

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ
ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г.

№ 3

Москва 2018

ОБЩИЙ РАЗДЕЛ

УДК 001.102

А.В. Нестеров

Соотношение категорий отражения и отображения, а также их связи с категорией информации*

Тема соотношения категорий отражения и отображения, а также их связи с категорией информации, несмотря на достаточно длинную историю до сих пор не имеет общепризнанного решения. Актуальность ей добавил интерес к системам искусственного интеллекта, так как содержание напрямую связано с пониманием категорий понятия и образа. Приведена аргументация соотношения указанных категорий для действительных, знаковых и/или идейных элементов в природной, общественной и/или психической сфере, на основе продуцентного подхода.

Ключевые слова: отражение, отображение, содержание, форма, знак, идея, категории, понятие, образ, значение, значимость, смысл, сведения, информация, идеалства

ВВЕДЕНИЕ

Несмотря на более чем 60-летнюю историю обсуждения темы соотношения категорий отражения и отображения, единообразного понимания этих категорий нет.

Во многих публикациях утверждается, что существует идеальное отражение. В частности П.К. Анохин [1] считает, что в процессе эволюции организмов сформировалась идеальная форма отражения – сознание. К.К. Колин в публикации [2] декларирует, что с точки зрения информатики, можно выделить идеальную реальность трех родов как отражения. Необходимо отметить, что иногда слова «отражение» и «отображение» отождествляются [3]. Таким образом,

* Публикация подготовлена при поддержке Программы РУДН «5-100»

утверждается, что отражение как продукт имеет идеальный характер.

Обычно элемент, который отражается, называют отражаемое, а то, что его отражает, обозначают как отражающее. При этом часто их взаимодействие называют отражением. В настоящем тексте под словом «отражение» понимается отражающее, а способность отражать обозначена как отражаемость, т.е. способность элементов наблюдаемого мира продуцировать отражение.

КАТЕГОРИИ СЛЕДОВ В КРИМИНАЛИСТИКЕ

Категория отражения в естествознании отличается от ее философского понимания, и имеет материалистическое значение. Тем не менее, в криминалистике известно понятие идеальных следов.

В криминалистике выделяют материальные и идеальные следы (отражения) [4]. С естественнонаучной точки зрения идеальные следы в действительном мире отсутствуют. Все наблюдаемые (чувственно воспринимаемые) следы (действительные отображения и/или отражения) имеют материально-вещественные носители.

Однако, носители следов могут меняться, а их содержание, называемое информацией, может оставаться без изменения.

Поэтому важно понимание следа как продукта, продуцируемого производителем. При продуцировании на продукте остаются следы продуцирования и/или производителя. К таким следам продуцирования относятся следы канала связи (связи) и/или взаимодействия продуцирования. Кроме того, продукт как результат, процесс и/или окружение продуцирования может еще обладать свойствами, отображающими следы процесса и/или окружения продуцирования.

Поэтому в криминалистике выделяют следы-вещества, следы-предметы и следы-отображения. Фактически при исследовании следов-веществ исследуются свойства действительного элемента, который отображает сам себя, т.е. его вещественные свойства (носителя) без учета его внешней формы. При исследовании следов-предметов важным становится их форма, а при исследовании следов-отображений исследуют непосредственно след как отображение.

К идеальным следам в криминалистике относят отражения, которые остались в памяти участников и/или очевидцев события. Если исходить из того, что эти отражения зафиксировались в нейронах мозга, то они относятся к отображению, т.е. имеют действительный, а не идеальный характер.

Это отображение содержит объективную, субъективную и/или интуитивную, в частности, эмоциональную составляющую, поэтому такое отображение лучше называть умственным отображением.

В отличие от идеи (идейного), идеал (идеальное) может обладать вневременными, внепространственными и/или нематериально-вещественными свойствами. Например, абсолютно черное тело является идеальным элементом, так как обладает свойствами, отсутствующими в действительном мире. Поэтому говорить об идеальном отражении не корректно.

Обратим внимание на то, что в публикации [5] отмечено, что «Материальный след релевантен лишь

в том случае, когда ведет к открытию умостигаемого плана реальности». Этот вывод можно интерпретировать так, что отображение действительного элемента как знакового элемента будет ему релевантно, если оно будет отображать идею (план) этого элемента.

В связи с этим, вопрос соотношения отражения и отображения имеет не только теоретический, но и практический характер.

Следующий вопрос: возможно ли отражение в природной сфере? Например, отражение медведя в водоеме является отражением или отображением?

Для ответа на этот вопрос необходимо вернуться к продуцентному подходу [6]. Любой след имеет причину и/или источник. Если источник выступает в качестве продуцента, который продуцирует причину, то эту причину можно рассматривать как продуцирование следа. Категория продуцирования состоит из категорий связи, взаимодействия и/или продукта продуцирования. Для продуцирования дуба необходим жёлудь. Однако для появления дуба должны быть сопродуценты, обеспечивающие достаточные условия сопродуцирования дуба, в частности, почва как связь и/или питание как взаимодействие. Таким образом, необходимое свойство для продуцирования продукта можно считать существенным свойством.

Любой элемент наблюдаемого мира обладает носителем, формой и/или содержанием. Это заметил еще Аристотель. Поэтому для действительного элемента природной сферы нужно выделить вещественные носители (элементы вещественной среды) и/или материальные носители (элементы материального поля). Например, в качестве элементов поля могут выступать световые лучи в виде видимого диапазона электромагнитных волн.

Продукт взаимодействия действительных природных элементов может оставлять следы в виде отображения, которое может носить видимый или невидимый характер, а также – в виде отражения, которое перестает существовать после окончания взаимодействия и/или исчезновения канала связи.

Например, рассмотрим световое отражение медведя в водоеме. Пасмурная погода и/или отсутствие медведя на берегу водоема, не смогут сопродуцировать отражение медведя. Однако если медведь пойдет по влажной почве, то его следы могут отобразиться на ней.

Появление отображения зависит от уровня взаимодействия, свойств отображающего элемента, а также условий окружающей среды. Воздействие элемента на отображающий элемент может иметь низкий уровень, что может привести к отсутствию следа-отображения, но если на отображающем элементе находится тонкий слой другого элемента, например, пыль, то тогда на поверхности отображающего элемента может остаться невидимый, но присутствующий след-отображение. Аналогично, на гладкой поверхности можно выявить следы пальцев человека.

То, что в природной сфере наблюдается отражение, говорит о существовании отражения, которое отражает причину и/или источник, продуцирующий отражения. Поэтому элементы отражения можно называть элементами существующего мира.

В общем случае, категорию элементов наблюдаемого мира можно рассматривать как категории элементов действительного мира, которые отображают сами себя и/или элементы существующего мира, состоящие из отражений продуцирования и/или продуцента. При этом отражение продуцирования отражает свойства связи и/или взаимодействия как причины, а отражение продуцента отражает его свойства как источника продуцирования.

Поэтому отражение медведя в водоеме отражает не только внешние свойства медведя, но и световой поток, отражающийся от поверхности воды. Таким образом, природные элементы могут продуцировать действительные отображения и/или существующие отражения.

То, что категории действительных и существующих элементов могут логически пересекаться, позволяет утверждать, что оба вида отражения могут одновременно отобразиться, т.е. оставить следы в действительном мире. Например, молния может отразиться в водоеме и одновременно отобразиться на дереве, в которое она попала, в виде обожженного участка его ствола.

Для элементов общественной и/или психической сферы классификация элементов существующего мира имеет свои особенности.

ОТРАЖЕНИЕ В ПСИХИКЕ ЧЕЛОВЕКА И ОТОБРАЖЕНИЯ В ОБЩЕСТВЕ

Любое воздействие на человека чувственно воспринимается сенсорами и отражается в виде реакции. Например, световой поток отражается на сетчатке глаз. Но психика человека откликается не на весь поток отражения, а только на существенную для него часть, на которой он сфокусировал свое внимание. Таким образом, часть потока отражения временно сохраняется и затем замещается следующим потоком отражения, а существенная часть может запомниться в памяти в виде отображения на достаточно долгий срок.

К. Лоренц считал, что физиологическая отражаемость в виде восприятия краткосрочной (мгновенной) информации «не должно оставлять никаких следов в воспринимающем её физиологическом аппарате, потому что свою главную функцию – держать организм в курсе быстро меняющейся окружающей обстановки – этот аппарат может выполнять лишь при условии, что он всегда в состоянии заменить только что принятое сообщение другим, часто противоположным» [7, с. 359].

При этом наличие памяти в психике человека позволяет существенные мгновенные сведения запоминать в виде отображений и превращать их в то, что называется опытом. Однако наша генетическая структура также хранит доопытные отображения, которые К. Лоренц называет доктринерскими гипотезами, стереотипно отображающими картину мира.

Естественно, знак и/или идея могут отражать идеальные и/или действительные свойства. Например, изображение кентавра содержит идеальное соединение иконических знаков действительных свойств лошади и человека.

Таким образом, человеческая память обладает свойством забывания и/или воспоминания, но количество и/или качество отображения зависит от многих факторов, связанных с субъективными, а также интуитивными (креативными, иррациональными и/или эмоциональными) свойствами человека.

Кроме того, воспринятый поток может содержать дефекты, которые возникли случайно и/или преднамеренно. Поэтому наблюдатель может получить искаженный поток отражения и с дефектами запомнить умственные отображения.

В обществе широко используются документы, которые содержат сведения. Сведения можно представить как документально зафиксированные отображения элементов наблюдаемого мира. При этом сведения могут отображать не только чувственно воспринятые элементы действительного мира, но, и субъективно ощущаемые элементы внутреннего мира. Элементы внутреннего мира, пока они не воплощены (выражены) рассматриваются как психические отражения, которые существуют только в этом мире.

Люди в обществе воспринимают сведения в виде знаков и/или идей, поэтому в общественной и/или психической сфере существующий мир состоит из мира знаков и/или мира идей. Знаки и/или идеи как отражения отражают на нейронном уровне (микроуровне) нечто отличное от отражаемого, в частности, связи и/или взаимодействия. Эти отражения в виде мыслей, состоят из мысленных действий модулированных знаками и/или идеями. Например, цвет света нейроны воспринимают как спектр электромагнитных волн определенной частоты, сетчатка глаз воспринимает знаки в перевернутом виде как в камере обскуре. Более того, психика может окрасить черно-белое изображение клубники, так как люди привыкли, что она красная. Это показал А. Китаока (https://motherboard.vice.com/en_us/article/wkq5n/this-picture-has-no-red-pixelsso-why-do-the-strawberries-still-look-red), опубликовав в твиттере фотографию с оптической иллюзией.

На макроуровне психики, человек оперирует образами и/или понятиями, поэтому знаки и/или идеи трансформируются в умственные образы и/или понятия. Психические отражения можно называть умственными отражениями, которые могут отражать, как действительные, так и существующие элементы. В частности, психика человека может отражать образы и/или понятия, как знаков, так и идей, в том числе идеальных идей. Например, образ кентавра базируется на идеальной идее.

Так как категории действительного, знакового и идейного мира могут логически пересекаться, то знаки и/или идеи могут отображаться на/в действительных элементах. В этой связи, категория элементов действительного мира в общественной и/или психической сфере состоит из категорий действительных элементов, отображающих сами себя, категорий знаковых элементов (отображений знаков) и/или идейных элементов (отображений идей).

Важной проблемой также является соотношение категорий информации и отображения.

СООТНОШЕНИЕ ОТОБРАЖЕНИЯ И ИНФОРМАЦИИ

Хотя категория информации стала обсуждаться в середине прошлого века и связывалась с категорией отражения, до сих пор у ученых нет единого мнения по этому вопросу. Известны публикации того периода зарубежных авторов К. Шеннона, Н. Винера, Л. Бриллюэна, У. Эшби, а также отечественных ученых [8, 9].

Известны два подхода к объяснению категории информации: естественный подход для элементов в неживой природе и кибернетический подход для элементов в живой природе, обществе и кибернетических устройствах. Различные ученые рассматривают информацию как отражение, разнообразие, неопределенность и т.д.

А.Д. Урсул считает, что понятие отражения акцентирует внимание на воспроизведении содержания в целом, а понятие информации – на воспроизведении одной его стороны – разнообразия. Так же известна другая позиция, в частности, Д.С. Чернавский в монографии, посвященной динамической теории информации [10], пишет, что: «Информация есть отражение отображения наших соображений».

Таких различных мнений достаточно много, а дискуссия вращается вокруг следующего вопроса. Обладает ли свойство отражаемости нейтральностью или содержит все три элемента: пассивность (нейтральность), активность, негативность? Всеобщий характер этого свойства позволяет утверждать, что последнее должно проявляться.

