычисления убедительности.

С помощью предложенной методики и разработанных инструментов было собрано 112 корпусов, включающих 2 360 текстов научно-популярного жанра, 109 из которых было размечено и для них построено 146 графов аргументации. Созданные корпуса планируется сделать открытыми для других исследователей.

Одним из важнейших направлений развития предложенного инструментария является включение в него средств для расширения набора типовых схем аргументации, что будет способствовать развитию теории аргументации.

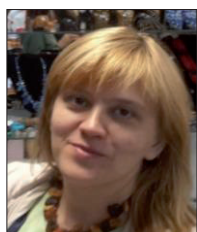
Благодарности

Статья подготовлена по итогам исследования, проведённого в рамках проекта Российского фонда фундаментальных исследований № 18-00-01376 (18-00-00889).

Список источников

- [1] **Lawrence, J.** Argument mining: A survey / J. Lawrence, C. Reed // *Int. J. of Computational Linguistics*. – 2019. – Vol. 45(4). – P.765-818.
- [2] **Reed, C.** Araucaria: Software for argument analysis, diagramming and representation / C. Reed, G. Rowe // *Int. J. on Artificial Intelligence Tools*. – 2004. – Vol. 13(4). – P.961-979.
- [3] **Stab, C.** Identifying Argumentative Discourse Structures in Persuasive Essay / C. Stab, I. Gurevych // *Empirical Methods in Natural Language Processing (EMNLP): Proc. of the Int. Conf. (Doha, Qatar)*. – 2014. – P.46–56.
- [4] **Peldszus, A.** An annotated corpus of argumentative microtexts / A. Peldszus, M. Stede // *Argumentation and Reasoned Action: Proc. of the 1st European Conference on Argumentation*. – London: College Publications, 2016. – Vol. 2. – P.801–816.
- [5] **Bex, F.** ArguBlogging: An application for the argument web / F. Bex, M. Snaith, J. Lawrence, C. Reed // *Int. J. of Web Semantics: Science, Services and Agents on the World Wide Web*. – 2014. – Vol. 25. – P.9-15.
- [6] **Walton, D.** Argumentation schemes / D. Walton, C. Reed, F. Macagno. – Cambridge: Cambridge University Press, 2008. – 443 p.
- [7] Корпус *AIFdb*. – <http://corpora.aifdb.org/>.
- [8] **Bex, F.** Implementing the argument web / F. Bex, J. Lawrence, M. Snaith, C. Reed // *Int. J. of Communications of the ACM*. – 2013. – Vol. 56(10). – P.66–73.
- [9] **Bex, F.** Dialogue templates for automatic argument processing / F. Bex, C. Reed // *Computational Models of Argument: Proc. of the 4th Int. Conf. COMMA 2012 (Vienna)*. – IOS Press, 2012. – P.366-377.
- [10] **Van Gelder, T.** The rationale for rationale / Tim van Gelder // *Int. J. of Law, Probability and Risk*. – 2007. – Vol. 6(1–4). – P.23-42.
- [11] **Gordon, T.F.** The Carneades model of argument and burden of proof / T.F. Gordon, H. Prakken, D. Walton // *Int. J. of Artificial Intelligence*. – 2007. – Vol. 171(10). – P.875–896.
- [12] **Kirschner, C.** Linking the thoughts: Analysis of argumentation structures in scientific publications / C. Kirschner, J. Eckle-Kohler, I. Gurevych // *Argumentation Mining: Proc. of the 2nd Workshop*. – Denver, CO. - 2015. - P.1-11.
- [13] **Rahwan, I.** Representing and classifying arguments on the semantic web / I. Rahwan, B. Banihashemi, C. Reed, D. Walton, S. Abdallah // *The Knowledge Engineering Review*. – 2011. – Vol. 26(4). – P.487-511.
- [14] **Chesñevar, C.I.** Towards an argument interchange format / C.I. Chesñevar, J. McGinnis, S. Modgil, I. Rahwan, C. Reed, G. Simari, M. South, G. Vreeswijk, S. Willmott // *The knowledge engineering review*. – 2006. – Vol. 21(4). – P.293-316.
- [15] **Загорюлько, Ю.А.** Моделирование аргументации в научно-популярном дискурсе с использованием онтологий / Ю.А. Загорюлько, Н.О. Гаранина, О.И. Боровикова, О.А. Доманов // *Онтология проектирования*. – 2019. – Т. 9, № 4(34). – С.496-509. – DOI: 10.18287/2223-9537-2019-9-4-496-509.
- [16] **Musi, E.** A Multi-layer Annotated Corpus of Argumentative Text: From Argument Schemes to Discourse Relations / E. Musi, T. Alhindi, M. Stede, L. Kriese, S. Muresan, A. Rocci A. // *Language Resources and Evaluation (LREC'2018): Proc. of the 11th Int. Conf. (Miyazaki, Japan)*. – 2018. – P.1629-1636.
- [17] **Kononenko, I.S.** Comparative analysis of rhetorical and argumentative structures in the study of popular science discourse / I.S. Kononenko, E.A. Sidorova, I.R. Akhmadeeva // *Computational Linguistics and Intellectual Technologies: Proc. of the Int. Conf. "Dialogue"*. – 2020. – Vol. 19 (26). – P.432-444.
- [18] **Сидорова, Е.А.** Подход к моделированию процесса извлечения информации из текста на основе онтологии / Е.А. Сидорова // *Онтология проектирования*. – 2018. – Т.8, №1(27). – С.134-151. – DOI: 10.18287/2223-9537-2018-8-1-134-151.
- [19] **Ахмадеева, И.Р.** Подход к построению шаблонов индикаторов для извлечения аргументов из научно-популярных текстов / И.Р. Ахмадеева, И.С. Кононенко, Н.В. Саломатина, Е.А. Сидорова // *Знания - Онтологии - Теории: Труды международной конф. (ЗОИТ-2019)*. – Институт математики им. С.Л. Соболева СО РАН, Новосибирский государственный университет, 2019. – С.24-32.

