

МЕТОДИКА ПРЕПОДАВАНИЯ | METHOD OF TEACHING

Научная статья | Original paper

УДК 378.147:004.946

Метод формализации учебных текстов

М.В. Воронов ✉, В.Т. Строев

Московский государственный психолого-педагогический университет

Москва, Российская Федерация

✉ mironovov@yandex.ru

Резюме

Контекст и актуальность. Формализация учебных текстов необходима для создания средств содержательного анализа, контроля логики изложения и компьютерной поддержки обучения, однако произвольные вербальные тексты плохо поддаются полной автоматической формализации без учета жанровых и функциональных особенностей. **Цель.** Разработать метод преобразования учебного текста в формальную модель, максимально сохраняющую содержание и последовательность изложения учебного материала. **Гипотеза.** Если учебный текст рассматривать как регулятивный текст, состоящий из последовательности денотативных ситуаций, то каждое предложение может быть представлено через предикат, актанты, сирконстанты и результаты описанного действия, а связи между предложениями — в виде ребер ориентированного графа. **Методы и материалы.** Использованы положения семантического синтаксиса, понятие пропозиции, фреймовые структуры и представление знаний в виде графов; материалом является вербальный учебный текст как объект содержательной формализации. **Результаты.** Предложен алгоритм, в котором предложения последовательно преобразуются в группы фреймов-экземпляров, образующие подграфы, а затем объединяются в единый граф на основе совпадения содержательных компонентов и логики их введения. **Выводы.** Метод создает основу для построения онтологий понятий предметной области, выявления нарушений связности и разработки средств компьютерной поддержки контроля результатов обучения и самоподготовки обучающихся.

Ключевые слова: учебный текст, модель, формализация, регулятивный текст, семантическая модель текста, граф знаний



Для цитирования: Воронов, М.В., Строев, В.Т. (2026). Метод формализации учебных текстов. *Моделирование и анализ данных*, 16(3), 259—273. <https://doi.org/10.17759/mda.2026160313>

A method for formalizing educational texts

M.V. Voronov ✉, **V.T. Stroeve**

Moscow State University of Psychology and Education, Moscow, Russian Federation

✉ miroonovov@yandex.ru

Abstract

Context and relevance. The formalization of educational texts is necessary for content analysis, checking the logic of exposition, and developing computer-based support for learning; however, arbitrary verbal texts are difficult to formalize automatically without taking their genre and functional features into account. **Objective.** The study aims to develop a method for transforming an educational text into a formal model that preserves the content and sequence of presentation of the educational material as fully as possible. **Hypothesis.** If an educational text is considered as a regulatory text consisting of a sequence of denotative situations, each sentence can be represented through a predicate, actants, circumstances, and the results of the described action, while relations between sentences can be represented as edges of a directed graph. **Methods and materials.** The study draws on semantic syntax, the concept of proposition, frame structures, and graph-based knowledge representation; the material is a verbal educational text considered as an object of content formalization. **Results.** An algorithm is proposed in which sentences are sequentially transformed into groups of frame instances that form subgraphs and are then combined into a single graph based on matching content components and the logic of their introduction. **Conclusions.** The method provides a basis for constructing ontologies of subject-domain concepts, identifying coherence violations, and developing computer-based tools to support assessment of learning outcomes and learners' self-study.

Keywords: educational text, model, formalization, regulatory text, semantic text model, knowledge graph

For citation: Voronov, M.V., Stroeve, V.T. (2026). A method for formalizing educational texts. *Modelling and Data Analysis*, 16(3), 259—273. (In Russ.). <https://doi.org/10.17759/mda.2026160313>

Введение

Учебный текст создается в интересах обучения. Его назначение заключается в том, чтобы обеспечить передачу сведений о конкретной предметной области



и способствовать организации учебного процесса в интересах усвоения учебного материала. В него обычно включают определения, классификации, правила, теоремы, алгоритмы, примеры, пояснения, контрольные вопросы и задания. Эти элементы имеют разную функцию, но благодаря структуре своего построения в совокупности отображают заданную автором траекторию изложения учебного текста.

Применение современных информационных технологий для анализа и содержательного оперирования представленными в вербальном виде текстами (кроме собственно их чтения) весьма затруднено. В этой связи представляется актуальной следующая задача.