Автор публикации [11], исходя из философского тезиса разделения понятий отражения, характерного неживой природе, и отображения, характерного живой природе, считает, что «отображение мыслится как способность живых объектов фиксировать в изменениях специализированной структуры целостности признаки, свойства и отношения отображаемого. Своеобразие процесса отображения состоит в том, что живой объект не меняет своих физических свойств в результате такого взаимодействия, а изменение отображающей структуры соотносено (не физически) с характеристиками отображаемого благодаря механизму кодирования признаков».

Сложно согласиться с такими интерпретациями категорий «отражения» и «отображения» в связи с категорией информации. Рассмотрим еще одну прагматическую модель такого соотношения.

Как правило, считается, что информация с количественной точки зрения представляет собой отражение (отображение) разнообразия. Простейшим примером разнообразия является наличие или отсутствие какого-либо знакового значения в материально-вещественном знакоместе. В частности, знаки разнообразия могут быть выражены в виде 0 или 1.

Однако кроме количественного разнообразия возможно качественное разнообразие, которое проявляется в виде качественного различия идей ячеек.

Поэтому в некоторых публикациях [12] под информацией в общественной и/или психической сфере понимают содержание сведений, а не сведения, как установлено в Федеральном законе РФ «Об инфор-

мации, информационных технологиях и о защите информации».

Кроме разнообразия известна категория индивида, которая связана с категорией разнообразия.

Категория индивида

Категория индивида [13] подразумевает, что как минимум один из однородных элементов рассматриваемой субстанции (среды/поля) выделяется из нее за счет сил, действующих в ней в виде продуцирования в таком элементе хотя бы одного индивидуального свойства. Именно наличие индивида в его окружении приводит к появлению свойства отражения индивида, как в его окружении, так и в самом себе.

Естественно, разрушение индивида приводит к его разложению на однородные элементы, и разнообразие, как совокупность индивидов исчезает. Эти однородные элементы могут обладать формой, поэтому не форма определяет разнообразие, а индивиды.

В природной сфере элемент состоит из носителя, формы и/или содержания. Если исходить из продуцентного подхода, то любой элемент наблюдаемого мира как продукт продуцирования продуцентом может обладать формой, которая модулирует носитель под действием сил, действующим в таком мире, и отображает его свойства как содержание продуцента. Например, желудь отображает дуб. Таким образом, элемент действительности можно рассматривать как действительный элемент, отображающий самого себя как продукт продуцирования, в виде: 1) носителя, 2) формы, отображающей продуцирование, и/или 3) содержания, отображающего продуцента.

Пример отображения и отражения в природной сфере

Облако в воздушной среде представляет собой природный продукт, который был продуцирован кругооборотом воды в природе, и как действительный элемент отображает сам себя.

Так, облако может отражаться в воде, т.е. отражать свойство продуцирования в виде связи и взаимодействия облака с лучами Солнца, а также отражать свойства продуцента в виде естественной логики (идеи) кругооборота воды в природе.

Естественное отражение, которое не оставляет следов, возможно на основе низкоэнергетического взаимодействия элементов и/или каналов связи с низким уровнем затухания (трения) в них.

Например, отражение облака возможно при определенном количестве облаков в воздушной среде, так как полная облачность не даст возможности облаку отразиться в воде. Или еще один пример, сила удара по бильярдному шару не должна превышать определенного уровня, иначе шар может покинуть поверхность стола и канал связи в виде такого стола разорвется для него.

Отображение всегда связано с появлением следов взаимодействия отображаемого и отображения. Это может быть деформация взаимодействующих вещественных элементов или ожог на поверхности от действия солнечного потока.

Если исходить из того, что облако как действительный элемент обладает носителем, формой и содержанием, то носителем являются молекулы воды, форма отображает внешнюю морфологию облака, а содержание облака с технической точки зрения может быть выражено в виде количественных значений свойств воды в облаке. Таким образом, категория содержания как информация представляет собой количественное значение измеряемой величины.

В общественной и/или психической сфере категория содержания приобретает иной вид. Она рассматривается как субъективная значимость (значимость) и/или прагматический смысл (смысл). Естественно, в природной сфере нет знаков и/или идей, но в общественной и/или психической сфере они существуют.

СООТНОШЕНИЕ ЭЛЕМЕНТОВ ДЕЙСТВИТЕЛЬНОГО, ЗНАКОВОГО И ИДЕЙНОГО МИРА

Ранее нами уже отмечалось наличие категорий действительного и существующего мира. Здесь в категорию существующего мира для общественной и/или психической сфер входят категории знакового и/или идейного мира.

Так как категории действительного и существующего мира могут логически пересекаться, то в действительном мире можно выделить элементы действительного мира, отображающие сами себя, знаковые элементы, отображенные на действительных элементах, и/или идейные элементы, отображенные на действительных элементах.

При этом категория знака является элементом знакового мира (мира знаков), а категория идеи – элементом идейного мира (мира идей). Таким образом, категории знака и/или идеи существуют в массовом сознании и/или психике конкретного человека, и воплощаются (выражаются) в виде отчуждаемого носителя, несущего знаковые и/или идейные элементы.

Например, действительный элемент в виде электрического напряжения в розетке отображает сам себя как его действительное значение, но, кроме того, рядом с розеткой может быть расположен документ с буквенно-цифровым отображением значения этого напряжения. Документ как действительный элемент содержит сведения в виде знаков, несущих определенные идеи, а их содержание представляет собой информацию.

Категорию содержания сведений (информации) можно разложить на категории объективных значений, субъективной значимости и/или прагматического смысла сведений. При этом необходимо понимать, что категория психического отклика человека на воспринимаемые сведения состоит из категорий восприятия (синтаксического получения), семантического понимания (интерпретации) и/или прагматического использования воспринятых сведений.

ЗАКЛЮЧЕНИЕ

Объяснить категорию отражения можно не прибегая к идеалистическому подходу. Отражение, с точки зрения продуцентного подхода, обладает материально-вещественным носителем и существует за счет свойств отражаемости продуцента и продуцирования. Информация как содержание отображения, в частности, сведений, может иметь, как количественные, так и качественные свойства, поэтому она отображает свойства элементов, как действительных, так и существующих миров в природной, общественной и/или психической сфере.

СПИСОК ЛИТЕРАТУРЫ

1. Анохин П.К. Теория отражения и современная наука о мозге. – М.: Знание, 1970. – 46 с.
2. Колин К.К. Философские проблемы информатики. – М.: БИНОМ, 2010. – 264 с.
3. Тюхтин В.С. Отражения, системы, кибернетика. Теория отражения в свете кибернетики и системного подхода. – М.: Наука, 1972. – 251 с.
4. Суворова Л.А. Идеальные следы в криминалистике. – М.: Юрлитинформ, 2006. – 200 с.
5. Минина М.Л. Теория отражения и проблема интерпретации: автореф. диссерт. канд. филос. наук. – Чебоксары: Чувашский государственный университет, 2007.
6. Нестеров А.В. Экспертика: Общая теория экспертизы. – М.: Тип. НИУ ВШЭ, 2014. – 261 с.
7. Лоренц К. Обратная сторона зеркала. – М.: Республика, 1998. – 635 с.
8. Урсул А.Д. Природа информации. Философский очерк. – М.: Политиздат, 1968. – 288 с.
9. Михайлов А.И., Черный А.И., Гиляревский Р.С. Основы научной информации. – М.: Наука, 1965. – 655 с.
10. Чернавский Д.С. Синергетика и информация: Динамическая теория информации. – М.: Наука, 2001. – 244 с.
11. Денискин С.А. Понятие информации в контексте категорий 'отражение', 'отображение', 'связь' // Вестник Челябинского государственного университета. – 2014. – № 11, Вып. 32. – С. 13–17.
12. Нестеров А.В. Об информационных объектах и их юридических свойствах // Научно-техническая информация. Сер. 1. – 2014. – № 2. – С. 28–31.
13. Нестеров А.В. Тензорный подход к анализу и синтезу систем // Научно-техническая информация. Сер. 2. – 1995. – № 9. – С. 26–31.

Материал поступил в редакцию 13.09.17.

Сведения об авторе

НЕСТЕРОВ Анатолий Васильевич – доктор юридических наук, кандидат технических наук, профессор, профессор Российского университета дружбы народов, профессор Российской таможенной академии
e-mail: nesterav@yandex.ru

Методологические основы онтологического моделирования документальной информации

Обсуждаются некоторые свойства онтологий, определяющие возможности для структурированного представления смысла документальной информации и его обработки. На основе анализа процессов жизненного цикла объекта и функционально-кибернетической модели элементарного процесса предложены концептуальные каркасы для построения типизированной таксономии объектов и отношений онтологии верхнего уровня.

Ключевые слова: *информационное моделирование; онтологии; семантика; текстовые документы; модель деятельности; концептуальный каркас; общая теория систем; системный анализ*

ВВЕДЕНИЕ

Построение онтологий не является сегодня самоцелью. Они используются для решения практических задач в таких областях, как информационный поиск, анализ текстов и машинный перевод, экспертные системы, системы сохранения знаний.

В области создания и исследования онтологий имеется большое количество академических и Internet публикаций. В данной статье приведены ссылки только на некоторые из них. Однако влияние на эти исследования оказали и многие другие работы.

Еще недавно сохранение и управление знаниями (и онтологии, как средство их представления) связывалось, преимущественно, с текстовыми формами информации (слабоструктурированная информация). Они характеризовались сравнительно компактным множеством ситуативных схем и понятий, а также практически не типизированными наборами отношений.

Распространенной формой существования знаний являются автоматизированные системы и базы данных, представляющие часть знаний, относящихся преимущественно к практике их применения. Это, в основном, хорошо структурированная фактографическая информация, представляющая бесконечное разнообразие объектов и связей действительности, причем непрерывно меняющейся. Тем не менее, семантика таких баз данных вполне определена соответствующими инфологическими моделями. Для их построения применяются обстоятельно проработанные методологии (в частности, SADT, UML).

Одной из первых хорошо формализованных моделей унифицированного представления данных предметной области (ПрО) является модель «сущность-

связь» [1], основанная на теоретико-множественном подходе, где семантика принадлежности определяется предикатом. Но необходимо отметить, что базы данных разрабатываются в основном для бизнес-приложений, ПрО которых имеет устойчивую и ограниченную структуру связей. Поэтому ситуативность в этом случае достаточно адекватно выражается свойствами связи, которые задаются в виде атрибутов. Это, в некотором смысле, предопределяет и то, что только сущности (но не связи) ассоциируются с понятием тип и супертип.

Задачи обработки как структурированной, так и слабо структурированной информации связаны с обработкой знаний. При этом онтологии (или онтологические принципы, как в случае БД) используются в качестве семантического или метаинформационного обеспечения.

Объединение этих подходов предложено в системе стандартов ISO15926¹, общая концепция представления предметной области в которой основана на классах, а концептуальные и технологические решения по практическому построению семантических сетей (онтологии) по текстовым документам представлены в [2]. Как отмечается в [3], «...этот подход позволил сформулировать удовлетворительный ответ

¹ Стандарт «Промышленные автоматизированные системы и интеграция. Интеграция данных жизненного цикла для перерабатывающих предприятий, включая нефтяные и газовые производственные предприятия» определяет основные типы объектов и отношений (задает онтологию), нормализует терминологию, а также определяет принципы расширения стандартной терминологии через механизм федеративных библиотек справочных данных.

на вопрос о том, как соотносятся сущность и атрибут: в терминах принадлежности объекта к классу объектов, обладающих данным атрибутом. Таким образом, существенными чертами модели стали отказ от единого иерархического дерева и допущение множественности классификаций, в том числе равноправных классификаций объектов и атрибутов. Реификация отношений (выделение отношений как сущностей первого уровня, способных, в свою очередь, быть сторонами в отношениях) позволила распространить на отношения шаблоны работы с объектами. Теоретико-множественный язык оказался хорошо совместим с четырехмерным представлением о пространстве-времени, позволившем перейти к единообразному описанию пространственно-временных свойств объектов, решающему проблемы моделирования течения времени».

Сегодня создание систем управления знаниями (СУЗ) промышленного масштаба (где онтологии являются инструментом экономики, основанной на знаниях) возможно только путем интеграции информационных процессов и ресурсов науки, технологии, промышленности и управления. Такой подход должен учитывать обусловленные разделением труда особенности создания и использования знаний.

Задокументированные знания (называемые явными) различаются по характеру: теоретические, практические, технологические, полученные аналитически или экспериментально, относящиеся к фундаментальным или прикладным результатам, предназначенные для изучения, применения или анализа и т.д. Это разнообразие приводит к появлению еще большего разнообразия решений: способов и форм их фиксации (документирования), а также методов обработки, связанных с перемещением знаний в вычислительную среду.

Кроме того, СУЗ должны быть ориентированы на решение новых задач, когда данные используются для синтеза нового знания или поиска новых форм и условий применения старых знаний, обеспечивающих новое качество. Это предопределяет, что «горизонт» объектов и их взаимосвязей (как данных, так и их ролей) для СУЗ должен быть существенно широким, чтобы обеспечить достаточные возможности для построения новых взаимосвязей и «контекстов» на «старых» знаниях.

