

М.В. ВОРОНОВ, В.И. ПИМЕНОВ  
**ФОРМАЛИЗАЦИЯ РЕГУЛЯТИВНЫХ ТЕКСТОВ**

*Воронов М.В., Пименов В.И. Формализация регулятивных текстов.*

**Аннотация.** Современные информационные технологии с высокой эффективностью обеспечивают манипулирование текстами. В первую очередь это хранение, редактирование и форматирование текстов и их составляющих. Добившись значительных успехов в разработке средств бессодержательной компьютерной обработки текстов, исследователи столкнулись с проблемами их содержательной обработки. Поэтому дальнейшие шаги в этом направлении связаны с созданием, в том числе, методов автоматизированного целенаправленного манипулирования текстами с учетом заложеного в них содержания. Проведен анализ работ, посвященных исследованию проблем формального представления текстов и их последующего использования. Несмотря на ряд успешных проектов, задачи по разрешению проблемы соотношений между содержанием текста и его смыслом остаются актуальными. По-видимому, формализация текста общего назначения при сохранении его семантики на данном этапе развития науки неосуществима. Вместе с тем существуют типы текстов, для которых формализация с сохранением части их семантики представляется возможной. Одним из них является регулятивный тип текста, который по существу представляет собой словесно выраженный алгоритм последовательности целенаправленных действий. Его отличают логичность и точность (отсутствие иносказаний), связность и цельность, ясность, понятность (за счет отсутствия эмоциональной окраски и образных средств), доступность (за счет использования специфической терминологии). Иначе говоря, при разработке регулятивных текстов обычно стараются как можно понятнее отображать механизмы описываемых действий. Цель: разработка метода формализации регулятивного текста с сохранением его семантики. Методы: структурная лингвистика, представление объектов в виде онтологии, конструктивные алгоритмы. Использование данного метода демонстрируется на примере описания решения системы алгебраических уравнений. Результаты: метод построения математической модели регулятивного текста. Практическая значимость: применение разработанного метода обеспечивает возможность разработки программных комплексов для построения библиотек отдельных предметных областей, разработки средств оценки регулятивных текстов на предмет их определенности, полноты, связности и других характеристик, а также тренажеров и средств самообучения.

**Ключевые слова:** регулятивный текст, понятие, структурная лингвистика, искусственный интеллект, математическая модель, конструктивный алгоритм, онтология, фрейм

**1. Введение.** Результаты своей мыслительной деятельности человек стремится зафиксировать на каком-либо материальном носителе. Для этого чаще всего используются определенные последовательности символов, называемые текстами. Изображение текста представляет собой определенную форму (упорядоченную последовательность символов), отображающую вложенное в него автором содержание и является попыткой автора фиксировать и объективировать последнее.

Представители многих областей знаний рассматривают тексты в качестве объектов своих исследований, поскольку они предназначены, в первую очередь, для переноса в пространстве и во времени заложенных в них автором результатов речемыслительной деятельности авторов. Иными словами, текст исполняет роль почтальона, цель которого – передать получателю вложенное автором содержание, причем без искажений.

Вычленив же из текста находящееся в нем содержание сегодня способен только человек, так как содержание текста не существует вне процесса его восприятия и понимания. Эти же процессы «включаются» под воздействием сигналов, формирующихся при выявлении составляющих текст символов. При этом успех процессов восприятия и передачи смысла (того, что может быть понято) обусловлен тем, в какой мере совпадают возможности восприятия информации данного контекста у автора и читателя что в свою очередь зависит и от того, как построен сам текст, то есть из каких компонентов и связей между ними он состоит.

Современные информационные технологии, оперируя с текстами в основном как с бессодержательной структурой, с высокой эффективностью обеспечивают процессы такого манипулирования ими. В первую очередь это хранение, редактирование и форматирование текстов и их составляющих. Но не только. Отдельные программы в определенной мере уже «умеют», например, выделять факты из текста, распознавать сарказм и т.п. Однако они «не понимают слова» как это делают люди, а используя, как правило, вероятностные модели и нечеткие алгоритмы, попросту угадывают значения слов, словесных оборотов и даже отдельных предложений. Дальнейшие шаги в этом направлении связаны с созданием, в том числе, методов автоматизированного целенаправленного манипулирования текстами с учетом заложенного в них содержания (исходя из этого содержания и руководствуясь им), что является предметом нашего исследования. Таким образом, несмотря на прогресс в области информационных технологий, задачи по разрешению проблемы соотношений между содержанием текста и его смыслом остаются актуальными [1].

Ставится задача формализации текста со стремлением сохранить исходное содержание при передаче, делается попытка открыть новые возможности формирования баз знаний для перспективных систем искусственного интеллекта.

**2. Анализ проблем формализации текстов.** Использование в текстах слов, представляющих собой имена некоторых сущностей денотата, обеспечивает читателю возможность соотносить с

конкретным словом свое представление о соответствующих объектах. Например, отдельно взятое слово «стол» практически у каждого читателя будет ассоциироваться с соответствующим элементом мебели. Однако о каком конкретно столе идет речь и, следовательно, можно ли получить о нем какие-либо дополнительные сведения, данная совокупность букв не сообщает. Иначе говоря, считанное изолированное слово передает весьма обобщенные и, как правило, крайне скудные сведения. Более того, интерпретация слова в существенной мере зависит от воспринимающего его субъекта. Дело в том, что передаваемое данным словом содержание для конкретного лица в значительной мере обусловлено его индивидуальными представлениями об этом объекте (свойствах, связях и отношениях с другими объектами и явлениями). В таких случаях посредником между использованным словом и действительностью выступает концепт, включающий в себя некоторый потенциал данного слова [2]. Концепт, как выраженная словом сущность, и выступает основной единицей, передающей определенный компонент содержания текста. В этой связи в процессе формализации для каждого использованного в тексте слова целесообразно фиксировать отображаемый им концепт.

Для конкретизации и уточнения передаваемой отдельными словами информации используются их определенные сочетания. Именно добавление структуры в данную совокупность отдельных слов обеспечивает передачу существенных с позиции автора законченных мыслей. В таких случаях эту упорядоченную совокупность слов называют предложением. Важно заметить, что образующие предложения слова, объединенные определенной лексической семантикой, обеспечивают эффект так называемой семантической эмерджентности: для данной ситуации это объединение более информативно значимо, чем все слова этого предложения, взятые в отдельности. Именно поэтому предложение считают основной единицей содержания в тексте.

Содержательно обусловленное соединение предложений образует текст, представляющий собой некую целостность, причем подмножество составляющих его предложений не может в полной мере отображать эту целостность. Иначе говоря, из отдельных слов и даже предложений невозможно в полной мере воспринять смысл данного текста.

Автор, закладывая в формируемый им текст определенное содержание, стремится к тому, чтобы восприятие читателя текста было в достаточной мере адекватно его мыслям. К сожалению, на пути от автора через текст к читателю (адресату) зафиксированные результаты

мыслительной деятельности автора подвергаются различного рода искажениям. В какой мере результаты деятельности читающего текст будут соответствовать результатам мыслительной деятельности автора, зависит, как минимум, от трех составляющих.

Во-первых, от личностных особенностей автора, ведь часто думая одно, он отображает в тексте несколько иное. Так, в работе В.Д. Орехова, например, отмечается, что точность такой передачи составляет порядка 14% [3].

Во-вторых, от читателя. Основная цель читающих данный текст людей заключается в извлечении из него той информации, которую в него заложил автор. Различные группы читателей могут по-разному воспринимать заложенное в данном тексте содержание, причем даже при условии, что автор и читатель текста, что называется, «говорят на одном языке». Влияние множества разнородных факторов на восприятие текста столь существенно, что позволяет ряду авторов утверждать: передача смысла в принципе не может быть автоматизирована [4].

В-третьих, от того, каким образом трансформировался текст на пути от автора к читателю. Дело в том, что после своего появления текст начинает свою «самостоятельную жизнь», перенося в пространстве и во времени зафиксированные в нем мысли автора, обеспечивает потенциальную возможность воспользоваться ими другим людям. При этом на своем «жизненном пути» текст может подвергаться различного рода трансформациям, причем как случайным, так и целенаправленным.

Стремление использовать компьютерные технологии для работы с текстами вызывает необходимость разработки методов, позволяющих трансформировать исходный вербальный текст в текст, представленный на искусственном (машинном) языке. Задача при этом состоит в том, чтобы при прочих равных условиях смысл, воспринимаемый при чтении текста в формализованном виде, в максимальной мере соответствовал смыслу, воспринимаемому при чтении этого же текста в вербальной форме.