Задан представленный в вербальной форме исходный учебный текст. Требуется трансформировать его в формализованный аналог, в максимальной мере сохраняющий его учебно-методическое содержание и структуру. Это, в частности, означает, что он должен передавать структуру учебного материала: какие объекты и как вводятся, какие свойства им приписаны, какие действия и как с ними выполняются, какие результаты получаются, а также какие зависимости и в каком порядке между этими компонентами в учебном материале установлены.

Теоретическая основа

Задача формализации вербальных текстов не нова, ее постановка восходит к работам А.П. Ершова (Ершов, 1985). Однако она до сих пор далека от удовлетворительного решения. Более того, имеют место достаточно убедительные доводы в пользу следующего тезиса: разработка методов формализации произвольных вербальных текстов с передачей их содержания в обозримом будущем, скорее всего, невозможна (Walter et al., 2017). Вместе с тем по мере развития информационных технологий потребность в разработке соответствующих средств постоянно повышается, о чем свидетельствуют многочисленные исследования в этом направлении (Большакова и др., 2017; Воронов, 2022; Зацман, 2009; Lochter et al., 2018; Пешкова, 2015; Пименов, Воронов, 2021).

В сложившейся ситуации стали разрабатываться подходы к решению такого рода задач применительно к особым видам текстов. Так, перспективной представляется разработка методов формализации текстов, обладающих достаточной степенью регулятивности.

К регулятивным (вносящим порядок) относят тексты, которые по своему назначению и построению явно ориентированы на управление интерпретационной, следовательно, и поведенческой деятельностью читателя (Сидоров, 1987). Важно отметить, что регулятивный текст обладает такими ярко выраженными качествами, как информативность, структурность, интегративность, целенаправленность и др. (Болотнова, 2011, с. 34). Подчеркнем, что центральная функция регулятивных текстов заключается не только в передаче сообщений, в которых задаются сущности, условия, порядок выполнения операций и фиксируются результаты,



но и в ориентации читателя на его последующие действия; в такие тексты в большом количестве включены глагольные конструкции, выражающие повеление (императивы). Наиболее характерными представителями регулятивных текстов являются приказы, указания, инструкции, планы (Хорохордина, 2013). Ярко выражены такого рода свойства в описаниях технологий (Воронов, 2023).

Важно отметить, что регулятивным текстам присущи: логичность изложения, отсутствие иносказаний, минимум эмоциональной окраски и образности, при широком использовании специфической, предметно ориентированной терминологии. Становится ясно, что существенно выражены свойства регулятивности и у учебных текстов. Достаточно отметить, что они по самой своей сути ориентированы на целеизъявление автора направить мысли и действия обучаемого на достижение по вполне конкретной логике конкретных целей. Существенно, что учебный материал излагается, как правило, во вполне определенном порядке. Так, фрагменты знаний вводятся последовательно и методично с опорой на уже введенные (или полагаемые априори известными) сущности. В этом плане учебный текст в известной мере может рассматриваться как представленный в словесном виде алгоритм. По нашему мнению, именно наличие свойств регулятивности открывает возможности в достаточной мере успешного решения задачи формализации такого рода текстов в виде графов (Qu et al., 2024).

Подчеркнем: в поставленной задаче важнейшим является требование в максимальной степени сохранить в модели содержание именно данного исходного вербального текста, не допуская никакой его корректуры. В этой связи для ее решения представляется полезным привлечь некоторые результаты исследований лингвистов. В регулятивных текстах ключевое положение занимает отображение описываемой событийности. В этой связи следует обратиться к понятию «денотативная ситуация», которая представляет собой отображение текстового фрагмента, связанного с определенным событием. Как правило, таким фрагментом выступает предложение, в котором производится описание этого события.

С лингвистических позиций денотативная структура представляет собой определенное множество денотатов и отношений между ними (при ведущей роли последних). Следовательно, можно сказать, что содержание исходного текста заключено, главным образом, в его денотативных ситуациях. В этой связи ключевым может стать решение вопроса формализации отдельных денотативных ситуаций, каждая из которых в регулятивных текстах обычно представлена в обособленных предложениях.

В лингвистике исследование организации предложения с целью выявления его семантики осуществляется в рамках семантического (функционального) синтаксиса. При этом делаются попытки выявления смысловой организации предложения. Вместе с тем «пока нет четко сформулированных критериев выделения семантических моделей предложения и его компонентов» (Галимова, 2021, с. 255). Однако доказано, что содержание данного предложения может быть отображено в формализованной структуре (семантической конфигурации), образованной соответствующим множеством денотатов и отношений между ними.