Сведения об авторах



Сидорова Елена Анатольевна, 1977 г. рождения. Окончила Новосибирский государственный университет в 2000 г., к.ф.-м.н. (2006). Старший научный сотрудник лаборатории искусственного интеллекта Института систем информатики им. А.П. Ершова СО РАН, старший преподаватель кафедры программирования Новосибирского государственного университета член Российской и Европейской ассоциаций искусственного интеллекта. В списке научных трудов более 130 работ в области компьютерной лингвистики, мультиагентных систем, представления знаний и онтологического инжиниринга. Author ID (РИНЦ): 146000; ORCID: 0000-0001-8731-3058; Author ID (Scopus): 41961707000; Researcher ID (WoS): K-2432-2018.

lsidorova@iis.nsk.su.

Ахмадеева Ирина Равильевна, 1991 г. рождения. Окончила Новосибирский государственный университет в 2015 г. Младший научный сотрудник Института систем информатики им. А.П. Ершова СО РАН, ассистент кафедры программирования Новосибирского государственного университета. В списке научных трудов более 20 работ в области искусственного интеллекта, разработки интеллектуальных систем и компьютерной лингвистики. Author ID (РИНЦ): 874172; ORCID: 0000-0002-7371-1087; Author ID (Scopus): 57188681471; Researcher ID (WoS): K-3145-2018. *i.r.akhmadeeva@iis.nsk.su.*



Загорюлько Юрий Алексеевич, 1957 г. рождения. Окончил Новочеркасский политехнический институт им. С. Орджоникидзе (1979), к.т.н. (1989). Заведующий лабораторией Института систем информатики им. А.П. Ершова СО РАН, доцент кафедры программирования и кафедры систем информатики Новосибирского государственного университета. Член Российской и Европейской ассоциаций искусственного интеллекта. В списке научных трудов более 270 публикаций в области искусственного интеллекта, разработки интеллектуальных систем, инженерии знаний, онтологического моделирования и компьютерной лингвистики. Author ID (РИНЦ): 4015; ORCID: 0000-0002-7111-6524; Author ID (Scopus): 23394231500; Researcher ID (WoS): R-1826-2016. *zagor@iis.nsk.su.*

Серый Алексей Сергеевич, 1987 г. рождения. Окончил Новосибирский государственный университет в 2010 г. Младший научный сотрудник лаборатории искусственного интеллекта Института систем информатики им. А.П. Ершова СО РАН. В списке научных трудов более 20 работ в области представления знаний и компьютерной лингвистики. Author ID (РИНЦ): 714554; ORCID: 0000-0001-8275-4700; Author ID (Scopus): 56403204900; Researcher ID (WoS): K-1557-2018. *alexey.seryj@iis.nsk.su.*



Шестаков Владимир Константинович, 1986 г. рождения. Окончил Новосибирский государственный университет в 2009 г. Младший научный сотрудник лаборатории искусственного интеллекта Института систем информатики им. А.П. Ершова СО РАН. В списке научных трудов более 20 работ в области разработки информационных систем, в том числе с использованием вики-технологий и онтологий. Author ID (РИНЦ): 711994; Author ID (Scopus): 56439120800; Researcher ID (WoS): J-8288-2018. *shestakov@iis.nsk.su.*



Поступила в редакцию 26.10.2020, после рецензирования 11.12.2020. Принята к публикации 17.12.2020.