*Примечание.* В данной работе авторы, стремясь минимизировать повторение фраз, используют категорические суждения типа «основной», «сохранить», «не может быть» и др., понимая, что более корректно дополнять их словами «как правило», «в основном» и т.п.

Тематика исследований, связанных с проблемами формального представления текстов и их последующего использования, весьма широка [5, 6]. При этом важное место в них обычно занимают вопросы

передачи смыслового содержания текстов [7]. Несмотря на широкий спектр подходов, оптимальным остается подход, позволяющий на основе выявления концептуального состава текста определять его смысловую структуру [8].

Одним из заслуживающих внимания подходов к решению поставленной выше задачи является методика, суть которой заключается в замене лексико-грамматической структуры текста денотативной структурой графа, который призван отображать присутствующие в тексте семантические связи [9]. В процессе его построения выделяются объекты (денотаты), о которых идет речь в тексте (будущие вершины графа) и связи между ними (будущие ребра графа). При этом все операции построения графа обусловлены передаваемым текстом смыслом и представляют собой некий способ материализации результата понимания [10]. К сожалению, методика построения денотатного графа (как и методики, построенные на использовании триединой конструкции «Узел-Функция-Объект» [11]) включает большое число процедур творческого характера, которые каждый индивидуум способен реализовывать по-своему, что может приводить к неоднозначности результата. Необходима же разработка методов, включающих минимум творческой составляющей субъектов.

Во многих современных приложениях, посвященных вопросам автоматизированного сравнения текстов, используется идея их представления в виде векторов (построение Vector Space Model) [12]. Координаты этих векторов формируются на основе частот вхождений каждого термина в данный текст. Тем самым исходная задача решается средствами векторной алгебры. Однако имеющие место несовпадения семантической и лексикографической схожести далеко не всегда гарантируют удовлетворительное решение поставленной задачи [13].

При решении вопросов формализации передаваемых текстами смыслов все большее значение приобретают достижения лингвистики. Дело в том, что объектом изучения лингвистики является язык, который самым тесным образом связан со всеми аспектами мыслительной и познавательной деятельности человека. Поэтому лингвистика в целом и когнитивная лингвистика в особенности контактируют с широким спектром наук и во все большей мере с проблемами информатики [14]. В этом плане представляет интерес анализ исследований по вычислительной лингвистике В.З. Демьянкова [15].

Со многими трудностями, характерными для задач формализации текстов, встречаются исследователи сферы переводов текстов с языка на язык (задачу формализации текста в известной

мере можно сопоставлять с задачами его перевода с языка на язык). В теории перевода формулируются лексические и грамматические правила, которые помогают анализировать текст на входном языке и синтезировать его на выходном. Однако переводчики, как правило, воспринимают текст в виде совокупности отдельных предложений. В итоге применение этих правил не всегда обеспечивает сохранение смысла, заложенного его автором в первоисточнике [16].

Информацию о представленных в тексте фактах лингвисты обычно связывают с его содержанием, рассматривая семантическое пространство, которое наделяют определенной структурой [17, 18]. Для нас важно, что основными компонентами этой структуры выступают пропозиции, как объективизированные средства содержания мысли [19]. Введение в анализ текстов структуры – существенный шаг в направлении решения вопросов их формализации. Следующим шагом в этом направлении было введение пресуппозиции, трактуемой как суждение, выводимое из данного высказывания по правилам истинности или уместности [20]. Казалось бы, здесь лингвисты вплотную приблизились к возможности применения математической логики. Однако основные их усилия направлены на развитие теории речевой коммуникации и дискурсивной психологии, поэтому полученные при этом результаты не всегда находят прямое применение при разработке формализованных методов передачи содержания текстов. Вместе с тем результаты лингвистических исследований являются базой, на которой строятся многие методы формализации текстов в различных сферах деятельности людей.

Проблематикой формализации передаваемых текстами смыслов занимаются и со стороны информатики. Здесь, в рамках рассматриваемой тематики, следует обратить внимание, например, на работу [21], где формирование и использование ассоциативно-онтологического представления данных обеспечивает получение средств автоматического реферирования научных и технических документов. Особо следует отметить работы, проведенные в рамках когнитивной информатики, которые связаны с изучением отображения в информационных системах процессов эволюции знаковых систем, в частности построения взаимно однозначных отношений между концептами и компьютерными кодами [22].

### **3. Основания построения моделей регулятивных текстов.**

Представляется перспективным вести поиск путей разрешения проблемы, возникающей при формализации текста с сохранением передаваемого им содержания, базируясь на следующем утверждении: содержание любых формальных преобразований заключено в причинности [23]. Причина

же становится ясной, если понятен механизм ее действия. Отсюда следует, что при построении модели текста максимум внимания целесообразно сосредоточить на выявлении и передаче содержащихся в нем причинно-следственных механизмов.

К сожалению, различные тексты в неодинаковой мере содержат явные и четко описанные механизмы действий. Именно поэтому многие авторы считают, что разработать методику формализации произвольного текста при достаточно высоком уровне сохранения его содержания вряд ли возможно [24]. По-видимому, на данном этапе развития науки — это так.

Однако существуют тексты, предназначенные для выполнения координационной (регулятивной) функции для обеспечения определенной деятельности людей, что обуславливает их специфическое построение и, как следствие, открывает достаточно возможности для их формализации с сохранением исходной семантики.

Вместе с тем замечено, что любой текст в той или иной мере обладает свойством регулятивности, под которым понимается способность текста управлять познавательной и интерпретационной деятельностью читателя [25]. Оказалось, что этот феномен регулятивности, благодаря целенаправленной организации текста, позволяет более эффективно реализовывать его авторскую интенцию. Поскольку намерения автора всегда направлены на адекватное восприятие его текста читателем.

Сопоставление различных типов текстов позволяет выявить образцы, отличающиеся друг от друга примененными в них неодинаковыми регулятивными стратегиями и спектрами используемых регулятивных средств (лексических, морфологических, синтаксических, композиционных, логических и др.).

Наибольшую активность в разработке теории и приложений регулятивности текстов проявляют лингвисты. Так, реализована весьма успешная попытка построения эмпирической теории регулятивности художественного текста [26]. В рамках этой теории описываются механизмы воздействия текста на читателя, а также формирования в сознании адресата смысла текста, адекватного замыслу автора. Подчеркнуто, что в художественных текстах регулятивные функции выполняют лексические и композиционные регулятивные средства, причем проявляются они не жестко и часто обусловлены образной направленностью текста. Авторы этой монографии отмечают фактор смысловой избыточности этого типа, заключающийся в многократной актуализации одного и того же смысла, представляемого с помощью

разных лексических средств, правда, с некоторыми вариациями. Заметим, эти факты, как и некоторые другие специфические черты художественных текстов, препятствуют разрешению проблем их формализации.

Регулятивность текстов достаточно часто является объектом рассмотрения психологов, юристов, социологов и педагогов [27–29]. К сожалению, работ в сфере информатики, базирующихся на использовании регулятивных свойств текстов, нами не отмечено.

Вместе с тем существует множество текстов, которые обладают характерным набором явно выраженных регулятивов, способствующих разработке методов их формализации. К таковым относятся тексты, описывающие технологии, различные распорядительные документы (приказы, инструкции, распоряжения и т.п.), в значительной мере это и учебные тексты, особенно математические. От текстов других типов их отличают логичность и точность (отсутствие иносказаний), связность и цельность, ясность, понятность (за счет отсутствия эмоциональной окраски и образных средств), доступность (за счет использования специфической терминологии). Они априори ориентированы на отображение регулятивных действий, интерпретация которых должна ориентировать читателей на адекватную организацию и воспроизведение описываемых в них процессов. Для приобретения таких качеств эти тексты должны содержать все необходимые для их интерпретации слова (при необходимости дополненные нужными уточнениями) и иметь схемы построения предложений, способствующие их быстрому и однозначному пониманию. Кроме того, в этих текстах широко используются специальные регулятивы (регулятивные структуры), такие как предписания (надо), запреты (нельзя), разрешения и рекомендации (можно) [30]. Тексты, структура построения и содержание которых явно ориентированы на регулирование поведения людей, будем называть регулятивными. Именно в силу своей специфики такие тексты в наибольшей мере подходят для разработки методов их записи на формальном языке, обеспечивая при этом достаточно высокий уровень сохранения содержания текста. Почему?