При этом основным предметом рассмотрения выступает пропозиция, как языковое воплощение описываемого события, заключенное в семантической структуре предложения. По существу, пропозиция представляет собой отраженную модель описываемой в предложении ситуации (Арутюнова, 1976, с. 7). Ее составляющие — предикат, актанты, сирконстанты и связывающие их отношения. Подчеркнем, основная семантическая роль в предложении, согласно вербоцентрической теории Теньера, принадлежит предикату, как вершине пропозиции.

Согласно этой теории, предикат (с позиции лингвистики — сказуемое) является центральным и единственным главным компонентом пропозиции. Чаще всего он выражен глаголом, иногда причастием, деепричастием или отглагольными существительными словами категории состояния. Для нас важно то, что именно предикат, обозначая действие, состояние, свойство или качество предмета, определяет места и роль остальных участников описания ситуации.

Актанты — это имена действующих лиц данной пропозиции. В их число входят подлежащее и дополнения. Как элементы событийной пропозиции они заложены в ролевую структуру сказуемого (входят в его базовую валентность). Описание актантов обычно дополняют и уточняют соответствующие для каждого из них атрибуты.

Для отражения факультативных валентностей использованного глагола в предложении могут присутствовать функциональные единицы, называемые сирконстантами (адъюнктами). В содержательном аспекте это обстоятельства или определения, как правило соответствующие времени, месту и способу описываемого в предложении действия.

В качестве результатов описываемого события выступает вводимая новая сущность, описание ее свойств или новое состояние участников описываемого события.



Рис. 1. Семантическая структура предложения

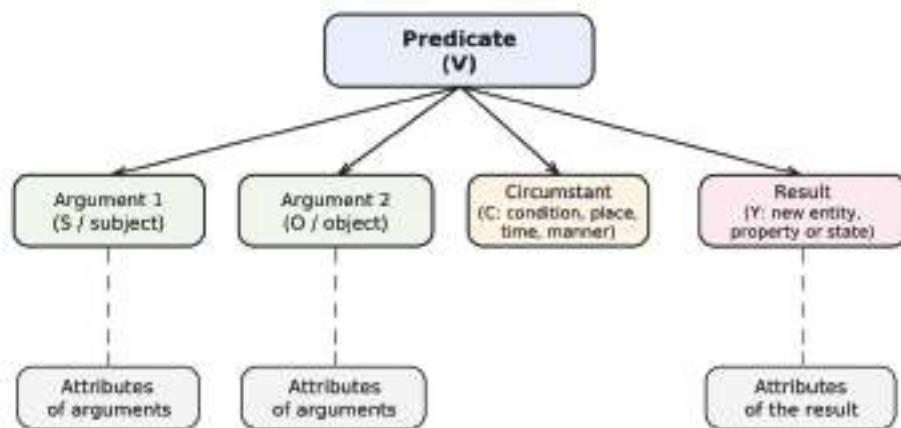


Fig. 1. Semantic structure of the sentence

Заметим, что содержательно «предикат и актанты — это компоненты смысла языкового выражения (предложения, словосочетания), а не слова, составляющие предложение, не члены предложения» (Прияткина, 1990, с. 60).

Начало научного рассмотрения проблематики формализации семантики текста связано, пожалуй, с работами А.И. Новикова (Новиков, 1983), в которых, в частности, высказана идея отображения денотативного уровня текста в виде денотатного графа, который трактуется как «свернутое эксплицитное отображение структуры содержания текста, которой могут соответствовать различные языковые формы» (Новиков, Нестерова, 1991, с. 61). Предлагаемые в этих работах подходы построения денотатного графа базируются на иерархическом выделении так называемых «ключевых» компонентов, подтем и раскрывающих их содержание субподтем. Иначе говоря, элементами формируемого графа выступают самые различные фрагменты учебного текста, часто входящие в состав его оглавления. Такой подход активно используется преподавательским составом, поскольку, отражая в целом логику изложения учебного материала, способствует организации работы и обучающего, и обучаемого.

Исследования показывают, что успех поставленной нами задачи в существенной степени зависит от способа членения текста на составляющие единицы (Пименов, Воронов, 2021). Фрагменты, на которые вполне обоснованно разбивается учебный текст с педагогических позиций, не могут рассматриваться в качестве неделимых единиц (учебных квантов), ибо представляют собой, как правило, несколько предложений. Отдельное же предложение — это связанная по смыслу группа слов, выражающая законченную мысль, в данном контексте ключевое значение в этом определении имеет термин «законченную мысль». Именно в этой связи в качестве содержательной единицы учебного текста примем его отдельное предложение.