Современные многочисленные проекты создания онтологий на текстах используют два подхода: лингвистический и статистический. Лингвистические методы, основанные преимущественно на использовании словарей, позволяют достигать высокой точности идентификации понятий и отношений, но они очень ресурсоемки. Кроме того, создание и актуализация словарей требуют участия экспертов не только в лингвистике, но и в предметной области. Безсловарные методы, основанные на использовании статистических зависимостей, плохо выявляют новые связи и понятия и не позволяют формировать семантические графы.

Целью данной статьи является обоснование подходов к онтологическому моделированию, а также к

построению процедур выделения сущностей и отношений из текстов. В основу разрабатываемой на базе системного подхода унифицированной структурно-информационной модели предметной области и стратифицированной онтологической модели ее представления положена гипотеза о схематизме рациональной деятельности и типизированности объектов предметной области [4–6].

ДЕЯТЕЛЬНОСТЬ, ЗНАНИЯ, ОНТОЛОГИИ

О назначении и использовании онтологий

Существуют многочисленные определения онтологий. Это объясняется тем, что каждое из них имеет собственное предназначение и способ применения. Используя определения, данные [7–10] в проектах WordNet, MikroKosmos, PyTез, будем содержательно рассматривать онтологию как *спецификацию концептуализации на уровне эксплицитных знаний, зависящую от предметной области или задачи, для которой она предназначена, и ориентированную на совместное многократное использование*.

Это сравнительно несложное по своей конструкции определение, тем не менее, является очень емким и нуждается в уточнении.

1. Под «концептуализацией» понимается строгое (т.е. абстрактное и, следовательно, упрощенное) описание *системы* понятий, объектов и других сущностей, а также отношений, связывающих их друг с другом. Таким образом, концептуализация *расчленяет* целостную область знаний, выделяя из нее отдельные объекты и отношения. Причем членение должно быть адекватным операбельным объектом, заменяющим саму предметную область в процессе достижения определенной цели.

Существует два подхода к процессу концептуализации: интенциональный и экстенциональный. При интенциональном подходе сущности идентифицируются через их свойства. При экстенциональном – задаются перечислением. На практике ни один из подходов в чистом виде не применим. Невозможно заранее составить множество всех объектов ПрО, равно как и свойств – признаков членения. Эта задача решается путем выделения относительно самостоятельных форм представления, отражающих суть обоих подходов, которые диалектически взаимодействуют в процессах синтеза и использования знаний.

2. Что касается форм представления онтологий (связанных с технологиями их создания), то здесь выделяют прикладные онтологии, онтологии области знания, онтологии метауровня. Отдельно выделяют лингвистические онтологии (WordNet, MikroKosmos, PyTез и другие), фиксирующие понятия (слова) вместе с их языковыми свойствами и отношениями (синонимии, гипонимии и т.п.). Показателен подход, предложенный в [11], где онтология – это наборы метаданных, идентифицирующих содержания ресурса и сервисов в нескольких аспектах: описание объектов, их свойств и отношений в предметной области (Domain ontologies); описание задач и процессов, их свойств и отношений (Task ontologies); описание атрибутов знания (Quality ontolo-

gies); описание причин и оценок использования (Argumentation ontologies), а также характеристика значимости контента (Value ontologies) и история обращений отдельного субъекта к информационным ресурсам (Personalization ontologies).

3. Совместное многократное использование знаний предполагает *интероперабельность* в случае использования онтологии в разных областях или применения несовместимых инструментальных средств. Также *многократность* имеет временной аспект – предметная область со временем изменяется, и ее онтология должна развиваться. Причем это будет диалектическое взаимодействие, поскольку и система представления ПрО, и понятийная система по существу самостоятельны², что и является основой устойчивого совершенствования.

4. В [12] онтологии рассматриваются, прежде всего, как средство идентификации знаний, обеспечивающее более высокую полноту и точность информационного поиска. Это достигается за счет того, что смысловое содержание документа будет обозначено не только основными существенными и отличительными понятиями, но и ситуативными связями между ними³.

Однако внедрение инструментария онтологий (в том числе и построение онтологий объектов, свойств и процессов на основе текстов научной и технической документации) как одного из видов информационного моделирования в область системной инженерии существенно расширяют функций онтологий, т.е. онтология становится «действующей» моделью ПрО. В терминах прямого/обратного преобразования это означает, что онтологическое представление должно по максимуму обеспечить возможность построения по этой онтологии текста, эквивалентного тому, по которому она сама была построена.

Исходный для построения онтологии текст сам является образом некоторого объекта деятельности (вещи, ситуации, теории и т.п.). Поэтому, правильнее утверждать, что должна быть обеспечена возможность воспроизводства этого объекта, а не описывающего его текста. Более того, если принять, что онтология в некотором смысле является аналогией сборочного чертежа, то это «сборочный чертеж» не самого объекта, а нашего представления о нем – ситуационного знания, обусловленного контекстом текущего состояния и конечной цели (в том числе и субъекта). Поэтому «точность» онтологии должна определяться не степенью совпадения исходного и восстановленного текста, а точностью «схемы» воспроизведения информационного образа объекта.

Отсюда следует, что онтология не имеет цели заменить документацию, принятую в конкретной об-

ласти. Ее функциональное назначение – быть средством для формирования виртуальных документов – подборок документов (результатов поиска) или их фрагментов, содержание которых связано с отдельным понятием, свойством объекта, или фрагментом его «жизненного пути». Последнее, в частности, может быть реализовано через навигацию в массивах задокументированного знания формированием контекстно обусловленной траектории (например, посредством динамически генерируемых гипертекстовых ссылок). Это и есть истинно семантический информационный поиск, поскольку данные (записи) будут непосредственно связаны (фактом выбора конкретной ссылки) с указанным пользователем контекстом – существом решаемой им задачи.

5. Онтология является результатом взаимодействия [13] субъекта и объекта познания, происходящего в контексте предмета познания, и представляет конкретное знание. Она имеет информационную природу. Поэтому её взаимодействие с другими информационными объектами будет основой для синтеза нового знания. Таким образом, поскольку процесс синтеза в своей основе имеет комбинаторную природу (взаимодействуя, в том числе с вновь построенными объектами) он также будет бесконечным. Но уже с возможностью управления или, по крайней мере, с возможностью выбора альтернатив.

Процесс познания бесконечен и непрерывен. Онтология же дискретна и конечна. В общем случае любой процесс представим в дискретной описательной форме, в том числе текстовой. Это и есть следствие того, что любое изменение параметров физических процессов может быть дискретизировано, как для отображения, так и для последующего восстановления исходного процесса. Допустимость дискретизации при формировании образа физических процессов следует из принципа параметрической локальной однородности ПрО, предопределяющего возможность сведения континуума квантовых значений физического параметра к ограниченному числу образов [15].

Именно положение о сводимости к дискретной форме имеет важное следствие. Непрерывно изменяющиеся и имеющие преходящий характер процессы (действие существует только до момента своего завершения) представляются отдельными, стабильными во времени объектами-сущностями (множеством свойств, типом связи, результатами и т.д.).

Онтология и текст

Существенной особенностью формирования информационного представления объекта на основе описывающих его текстов является его семиотическая природа: свойства объекта и его поведение должны быть представлены знаковой системой. Текст, который представляет знания в форме овеществленной информации, обобществляющей и отчуждающей знания, может рассматриваться как информационная модель знания, задающего интенциональную семантику текста, в то время как предмет знания определяет референцию текста. Текст может служить инструкцией для осуществ-

² Относительная изолированность является следствием принципа локального детерминизма, а также организационной основой эффективности технологий их поддержки за счет специализации деятельности.

³ Следует отметить, что такой подход далеко не нов. Еще в 60-х годах прошлого века были предложены языки [14] и разработаны ИПС, ориентированные на точный поиск не только понятий, но и фактов – понятий, связанных функциональными отношениями.

ления действий и использоваться для творческой реконструкции заложенных в него знаний. В первом случае информативность текста определяется снимаемой неопределенностью выбора, во втором – создаваемыми новыми возможностями действий [16].

Следовательно, с точки зрения идентификации можно различать⁴ описание собственно знания (как устроен объект, как функционирует, как достигаются целевые показатели и т.п.) и описание–опознаватель, представляющее знание (экземпляр носителя знания) с точки зрения поиска.

Если функциональное описание представляет существо объекта «изнутри» с полнотой и точностью, обусловленной уровнем требований к пониманию и использованию свойств и поведения объекта, или к воспроизводимости объекта, или его состояния в тех или иных ситуациях, то идентификатор представляет объект «извне». При этом его основные свойства и признаки обеспечивают в конкретных условиях узнавание и выделение такого объекта из множества других.

Таким образом, задача построения онтологического описания ПрО формально может быть определена как выбор такого способа представления, который при минимальных затратах усилий на восприятие (чтение) позволил бы воспроизвести все моменты наблюдения или процесса построения в той последовательности, обеспечивающей эквивалентность образа оригиналу, а также возможность манипулирования образами для моделирования предметных областей и аспектов действительности.

О системности и схематизме

Формирование информационного представления в виде системы, т.е. построения совокупности объектов, признаков и отношений, основывается на двух взаимодополняющих направлениях анализа: 1) структуры и поведения самого объекта и 2) его образа, обеспечивающего возможность для моделирования – опосредованного изучения объекта, прогнозирования его поведения, управления и т.д.

Механизм построения такого представления основан на использовании системного подхода, представляющего любой объект как систему – совокупность типизированных элементов, связанных некоторыми отношениями, в совокупности образующими единство, для которого характерно появление свойств, не присущих составляющим.

Такое представление в рамках общей теории систем [4] позволяет сформулировать методологические основы синтеза нового знания. В основе модели син-

теза знаний как самоорганизующегося процесса лежит структурная особенность системы – сложная система может быть описана набором относительно независимых аспектных представлений. В процессе декомпозиции не только выделяются и связываются составляющие, но и формируется *схема* – система характеристических признаков, в соответствии с которой и проводится декомпозиция.

Если представлять объект познания в виде совокупности однородных образующих единство взаимосвязанных элементов, а систему этих однородных объектов в виде классификации, то можно выделять в явной форме новые характеристические признаки, определять способы организации подсистем, и на основе свойств соответствия и симметрии обнаруживать связи с другими системами классификации.

С точки зрения природы связанных со знаниями операционных объектов в сложных социально-технических системах выделяются блоки двух типов: «практика» и «моделирование». «Практика», включая промышленное производство, – это процесс непосредственного энерго-материального взаимодействия со средой⁵. «Моделирование» – имитация практики – то, что позволяет на основе модельных оценок⁶ прогнозировать, анализировать, оптимизировать практическую деятельность.

Назовем *действительностью* ту часть реальности⁷ (мира физических и абстрактных вещей, существующих независимо от субъекта), которая непосредственно или опосредованно (например, через измерения) взаимодействует с субъектом. Его действия и представления, при условии рациональности деятельности, всегда соответствуют некоторой модели (цели, технологии и т.п.). Действительность проявляется через *свойства* – информационный образ результата взаимодействия между объектом реальности и средствами измерения, построенными согласно модели. Отметим, что именно информационная природа свойств обеспечивает автономность моделирования и в результате – оптимизацию процесса целенаправленного изменения действительности.

«Технология» соотнесения модели и практики как диалектическая спираль перехода адекватно и просто изложена в [5]. «Развернуть предмет некоторой науки – это значит построить, создать особые структурные схемы, которые бы задавали модель, или систему моделей этого предмета. Затем нужно оправдать эти схемы как задающие некоторую онтологию развертываемого предмета с точки зрения более широкого целого. ... После этого нужно построить собст-

⁴ Это аналогично тому, что по своему функциональному назначению модели можно подразделить на описательные и преобразовательные. Описательные модели представляют оригинал как статичный объект, свойства которого имеют структуру или поведение (состояния) объекта или объясняют их. Например, через такие статичные образы, как схемы или алгоритмы. Модели преобразовательного типа отражают процессы: либо внутреннее существо – изменение структуры и поведения объекта, либо «внешнее» взаимодействие оригинала и модели или другого оригинала, включая наблюдения, измерения, экстраполяцию в обоих направлениях.

⁵ Сюда относятся и мыслительные процессы, т.к. субъекты процесса – человек и социум – в своей биофизической основе могут существовать и развиваться только за счет этой среды.

⁶ Модель – это такая мысленно или материально реализованная система, которая, отображая или воспроизводя объект исследования, способна замещать его так, что ее изучение дает информацию об объекте (оригинале) [17]

⁷ Реальность непрерывна. Поэтому выделение, как и именование частей, систем или классов подчиняется принципу неопределенности и может быть только условным (сопровождаться условиями).