Основу лингвистической структуры регулятивных текстов образует упорядоченная последовательность побудительных предложений. В них каждое предложение (или логически агрегированная их группа) описывает конкретное обособленное (частичное) действие, приводящее к определенному результату, которое

можно трактовать как описание элементарной технологии [31]. В этой связи их часто называют императивными [32].

Как правило, такое действие подготовлено предшествующими действиями (по контексту понятно, что его следует начинать только после окончания (или начала) всех предшествующих ему частичных действий). После описания (читай выполнения) данного частичного действия соответствующая локальная цель предполагается достигнутой и переходят к описанию следующих частичных действий, реализация которых возможно только после выполнения ряда определенных предваряющих их частичных действий. Фактически идет описание алгоритма действия читателя, что является еще одним проявлением регулятивности такого рода текстов.

Процессная сущность регулятивных текстов обуславливает целесообразность более внимательного взгляда на синтаксис составляющих их предложений с позиций структурной лингвистики, в основу которой положена вербоцентрическая теория структурного синтаксиса [33]. Согласно этой теории, единственным главным членом предложения является глагол (глагольный предикат), при котором указаны актанты – участники описываемого действия, обстоятельства и условия его реализации (сирконстанты), а также отношения к этому действию (через модальность). Большинство предложений наиболее характерных представителей регулятивных текстов построены именно так: имеется предикативное ядро, обычно представляемое в виде императивных глагольных словоформ (включить, проверить, повернуть и т.п.), непосредственно участвующие в соответствующем действии объекты и ряд уточняющих обстоятельств и дополнений.

В качестве примера приведем фрагмент описания технологии пошива накладных карманов [34]:

- 1) на изнаночной стороне, на расстоянии 3 см от среза припуска на обработку верхнего края кармана наметить линию перегиба верхнего края кармана;
- 2) накладывая клеевую прокладку клеевой стороной на изнаночную сторону припуска, располагая край по намеченной линии, продублировать припуск на обработку верхнего края кармана;
- 3) вместе с прокладкой, выполняя строчку с лицевой стороны детали, обметать срез припуска на обработку верхнего края кармана.

Позиционирование в рамках структурного синтаксиса глагола как единственного главного члена предложения устраняет ряд трудностей, возникающих при попытках построения формальных

моделей предложения. В частности, предложение регулятивного текста может быть представлено упорядоченной гроздью слов, где формирующим центром выступает глагольный предикат [35]. К нему через «входящие» связи сходятся все участвующие в данном действии объекты с указанием дополнений и обстоятельств, а также «исходящие» из центра связи к результатам этого действия (рис. 1). Это еще один результат лингвистики, который использован нами при решении поставленной задачи.

*Примечание.* В регулятивных текстах встречаются и так называемые описательные предложения. У них более простая, чем у предложений иного типа, структура. Обычно они представлены в форме предложения о состоянии, где в роли предиката выступают слова категории «состояния». При этом подразумевается наличие отношения, неявно выражаемое глаголами: быть, иметь, являться. Например, «расстояние от кромки припуска до пуговицы 3 см». В этой связи их формализация при помощи излагаемого ниже метода реализуется значительно проще, чем в случае предложений иного типа, а именно: указанные в тексте значения вводятся в соответствующие места онтологии, относящейся к описываемому объекту.

Все вышеизложенное позволяет сформулировать суть предлагаемого метода формализации регулятивных текстов: обеспечение передачи предметной сущности используемых слов и механизмов причинно-следственных связей между ними, отображаемых соответствующим конструктивным алгоритмом.

Опираясь на специфический характер построения регулятивных текстов и ряд идей структурного синтаксиса, можно предположить нижеследующий метод формализованного представления такого рода текстов.

По существу, регулятивный текст можно рассматривать как вербальную запись алгоритма действий, направленных на обеспечение перехода некоего объекта из заданного исходного фазового состояния  $S(X_0)$  в определенное конечное состояние  $S(Y_f)$ . Задача состоит в том, чтобы представить данный алгоритм в формализованном виде:

$$W: S(X_0) \rightarrow S(Y_f).$$

Этот переход осуществляется в результате реализации упорядоченной последовательности ряда конкретных действий, рассматриваемых в данном тексте как элементарные. В этой связи алгоритм построения формальной модели регулятивного текста предлагается представлять в виде конструктивного процесса упорядо-

чивания моделей отдельных действий. Тем самым решение поставленной задачи распадается на два этапа: построение моделей отдельных действий и конструирование из них соответствующего исходному тексту семантического графа.

**4. Модель предложения регулятивного текста.** Как правило, в регулятивных текстах каждое действие описывается одним предложением (условно назовем его содержательным), которое может дополняться и уточняться. Структурно такие части текста часто имеют нумерацию или выделены в виде абзацев.

Первый этап метода формализации регулятивного текста представляет собой автоматизированную процедуру построения модели одного содержательного предложения (описывающего одно элементарное действие как одноактное действие [36]). Для ее реализации по мере рассмотрения (чтения) исходного текста каждому очередному содержательному предложению приписывается свой номер  $q$  ( $q \in [1, n]$ ). Пусть описываемое в этом предложении действие  $d_q$  по переводу объекта рассмотрения из фазового состояния  $S(X_q)$  в  $S(Y_q)$  описывает оператор  $W_q$ :

$$W_q: S(X_q) \rightarrow S(Y_q). \quad (1)$$

Требуется сформировать формализованное представление компонентов, входящих в (1).

В рассматриваемом предложении выявляется описываемое действие ( $d_q$ ) и фиксируется обозначающий его глагол (глагольный предикат). Для него формируется соответствующий фрейм-экземпляр с именем  $FD(q)$ . В этот фрейм, имя которого соответствует выявленному действию, заносятся имена участвующих в данном действии акторов и сведения об описывающих его сирконстантах. По существу, он в конструктивной форме описывает реализацию операций, осуществляемых оператором  $W_q$ .

*Примечание.* Если при анализе данного предложения выявлены отношения с акторами и обстоятельства их реализации, не зафиксированные на предыдущих шагах, то они вводятся и в формируемый протофрейм данного глагола, где нарастающим итогом формируется структурный облик последнего.

Выявленные акторы данного действия разбиваются на два множества:  $X_q$  и  $Y_q$ . К первому (рис. 1) относятся объекты  $x_{iq} \in X_q$ , участвующие в осуществлении этого действия (назовем их входными), ко второму – результирующие объекты  $y_{qj} \in X_q$  (назовем их выходными).

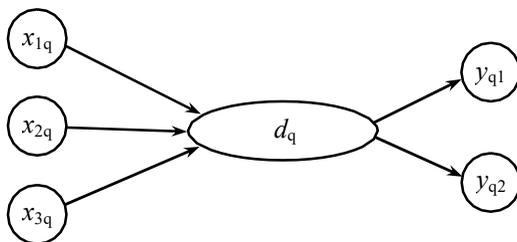


Рис. 1. Структура модели предложения регулятивного текста

Ряд авторов полагает, что текст семантически эквивалентен данному отображению, если совпадают их предикатные выражения [37]. Несомненно, в целом последовательность предикатных выражений обеспечивает возможность уяснить суть данного текста. Однако в регулятивных текстах весьма важны представленные в них описания условий, в которых описываемые действия протекают. В этой связи, помимо фиксации в формализованной форме сведений о собственно действиях, требуется отображать все сведения об участвующих в этих действиях объектах, причем по принципу «здесь и сейчас», и вот почему.

При изложении одних и тех же мыслей разные авторы используют разные слова, применяют синонимию, омонимию, иносказания, другие возможности живого языка. Эти обстоятельства обуславливают значительные трудности при формализации текстов. Сложностей добавляет использование знаков препинания и различных структур построения предложений, способных изменить их семантику. Отметим, что регулятивные тексты, как правило, используются в конкретных предметных областях, где присутствует соответствующая устоявшаяся лексика. Однако в излагаемом подходе предлагается решать вопросы совпадения смыслов рассматриваемых слов не столько по написанию имен сущностей, сколько по использованию интенционального метода построения совокупности формализующих их фреймов в виде родовидовой структуры [38].