В учебных текстах практически каждое предложение несет на содержательном уровне конкретную единицу учебной информации, отображая введение новой сущности, ее свойств или процедуру изменения состояния рассматриваемого объекта. Логически упорядоченная последовательность такого рода содержательных предложений и передает содержание учебного текста в целом.

Вышеизложенное позволяет разработать конструктивный метод формализации предложений учебного текста.

Результаты

Предлагается метод построения формальной модели, отображающей содержание и логику данного учебного текста, осуществляемый в виде конструктивного процесса. Этот процесс разобьем на два этапа. На первом в автоматизированном режиме каждое предложение (в строгом порядке их следования по тексту) отображается в соответствующую совокупность фреймов, которая трактуется как граф предложения. На втором в автоматическом режиме построенные графы предложений объединяются в единый граф, который и представляет собой формальную модель рассматриваемого текста.

Такого рода алгоритм представим в следующем виде.

1. Организуется строго занумерованная последовательность просмотра очередного i -го предложения данного текста $i \in (1-n)$. Если все предложения просмотрены, то переход к п. 7.
2. С одинаковым номером i формируются следующие три фрейма-прототипа:
 - 2.1 Во фрейме с именем D_i фиксируется информация о содержании описываемого в данном предложении предиката

$$D_i = \{ Pf_i, b, \delta_{i,b}, \{h, B\delta_{b,h}\}, \},$$

где:

b — номер присутствующего в данном предложении предиката. В предложении (например, в сложном предложении) может быть более одного сказуемого, поэтому $b \in (1-b_0)$;

$\delta_{i,b}$ имя предиката (состоит из слов, отображающих в i -м предложении описание действия);

$\{h, B\delta_{b,h}\}$ — множество относящихся к данному предикату номеров и соответствующих имен сирконстантов (слов, указывающих на время, место, образ действия и другие обстоятельства ситуации). Их у предиката может быть несколько, поэтому $h \in (1-h_0)$;

Pf_i — файл исходного текста предложения (выполняет служебную роль).

- 2.2 Во фрейме с именем A_i фиксируется информация об описании участвующих в описываемой ситуации i акторов

$$A_i = \{ a, X_{i,a}, \{r, AX_{a,r}\}, \}$$



где:

a — номер очередного актора. В каждом действии обычно участвует несколько акторов, поэтому $a \in (1-a_0)$;

$X_{i,a}$ — имя актора a ;

$\{r, AX_{a,r}\}$ — множество пар номеров и имен атрибутов, дополняющих и уточняющих описание соответствующего актанта. При каждом актанте может находиться по несколько атрибутов, поэтому $q \in (1-q_0)$.

2.3 Во фрейме Y_i фиксируется информация о результатах описанного в предложении действия (действиях)

$$R_i = \{c, Y_{i,c}, \{k, RY_{c,k}\}\}$$

где:

c — номер результата. В каждом действии могут фиксироваться несколько результатов, например, когда описаны новые состояния нескольких сущностей, поэтому $c \in (1-c_0)$;

$Y_{i,c}$ — имя результата c (слова, описывающие в данном предложении результат под номером c);

$\{k, RY_{c,k}\}$ — множество пар номеров и имен результатов, дополняющих и уточняющих описание соответствующего результата. В предложении может фиксироваться несколько результатов, поэтому $k \in (1-k_0)$.

3. В автоматизированном режиме осуществляется заполнение этих фреймов, и они превращаются в группу фреймов-экземпляров, в которых зафиксировано содержание и структура соответствующего предложения.
4. Формализмом, отображающим содержание конкретного i -го предложения, становится заполненная именованная тройка фреймов-экземпляров

$$KV_i = \langle A_i, D_i, Y_i \rangle$$

Примечание. Еще раз подчеркнем важное требование: в KV_i фиксируются все, но только те элементы содержания, которые представлены в данном предложении рассматриваемого текста.

5. Выполняется содержательно-логическое объединение построенных фреймов-экземпляров в единую модель, представляющую формальную модель содержания рассматриваемого текста. Если учебный текст написан корректно, то представленные в данном предложении a акторы $\{X_{i,a}\}$ уже были введены ранее в качестве результатов описанных действий или относятся к так называемым базовым, которые автором текста предполагаются априори известными (такого рода факты целесообразно специальным образом фиксировать). В этой связи для каждого присутствующего в данном предложении i актора a можно найти эквивалентный ему по содержанию фрейм (поиск производится по совпадению имен, ибо предполагается, что в учебном тексте омонимы не встречаются), а, следовательно,



и номер предложения j , в котором данная сущность была введена в текст выше в качестве соответствующего результата (результат под номером c).