венно знания, описывающие разные стороны этой действительности, знания, которые мы могли бы использовать в каких-то практических ситуациях. Подобная процедура отнесения схем развертываемого предмета к более широкому целому выступает как построение онтологии этих схем и как обоснование этой онтологии и, вместе с тем, как средство более точного и тонкого ограничения эмпирического материала.... Таким образом, мы получаем некоторое изображение, полученное в результате наложения неадекватных средств. Причем, это изображение приобретает, фактически, реалистическое существование, т.е. в него заложен некоторый смысл, мы его особым образом понимаем. Это – неадекватная картинка данного эмпирического материала. ...Чтобы правильно построить схему-модель мы, как обычно в подобных случаях, преодолеваем этот парадокс методом последовательных челночных приближений. Мы попеременно отдаем преимущество и определенную роль то схемам – и тогда по ним корректируем область эмпирического материала, – то эмпирическому материалу – и тогда по нему корректируем и перестраиваем сами схемы».

СТРАТИФИЦИРОВАННАЯ МОДЕЛЬ ДЕЯТЕЛЬНОСТИ

Одна из особенностей онтологического подхода, порождающая трудности в практике применения онтологий, состоит в том, что при создании последних возможно использовать любые отношения. В определенных случаях эта возможность бывает полезной. Однако она способна создавать и сложности. Например, разработчик может ввести в свою онтологию отношение, которое либо аналогично уже имеющемуся, либо определено комбинацией существующих, либо в силу специфичности порождает неадекватность восприятия информации об объекте. Все это затрудняет построение выводов, основанных на формальных правилах.

Возможное решение проблемы состоит в использовании типизированной системы отношений (имеющей концептуальный каркас). Построение такой унифицированной «естественной» системы отношений может основываться на анализе природы ситуативных связей объектов предметной области, а в итоге привести к систематизации и типизации процессов. Некоторые подходы, построенные на принципе схематизма, представлены в ряде работ, например [18–20]. Но, в этих работах подход обусловлен спецификой конкретной ПрО и систематизации сущностей и отношений уделено недостаточно внимания.

Ориентируясь на принцип схематизма, составляющего базис рациональной специализированной деятельности человека, рассмотрим модели, отражающие закономерности этой деятельности на следующих четырех уровнях: 1) интегрированного в рамках жизненного цикла (ЖЦ) процесса целенаправленной деятельности; 2) целостного процесса формирования и решения проблемы/задачи; 3) отдельного функционального процесса; 4) операции функциональной обработки.

Структурно-информационная модель процесса деятельности

Для того чтобы система управления знаниями была жизнеспособна и эффективна, она должна интегрировать информационные потоки и ресурсы, относящиеся ко всем видам и этапам сознательной деятельности: от НИР до вывода из эксплуатации, от стратегического планирования до реализации и оценки результатов, от формирования образовательных программ до подготовки и переподготовки кадров.

Обобщенная функционально-ориентированная схема интегрированного процесса целенаправленной деятельности представлена на рис.1. В общем случае она включает научно-исследовательские, опытно-конструкторские работы, производство, которые сопряжены с другими видами деятельности (организация производства, сбыта и т.д.). Укрупненная функциональная модель, представленная в верхней части рис. 1, – это последовательность, которая начинается с исследования и заканчивается производством и коммерциализацией. В общем случае процессы не обязательно являются последовательно выполняемыми, а границы между этапами не выражены явно. Информационная структура такой системы включает как теоретические знания, так и промышленные технологии. Последние взаимосвязаны с системами понятий и знаков инструментами познания и материальными средствами представления знаний, соответственно. В совокупности с образовательными и информационными процессами схема на рис. 1 представляет архитектуру СУЗ, интегрирующую посредством информационных потоков и технологий научные и практические знания.

Таким образом, весь процесс может быть представлен как продолжающиеся и повторяющиеся взаимодействия, а также обратные связи между наукой, технологией и производством. Причем функциональные блоки на отдельных этапах жизни имеют одинаковую структуру: получение исходного материала, управляемое параметрами выполнение целевой функции, оценка и передача результата.

Необходимо отметить, что каждый блок (этап, работа) заканчивается не только целевым результатом, но и его оценкой. Результаты будут иметь документальное представление, что обеспечивает возможность разделения труда и управления. Такое представление является отображением объектов и процессов этапа на систему понятий и отношений соответствующей модели деятельности. Отсюда, в частности, следует, что для представления знаний должны существовать, по крайней мере, три таксономии: таксономия модели (теория), таксономия целевого объекта (состав и структура) и таксономия форм представления (видов документации). Отметим, что система видов документации является следствием разделения труда по видам деятельности. Это, в частности, предопределяет, что соответствующие понятийно-лексические системы также будут различаться, что, в свою очередь, потребует средств, обеспечивающих их смысловое соответствие.

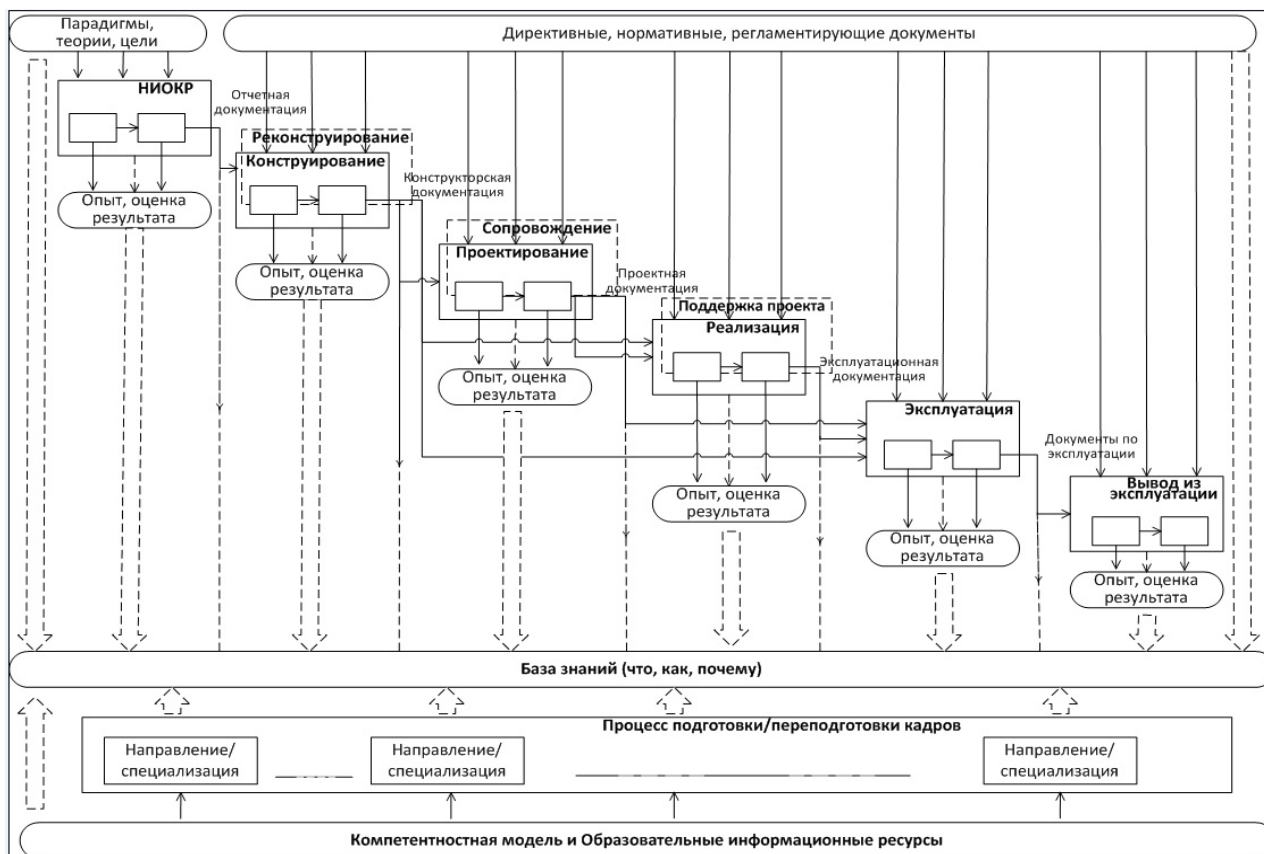


Рис. 1. Структурно-информационная модель деятельности, основанной на управлении знаниями

Модель процесса формирования и решения задачи

Соотношение модельной и практической составляющей в научно-исследовательской работе схематично представлено на рис. 2, что может рассматриваться как общий случай, так как этот тип процесса имеет наибольшее разнообразие объектов, решений и исходов: для НИР характерна существенная вариативность⁸ решений и результатов, включая отрицательные, которые тоже так или иначе обладают свойством полезности.

Решение проблемы или задачи – это отыскание форм и условий взаимодействия нового элемента знания (стимула, породившего неопределенность) с ранее существующим. При этом собственно решение проблемы основывается на её декомпозиции⁹ на простые задачи.

Проблема выбора одного из методов решения задачи тесно связана с проблемой определения *критерия*

⁸ Особенностью является то, что абстрактно-конкретные системы динамически совмещают объекты абстрактной (цели, гипотезы, теории, модели и т.д.) и материальной природы (машины, технологии и т.д.). Единство методов также условно, т.к. определяется парадигмой, выбираемой на предпроектной стадии.

⁹ Разложимость, т.е. возможность раздельного принятия решений по независимым группам позволяет ориентироваться в мире, обладающем, на первый взгляд, бесконечной сложностью. В математическом плане свойство разложимости означает, что соответствующая функция, описывающая «поведение» системы, может быть представлена как суперпозиция функций от меньшего числа переменных [23].

рия оценки альтернатив, где существуют различные точки зрения и основания. Существенной особенностью в данном случае является то, что «выбираемые основания – это еще и средство организации наличного знания субъекта с целью *удобного* получения следствий» [21, 22].

Причем взаимосвязь принципов, целей и задач при их выборе практически всегда обуславливается требованием «конструктивности» (разрешимости задач и достижимости цели), что допускает их вариантность и комбинативность¹⁰.

¹⁰ Согласно [24], «Цель» – направление движения науки, определяется ее возможностями. Но при этом выбор конкретного «возможного» действия (метода) определяется целью, что приводит к порочному кругу. Выход из этого парадокса заключается в том, что при описании научного познания к понятиям «цель» и «метод» следует относиться конструктивно. В познании понятие «цель» не имеет того абсолютного смысла, как в традиционной схеме нахождения решения при заданной постановке задачи: дано, требуется доказать (решить), доказательство (решение). Цель в случае процесса познания не первична, как в обычной человеческой интерпретации, и не вторична, как в причинной механической интерпретации, в нее вкладывается иной смысл – относительный. Вначале цель и метод дополнительны, одновременны. Иерархия между ними устанавливается лишь в процессе формирования целостного представления о данной задаче. Кроме того, в общем случае решение задачи понимается не как решение некоторой конкретной изначальной задачи, а как решение той задачи, которая поддалась решению.