Формируемая онтология понятий представляется в виде кортежа

$$Q \langle =C, M, R \rangle,$$

где  $C = \{c_j\}$  – множество понятий, образующих онтологию  $Q$ ,  $i = \overline{1, I}$ , то есть  $C \models I$ ;

$M_i = \{m_{i_1}, \dots, m_{i_{d_i}}\}$  – множество атрибутов понятия  $c_i$ ,  $d_i$  – количество атрибутов, описывающих данное понятие;

$R \subseteq C \times C$  – отношение непосредственного наследования.

Отношение  $R$  может быть задано матрицей размером  $I \times I$ . Если понятие  $c_k$  непосредственно наследует понятию  $c_i$ , то есть  $(c_j, c_k) \in R$ , то вводится элемент матрицы наследования  $r_{ik}$ , который принимает значение единица ( $r_{ik}=1$ ), в противном случае, если  $(c_i, c_k) \notin R$ , то  $r_{ik}=0$ .

Онтология в общем случае поддерживает множественное наследование понятий, то есть:

$$\exists c_i \exists c_k \exists c_l : ((r_{il}=1) \wedge (r_{kl}=1)).$$

Отсюда нахождение двух понятий в отношении непосредственного наследования означает, что между ними нет ни одного промежуточного понятия:

$$\exists c_i \exists c_k : ((c_i, c_k) \in R \Rightarrow \exists c_h (((c_i, c_h) \in R) \wedge ((c_h, c_k) \in R))).$$

Дочерние понятия наследуют все атрибуты, входящие в состав родительского понятия, расширяя его состав собственными атрибутами:

$$\forall c_i \forall c_k : (r_{jk}=1) \Rightarrow (M_i \subset M_k).$$

Если множества атрибутов двух понятий совпадает, то понятие  $c_i$  эквивалентно понятию  $c_j$ , то есть:

$$M_i = M_j \Rightarrow c_i \sim c_j.$$

Данный метод автоматизированного построения онтологии предметной области позволяет, в частности, автоматически определить отношение, в котором находится данная пара понятий.

Такой подход в значительной мере позволяет снизить актуальность предварительной нормализации исходного текста.

В этой связи описываемое в данном предложении состояние каждого  $i$ -го входного актора  $x_{iq}$  фиксируется в соответствующем фрейме-экземпляре этого объекта  $FX(x_{iq}) \equiv FX_{iq}$ , а для каждого  $j$ -го выходного актора  $y_{qj}$  во фрейме-экземпляре  $FY(x_{qj}) \equiv FY_{qj}$ . Подобные фреймы строятся для всех входных акторов данного действия, что обеспечивает формирование множества  $\{FX_{iq}\}$ , описывающего состояние рассматриваемого объекта  $S(X_q)$  к началу действия  $d_q$ . Аналогичным образом формируется множество  $\{FY_{qj}\}$ , описывающее состояние этого объекта на момент окончания действия  $S(Y_q)$ .

В рассматриваемом тексте данный объект, находясь в разных состояниях и отношениях с другими объектами и действиями, может упоминаться несколько раз. Следовательно, содержание фреймов-экземпляров для одного и того же объекта, упомянутого в разных предложениях, обычно не совпадает. Для обеспечения же возможностей по использованию модели текста, сведения о каждом объекте данной предметной области целесообразно накапливать, что осуществляется в соответствующих протофреймах.

Таким образом, моделью каждого содержательного (описывающего отдельное действие регулятивного текста) предложения выступает именованная тройка:

$$M_q = \langle FD(q), \{FX_{iq}\}, \{FY_{qj}\} \rangle, \quad (2)$$

где  $FD(q)$  – фрейм-экземпляр, описывающий концепт-действие данного предложения  $d_q$ ;  $\{FX_{iq}\}$  – множество фреймов-экземпляров, описывающих участвующих в данном действии акторов  $X_q = \{x_{iq}\}$ ;  $\{FY_{qj}\}$  – множество фреймов, описывающих множество результирующих акторов  $Y_q = \{y_{qj}\}$ . Применение регулятивного текста предполагает, что читателю известно исходное состояние описываемых в рассматриваемом тексте объектов  $X_0 = (x_{01}, x_{02}, \dots, x_{0r})$ , либо дается их описание. Такого рода сведения фиксируются в виде соответствующих фреймов-экземпляров  $\{FX_0\}$ .

*Примечание.* Основным отличительным моментом модели (2) является то, что ее ядром является выступающий в роли оператора  $W_q$  фрейм, описывающий отображение отдельного элементарного перехода рассматриваемого объекта из одной точки его фазового

состояния  $S(X_q)$  в другую  $S(Y_q)$ . При этом, поскольку каждое содержательное действие исходного текста рассматривается как элементарное, построенная описанным способом модель адекватна содержанию, передаваемому соответствующим предложением исходного текста (фиксируется только то, что непосредственно представлено в тексте, так как задача заключается в сохранении смысла именно данного текста).

Все сформированные таким образом модели вида (2) образуют множество вершин описывающего текст семантического графа  $M = \{M_q\}$ .

**5. Формализация поиска связей между действиями.** Целью следующего этапа на пути построения математической модели регулятивного текста является установление на множестве  $\{M_q\}$  формальных связей, отражающих описанные в тексте отношения между действиями. Ключевым моментом для достижения этой цели является следующее обстоятельство: в регулятивном тексте выходные объекты каждого действия (кроме так называемых заключительных, то есть  $\{Y_{\beta}\}$ ) выступают в качестве входных для других, последующих действий. Тем самым, крайне важно подчеркнуть, что эти связи неявно уже присутствуют во множествах  $\{FX_{iq}\}$  и  $\{FY_{qj}\}$ . Задача заключается в том, чтобы выявить связи непосредственного следования и зафиксировать их в явном виде. С этой целью на данных множествах необходимо сформировать систему переходов Келлера, иными словами, построить бинарную матрицу, отображающую искомые отношения в явном виде.

Непосредственное предшествование действия  $d_k$  ( $k \in [1, n]$ ) действию  $d_q$  означает, что среди выходных объектов действия  $d_k$  существует аналогичный входному объекту действия  $d_q$ . Если исходный текст с точки зрения описываемых процессов корректен, такого рода отношения должны находиться во множестве  $\{M_q\}$ . Иначе говоря, для непосредственно следующих друг за другом действий  $d_k$  и  $d_q$  во множестве  $\{M_q\}$  найдется пара фреймов  $FY_{ki}, FX_{qj} \in \{M_q\}$ , описывающих один и тот же объект, причем находящийся в одинаковых фазовых состояниях. Методы автоматического сравнения фреймов представлены в работе [39]. Найденная на этой основе пара фреймов

(результат одного действия и актора другого действия) находится в отношении эквивалентности  $FY_{ki} \leftrightarrow FX_{qj}$  (см. рис. 2).

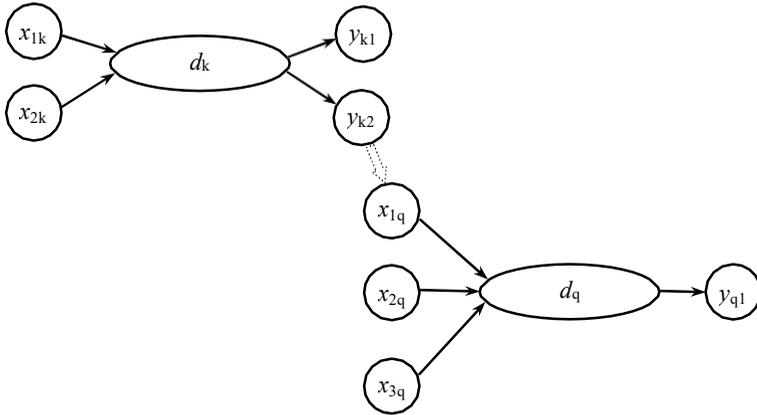


Рис. 2. Установление связей непосредственного следования

*Примечание.* Для удобства выполнения процедуры проверки условия эквивалентности фазовых состояний данного объекта структуры всех его фрейм-экземпляров должны быть одинаковыми. Носителями этой структуры и призваны быть соответствующие фрейм-прототипы упомянутых в тексте объектов.