Для выявления этого соответствия реализуется следующая процедура:

$$\forall i, a, c, \forall j < i, \exists (j, c): X_{i,a} \sim Y_{j,c} \Rightarrow g_{i,j} = 1.$$

6. Выявленные отношения эквивалентности $X_{i,a} \sim Y_{j,c}$ фиксируются в формируемой матрице смежностей

$$g_{i,j} \in MG = g_{i,j}$$

7. Построенный ориентированный граф, отображающий рассматриваемый учебный текст в целом, представляет собой объединение $(n+1)$ графов:

$$G = \{KV_i\} \cup MG$$

вершины которого соответствуют элементам множества $\{KV_i\}$, а ребрам — элементы матрицы MG .

Обсуждение результатов

Полученные результаты подтверждают перспективность перехода от общей задачи формализации произвольных текстов к работе с регулятивными текстами. В отличие от методов построения денотатного графа на основе крупных тематических фрагментов (Новиков, 1983; Новиков, Нестерова, 1991), предлагаемый подход использует предложение как минимальную содержательную единицу. Это позволяет сохранять порядок введения понятий, отношения между содержательными составляющими предложения и проследить связи между результатами одних предложений и акторами последующих. По сравнению с обобщенными подходами к образовательным графам знаний (Qu et al., 2024), метод ориентирован не на интеграцию внешних ресурсов, а на формальное воспроизведение логики конкретного учебного текста.

В результате применения изложенного метода появляется возможность построения ориентированного графа, представляющего собой формальную модель рассматриваемого учебного текста в целом. Каждый его подграф отображает содержание отдельного предложения, выступающего в качестве кванта исходного учебного текста. Практически в каждом предложении текста вводится новая сущность или процедура, выполнение которой изменяет состояние соответствующей группы сущностей. Эта информация фиксируется в ряде фреймов, причем в такой форме, чтобы имелась возможность адекватного воспроизведения содержания, передаваемого данным предложением. Ключевым моментом в достижении этого эффекта является декомпозиция предложения, следующая его семантической структуре (рис. 1).



Для подтверждения работоспособности разработанного метода были проведены эксперименты по сравнению содержания исходного и восстановленного из модели текстов. Для этого были выбраны фрагменты текста из известного учебника «Математика и информатика» (Турецкий, 2002). Экспертиза признала высокий уровень совпадения содержания сравниваемых текстов и практически адекватное совпадение передаваемого ими содержания.

Все включенные в подграфы содержательно связаны, причем следуя точно по той же логике, что и содержательные элементы исходного текста. Поскольку построенный граф адекватно отображает установленные в тексте между этими элементами отношения, открываются перспективы построения компьютерных методик для эффективного решения широкого ряда прикладных вопросов. Заметим, что графовые модели очень удобны для построения соответствующих программных продуктов.

Перспективы использования метода

Построенный граф, отображающий содержание учебного текста, — это не конечный продукт, а основа для разработки множества приложений. Вот только некоторые из них.

Важным свойством учебного текста является полнота вводимых сущностей и порядок их введения. Критерием полноты вводимых сущностей может выступать результат проверки связности построенного графа GT.

Может быть поставлен и решен вопрос об интенсивности использования в данном учебном материале той или иной сущности; на базе выбранной метрики можно оценить ее «важность» и вынести более обоснованные учебно-методические рекомендации по изучению данного учебного курса.

На многие вопросы, связанные с оценкой учебного текста, можно найти ответы, используя результаты теории оценки целостности графа (Быкова, 2014). В частности, оценивая степень возможного «повреждения графа», можно судить об уровне влияния фактов невведения в учебный текст предметных сущностей и выявлять другие факторы «повреждения» логики изложения учебного материала.

В открытом доступе имеются библиотеки компьютерных процедур, обеспечивающих решение широкого спектра так называемых «задач на графах». Среди них важное место занимают задачи «построения путей на графе». Такого рода инструменты активно используются при построении средств поддержки процессов самоподготовки студентов. В ходе этих процессов часто возникают ситуации, когда студент испытывает затруднения при выполнении конкретного учебного задания. При этом обычно возникает вопрос о том, что лежит в основе возникших трудностей именно для данного студента, причем «здесь и сейчас». Наличие возможности построения «обратных» путей от данной учебной ситуации к ее истокам, причем строго следуя логике изложения учебного материала в данном учебном тексте и учитывая логику использования «предпосылок», позволяет найти причины возникших затруднений и сформировать актуальный для данной ситуации вариант адекватных рекомендаций обучающемуся.