Research platform for the study of argumentation in popular science discourse

E.A. Sidorova, I.R. Akhmadeeva, Yu.A. Zagorulko, A.S. Sery, V.K. Shestakov

A.P. Ershov Institute of Informatics Systems of Siberian Branch of RAS, Novosibirsk, Russia

Abstract

The paper discusses a software system designed to support the study of argumentation in Russian-language popular science texts. This system is based on an ontology built on modern principles of argumentation modeling. In particular, this ontology contains formal descriptions of typical reasoning schemes that are used for annotating texts, analyzing the

arguments presented in them, and assessment of its persuasiveness relative to a given audience. A method of argumentative marking of a text is proposed, which provides the allocation of statements and the construction on their basis of an argumentation graph using knowledge about typical reasoning schemes. The paper also describes a set of web tools that provide the creation of thematic corpora, visualization of the argumentation ontology used, the construction of the argumentation graph, the selection of argumentation indicators in the texts, as well as the search for various entities in the text corpora in ontology terms. Analytical tools are presented by means of collecting statistical information on the occurrence of typical elements of argumentation in the body of texts, by means of researching indicators of argumentation and by means of analyzing the persuasiveness of argumentation. The novelty of the work consists in the development of an original methodology for studying argumentation in popular science discourse, based on the ontology of argumentation and supported by a specialized web platform.

Key words: popular science discourse, ontology of argumentation, argumentative markup of text, argumentation scheme, indicator of argumentation, persuasiveness of argumentation.

Citation: Sidorova EA, Akhmadeeva IR, Zagorulko YuA, Sery AS, Shestakov VK. Research platform for the study of argumentation in popular science discourse [In Russian]. *Ontology of designing*. 2020; 10(4): 489-502. DOI: 10.18287/2223-9537-2020-10-4-489-502.

Acknowledgment: The paper was prepared based on the results of a study conducted as part of the project of the Russian Foundation for Basic Research No. 18-00-01376 (18-00-00889).

List of figures

- Figure 1 - Example of the argument scheme *Example_Inference* and its meta description
- Figure 2 - The architecture of the platform for argumentation study
- Figure 3 - Creating a text corpus
- Figure 4 - Argumentative markup of text
- Figure 5 - Search for statements by a given role in the structure of arguments
- Figure 6 - Search for indicators of argumentation
- Figure 7 - An example of the persuasiveness of the argumentation

References

- [1] Lawrence J, Reed C. Argument mining: A survey. *Int. J. of Computational Linguistics* 2019; 45(4): 765-818.
- [2] Reed C, Rowe G. Araucaria: Software for argument analysis, diagramming and representation. *Int. J. on Artificial Intelligence Tools* 2004; 13(4): 961-979.
- [3] Stab C, Gurevych I. Identifying Argumentative Discourse Structures in Persuasive Essay. *Empirical Methods in Natural Language Processing (EMNLP): Proc. of the Int. Conf. (Doha, Qatar)*; 2014: 46–56.
- [4] Peldszus A, Stede M. An annotated corpus of argumentative microtexts. *Argumentation and Reasoned Action: Proc. of the 1st European Conference on Argumentation*. London: College Publications; 2016; 2: 801–816.
- [5] Bex F, Snaith M, Lawrence J, Reed C. ArguBlogging: An application for the argument web. *Int. J. of Web Semantics: Science, Services and Agents on the World Wide Web*; 2014; 25: 9-15.
- [6] Walton D, Reed C, Macagno F. *Argumentation schemes*. Cambridge: Cambridge University Press; 2008.
- [7] Corpus AIFdb. Source: <http://corpora.aifdb.org/>.
- [8] Bex F, Lawrence J, Snaith M, Reed C. Implementing the argument web. *Int. J. of Communications of the ACM* 2013; 56(10): 66–73.
- [9] Bex F, Reed C. Dialogue templates for automatic argument processing. *Computational Models of Argument: Proc. of the 4th Int. Conf. COMMA 2012 (Vienna)*. IOS Press; 2012: 366-377.
- [10] Van Gelder T. The rationale for rationale. *Int. J. of Law, Probability and Risk* 2007; 6(1–4): 23-42.
- [11] Gordon TF, Prakken H, Walton D. The Carneades model of argument and burden of proof. *Int. J. of Artificial Intelligence*; 2007; 17(10): 875–896.
- [12] Kirschner C, Eckle-Kohler J, Gurevych I. Linking the thoughts: Analysis of argumentation structures in scientific publications. *Argumentation Mining: Proc. of the 2nd Workshop*. Denver, CO; 2015: 1-11.
- [13] Rahwan I, Banihashemi B, Reed C, Walton D, Abdallah S. Representing and classifying arguments on the semantic web. *The Knowledge Engineering Review* 2011; 26(4): 487-511.