Построенное множество фреймовых моделей типа (2) позволяет формальным образом (следовательно, автоматически) сформировать бинарную квадратную матрицу порядка  $n \times n$  непосредственно предшествующих действий  $GD = \parallel g_{k,q} \parallel$ . Для этих целей необходимо, рассматривая все возможные пары действий, установить те из них, которые имеют эквивалентные пары акторов и зафиксировать эти факты:

$$\forall q, k, \exists(ki, qj) : FY_{ki} \leftrightarrow FX_{qj} \rightarrow g_{kq} = 1. \quad (3)$$

Поскольку каждое действие может иметь несколько выходных  $\{y_{ki}\}$  и входных объектов  $\{x_{qj}\}$ , требуется зафиксировать именно те из них, для которых выполнено условие (3). Эту фиксацию удобно осуществлять непосредственно в модифицированной матрице  $GD M$ ,

в которую вместо символа «1» вводить, например, вектор, описывающий пары номеров фреймов, описывающих эквивалентные понятия:

$$\forall q, k, \exists(ki, qj) : FY_{ki} \leftrightarrow FX_{qj} \rightarrow g_{kq} := \{(k; i), (q; j)\}.$$

Построенную модель рассматриваемого текста:

$$T = \langle M, GDM \rangle,$$

где  $GDM = \|(k; i), (q; j)\|$ , можно интерпретировать как ориентированный граф (семантическую сеть). Вершинами построенного графа выступают фреймы, описывающие отдельные действия, и участвующие в них акторы. В этом графе присутствуют три вида дуг: дуги, связывающие акторов с действием, дуги, связывающие действие с его результатом и дуги, связывающие результаты и акторы, являющиеся эквивалентными. Построенный граф обладает одной особенностью, он состоит из ярко выраженных подграфов, каждый из которых отображает отдельные содержательные предложения исходного текста (см. рис. 2).

В качестве примера рассмотрим следующий фрагмент текста из учебника по математике, где приводится описание решения линейных матричных уравнений. Пусть задана система линейных алгебраических уравнений в матричной форме  $A \times X = B$ . Если матрица коэффициентов системы  $A$  обратима, то, используя свойство ассоциативности умножения матриц и свойства единичной матрицы, можно показать, что решение имеет вид  $X = A^{-1}B$ . Для этого необходимо умножить левую и правую части уравнения слева на обратную матрицу  $A^{-1}$ , а затем применить свойства единичной матрицы:  $A^{-1} \cdot A = E$  и  $E \cdot A = A$ .

Описанный выше подход формализованного представления такого рода текстов позволяет формализовать это описание следующим образом.

Пусть в ходе формализации предыдущих параграфов данного учебника были зафиксированы входящие в условия задачи свойства матриц, представленные на рисунке 3 в виде последовательности стилизованных равенств.

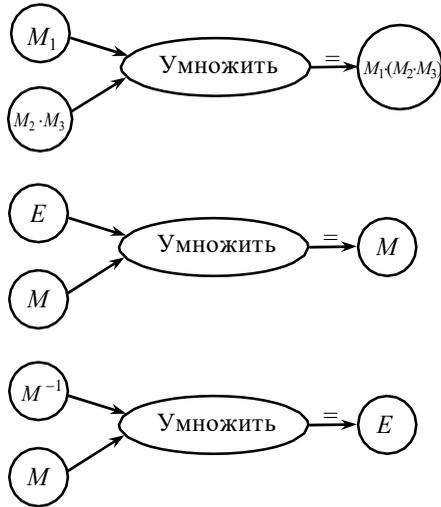


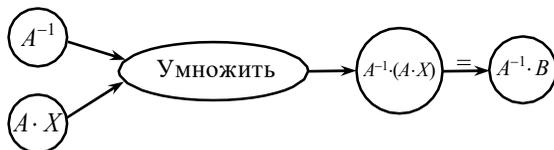
Рис. 3. Структурное представление использованных свойств

Сформируем фреймы акторов  $A, X, B$  и представим исходное уравнение  $A \cdot X = B$  в виде следующей структуры (см. рис. 4).



Рис. 4. Структура исходного уравнения

Отобразим факт умножения левой и правой части исходного уравнения слева на обратную матрицу  $A^{-1}$  в виде структуры, представленной на рисунке 5.


 Рис. 5. Структура уравнения после его умножения слева на  $A^{-1}$ 

Расположенный в левой части уравнения результат есть уже представленная по условию структура, описывающая свойство ассоциативности умножения матриц  $[M_1 \cdot (M_2 \cdot M_3) \equiv (M_1 \cdot M_2) \cdot M_3]$ . После нахождения совпадений соответствующих структур, алгоритм, как показано на рисунке 6, меняет левую часть уравнения.

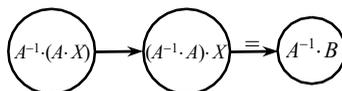


Рис. 6. Использование свойства ассоциативности умножения матриц

Поскольку искомый результат не получен, алгоритм продолжает поиск структур, аналогичных тем, что стоят теперь в левой и правой части уравнения. Структура левой части может активизировать имеющееся в условии основное свойство (часто рассматриваемое как определение) единичной матрицы  $M^{-1} \times M = E$ . Его применение приводит к новому виду уравнения, представленного на рисунке 7.

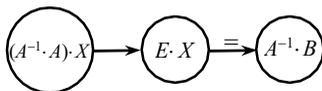


Рис. 7. Использование определения единичной матрицы

Полученная в левой части уравнения структура аналогична той, что отвечает другому свойству единичной матрицы  $(E \cdot M = M)$  – свойству умножения единичной матрицы на матрицу. Активация этого свойства, как показано на рисунке 8, изменяет левую часть уравнения, приводя к виду, полностью аналогичному искомому результату.

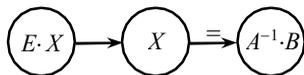


Рис. 8. Использование свойства умножения единичной матрицы на матрицу

Объединение всех описанных шагов позволяет получить формализованную структуру, которая отражает содержание изложенного в тексте фрагмента учебника.

*Примечание.* Анализ изложения описанного фрагмента свидетельствует, что в разных учебниках при решении описанной задачи не акцентируется внимание на том, какие конкретно свойства матриц используются. Чаще всего опускается явное указание на необходимость использования свойства ассоциативности умножения матриц. Авторы текстов считают это необязательным, надеясь на память читателя. Это обстоятельство приводит к невозможности построения формального представления данного текста изложенным методом, что свидетельствует об изъянах рассматриваемого текста.

Нужно понимать, что на практике рассматриваемые тексты всегда конечны. Следовательно, они содержат слова, которые именуют понятия, предполагаемые априорно общеизвестными и в достаточной мере одинаково трактуемые читателями. Иначе говоря, любой конкретный текст формально не является замкнутым относительно использованных в них понятий.

В этой связи, во-первых, всегда следует понимать, что в определенной предметной области присутствует достаточно устоявшаяся лексика, не требующая. Во-вторых, каждый регулятивный текст предполагает, что читатель имеет определенный уровень подготовки (образования). Таким образом, для использования изложенного метода исходное состояние описываемых в рассматриваемом тексте объектов читателю должно быть известно и зафиксировано в виде упомянутых выше соответствующих фреймов-экземпляров  $\{FX_0\}$ .

**6. Заключение.** Регулятивный текст, по существу, представляет собой словесно выраженный алгоритм последовательности целенаправленных действий, что открывает возможность построения математической модели, адекватно отображающей семантику этого текста.

Процедура построения такой модели осуществляется в два этапа. Вначале в автоматизированном режиме производится формирование и

наполнение слотов фреймов, адекватно отображающих описанные в тексте действия, и участвующих в них акторов. При этом применяется интенциональный способ построения совокупности формализующих их фреймов в виде родовидовой структуры, что в значительной мере позволяет снизить актуальность предварительной нормализации исходного текста. Затем в автоматическом режиме формируется соответствующий тексту семантический граф.

Важным преимуществом представленного метода является возможность существенного сокращения количества актов принятия творческих решений при построении формальной модели рассматриваемого текста. По существу, субъект, работающий над формализацией текста, выполняет в основном рутинные действия: выделяя из текста отдельные слова (устоявшиеся фразы) он разносит их по соответствующим фреймам, если необходимо, вводя в них при этом новые пары слотов. Тем самым результат построения формализованного представления исходного текста становится слабо зависящим от личностных характеристик исполнителя.

Рассмотренный подход обеспечивает возможность разработки программных комплексов для построения баз знаний отдельных предметных областей, предприятий и организаций, активное использование которых обеспечит возможность эффективного решения расширенного списка задач. Поясним сказанное на следующем примере.