Разработанный метод позволяет строить методики формирования родовидовой структуры используемых в данном учебном материале сущностей и на их основе разрабатывать программные продукты, реализующие эти методики для конструирования таксономической онтологии рассматриваемой предметной области (Маторин, Михелев, 2021).

Заключение

Изложенный метод построения модели, отображающей содержание учебного текста, базируется на результатах семантической лингвистики, методах формирования баз знаний и идеях конструктивной математики. В результате его применения представленный в вербальной форме учебный материал отображается в ориентированный граф, в котором сохраняется содержание и логика построения исходного учебного материала.

Благодаря этому сформированный граф может использоваться в качестве базы для создания ряда качественно новых программных средств. Так, в результате анализа построенной для данного учебного текста модели может быть разработана онтология вводимых понятий, произведена оценка содержательно-логической связности представленного текста, выявлены случаи нарушения логики изложения учебного материала и др.

Если на основе данного учебного материала построена программа освоения соответствующей учебной дисциплины, удобно строить средства компьютерной поддержки процессов подготовки контрольно-измерительных материалов, проведения собственно контроля и самоподготовки конкретного обучаемого.

Объединение моделей, построенных для совокупности книг по данной предметной области, открывает возможность построения открытой для наращивания соответствующей единой базы знаний, более обоснованно формировать рекомендации по методическому обеспечению данной учебной дисциплины и даже образовательной программы в целом.

Ограничения. Несомненно, предложенный метод имеет ряд ограничений. Во-первых, желательна высокая регулятивность текста, в первую очередь его логичность, терминологическая устойчивость, минимальная образность и отсутствие омонимии. В этой связи наиболее подходящими являются учебные материалы для математики, естественно-научных и профильных дисциплин инженерного и технологического направления.

Во-вторых, эффективность метода существенно зависит от правильности интерпретации предикатно-актантной структуры предложения, ибо корректное определение того, какие элементы участвуют в описываемой ситуации и какую роль они в ней играют, обеспечивает уровень адекватности отображения семантики данного исходного предложения при его формализации.



В-третьих, метод описан для учебных текстов, написанных на русском языке, и, по-видимому, требует адаптации к особенностям различного, вообще говоря, литературного почерка авторов, а также при рассмотрении переводных текстов.

Список источников / References

1. Арутюнова, Н.Д. (1976). *Предложение и его смысл: Логико-семантические проблемы*. М.: Наука.
Arutyunova, N.D. (1976). *Predlozhenie i ego smysl: Logiko-semanticheskie problemy* [Sentence and its meaning: Logical-semantic problems]. Moscow: Nauka. (In Russ.).
2. Болотнова, Н.С. (2011). О типологии регулятивных структур в тексте как форме коммуникации. *Вестник Томского государственного педагогического университета*, 3(105), 34–40.
Bolotnova, N.S. (2011). O tipologii regulativnykh struktur v tekste kak forme kommunikatsii [On the typology of regulative structures in the text as a form of communication]. *Vestnik Tomskogo gosudarstvennogo pedagogicheskogo universiteta*, 3(105), 34–40. (In Russ.).
3. Большакова, Е.И., Воронцов, К.В., Ефремова, Н.Э., Клышинский, Э.С., Лукашевич, Н.В., Сапин, А.С. (2017). *Автоматическая обработка текстов на естественном языке и анализ данных: Учебное пособие*. М.: Изд-во НИУ ВШЭ.
Bolshakova, E.I., Vorontsov, K.V., Efremova, N.E., Klyshinsky, E.S., Lukashevich, N.V., Sapin, A.S. (2017). *Avtomaticheskaya obrabotka tekstov na estestvennom yazyke i analiz dannykh: Uchebnoe posobie* [Natural language processing and data analysis: Textbook]. Moscow: HSE Publ. (In Russ.).
4. Быкова, В.В. (2014). О мерах целостности графов: обзор. *Прикладная дискретная математика*, (4(26)), 96–111. <https://doi.org/10.17223/20710410/26/11>
Bykova, V.V. (2014). O merakh tselostnosti grafov: obzor [On measures of graph integrity: A survey]. *Prikladnaya diskretnaya matematika*, (4(26)), 96–111. (In Russ.). <https://doi.org/10.17223/20710410/26/11>
5. Воронов, М.В. (2022). Моделирование структуры представленных в учебных текстах знаний. *Педагогика информатики*, (1–2), 99–108.
Voronov, M.V. (2022). Modelirovanie struktury predstavlennykh v uchebnykh tekstakh znaniy [Modeling the structure of knowledge presented in educational texts]. *Pedagogika informatiki*, (1–2), 99–108. (In Russ.).
6. Воронов, М.В. (2023). Метод формализации описания технологических процессов. *Информатика и кибернетика*, 3(33), 50–57.
Voronov, M.V. (2023). Metod formalizatsii opisaniya tekhnologicheskikh protsessov [A method for formalizing the description of technological processes]. *Informatika i kibernetika*, 3(33), 50–57. (In Russ.).
7. Галимова, Х.Н. (2021). Семантический синтаксис и пропозиция как единица пропозиционального анализа текстов. *Балтийский гуманитарный журнал*, 10(3(36)), 255–258. <https://doi.org/10.26140/bg3-2021-1003-0062>
Galimova, Kh.N. (2021). Semanticheskii sintaksis i propozitsiya kak edinitsa propozitsionalnogo analiza tekstov [Semantic syntax and proposition as a unit of propositional analysis of texts]. *Baltiiskii gumanitarnyi zhurnal*, 10(3(36)), 255–258. (In Russ.). <https://doi.org/10.26140/bg3-2021-1003-0062>