Рис. 2. Модель формирования и решения проблем в научно-исследовательской сфере

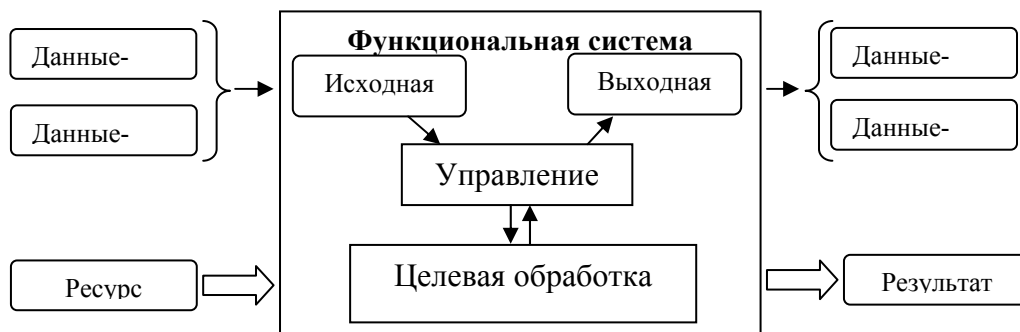


Рис. 3. Обобщенная схема процесса с точки зрения управления данными

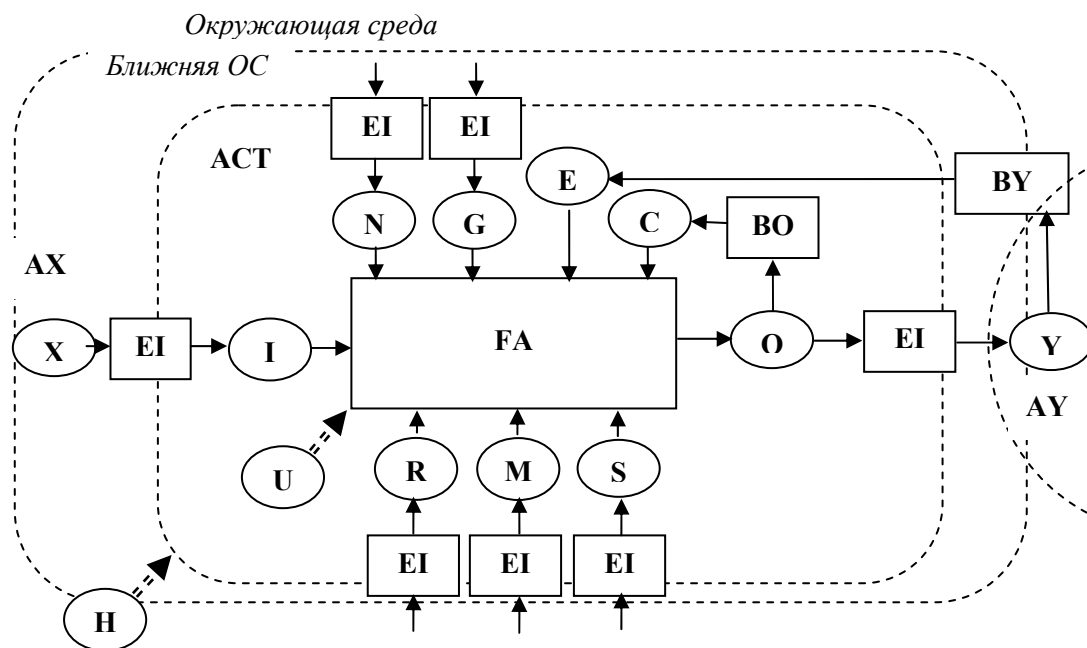


Рис. 4. Функциональная диаграмма реализации операции функциональной обработки,

где: I – «вход» операции FA , исходный материал; X – причина; AX – источник; G – управляющий сигнал; C, E – сигналы регулирования по обратной связи внутреннего и внешнего контуров; N – помеха; M – «механизм» – неизменяемые ресурсы, с помощью которых реализуется операция (устройство, метод, способ); R – материально-энергетические ресурсы, обеспечивающие выполнение операции; S – «субъект», осуществляющий управление в автоматизированном режиме, чаще всего – люди; O – «выход функциональный» – результат операции; FA, Y – «практический выход», следствие – то, что получено приемником AY в результате элементарного процесса ACT и интерфейсной операции (преобразования формы) EI . U – методы и средства, на основе которых построен преобразователь; H – условия и теории построения и функционирования элементарного процесса.

При этом существенно, что наряду с получением целевого практически значимого результата появляются информационные образы, не только характеризующие объекты и операции, но и сводящие этот результат к процессу, управляемому данными.

Модель функционального процесса с точки зрения управления данными

Любой элемент деятельности может быть представлен в виде функционального процесса, управляемого данными (рис. 3.). Это позволяет явно зафиксировать, что помимо фактов, определяющих реализацию решения, отдельно выделяется класс данных, представляющих задачи, проблемы или гипотезы. В целенаправленной деятельности человека и общества именно формулировки задач выступают в роли данных, инициирующих (определяющих) функциональный процесс, где в качестве ресурса уже могут выступать как материя или энергия, так и информация.

Таким образом, система преобразования исходного ресурса в целевой результат обусловлена контекстом – данными, представляющими целевую задачу, и появляющимися в результате ее решения. При этом потенциально полезные данные, выделенные из общего множества в соответствии с контекстом задачи (исходная информация), в результате использования порождают выходную ин-

формацию – актуализированные данные, подтверждающие или отрицающие действительность выбранных исходных данных и методов решения задачи.

Функционально-кибернетическая модель элементарного процесса

Акт процесса (реализация операции функциональной обработки) целенаправленной и управляемой (познавательной, производственной и т.д.) деятельности рассматривается с точки зрения теории систем и может быть представлен средствами функционального моделирования¹¹. На рис. 4 помимо объектов (вход, выход, управление и т.д.) и операции преобразования можно выделить¹² ближнюю среду – источник и цель деятельности, «интерфейсные» блоки – средства получения/передачи информации из/в внешней среды, контур управления – блоки, обеспечивающие контроль соответствия получаемых результатов установленным критериям. Такая схема

¹¹ Аналогичный подход представлен в [25], однако, там учитываются в основном «направления» функционального моделирования: вход – выход – управление – ресурс.

¹² Такая схема выделения компонентов соответствует принципу наименьшего действия, выражающегося в стремлении к однородности объектов или преобразований, что обеспечивает «однофакторность» преобразования, т. е. каждый выделяемый компонент будет иметь один вход, один выход, один управляющий вход и т.д.

представляет собой «моделирование», где объекты и процессы четко разграничены, хотя в реальности любой объект находится в непрерывном взаимодействии и, так или иначе, изменяется.

Операция преобразования представлена *Элементарным процессом (ACT)*, который изменяет объект, поступающий через входной компонент (интерфейс, канал).

К *ближней среде* относится та часть среды, которая существенно воздействует на систему и/или подвергается существенному воздействию со стороны системы. Также средой могут быть другие процессы или системы, в том числе более высокого уровня. С этой точки зрения внешний вход (X) трактуется как причина, а внешний выход (Y) – как следствие взаимодействия. Выделение «ближней среды» среди всех объектов ПрО с точки зрения общей теории систем есть представление выделенного подмножества в виде образования, целостного по отношению к преобразующей составляющей.

Интерфейсные операции (*EI*) создают условия для согласования и «доставки» (передачи), реализуя восприятие или изменение среды. Это может быть активное воздействие на систему, или активное изменение системой параметров, характеризующих состояние среды, или приведение к необходимой форме ресурсов и сигналов.

Блоки контура управления на основе принципа обратной связи обеспечивают измерение получаемых результатов, установление их соответствия заданным критериям, а также выработку управляющего воздействия. Они реализуют:

1) оперативное (*C*) управление, обеспечивающее устойчивость процесса получения «выхода» – целевого результата (*O*) операции (*FA*);

2) стратегическое (*E*) управление, обеспечивающее получение практического результата (*Y*).

Необходимость¹³ и достаточность двух¹⁴ контуров управления объясняется с позиции принципа наименьшего действия. Первый контур «оперативного» управления замыкает преобразование, «изолируя» его от изменяющегося во времени воздействия среды. Второй контур «стратегического» управления через выбор базиса влияет на взаимодействие со средой, определяет степень и границы применимости выбранного решения, а также анализирует актуальные или потенциальные противоречия внедрения получаемого объекта в структуру существовавших ранее.

Отметим, что приведенные на рис. 4 методы и средства (*U*), на основе которых построен преобразователь, а также условия и теории построения и функционирования элементарного процесса (*H*) относятся

¹³ В логике (анализа парадоксов) Д.А. Бочваром еще в 40-х гг. XX в. была предложена идея различия внутренних и внешних связей и, следовательно, необходимость двух уровней языков – внутреннего, в котором выражаются утверждения, но нет средств доказательства, и внешнего, в котором доказываются утверждения.

¹⁴ Здесь не рассматриваются многоуровневые (иерархические) системы управления, связанные с задачами оптимального распределения и координации целей.

к цепочке жизненного цикла создания преобразователя (*FA* как результата деятельности), а не жизненного цикла создания целевого продукта, в рамках которого преобразователь используется как функция.

Семантическая структура представления факта

Описывая машинное представление онтологий и их построение на основе текстовых форм информации, необходимо понимать, что они должны быть независимы от языка документа. Более того, связи и сущности отображаемой действительности при всем их разнообразии должны быть представлены в хорошо структурированном виде и в форме, приемлемой для машинной обработки (т.е. заданы хорошо определенными *сопоставимыми* величинами).

Будем считать минимальным событием (элементом действия, поведения) *факт*. Определим его как акт (элементарный процесс) взаимодействия двух и более объектов, зафиксированный через значения свойств объектов и связей. Обобщенная семантическая структура представления факта, данная в Gellish [2], приведена на рис. 5, где *роль* – это функциональное назначение, которое объект (нечто, anything) выполняет в данной ассоциации по отношению к связи, или наоборот, связь по отношению к объекту [26, 27]. Кроме того, поскольку у пары объектов вследствие их комплексности может быть несколько видов связей даже в одном процессе, назначение роли – это конкретизация функциональности взаимосвязи, исходя из выбранного аспекта рассмотрения (предмета, модели).

Роль выступает как форма *объективизации* взаимосвязи. Это особенно заметно, например, в случае отношения «вход» (см. рис 4), когда «поглощение» входного объекта для обработки является скорее событием, чем действием, и его очевидно удобнее идентифицировать ролью «исходный материал».

Схема на рис.5 является обобщением ER-модели. Это объясняется, во-первых, тем, что один и тот же объект (как макрообъект действительности и учетная единица в информационной модели) может участвовать более чем в одном элементарном акте, но, скорее всего, разными сторонами-свойствами и разными способами; во-вторых, представлен не только вещный уровень, но и уровень *знаний* – концепты и их взаимосвязи, соответствующие объектам вещного уровня.

Представление (схема определения) факта включает контекст, позволяющий интерпретировать конкретную взаимосвязь объектов, составляющих данный факт ПрО. Такой контекст образуют существенные (для объекта и ситуации) признаки, определяемые указанием на природу объектов через их соотнесение с классами, каждый из которых, в свою очередь, может соотноситься с суперклассом. Отметим, что указание на класс отношения (когда отношение является объектом рассмотрения) позволяет верифицировать его корректность через соотнесение свойств связываемых объектов, а также выявлять другие возможные свойства через взаимосвязи классов.

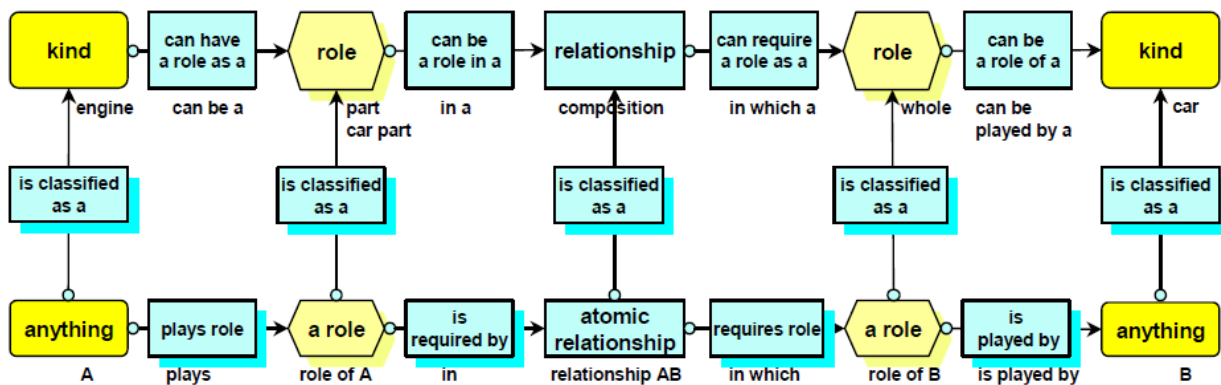


Рис. 5. Обобщенная семантическая структура представления факта [2].

К ОПРЕДЕЛЕНИЮ ОНТОЛОГИЧЕСКОГО ПРЕДСТАВЛЕНИЯ

Отметим несколько положений, важных с точки зрения конструктивности определения онтологий.

1. Практически все известные авторам определения включают следующие категории: объект, понятие, отношение, атрибут, класс. Однако при этом они имеют разную структуру и аксиоматику. И, как ни странно, имея информационную природу (реализация отображения «оригинал -> образ») и знаковую форму, ни одно определение не ассоциируется явно с семиотическим треугольником.

2. Формальные определения онтологий имеют алгебраическую форму, что неявно предполагает использование закрытых систем. Но онтология, как средство *разделения и повторного использования знаний для синтеза нового знания*, должна быть открытой системой, т.е. есть определение будет основываться на понятии системы как целого, обладающего свойством эмерджентности и допускающего разные членения.

3. «Архитектура» онтологии призвана обеспечивать технологичность процессов ее создания, развития и адаптации.

Исходя из этих положений, в [28] онтология (точнее, онтологическое пространство) была определена как совокупность трех взаимосвязанных систем:

$$O = \langle S_f, S_c, S_t \equiv \rangle,$$

где: S_f – функциональная система (объекты и связи действительности), определяемая как $S_f = \langle M_f, A_f, R_f, Z_f \rangle$, где M_f – множество объектов¹⁵ ПрО, A_f – множество характеристических атрибутов, R_f – множество функциональных отношений, Z_f – закон композиции – правила упорядочения (таксономия) объектов ПрО;

S_c – понятийная система, определяемая как $S_c = \langle M_c, A_c, R_c, Z_c \rangle$, где M_c – множество понятий ПрО, A_c – множество атрибутов, R_c – отношения

родо-видовые и эквивалентности, Z_c – закон композиции (концептуальный каркас);

S_t – терминологическая система, определяемая как $S_t = \langle M_t, A_t, R_t, Z_t \rangle$, где M_t – множество терминов ПрО, A_t – множество атрибутов, R_t – отношения эквивалентности и включения, лингвистические отношения, Z_t – закон композиции – правила и ограничения построения выражений;

\equiv – операция сопоставления¹⁶ элементов различных систем, обеспечивающая их тождество в функциональной, понятийной и терминологической системах на уровне знаков.