- [14] *Chesñevar CI, McGinnis J, Modgil S, Rahwan I, Reed C, Simari G, South M, Vreeswijk G, Willmott S.* Towards an argument interchange format. *The knowledge engineering review* 2006; 21(4): 293-316.
 - [15] *Zagorulko YuA., Garanina NO., Borovikova OI., Domanov OA.* Argumentation modeling in popular science discourse using ontologies [In Russian]. *Ontology of designing.* 2019; 9(4): 496-509. DOI: 10.18287/2223-9537-2019-9-4-496-509.
 - [16] *Musi E, Alhindi T, Stede M, Kriese L, Muresan S, Rocci A.* A Multi-layer Annotated Corpus of Argumentative Text: From Argument Schemes to Discourse Relations. *Language Resources and Evaluation (LREC'2018): Proc. of the 11th Int. Conf. (Miyazaki, Japan) 2018:* 1629-1636.
 - [17] *Kononenko IS, Sidorova EA, Akhmadeeva IR.* Comparative analysis of rhetorical and argumentative structures in the study of popular science discourse. *Computational Linguistics and Intellectual Technologies: Proc. of the Int. Conf. "Dialogue" 2020;* 19(26): 432-444.
 - [18] *Sidorova EA.* An approach to modeling the process of information extracting based on ontology [In Russian]. *Ontology of designing.* 2018; 1(27):134-151. DOI: 10.18287/2223-9537-2018-8-1-134-151.
 - [19] *Akhmadeeva IR, Kononenko IS, Salomatina NV, Sidorova EA.* An approach to building indicator templates for extracting arguments from popular science texts [In Russian]. *Knowledge-Ontology-Theory (KONT-19): Proc. of Russian Conf. Novosibirsk. Mathematics Institute of SB RAS; Novosibirsk State University;* 2019: 24-32.
-

About the authors

Elena Anatolievna Sidorova (b. 1977) graduated from the Novosibirsk State University in 2000, PhD (2006). She is a Senior Researcher of the Laboratory of Artificial Intelligence at the A.P. Ershov Institute of Informatics Systems of Siberian Branch of RAS, Senior Lecturer at Novosibirsk State University. She is a member of Russian and European Associations for Artificial Intelligence. Dr. Sidorova has about 130 peer-reviewed publications in the field of NLP Systems, Multi-agent Systems, Knowledge Representation, and Ontology Engineering. Author ID (RSCI): 146000; ORCID: 0000-0001-8731-3058; Author ID (Scopus): 41961707000; Researcher ID (WoS): K-2432-2018. *lsidorova@iis.nsk.su*.

Irina Ravilevna Akhmadeeva (b.1991) graduated from the Novosibirsk State University in 2015. She is a Junior Researcher at A.P. Ershov Institute of Informatics Systems of Siberian Branch of RAS, Assistant Lecturer at Novosibirsk State University. She is the author of more than 20 publications in the fields of AI, Intelligent System Development and NLP. ORCID: 0000-0002-7371-1087; Author ID (RSCI): 874172; Author ID (Scopus): 57188681471; Researcher ID (WoS): K-3145-2018. *i.r.akhmadeeva@iis.nsk.su*.

Yury Alekseevich Zagorulko (b.1957) graduated from the Novocherkassk Polytechnic Institute in 1979, PhD (1989). He is the Head of Laboratory at A.P. Ershov Institute of Informatics Systems of Siberian Branch of RAS, Associate Professor at Novosibirsk State University. He is a member of Russian and European Associations for Artificial Intelligence. He is the author of more than 270 publications in the fields of AI, Knowledge and Ontology Engineering, Intelligent System Development and Computational Linguistics. Author ID (RSCI): 4015; ORCID: 0000-0002-7111-6524; Author ID (Scopus): 23394231500; Researcher ID (WoS): R-1826-2016. *zagor@iis.nsk.su*

Alexey Sergeevich Sery (b.1987) holds a master's degree in mathematics from Novosibirsk State University (2010) and the position of Junior Researcher at the A.P. Ershov Institute of Informatics Systems of Siberian Branch of RAS. He is the author of more than 20 papers in the fields of NLP systems and Knowledge Representation. Author ID (RSCI): 714554; ORCID: 0000-0001-8275-4700; Author ID (Scopus): 56403204900; Researcher ID (WoS): K-1557-2018. *alexey.seryj@iis.nsk.su*

Vladimir Konstantinovich Shestakov (b. 1986) graduated from the Novosibirsk State University in 2009. He is a Junior Researcher of the Artificial Intelligence Laboratory at the A.P. Ershov Institute of Informatics Systems of Siberian Branch of RAS. He is the author of more than 20 publications in the fields of developing information systems, including using wiki technologies and ontologies. Author ID (RSCI): 711994; Author ID (Scopus): 56439120800; Researcher ID (WoS): J-8288-2018. *shestakov@iis.nsk.su*

Received October 26, 2020. Revised December 11, 2020. Accepted December 12, 2020.