На ряде производств своевременное обновление ассортимента выпускаемой продукции является одним из главных залогов их успешной деятельности. Так, например, владимирская обувная фабрика в год обновляет до 80 наименований своей продукции. При этом полный цикл смены одного наименования (включая маркетинговые исследования, работу конструкторов и технологов, отладку собственно производства, элементов снабжения и сбыта) занимает до полутора лет. В этой связи формирование системы, обеспечивающей поддержку процессов обновления ассортимента – весьма актуальная трудоемкая задача.

Для каждой выпускаемой модели технология ее производства на данном предприятии, описание используемых материалов и оборудования зафиксированы в соответствующих информационных базах в виде текстов. Поскольку применяемые технологические приемы, свойства материалов и подготовка новых моделей к выпуску занимают существенно различное время, а спектр возможных решений весьма широк, предприятие заинтересовано в создании автоматизированных средств поддержки процессов разработки

технологий производства новых моделей, которые в сложившихся условиях были бы технологически и ресурсно возможны и отвечали бы интересам предприятия. По существу, речь идет о создании современного инструментария решения крайне сложной задачи синтеза производственных процессов в условиях функционирования данного предприятия.

Эффективное решение этой задачи можно обеспечить при возможности целенаправленно, в автоматическом режиме оперировать представленными в формализованном виде описаниями всех возможных технологий изготовления продукции. Подготовку этих формализмов с учетом актуальных для предприятия ограничений и допущений обеспечивает изложенный метод.

Соответствующие алгоритмы синтеза новых технологий удобно формировать в автоматическом режиме, комбинируя имеющиеся фрагменты существующих технологий. Для этого целесообразно наличие специальным образом сформированных библиотек, компоненты которых позволяют формировать изложенный подход.

В итоге рассмотренный подход обеспечивает возможность разработки программных комплексов для построения библиотек регулятивных текстов отдельных предметных областей, разработки средств оценки такого рода текстов на предмет их определенности, полноты, связности и других характеристик, а также тренажеров и средств, используемых при самообучении.

## Литература

1. *Пеикова Н.П.* Семантика и смысл текста: >, <, =, #? (экспериментальный подход к теоретическим проблемам // Вестник Челябинского государственного университета. 2015. № 15 (370). Филология. Искусствоведение. Вып. 96. С. 69–77.
2. *Арапова Г.У.* Концепт, понятие и значение слова // Международный журнал прикладных и фундаментальных исследований. 2016. № 1. С. 591–593.
3. *Орехов В.Д.* Прогностика: от прошлого к будущему человечества // М., Сканбук, E-Books. 2016. 210 с.
4. *Барнинова И.А.* К вопросу о соотношении смысла и содержания текста при переводе // Вестник Вятского государственного университета. 2015. № 10. С. 69–72.
5. *Lochter Johannes V., Pires Pedro R., Bossolani Carlos, Yamakami Akebo, Almeida Tiago A.* Evaluating the impact of corpora used to train distributed text representation models for noisy and short texts // International Joint Conference on Neural Networks (IJCNN). 2018. pp. 315–322.
6. *Двойникова А.А., Карнов А.А.* Аналитический обзор подходов к распознаванию тональности русскоязычных текстовых данных // Информационно-управляющие системы. 2020. № 4(107). С. 20–30.
7. *Ali Ismael, Al Madi Naser, Melton Austin.* Using Text Comprehension Model for Learning Concepts, Context, and Topic of Web Content // IEEE 11th International Conference on Semantic Computing (ICSC). 2017. pp. 101–104.

8. *Li Jie, Chen Siming, Chen Wei, Andrienko Gennady, Andrienko Natalia.* Semantics- Space-Time Cube: A Conceptual Framework for Systematic Analysis of Texts in Space and Time // IEEE Transactions on Visualization and Computer Graphics. 2020. vol. 26. iss. 4. pp. 1789–1806.
9. *Sheridan Tom.* Models of Human-System Interaction [Book Review] // IEEE Systems, Man, and Cybernetics Magazine. 2017. vol. 3. iss. 2. pp. 56–C3.
10. *Митрофанова О.И., Романенко О.Н.* Работа с научным текстом на занятиях по русскому языку с иностранными военнослужащими // Научный альманах. 2016. № (22). С. 473–478.
11. *Levinkov Evgeny, Uhrig Jonas, Tang Siyu, Omran Mohamed, Insafutdinov Eldar, Kirillov Alexander etc.* Joint Graph Decomposition & Node Labeling: Problem, Algorithms, Applications // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. pp. 1904–1912.
12. *Xu LiHong, Sun ShuTao, Wang Qi.* Text similarity algorithm based on semantic vector space model // Proceedings of IEEE ACIS 15th International Conference on Computer and Information Science (ICIS). 2016. pp. 1193–1196.
13. *Бондарчук Д.В.* Алгоритмы интеллектуального поиска на основе метода категориальных векторов: дис. к-та физ.-мат. наук. Екатеринбург: УрГУПС. 2016. 141 с.
14. *Gaspers Judith, Cimiano Philipp, Rohlfing Katharina, Wrede Britta.* Constructing a Language from Scratch: Combining Bottom-Up and Top-Down Learning Processes in a Computational Model of Language Acquisition // Proceedings of IEEE Transactions on Cognitive and Developmental Systems. 2017. vol. 9. iss. 2. pp. 183–196.
15. Когнитивные аспекты лексикографии. Список публикаций В. З. Демьянкова. Статьи и монографии. URL: [www.infolex.ru/Dem\\_Publ.html](http://www.infolex.ru/Dem_Publ.html) (дата обращения: 25. 07. 2020).
16. *Демьянков В.З.* Трансфер знания и перевод // Язык. Культура. Перевод. Коммуникация: сб. научн. тр. М.: КДУ, Университетская книга. 2018. Вып. 2. С. 18–21.
17. *Pavlovskiy I.S.* The semantic models of large terminological texts // Tenth International Conference Management of Large-Scale System Development (MLSD). 2017. pp. 1–5.
18. *Kang Sungku, Patil Lalit, Rangarajan Arvind, Moitra Abha, Jia Tao, Robinson Dean, Ameri Farhad, Dutta Debasish.* Extraction of Formal Manufacturing Rules from Unstructured English Text // Computer-Aided Design. 2021. vol. 130: 102990.
19. *Fernández-Duque David, Nepomuceno-Fernández Ángel, Sarrión-Morrillo Enrique, Soler-Toscan Fernando, Velázquez-Quesada Fernando R.* Forgetting complex propositions // Logic Journal of the IGPL. 2015. vol. 23. iss. 6. pp. 942–965.
20. *Sauerland Uli, Yatsushiro Kazuko.* Remind-Me Presuppositions and Speech-Act Decomposition: Evidence from Particles in Questions // Linguistic Inquiry. 2017. vol. 48. iss. 4. pp. 651–678.
21. *Кулешов С. В., Зайцева А.А., Марков В.С.* Ассоциативно-онтологический подход к обработке текстов на естественном языке // Интеллектуальные технологии на транспорте. 2015. № 4. С. 40–45.
22. *Зацман И.М.* Семантическая модель взаимосвязей концептов, информационных объектов и компьютерных кодов // Информатика и ее применение. 2009. Том 3 №2. С. 65–81.
23. *Мареев С.Н.* Диалектика содержания и формы и проблемы формализации // Научные ведомости Белгородского государственного университета. Серия «Философия. Социология. Право». 2017. №42(24). С. 15–26.
24. *Walter Benedikt, Hammes Jakob, Piechotta Marco, Rudolph Stephan.* A Formalization Method to Process Structured Natural Language to Logic Expressions to Detect Redundant