Определенная таким образом онтология также обладает свойствами системы. Для отдельной предметной области (в том числе и отдельного её элемента) может быть построено столько онтологий, сколько может быть аспектов рассмотрения ПрО¹⁷.

Очевидно, что в качестве функционального компонента может выступать система понятий ПрО. В этом случае её понятийной системой будет система метапонятий (это отвечает, например, определению «метаонтологии» [29], где на верхнем уровне есть лишь три категории: «объект», «процесс» и «роль»).

Выделение трех отдельных целостных структур, каждая из которых имеет своё системное основание, выглядит избыточным по отношению к структуре информационной модели ISO-15926, где онтологии представляются категориями понятия, отношения, роли, класса, отдельного экземпляра. Действительно, система стандартов ISO-15926 обеспечивает хорошо определенные механизмы даталогического уровня. Унификация формы представления сущностей, связей и атрибутов позволяет хранить их как однотипные элементы данных, а теоретико-множественное основание позволяет использовать алгебраические конструкции для построения запросов и выводов. Однако всеобщая унификация и сведение разноплановой информации в одно логическое пространство кажется некорректной и не всегда конструктивной [30].

¹⁵ Объекты, понятия и термины онтологии представлены их знаковыми эквивалентами (образами). При этом множества знаков, обозначающих объекты и понятия, могут пересекаться, т.е. один и тот же знак может использоваться для обозначения как некоторого понятия, так и объекта функциональной системы.

¹⁶ Так как S_f, S_c, S_t являются развивающимися системами (могут иметь разные состояния в момент сравнения) в операциях над онтологиями они должны быть приведены к общему основанию. Для чего используются операции над онтологиями [28] и мера подобия, предложенная в [31].

¹⁷ Это определение расширяет определение, данное в [9].

Но также важно отметить, что «разведение» объектов разной природы автоматически создает условие (без фильтрации на понятия и экземпляры, которые могут иметь и одно имя) использования статистических методов поиска информации и выявления зависимостей.

Для учета семантики и прагматики определения (выделения) и использования сущностей и связей обычно вводятся соответствующие свойства (указатели роли и т.п.). Однако для построения выводов (особенно в случае их автоматического вычисления) необходимо, чтобы связи, используемые в выражениях, не только отражали функциональность, но и логически согласовывались¹⁸, т.е. относились бы к одному аспекту, были взаимодополняющими и т.д.

Чтобы обеспечить возможность построения целесообразных и неравноценных комбинаций на множествах (изначально равноценных) сущностей, связей, атрибутов, необходимо задать порядок. Это означает, что наряду с иерархиями классов надо зафиксировать классы признаков (определяющих содержание понятий), задающих соответствующие таксономии объектов (объемы понятий).

Исходя из основного свойства систем – их потенциальной членимости – любую совокупность реальных или мыслимых вещей (сущностей) будем рассматривать как множество множеств $\{M_i\}$, для каждой пары которых установлено много-многозначное отображение $R^{ab} : M_i^a \rightarrow M_i^b$, где M_i^b – выполняет роль диагностического множества¹⁹ (характеристических атрибутов в определении онтологии с точки зрения ОТС), т.е. при выборе пары мы получаем множество отношений пар элементов этих множеств. И, поскольку определение системы подразумевает не единственность членения, то может быть определено множество диагностических множеств, индуцирующих в системе множество базисных отношений. Некоторое выбранное членение системы с базисными отношениями будет составлять *представление системы*.

Членения будут однородными, если M_i^a и M_i^b имеют единую природу, и неоднородными, если M_i^a и M_i^b имеют разную природу. Первый случай соответствует ситуативным отношениям между объектами, второй – имманентным, где наиболее типичным являются отношения «род-вид» («класс-подкласс»). Аналогично, рассматривая свойства объектов, обычно имеем в виду их концепты (признаки, схемы, отражающие существенные свойства), т.е. диагностические множества M_i^b и множества объектов – членения M_i^a будут иметь разную природу. Соответственно, рассматривая действия над отношениями, необходимо иметь в виду операции над их концептами. Это означает, что может быть построена сигнатура, обладающая алгебраической структурой [32], предопределяющей возможность построения выводов, аналогично методам логики (индукция, дедукция, традукция, абдукция). Такое упорядоченное

¹⁸ Собственно логика согласования, так или иначе, соотносится со схемой построения знания, отражающей целенаправленную последовательность отбора и связывания объектов, что в итоге и обеспечивает синтез новых свойств.

¹⁹ Это соответствует приведенному выше принципу деления на «практику» и «моделирование».

множество диагностических признаков, характерных для предметной области, образует ее концептуальный каркас²⁰ или, согласно [33], представляет собой мерономию.

Вследствие локальной ограниченности процесса субъекта познания, «организация» конкретного процесса анализа/синтеза знания (рационального, в том числе и научного) основывается на принципе локального детерминизма. При таком подходе выделяется и принципиально фиксируется основание (парадигма, аксиоматика, методология и т.п.), которое остается неизменным²¹ во всей предметной области до появления неразрешимых противоречий, предопределяющих смену основания и пересмотр теории. В рамках семиотики таким основанием является система понятий, представляемая тезаурусом²². И именно устойчивость тезауруса, как определенным образом организованной информации (рассматриваемого как самоотображение системы), гарантирует для развивающейся ПрО «защиту» от случайных ошибок сочетания (связывания) элементов в процессе синтеза. При этом как сама система, так и среда не должны быть изолированы от изменений. В противном случае это исключит возможность адаптации.

Знаки, представляющие понятия тезауруса, как и сообщения об объектах ПрО – это реальные объекты. Они относятся к тому же пространству²³, что и описываемые предметы. Объекты и связи, относящиеся к сфере предметной области и к понятийной системе, представлены знаками одного универсума. Поэтому для разграничения необходимо либо непосредственно идентифицировать их принадлежность к сфере отдельным свойством, либо выделять и использовать разные пространства их существования.

Необходимо отметить еще один фактор, принципиальный с точки зрения когнитологии. Сведение объектов и связей, относящихся к сфере ПрО и к понятийной системе, в одно пространство без возможности их разграничения фактически уничтожает диа-

²⁰ Для определения концептуального каркаса в [34] используется следующий подход. Пусть $B(F(di)) = \{\emptyset, 2^{F(di)}\}$ – булеан содержания $F(di)$ (множества признаков) некоторого понятия $di \in d$, имеющего объем $V(di) = \{v_{iq}\}$, где v_{iq} – объекты, имеющие признаки $F(di)$, $2^{F(di)}$ – множество всех подмножеств содержания $F(di)$. Элементы булеана образуют частично упорядоченное множество по включению его элементов, т.е. решетку $K(di) = (B(F(di)), \leq, \wedge, \vee)$. Любой элемент $F(di^H) \in B(F(di))$, $H \in [1, 2^{F(di)}]$ полученной решетки формально есть содержание некоторого понятия di^H , обобщающего понятие di , если $F(di^H) \subseteq F(di)$.

Решетку всех подмножеств содержания понятия di будем называть концептуальным каркасом.

²¹ Именно неизменность основания позволяет сделать процесс контролируемым – получать отдельные блоки, результаты которых могут быть оценены на состоятельность, и таким образом обеспечивать конечность и рациональность процесса.

²² Тезаурус понимается в широком смысле, хотя обычно ассоциируется с информационно-поисковым тезаурусом.

²³ Это достаточная причина для естественного использования единого пространства.

лектический дуализм методологической основы синтеза нового знания.

Аналогично, для определения целесообразности «разграничения» парадигматических и синтагматических отношений, традиционно фиксируемых в тезаурусе, рассмотрим их природу. Парадигматические связи отражают схему соотношения понятий как абстрактных логически взаимосвязанных объектов. Процесс построения понятийной системы осуществляется по схеме выделения подмножеств. Он основан на различении сходных по характеристическим признакам объектов-понятий, принадлежащих классу.

Синтагматические связи отражают свойства взаимодействия (или соотношения), по крайней мере, двух объектов между собой. Связи проявляются через «соединение по валентным свойствам». Выявляется признак – факт наличия соотношения некоторого типа между объектами, необязательно относящимися к одному классу.

Отсюда следует разноплановость («ортогональность») использования связей при расширении (обогащении) поисковых запросов, основанном на свойстве транзитивности. Парадигматические отношения обеспечивают общность признаков самого понятия, получая множество более или менее точных понятий, соответствующих одному объекту. Синтагматические же связи в лучшем случае могут обеспечивать общность, «валентных» признаков, определяющих, скорее тип связи, но не свойства объекта.

Процесс построения классификационных признаков, лежащих в основе парадигматики, гомоморфен процессу, ориентированному на выявление общей основы и универсальных законов синтеза рациональных знаний. Поэтому можно утверждать, что он имеет тенденцию к сходимости.

Напротив, комбинативная основа возникновения ситуативных связей (когда взаимодействия порождают новые объекты и, соответственно, потенциально новые связи) приводит к росту их числа и разнообразия. Это обуславливает расходящийся характер процесса появления новых экземпляров и типов ситуативных (и, соответственно, синтагматических) связей.

Таким образом, основу понятийной системы должны составлять парадигматические отношения, обеспечивающие устойчивость. Синтагматические же связи полезны и необходимы при использовании тезауруса, а также для его развития.

Следует учитывать и то, что выявление и фиксирование связей осуществляется на разных этапах, связанных с разными видами деятельности. Парадигматика возникает в процессе аналитической деятельности, связанной с синтезом понятий. Синтагматика же отражает практику использования их самих или соответствующих им объектов и может включаться в состав тезауруса после выбора основных понятий и определения их парадигматики.

Исходя из этого, можно сказать, что при машинном хранении и использовании тезаурусов (обычно в виде *n*-арных отношений) нецелесообразно вместе хранить парадигматические и синтагматические связи дескриптора. Хотя при ручном использовании и

экспертном обновлении тезаурусов их «совмещенное» представление связей в составе дескрипторных статей, безусловно, удобно.

КОМПОНЕНТЫ ОНТОЛОГИИ ВЕРХНЕГО УРОВНЯ

Онтологии верхнего уровня строятся на восходящем к Аристотелю, Канту, Гегелю, Пирсу, Уайтхеду гносеологическом подходе, включающем основные категории реальности (бытия). Традиционно к философским общесистемным категориям познания относят следующие пары взаимодополняющих компонентов (категорий): *материя – энергия, пространство – время, дискретное – непрерывное, сущность – явление, конкретное – абстрактное, сознательное – бессознательное, содержание – форма, общее – частное, всеобщее – особенное, единичное – множественное, простое – сложное, необходимость – случайность, количество – качество*²⁴.

Построение (выделение и упорядочение) таксономии сущностей (иерархии классов объектов) основывается на системе характеристических признаков – существенных, имманентных свойств сущностей, представляемых, в частности, концептуальной решеткой.

Исходя из соотношения «модель-действительность», расширим концептуальную решетку базовых абстракций онтологии верхнего уровня Дж. Соуи [35]. Введем категорию *Осознаваемое*²⁵, рассматриваемую как информационный образ результата отображения конкретного объекта на некоторую модель, что позволяет вводить измерения.

В предлагаемом варианте используются следующие категории-признаки: *Реальное/Абстрактное/Осознаваемое, Отдельное/Односвязное/Многосвязное, Статика(покой)/Динамика(движение)*, которые образуют содержание начального понятия «сущее» и образуют концептуальный каркас онтологии верхнего уровня (табл. 1).

Таким образом, появляются объекты (свойства, ассоциации, поведение и т.д.), имеющие свойства информации – образы, которые при своей изначальной условности имеют реальную форму, позволяющую их сохранять, находить и на их основе создавать новые, в том числе реальные объекты.

Таксономия сущностей

На основе концептуального каркаса была создана таксономия категории СУЩЕЕ (TOTALITY). Она основана на обобщении и развитии результатов практических исследований по построению онтологий верхнего уровня, в первую очередь – это проекты BFO, DOLCE, SUMO, SOWA, GELLISH, Hozo, а также многочисленные публикации, среди которых необходимо выделить [39–43]

²⁴ «Предположение о том, что существует конечный перечень категорий, является мифом [36, с. 323-338]. Необходимо обоснование системы категорий бытия и познания [37].

²⁵ Это, в частности, отвечает трем основным абстракциям: «вещь», «свойство», «отношение» [38].