8. Ершов, А.П. (1985). Машинный фонд русского языка (Внешняя постановка вопроса). *Вопросы языкознания*, (2), 51–53.
Ershov, A.P. (1985). Mashinnyi fond russkogo yazyka (Vneshnyaya postanovka voprosa) [Machine fund of the Russian language (External statement of the problem)]. *Voprosy yazykoznanii*, (2), 51–53. (In Russ.).
9. Зацман, И.М. (2009). Семиотическая модель взаимосвязей концептов, информационных объектов и компьютерных кодов. *Информатика и ее применение*, 3(2), 65–81.
Zatsman, I.M. (2009). Semioticheskaya model vzaimosvyazei kontseptov, informatsionnykh obektov i kompyuternykh kodov [A semiotic model of relationships between concepts, information objects and computer codes]. *Informatika i ee primeneniye*, 3(2), 65–81. (In Russ.).
10. Маторин, С.И., Михелев, В.В. (2021). Системно-объектный детерминантный анализ. Построение таксономии предметной области. *Искусственный интеллект и принятие решений*, (1), 15–24. <https://doi.org/10.14357/20718594210102>
Matorin, S.I., Mikhelev, V.V. (2021). Sistemno-obektnyi determinantnyi analiz. Postroenie taksonomii predmetnoi oblasti [System-object determinant analysis. Constructing a taxonomy of a subject area]. *Iskusstvennyi intellekt i prinyatie reshenii*, (1), 15–24. (In Russ.). <https://doi.org/10.14357/20718594210102>
11. Новиков, А.И. (1983). *Семантика текста и ее формализация*. М.: Наука.
Novikov, A.I. (1983). *Semantika teksta i ee formalizatsiya* [Text semantics and its formalization]. Moscow: Nauka. (In Russ.).
12. Новиков, А.И., & Нестерова, Н.М. (1991). *Реферативный перевод научно-технических текстов*. М.: Наука.
Novikov, A.I., & Nesterova, N.M. (1991). *Referativnyi perevod nauchno-tekhnicheskikh tekstov* [Abstract translation of scientific and technical texts]. Moscow: Nauka. (In Russ.).
13. Пешкова, Н.П. (2015). Семантика и смысл текста: >, <, =, #? (экспериментальный подход к теоретическим проблемам). *Вестник Челябинского государственного университета. Филология. Искусствоведение*, 15(370), 96, 69–77.
Peshkova, N.P. (2015). Semantika i smysl teksta: >, <, =, #? (eksperimentalnyi podkhod k teoreticheskim problemam) [Semantics and meaning of text: >, <, =, #? (An experimental approach to theoretical problems)]. *Vestnik Chelyabinskogo gosudarstvennogo universiteta. Filologiya. Iskusstvovedenie*, 15(370), 96, 69–77. (In Russ.).
14. Пименов, В.И., Воронов, М.В. (2021). Формализация регулятивных текстов. *Информатика и автоматизация*, 20(3), 562–590. <https://doi.org/10.15622/ia.2021.3.3>
Pimenov, V.I., Voronov, M.V. (2021). Formalizatsiya regulyativnykh tekstov [Formalization of regulatory texts]. *Informatika i avtomatizatsiya*, 20(3), 562–590. (In Russ.). <https://doi.org/10.15622/ia.2021.3.3>
15. Прияткина, А.Ф. (1990). *Русский язык*. М.: Высшая школа.
Priyatkina, A.F. (1990). *Russkii yazyk* [Russian language]. Moscow: Vysshaya shkola. (In Russ.).
16. Сидоров, Е.В. (1987). *Проблемы речевой системности*. М.: Наука.
Sidorov, E.V. (1987). *Problemy rechevoi sistemnosti* [Problems of speech systematicity]. Moscow: Nauka. (In Russ.).
17. Турецкий, В.Я. (2002). *Математика и информатика* (3-е изд., испр. и доп.). М.: ИНФРА-М.
Turetskii, V.Ya. (2002). *Matematika i informatika* [Mathematics and informatics] (3rd ed., rev. and enl.). Moscow: INFRA-M. (In Russ.).