- Specification and Test Statements // IEEE 25th International Requirements Engineering Conference (RE). 2017. pp. 263–272.
25. *Болотнова Н.С.* О типологии регулятивных структур в тексте как форме коммуникации // Вестник Томского государственного педагогического университета. 2011. Вып. 3 (105). С. 34–40.
  26. *Болотнова Н.С., Бабенко И.И., Бакланова Е.А.* и др. Коммуникативная стилистика текста: лексическая регулятивность в текстовой деятельности // Томск: ТГПУ. 2011. 492 с.
  27. *Троуцук И.В.* Анализ текстовых данных в социологии: основания систематизации концептуальных моделей, методологических принципов и методических решений: дис. д-ра социол. наук. Москва: РУДН. 2014. 338 с.
  28. *Мясникова О.В., Фроловская М.Н.* Педагогические условия понимания текстов в процессе обучения // Известия АлтГУ. 2014. Вып. 2 (82). Том 2. С. 43–48.
  29. *Дьякова А.А.* Регулятивность текстов законов, адаптированных к условиям медийного дискурса // Грани познания. 2015. №1. С. 95–100.
  30. *Кармин А.С.* Культурология. Краткий курс // СПб.: Питер. 2009. 240 с.
  31. *Krupp Lars, Gruenerbl Agnes, Bahle Gernot, Lukowicz Paul.* Towards Automatic Semantic Models by Extraction of Relevant Information from Online Text // IEEE Inter- national Conference on Smart Computing (SMARTCOMP). 2019. pp. 481–483.
  32. *Zimmer Melanie, Al-Yacoub Ali, Ferreira Pedro, Lohse Niels.* Understanding Human Decision-making during Production Ramp-up using Natural Language Processing // IEEE 17th International Conference on Industrial Informatics (INDIN). 2017. pp. 337 – 342.
  33. *Zakharov Victor, Krassovitskiy Alexander, Meirambekkyzy Zhazirakhanym, Ualiyeva Irina, Khoroshilov Aleksander, Khoroshilov Alexey.* Automatic Creation Technologies of Declarative Tools for Clustering Media Documents // International Conference on Engineering Technologies and Computer Science (EnT). 2019. pp. 39–42.
  34. *Рослякова Т.А.* Школа шитья // Ростов-на-Дону: Феникс. 2000. 445 с.
  35. *Sithole Vusi, Marshal Linda.* Building Concept Hierarchies for the Internet of Things Patterns Using Domain-specific Dependency Knowledge // Open Innovations (OI). 2019. pp. 1–7.
  36. *Воронов М.В.* Модель технологического действия // Вестник Санкт-Петербургского государственного университета технологии и дизайна. Серия 1. Естественные и технические науки. 2016. № 1. С. 47–52.
  37. *Богатырев М.Ю., Коржук Н.Л.* Исследование семантической инвариантности концептуальных моделей текстовых данных // Известия ТулГУ. Технические науки. 2019. Вып. 12. С. 499–503.
  38. *Антонов И.В., Воронов М.В.* Методы анализа данных в задачах автоматизации построения онтологии предметной области // Дистанционное и виртуальное обучение. 2011. № 8 (50). С. 19–35.
  39. *Воронов М.В.* Система активной поддержки самоподготовки студентов // Образовательные технологии. 2018. № 3. С. 107–117.

**Воронов Михаил Владимирович** – д-р техн. наук, профессор, заведующий кафедрой, кафедра прикладной математики, Московский государственный психолого-педагогический университет. Область научных интересов: теория систем и системный анализ, управление организационными системами. Число научных публикаций – 400. mivoronov@yandex.ru, mgppu.ru/project/156; Открытое ш., д. 24, стр. 27, г. Москва, 107143, РФ; р.т.: +7(499)1676674.

**Пименов Виктор Игоревич** – д-р техн. наук, старший научный сотрудник, заведующий кафедрой, кафедра информационных технологий, Санкт-Петербургский государственный университет промышленных технологий и дизайна. Область научных интересов: распознавание образов, интеллектуальный анализ данных, когнитивные технологии. Число научных публикаций – 200. v\_pim@mail.ru, sutd.ru/universitet/staff/person/0\_0\_2398; Вознесенский пр., д. 46, СПбГУПТД, каф. ИТ, каб. 475, г. Санкт-Петербург, 190068, РФ; р.т.: +7(812)3122567.

M. Voronov, V. Pimenov  
**FORMALIZATION OF REGULATORY TEXTS**

*Voronov M., Pimenov V. Formalization of Regulatory Texts.*

**Abstract.** Modern information technologies provide text manipulation processes with high efficiency. First of all, this means storing, editing, and formatting texts and their components. Having achieved significant success in developing tools for content-free computer text processing, researchers faced problems with their content processing. Therefore, further steps in this direction are associated with the creation, among other things, of methods for automated purposeful manipulation of texts, taking into account their content. The analysis of works devoted to the study of the problems of formal presentation of texts and their subsequent use is carried out. Despite a number of successful projects, the challenges of solving the problem of the relationship between the content of the text and its meaning remain relevant. It seems that formalization of a General-purpose text while preserving its semantics is not feasible at this stage. However, there are types of texts that can be formalized while preserving their semantics. One of them is a regulatory text type, which is essentially a verbally expressed algorithm for a sequence of targeted actions. It is distinguished by logic and accuracy (lack of allegories), coherence and integrity, clarity, understandability (due to the lack of emotional coloring and figurative means), accessibility (due to the use of specific terminology). In other words, when developing regulatory texts, they usually try to display the mechanisms of the described actions as clearly as possible. Purpose: development of a method for formalizing a regulatory text while preserving its semantics. Methods: structural linguistics, representation of objects in the form of an ontology, constructive algorithms. The use of this method is demonstrated by describing the solution of a system of algebraic equations. Results: method for constructing a mathematical model of a regulatory text. Practical relevance: the application of the developed method makes it possible to develop software systems for building libraries of individual subject areas, develop tools for evaluating regulatory texts for their certainty, completeness, connectivity and other characteristics, as well as simulators and self-learning tools.

**Keywords:** Regulatory Text, Artificial Intelligence, Mathematical Model, Constructive Algorithm, Ontology, Frame

**Voronov Mihail** – Dr. Sci., Tech., Professor, Head of the Department, Department of Applied Mathematics, Moscow State University of Psychology & Education. Research interests: systems theory and system analysis, management of organizational systems. The number of publications – 400. mivoronov@yandex.ru, mgppu.ru/project/156; 24, Otkrytoye Hwy., Moscow, 107143, Russia; office phone +7(499)1676674.

**Pimenov Viktor** – Dr. Sci., Tech., Senior Researcher, Head of the Department, Department of Information Technologies, Saint-Petersburg State University of Industrial Technologies, and Design. Research interests: image recognition, data mining, cognitive technologies. The number of publications – 200. v\_pim@mail.ru, sutd.ru/universitet/staff/person/0\_0\_2398; 46, B. Morskaya St., St. Petersburg, 190068, Russia; office phone +7(812) 3122567.

### References

1. Peshkova N.P. [Text semantics and text sense: >, <, =, #? (experimental approach to theoretical problems)]. *Vestnik Cheljabinskogo gosudarstvennogo universiteta – Bulletin of Chebysinsk State University*. 2015. no. 15 (370). Philology. Arts. iss. 96. pp. 69–77. (In Russ.).