Концептуальный каркас онтологии верхнего уровня

	Реальное		Осознаваемое (абстрактно-конкретное)		Абстрактное	
	Статика	Динамика	Статика	Динамика	Статика	Динамика
Отдельное	Компонент		Проявление		Форма (базис)	
	элемент	действие	свойство, параметр	явление, событие	понятие	операция
Односвязное	Взаимосвязь		Акт		Утверждение	
	связка, узел	влияние, участие	факт, ассоциация	отд. преобразование	класс, отношение	функция, зависимость
Многосвязное	Конкретная система		Система		Теория	
	агрегат, блок	процесс, деятельность	структура, схема	поведение, технология	сущность	метод, алгоритм

Построенная таксономия, наряду с общепринятыми, включает классы абстрактно-конкретных сущностей, имеющих условную опосредованную природу, как например, *поведение*, а также условно определяемые сущности, как *граница тела* [44, 45]. Кроме того, исходя из дуализма, следует различать классы стабильных во времени объектов – *продолжающиеся* и *возникающие* – явления и процессы. Также различаются дискретные и непрерывные объекты [46].

Таксономия отношений

Исходя из цели построения информационной модели сложного объекта по текстам, необходимо отметить решающую роль идентификации отношений²⁶, т.е. следует определить классы отношений и далее выявлять и типизировать предикатную лексику, с помощью которой они могут быть выражены в текстах на естественном языке.

Однако в этом направлении пока нет однозначных решений. Прежде всего, здесь надо отметить достижения лингвистов (Ю.Д. Апресян, Л.М. Васильев, В.В. Виноградов, Ю.С. Маслов, Е.В. Падучева, З. Вендлер, У. Чейф, Ч. Филмор, У. Кук, Б. Левин). Например, в одной из первых работ в этом направлении [47] выдвигается тезис, что существует шесть актантных ролей, достаточных для описания все возможных отношений, а также любой деятельности, объединенных в три пары: *субъект-объект*, *отправитель-получатель*, *помощник-оппонент*. Однако, поскольку они ориентированы на специфику естественных языков, то в выделяемых классификациях отсутствуют четкие границы между классами. Ведь классы выделяются за счет обстоятельств, а не характеристических признаков.

С другой стороны, существует ряд практических онтологий (DOLCE, SUMO, YATO, Gellish, проекты

FrameNet, VerbNet), а также публикаций [46, 48, 52], которые представляют разные таксономии отношений. Однако ни одно из известных решений не может считаться обоснованным и пригодным для информационного моделирования, так как классификационные признаки явно не обозначены, разнообразны и не систематизированы. Это в итоге не позволяет «вычислять» с помощью операций над онтологиями возможные зависимости.

Как ранее отмечалось, возможное решение проблемы идентификации отношений состоит в типизации, основанной на предположении, что классы будут заданы в предопределенном контексте, т.е. будет задан концептуальный каркас.

Для выявления и типизации отношений будем рассматривать состав, структуру и поведение объекта в контексте его жизненного цикла от появления абстрактной идеи объекта до его материального воплощения и использования. Также будем учитывать такой существенный фактор, как дуализм «абстрактное-конкретное». Схема отношений «абстрактное-конкретное» представлена на рис. 6.



Рис. 6. Схема отношений «абстрактное-конкретное»

Конкретные объекты (вещи) ПрО находятся в обусловленной ситуации взаимосвязи (функциональной: взаимодействия, локативности, соотношения). При этом взаимодействия понятий (*абстрактное-абстрактное*) представляют теорию (модель ПрО), в соответствии с которой рассматриваются конкретные вещи. Связь между конкретным и абстрактным соответствует отношению «квалифицируется как», а в случае, когда объектом является понятие, связь будет соответствовать отношению «класс-суперкласс». Отметим, что семиотика элементарного факта (см. рис. 5) отвечает именно этой схеме.

²⁶ Под отношением будем понимать *логическую* (предикатно-ориентированную) форму спецификации взаимосвязи предметов, явлений и процессов в природе, обществе, мышлении. Естественной основой для возможности спецификации является тот факт, что взаимосвязь обнаруживается и потому может быть зафиксирована через свое проявление в форме силы, результата.

Концептуальный каркас онтологии отношений верхнего уровня

Основания деления		Реальное (Конкретное)	Абстрактно-конкретное	Абстрактное
Креативные	Отдельное	Создать / Уничтожить	Обнаружение / Потеря	Утверждение, Определение
Воздействие / Локативность	Отдельное – Отдельное	Воздействие, взаимодействие, перемещение	Зависимость, Сопоставление	Обозначение, Сравнение
	Отдельное – Целое	Присоединение / Размещение	Включение / Исключение	Принадлежность, Разложение
	Все части – Целое	Слияние / Разделение	Сборка / Разборка	Порядок, Распределение
	Целое – Целое	Сохранение / Изменение	Синтез / Декомпозиция	Отображение, Вывод

При этом функциональные отношения объекта деятельности (как для материальных, так и абстрактных сущностей) обобщенно представляются унифицированным функциональным блоком (см. рис. 4), связанным с сопряженными блоками и блоками других этапов жизненного цикла.

В табл. 2 представлена концептуальная решетка базовых абстракций для построения таксономии отношений верхнего уровня, где используются следующие категории-признаки:

- *Реальное/Абстрактное/Абстрактно-конкретное* – отражающие соотношение «реальность/модель»;
- *Отдельное/Целое*, отражающие соотношения *входа и выхода* (равные или неравные потоки), как комбинации соотношений *отдельного (часть)* и *агрегатного (целое)*.

Подразделение отношений на *креативные* и *трансформативные (воздействие / локативность)* достаточно условно и отражает лишь тот факт, что первые, в отличие от трансформативных, не имеют правой либо левой составляющей конструкции отношения (в соответствии со схемой «субъект – предикат – объект»).

Отметим, что выделение группы абстрактных отношений²⁷ связано с опосредованностью или системой понятий, т.е. искусственностью (абстрактностью) природы хотя бы одного объекта. Группа абстрактно-конкретных отношений представляет обусловленные (так или иначе, той или иной моделью) взаимодействия или соотношения. Сюда же включаются и аналитико-синтетические отношения, обусловленные той или иной моделью деятельности. Отношения группы *Целое – Целое* в отличие от отношений группы *Отдельное – Отдельное* соответствуют объектам, обладающим свойством эмерджентности (представляющими целостные системы). Это

предполагает, что взаимодействие может иметь и другие результаты помимо непосредственного.

Каждый подкласс (ячейка таблицы), в свою очередь, подразделяется по признаку «предмет»²⁸ (целевая направленность, форма проявления взаимосвязи), где выделяются следующие подклассы отношений:

- действие-ориентированные – воздействие на целевой объект (аналог силы, энергии);
- объект-ориентированные – действие для изменения состояния целевого объекта (аналог потока, материи, массы);
- результат-ориентированные – взаимодействие для получения «нового» объекта или состояния (аналог события, факта, рассматриваемого как соотношение).

В целом интенциональный подход, когда классы определяются через свойства, не только обеспечивает обоснованность, но и создает возможность использования свойства наследования, что позволяет перейти к *контролируемому* естественному языку.

ЗАКЛЮЧЕНИЕ

Согласно [53] «Информационная модель — модель объекта, представленная в виде информации, описывающей существенные для данного рассмотрения параметры и переменные величины объекта, связи между ними, входы и выходы объекта и позволяющая путём подачи на модель информации об изменениях входных величин моделировать возможные состояния объекта». Это определение уточняется и расширяется (хотя и неявно) в рамках методологии ISO-15926. При этом предполагается, что информационная модель объекта – это такой мультинабор данных и схем, который позволяет актуально характеризовать объекты и процессы проекта на всех этапах его жизненного цикла. Информационная модель объекта, согласно ISO-15926, Fiatech and POSC

²⁷ Примечательно, что группе абстрактных отношений не свойственна парность взаимобратных типов взаимосвязи, характерная для группы конкретных и абстрактно-конкретных.

²⁸ «Предмет» здесь по существу является моделью «осознания», определяющей характер восприятия связи (ее результата) субъектом.

Caesar Association, состоит (с точки зрения инструментария и технологий их построения) из двух составляющих: пространственно-временного расширения 3D-модели (структурно-функциональный подход) и диаграмм классов и схем (взаимосвязей), представляемых средствами OWL, RDFS (инфологический подход). По существу, ISO-15926 (включая GELLISH) предназначен для интеграции таких составляющих проекта, как техническая (информационные компоненты SAP, ERP-систем, АСУП-данные эксплуатации объекта, каталоги и т.п.), так и семантическая (документальная).

Таким образом, семантика – это не только концептуальные модели базы данных проекта, но и то, что является его «генетическим кодом: парадигмы и использованные теории, принятые допущения и судьбоносные случайности, которые явно не присутствуют в проектно-конструкторских документах, но могут быть найдены в научных статьях, учебниках и другой литературе. Именно эта информация позволяет выстраивать (восстанавливать) цепочки обоснований необходимых значений параметров тех или иных узлов и процессов (например, в случае нетождественной замены), а в пределе прогнозировать возможные запроектные состояния и связи.

Особенностью ситуации является то, что семантическая информация представлена по большей части в слабоструктурированной документальной форме. Семантической основой формализации смысла является хорошо определенная таксономия сущностей и связей. Как показано в работе, построение таксономии отношений возможно вследствие следующих положений, полученных на основе системного анализа процессов деятельности.

Первое положение. Эффективность рациональной деятельности достигается локализацией – специализацией видов деятельности (этапы, фазы и т.п.) жизненного цикла. При этом всем этапам свойственны типовые процессы и схемы, что в значительной степени, обусловлено общей структурой функциональной модели.

Второе положение. Исходя из функционально-кибернетического подхода разнообразие отношений, свойственных типовому элементарному процессу (факту), конечно и функционально определено.

Третье положение. Управляемость основывается на двухконтурности схемы рациональной деятельности, для которой характерны следующие признаки: 1) основной целевой функции и потокам соответствуют информационные объекты; 2) операции над информационными объектами позволяют формировать управляющие сигналы, изменяющие поведение процесса.

Таким образом, данное в статье обобщение подходов к представлению семантики сложных систем на основе структуризации и систематизации объектов и процессов ЖЦ позволяет не только обоснованно идентифицировать объекты и связи ПрО, но и соотносить по смыслу тексты, представленные разными вербальными выражениями²⁹, а также вычислять возможные состояния и зависимости ПрО.

Из статьи следует еще один важный вывод. Основной цифровой платформы любого проекта является датацентрический подход. Он означает, что с точки зрения семиотики информация, знания, факты, понятия – все, что связывается с понятием «смысл», в среде обработки представлено в форме знаков. Другими словами, если знак имеет смысл, то для этого знака есть конкретные данные – *контекст*, который, тоже представлен знаками. Причем такие конкретные данные устойчиво ассоциируют этот контекст с некоторым абстрактным или реальным объектом. Язык, основанный на знаковой системе, в своем существе комбинаторен и изменчив – один и тот же знак в разных ПрО или в разное время может иметь разное значение. Следовательно, для цифровых сред, где крайне нежелательна неединственность смысла знака, должны быть средства, обеспечивающие согласованное использование контекста этого знака во времени и пространстве, на всех стадиях и во всех видах деятельности. Такая возможность следует из взаимодополнительности представлений.

Подобно приведенному выше соотношению практической и модельной составляющих в целенаправленной деятельности, в информационной деятельности (формирующей семантическую среду) выделяются ситуационная и структурно-концептуальная (парадигматика) составляющие. Первая отражает комбинаторную природу языка, позволяющую выражать новые смыслы «старыми» знаками. Вторая – представляет базис, который обеспечивает устойчивость (преемственность) развития ПрО и возможность коммуникаций, в частности, «узнавания» нового за счет апелляции к общим смыслам. Первая формирует контекст через статистические связи совместного употребления знаков. Вторая – через парадигматику – иерархию понятий и соответствующих классу знаков. Она явно или неявно формируется на основе существенных свойств ПрО. При этом такие свойства представляют парадигму деятельности. Отсюда следует важная роль парадигматики – построение и дальнейшая актуализация понятийно-знаковой системы должна строго следовать архетипу (системе признаков) предметной области. И если статистика употребления знака (ситуативные связи) не соответствует парадигме (не «вписывается» в систему признаков), то это будет сигналом того, что парадигма уже не отвечает накопленным знаниям.

СПИСОК ЛИТЕРАТУРЫ

1. Chen P.P.-S. The Entity-Relation Model // ACM TODS. – 1976. – №1; Чен П. Пин-Шен. Модель «сущность-связь» // СУБД. – 1995. – № 3.
2. Van Renssen A. Gellish: A Generic Extensible Ontological Language. – Delft: Delft University Press, 2005. – 238 p.
3. Агроскин В., Левенчук А., Головков В. Онтология для инженерных данных // СУБД. – 2013. – № 6.

²⁹ Некоторые примеры соотнесения с помощью тезаурусов и типизированных отношений документов, тематически

близких, но практически не имеющих общей лексики, были приведены в [54].