18. Хорохордина, О.В. (2013). Инструкция как тип текста. *Мир русского слова*, (4), 7–14.
Khorokhordina, O.V. (2013). Instruktsiya kak tip teksta [Instruction as a type of text]. *Mir russkogo slova*, (4), 7–14. (In Russ.).
19. Lochter, J.V., Pires, P.R., Bossolani, C., Yamakami, A., Almeida, T.A. (2018). Evaluating the impact of corpora used to train distributed text representation models for noisy and short texts. In *2018 International Joint Conference on Neural Networks (IJCNN)* (pp. 1–8). IEEE. <https://doi.org/10.1109/IJCNN.2018.8489355>
20. Qu, K., Li, K.C., Wong, B.T.M., Wu, M.M.F., Liu, M. (2024). A survey of knowledge graph approaches and applications in education. *Electronics*, 13(13), Article 2537. <https://doi.org/10.3390/electronics13132537>
21. Walter, B., Hammes, J., Piechotta, M., Rudolph, S. (2017). A formalization method to process structured natural language to logic expressions to detect redundant specification and test statements. In *2017 IEEE 25th International Requirements Engineering Conference (RE)* (pp. 263–272). IEEE. <https://doi.org/10.1109/RE.2017.38>

Информация об авторах

Михаил Владимирович Воронов, доктор технических наук, профессор, заведующий кафедрой прикладной математики, факультет информационных технологий, Московский государственный психолого-педагогический университет (ФГБОУ ВО МГППУ), Москва, Российская Федерация, ORCID: <https://orcid.org/0000-0001-7839-6250>, e-mail: mironovov@yandex.ru

Владислав Тазирович Строев, аспирант, Московский государственный психолого-педагогический университет (ФГБОУ ВО МГППУ), Москва, Российская Федерация, ORCID: <https://orcid.org/0009-0002-8630-455X>, e-mail: liltot@bk.ru

Information about the authors

Mikhail V. Voronov, Doctor of Science (Engineering), Professor, Head of the Department of Applied Mathematics, Faculty of Information Technologies, Moscow State University of Psychology and Education, Moscow, Russian Federation, ORCID: <https://orcid.org/0000-0001-7839-6250>, e-mail: mironovov@yandex.ru

Vladislav T. Stroeov, Postgraduate Student, Moscow State University of Psychology and Education, Moscow, Russian Federation, ORCID: <https://orcid.org/0009-0002-8630-455X>, e-mail: liltot@bk.ru

Вклад авторов

Воронов М.В. — концептуализация, методология, научное редактирование, руководство исследованием;

Строев В.Т. — анализ предметной области, визуализация, написание первоначального варианта, оформление рукописи. Все авторы участвовали в разработке алгоритма, обсуждении результатов и согласовали окончательный текст рукописи.

Authors contribution

Mikhail V. Voronov — conceptualization, methodology, scientific editing, research supervision

Vladislav T. Stroeov — domain analysis, visualization, original draft preparation, manuscript formatting. Both authors contributed to the development of the algorithm, discussed the results, and approved the final manuscript.



Конфликт интересов

Авторы заявляют об отсутствии конфликта интересов.

Conflict of Interest

The authors declare no conflict of interest.

Поступила в редакцию 30.05.2026

Поступила после рецензирования 07.08.2026

Принята к публикации 10.08.2026

Опубликована 30.09.2026

Received 2026.05.30

Revised 2026.08.07

Accepted 2026.08.10

Published 2026.09.30