2. Arapova G.U. [Concept, comprehension and meaning of word]. *Mezhdunarodnyj zhurnal prikladnyh i fundamental'nyh issledovanij – International Journal of Applied and Fundamental Research*. 2016. no. 1. pp. 591–593. (In Russ.).
3. Orehov V.D. *Prognostika: ot proshlogo k budushchemu chelovechestva* [Prognostics: from the past to the future of humanity]. Moscow, Skanbuk, E-Books, 2016. 210 p. (In Russ.).
4. Barinova I.A. [On relation between sense and content in text translating]. *Vestnik Vyatskogo gosudarstvennogo gumanitarnogo universiteta – Bulletin of Vyatka state University*. 2015. no. 10. pp. 69–72. (In Russ.).
5. Lochter Johannes V., Pires Pedro R., Bossolani Carlos, Yamakami Akebo, Almeida Tiago A. Evaluating the impact of corpora used to train distributed text representation models for noisy and short texts. International Joint Conference on Neural Networks (IJCNN). 2018. pp. 315–322.
6. Dvojnjkova A.A., Karpov F.F. [Analytical review of approaches to Russian text sentiment recognition]. *Informacionno-upravliaiushchie sistemy – Information and Control Systems*. 2020. no. 4. pp. 20–30. (In Russ.).
7. Ali Ismael, Al Madi Naser, Melton Austin. Using Text Comprehension Model for Learning Concepts, Context, and Topic of Web Content. IEEE 11th International Conference on Semantic Computing (ICSC). 2017. pp. 101–104.
8. Li Jie, Chen Siming, Chen Wei, Andrienko Gennady, Andrienko Natalia. Semantics-Space-Time Cube: A Conceptual Framework for Systematic Analysis of Texts in Space and Time. *IEEE Transactions on Visualization and Computer Graphics*. 2020. vol. 26. iss. 4. pp. 1789–1806.
9. *Sheridan Tom*. Models of Human-System Interaction [Book Review]. *IEEE Systems, Man, and Cybernetics Magazine*. 2017. vol. 3. iss. 2. pp. 56–C3.
10. Mitrofanova O.I., Romanenko O.N. [Working with academic text in the Russian language classes with the foreign soldiers]. *Nauchnyj al'manah – Science Almanac*. 2016. no. 8–1(22). pp. 473–478. (In Russ.).
11. Levinkov Evgeny, Uhrig Jonas, Tang Siyu, Omran Mohamed, Insaftudinov Eldar, Kirillov Alexander etc. Joint Graph Decomposition & Node Labeling: Problem, Algorithms, Applications. IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. pp. 1904–1912.
12. Xu LiHong, Sun ShuTao, Wang Qi. Text similarity algorithm based on semantic vector space model. IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS). 2016. pp. 1193–1196.
13. Bondarchuk D.V. *Algoritmy intellektual'nogo poiska na osnove metoda kategorial'nyh vektorov. Dis. kand. fiz.-mat. nauk* [Intelligent search algorithms based on the categorical vector method. PhD. phys. and math. sci. diss.]. Yekaterinburg, UrGUPS Publ. 2016. 141 p. (In Russ.).
14. Gaspers Judith, Cimiano Philipp, Rohlfing Katharina, Wrede Britta. Constructing a Language from Scratch: Combining Bottom–Up and Top–Down Learning Processes in a Computational Model of Language Acquisition. *IEEE Transactions on Cognitive and Developmental Systems*. 2017. vol. 9. iss. 2. pp. 183–196.
15. Kognitivnye aspekty leksikografii. Spisok publikacij V. Z. Dem'yankova. Stat'i i monografii [Cognitive aspects of lexicography. List of publications by V. Z. Demyankov. Articles and monographs]. Available at: [www.infolex.ru/Dem\\_Publ.html](http://www.infolex.ru/Dem_Publ.html) (accessed 25.07.2020). (In Russ.).
16. Dem'jankov V.Z. [Knowledge transfer and translation]. *Yazyk. Kul'tura. Perevod. Kommunikaciya: Sb. nauchn. tr.* [Language. Culture. Translation. Communication: Collected papers]. Moscow: KDU, Universitetskaya kniga Publ, 2018. iss. 2. pp. 18–21. (In Russ.).

17. Pavlovskiy I.S. The semantic models of large terminological texts. Tenth International Conference Management of Large-Scale System Development (MLSD). 2017.pp.1–5.
18. Kang Sungku, Patil Lalit, Rangarajan Arvind, Moitra Abha, Jia Tao, Robinson Dean, Ameri Farhad, Dutta Debasish. Extraction of Formal Manufacturing Rules from Unstructured English Text. *Computer-Aided Design*. 2021. vol. 130: 102990.
19. Fernández–Duque David, Nepomuceno–Fernández Ángel, Sarrión–Morrillo Enrique, Soler–Toscan Fernando, Velázquez–Quesada Fernando R. Forgetting complex propositions. *Logic Journal of the IGPL*. 2015. vol. 23. iss. 6. pp. 942–965.
20. Sauerland Uli, Yatsushiro Kazuko. Remind-Me Presuppositions and Speech-Act Decomposition: Evidence from Particles in Questions. *Linguistic Inquiry*. 2017. vol. 48. iss. 4. pp. 651–678.
21. Kuleshov C.V., Zajceva A.A., Markov V.C. [Associative-ontological approach to natural language texts processing]. *Intellectual'nye tehnologii na transporte – Intellectual Technologies on Transport*. 2015 no. 4. pp. 40–45. (In Russ.).
22. Zacman I.M. [Semiotic model of interrelations of concepts, information objects and computer codes]. *Informatika i ee primenenie – Informatics and its application*. 2009. vol. 3. no. 2. pp. 65–81. (In Russ.).
23. Mareev S.N. [The dialectic of content and form, and the problem of formalization]. *Nauchnye vedomosti Belgorodskogo gosudarstvennogo universiteta. Seriya «Filosofija. Sociologi. Pravo» – Belgorod State University Scientific bulletin. Series' philosophy. Sociologists. Right*. 2017. no. 42 (24). pp. 15–26. (In Russ.).
24. Walter Benedikt, Hammes Jakob, Piechotta Marco, Rudolph Stephan. A Formalization Method to Process Structured Natural Language to Logic Expressions to Detect Redundant Specification and Test Statements. IEEE 25th International Requirements Engineering Conference (RE). 2017. pp. 263–272.
25. Bolotnova N.S. [Typology of regulative structures in text as a form of communication]. O tipologii regulativnyh struktur v tekste kak forme kommunikacii. *Vestnik Tomskogo gosudarstvennogo pedagogicheskogo universiteta – Tomsk state pedagogical university bulletin*. 2011. iss. 3 (105). pp. 34–40. (In Russ.).
26. Bolotnova N.S., Babenko I.I., Baklanova E.A. i dr. *Kommunikativnaja stilistika teksta: leksicheskaja reguljativnost' v tekstovoj dejatel'nosti* [Communicative stylistics of the text: lexical regularity in text activity] Tomsk: TSPU, 2011. 492 p. (In Russ.).
27. Trocuk I.V. *Analiz tekstovyh dannyh v sociologii: osnovanija sistematizacii konceptual'nyh modelej, metodologicheskikh principov i metodicheskikh reshenij. Dis. dokt. sociol. nauk* [Text data analysis in sociology: the basis for systematization of conceptual models, methodological principles and methodological solutions]. Moscow: RUDN, 2014. 338 p. (In Russ.).
28. Mjasnikova O.V., Frolovskaja M.N. [Pedagogical Conditions of Text Understanding in the Process of Studying]. *Izvestiya of Altai State University*. 2014. iss. 2 (82). vol. 2. pp. 43–48. (In Russ.).
29. D'jakova A.A. [Regulativeness of law texts adapted to media discourse conditions]. *Grani poznaniya – Facets of knowledge*. 2015. №1. pp. 95–100. (In Russ.).
30. Karmin A.S. *Kul'turologija. Kratkij kurs* [Cultural studies. Short course]. Saint Petersburg: Piter, 2009. 240 p. (In Russ.).
31. Krupp Lars, Gruenerbl Agnes, Bahle Gernot, Lukowicz Paul. Towards Automatic Semantic Models by Extraction of Relevant Information from Online Text. IEEE International Conference on Smart Computing (SMARTCOMP). 2019. pp. 481–483.
32. Zimmer Melanie, Al-Yacoub Ali, Ferreira Pedro, Lohse Niels. Understanding Human Decision-making during Production Ramp-up using Natural Language Processing. IEEE 17th International Conference on Industrial Informatics (INDIN). 2017.pp.337–342.
33. Zakharov Victor, Krassovitskiy Alexander, Meirambekkyzy Zhazirakhany, Ualiyeva Irina, Khoroshilov Aleksander, Khoroshilov Alexey. Automatic Creation Technologies

- of Declarative Tools for Clustering Media Documents. International Conference on Engineering Technologies and Computer Science (EnT). 2019. pp. 39–42.
34. Rosl'jakova T.A. *Shkola shit'ja* [School of sewing]. Rostov-on-Don: Feniks, 2000. 445 p. (In Russ.).
35. Sithole Vusi, Marshal Linda. Building Concept Hierarchies for the Internet of Things Patterns Using Domain-specific Dependency Knowledge. *Open Innovations (OI)*. 2019. pp. 1–7.
36. Voronov M.V. [Modeling of technological processes]. *Vestnik Sankt-Peterburgskogo gosudarstvennogo universiteta tekhnologii i dizajna. Seriya 1. Estestvennye i tekhnicheskie nauki – Bulletin of the Saint Petersburg state University of technology and design. Series 1. Natural and technical Sciences*. 2016. no. 1. pp. 47–52. (In Russ.).
37. Bogatyrev M.Y., Korzhuk N.L. [Study of semantic invariance of conceptual models of text data]. *Izvestiya TulGU. Tekhnicheskie nauki – News of TulSU. Technical science*. 2019. iss. 12. pp. 499–503. (In Russ.).
38. Antonov I.V., Voronov M.V. [Data analysis methods for automating the construction of domain ontology]. *Distancionnoe i virtual'noe obuchenie*. 2011. no. 8 (50). pp. 19–35. (In Russ.).
39. Voronov M.V. [Active support system for students self-training]. *Obrazovatel'nye tekhnologii – Educational technology*. 2018. no. 3. pp. 107–117. (In Russ.).