4. Урманцев Ю.А. Общая теория систем: Состояние, приложения и перспективы развития // Сб. «Система, Симметрия, Гармония». – М., Мысль, 1988. – С. 38-124.
5. Щедровицкий Г.П. Проблемы логики научного исследования и анализ структуры науки: из архива Г.П. Щедровицкого. Т. 7. – М., 2004.
6. Hubka V., Eder W.E. Theory of technical systems. - Springer-Verlag, 1988.
7. Gruber Th. What is an Ontology. – URL: <http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>
8. Guarino N. Understanding, Building, and Using Ontologies. – URL: <http://ksi.cpsc.ucalgary.ca/KAW/KAW96/guarino/guarino.html>
9. Van Heijst G., Wielinga, B.J. Using Explicit Ontologies in KBS Development. – URL: <http://ksi.cpsc.ucalgary.ca/KAW/KAW96/borst/nodel6.html>
10. Бениаминов Е.М., Болдина Д.М. Система представления знаний Ontolingua – принципы и перспективы // Научно-техническая информация. Сер. 2. – 1999. – № 10. – С. 26-32.
11. De Roure D., Jennings N.R., Shadbolt N.R. The Semantic Grid: a Future e-Science Infrastructure. – URL: www.semanticgrid.org/documents/semgrid-journal/semgrid-journal.pdf
12. Голицына О.Л., Куприянов В.М., Максимов Н.В. Информационные и технологические решения в задачах управления знаниями // Научно-техническая информация. Сер. 1. – 2015. – № 8. – С. 1-12.
13. Максимов Н.В. Информация и знания: природа, концептуальная модель // Научно-техническая информация. Сер. 2. – 2010. – № 7. – С. 1-10.
14. Скороходько Э.Ф. Лингвистические проблемы обработки тезисов в автоматизированных информационно-поисковых системах // Вопросы информационной теории и практики. – 1974. – № 25. – С. 5-120.
15. Бодякин В.И. Механизм автоматического формирования информационной модели в информационно-управляющей системе, построенной на базе нейро-семантической парадигмы // Тр. конф. «Нелинейная динамика в когнитивных исследованиях-2011». – Н. Новгород, 2011. – С. 16-19.
16. Шрейдер Ю.А. Текст как информационная модель знания // Анализ знаковых систем. История логики и методологии науки // Тезисы докл. 9-е Всерос. совещ. (Харьков, окт. 1986). – Киев: «Наукова думка», 1986. – С. 96–97.
17. Штофф В.А. Роль моделей в познании. – Л.: ЛГУ, 1963. – 128 с.
18. Kitamura Y., Sano T., Namba K., Mizoguchi R. A functional concept ontology and its application to automatic identification of functional structures // Advanced Engineering Informatics. – 2002. – Vol. 16(2). – P. 145-163.
19. Johansson I. et al. Functional anatomy: A taxonomic proposal // Acta biotheoretica. – 2005. – Vol. 53, № 3. – P. 153-166.
20. Максимов Н.В., Окропишин А.Е., Окропишина О.В., Передеряев И.И. Использование технологии автоматизированного формирования понятийной структуры предметной области научного исследования в задачах управления научными кадрами // Вестник РГГУ. Серия «Управление». – 2011. – № 4 (66). – С. 175-185.
21. Разумовский О.С. От конкурирования к альтернативам. Экстремальные принципы и проблема единства научного знания. – Новосибирск: Наука, 1983. – 223 с.
22. Пенроуз Р. Тени разума. В поисках науки о сознании. Часть 1. Понимание разума и новая физика. – М., Ижевск: Институт компьютерных технологий, 2003. – 368 с.
23. Яблонский А.И. Модели и методы исследования науки // Философы России XX века. – М.: Едиториал УРСС, 2001. – 400 с.
24. Наумова Н.Ф. Целеполагание как системный процесс. – М.: ВНИИСИ, 1982. – 66 с.
25. Kitamura Y. et al. An ontology-based annotation framework for representing the functionality of engineering devices // Proc. of ASME IDETC/CIE 2006. – 2006.
26. Kozaki K., Sunagawa E., Kitamura Y., Mizoguchi R. Role Representation Model Using OWL and SWRL // Roles'07. Proceedings of the 2nd Workshop on Roles and Relationships in Object Oriented Programming, Multiagent Systems, and Ontologies. – Berlin, 2007. – P. 39-46.
27. Mizoguchi R., Sunagawa E., Kozaki K. and Kitamura Y. A Model of Roles within an Ontology Development Tool: Hozo // J. of Applied Ontology. – 2007. – Vol. 2, № 2. – P. 159-179.
28. Голицына О.Л., Максимов Н.В., Окропишина О.В., Строгонов В.И. Онтологический подход к идентификации информации в задачах документального поиска // Научно-техническая информация. Сер. 2. – 2012. – № 5. – С. 1-9.
29. Nirenburg S., Raskin V. Ontological Semantics. – Cambridge, 2004.
30. Barry Smith. Against Ideosyncrasy in Ontology Development // Formal Ontology and Information System / eds. Preprint version in B. Bennet, C. Fellbaum – Amsterdam: IOS Press, 2006. – P. 15-26.
31. Голицына О.Л., Максимов Н.В., Федорова В.А. К определению семантической близости на основе связей объединенного тезауруса // Научно-техническая информация. Сер. 2. – 2016. – № 6. – С. 30-44.
32. Борщев В.Б., Хомяков М.В. Об информационной эквивалентности баз данных // Научно-техническая информация. Сер.2. – 1979. – № 7. – С. 14-21.
33. Панова Н.С., Шрейдер Ю.А. Принцип двойственности в теории классификации // Научно-техническая информация. Сер. 2. – 1975. – № 10. – С. 3-10.
34. Кулинич А.А. Концептуальные каркасы онтологий слабоструктурированных предметных областей // Искусственный интеллект и принятие решений. – 2015. – № 4. – С. 31-41.

35. Sowa J.F. Ontology. – URL: <http://www.jfsowa.com/ontology/>
36. Райл Г. Понятие сознания: пер. с англ. – М.: Идея-Пресс, Дом интеллектуальной книги, 2000. – 408 с.
37. Омельченко В.В. Основы систематизации. – М.: КД «ЛИБРОКОМ», 2012. – 480 с.
38. Райбекас А.Я. Вещь, свойство, отношение как философские категории. – Томск: Изд-во Томского ун-та, 1977. – 243 с.
39. Mizoguchi R. Yet another top-level ontology: Yato // Proc. of the Second Interdisciplinary Ontology Meeting. – 2009. – P. 91-101.
40. Henrichs M.A. Conceptual Framework for Constructing Distributed Object Libraries using Gellish: Master's Thesis in Computer Science. – Delft University of Technology, 2009. – 111 p.
41. Borgo S., Franssen M., Garbacz P., Kitamura Y., Mizoguchi R., Vermaas P.E. Technical artifacts: An integrated perspective // Applied Ontology. – 2014. – № 9. – P. 217–235.
42. Mizoguchi R., Kitamura Y. A Functional Ontology of Artifacts // The Monist. – 2009. – Vol. 92, № 3. – P. 387-402.
43. Bennett B. Space, time, matter and things // ACM. Proceedings of the international conference on Formal Ontology in Information Systems. – 2001. – Vol. 2001. – P. 105-116.
44. Smith B. Fiat Objects // Topoi. – 2001. – № 20 (2). – P. 131-148 (Preprint).
45. Smith B., Varzi A.C. 'Fiat and Bona Fide Boundaries' // Philosophy and Phenomenological Research. – 2000. – Vol. 60. – P. 401–420.
46. Smith B., Grenon P. The Cornucopia of Formal-Ontological Relations // Dialectica. – 2004. – Vol. 58, № 3. – P. 279-296.
47. Греймас А.-Ж. Размышления об актантных моделях / Французская семиотика: от структурализма к постструктурализму. – М., 2000. – С. 153-170.
48. Апресян Ю.Д. Фундаментальная классификация предикатов / Языковая картина мира и системная лексикография. – М.: Языки славянских культур, 2006. – С. 76-110.
49. Васильев Л.М. Теоретические проблемы общей лингвистики, славистики, русистики: Сборник избранных статей. – Уфа: РИО БашГУ, 2006. – С. 171-175.
50. Kitamura Y., Mizoguchi R. Ontology-based systematization of functional knowledge // Journal of Engineering Design. – 2004. – Vol. 15(4). – P. 327-351.
51. Kitamura Y., Koji Y., Mizoguchi R. An Ontological Model of Device Function: Industrial Deployment and Lessons Learned // Journal of Applied Ontology (Special issue on "Formal Ontology Meets Industry"). – 2006. – Vol. 1, № 3-4. – P. 237-262.
52. Wood K. L. et al. A Functional Basis for Engineering Design: Reconciling and Evolving Previous Efforts // None. – 2002.
53. ГОСТ 34.003-90 Информационная технология. Комплекс стандартов на автоматизированные системы. Автоматизированные системы. Термины и определения. – М.: Изд-во СТАНДАРТ-ИНФОРМ, 2009.
54. Голицына О.Л., Максимов Н.В., Окропишина О.В., Строгонов В.И. Онтологический подход к идентификации информации в задачах документального поиска: практическое применение // Научно-техническая информация. Сер. 2. – 2013. – № 3. – С. 1-8.

Материал поступил в редакцию 08.12.17.

Сведения об авторе

МАКСИМОВ Николай Вениаминович – профессор, доктор технических наук, профессор кафедры Системного анализа Национального исследовательского ядерного университета «МИФИ», Москва
e-mail: NV-MAKS@YANDEX.RU

Задача линейного оценивания и информация в системах Больших Данных

Рассмотрена проблема трансформации процедуры оптимального линейного оценивания так, чтобы отдельные фрагменты исходных данных могли обрабатываться независимо и параллельно. Предложена форма представления промежуточной информации, позволяющая алгоритму извлекать такую информацию параллельно из каждого набора исходных данных, объединять ее и использовать для получения результата. Показано, что на построенном информационном пространстве индуцируется упорядочение, отражающее понятие качества информации.

Ключевые слова: Большие Данные, линейное оценивание, каноническая информация, распределенные системы сбора и обработки данных, алгебра информации, информационное пространство

ВВЕДЕНИЕ

Специфика обработки информации в системах Больших Данных была рассмотрена автором в работе [1]. В таких системах данные нередко собираются и хранятся распределенно и могут постоянно пополняться. Было показано, что для эффективной обработки распределенных данных ключевую роль играет возможность введения промежуточной формы представления информации, обладающей определенными алгебраическими свойствами. В настоящей работе исследуется задача линейного оценивания с точки зрения распределенных систем сбора и обработки информации.

Подходы, развиваемые в этой статье, могут быть полезны для многих прикладных задач, например, при сборе и обработке информации в крупномасштабных экспериментах, где данные собираются в многочисленных исследовательских центрах, разбросанных по всему Земному шару. Однако для нас основной интерес представляют особенности информационных пространств, возникающих при необходимости распределенной обработки данных. На построенном информационном пространстве естественным образом порождается алгебраическая структура, описывающая композицию отдельных фрагментов информации, и согласованное с ней отношение предпорядка, отражающее феномен качества информации.

ЛИНЕЙНЫЙ ЭКСПЕРИМЕНТ И ЗАДАЧА ЛИНЕЙНОГО ОЦЕНИВАНИЯ

Рассмотрим схему линейного измерения как представлено в [2, 3]

$$y = Ax + v, \quad (1)$$

где $x \in \mathcal{D}$ – неизвестный вектор евклидова пространства – объект измерения, $y \in \mathcal{R}$ – результат измерения, $A: \mathcal{D} \rightarrow \mathcal{R}$ – линейное отображение, описывающее ис-

кажения измерительной системы, и $v \in \mathcal{R}$ – случайный вектор шума с нулевым средним $Ev = 0$ и заданным ковариационным оператором $Dv = S$.

Ковариационный оператор случайного вектора $\mu \in \mathcal{R}$ является многомерным обобщением понятия дисперсии и определяется как

$$(D\mu)(z) = E\langle \mu - E\mu, z \rangle (\mu - E\mu)^1.$$

Несложно проверить, что ковариационный оператор случайного вектора μ является самосопряженным неотрицательно определенным оператором и его матрица в ортонормированном базисе представляет собой ковариационную матрицу координат вектора μ в этом базисе.

Таким образом, вся информация об измерении – это модель измерения, описываемая парой (A, S) и результат измерения y . Будем рассматривать здесь лишь измерения, в которых оператор S положительно определен, $S > 0$ и, следовательно, обратим. По сути это означает, что шум v возможен во всех направлениях, т.е., не существует собственного подпространства $\tilde{\mathcal{R}} \subset \mathcal{R}$ такого, что $v \in \tilde{\mathcal{R}}$ с вероятностью единица.

Кратко приведем постановку и решение задачи линейного оценивания, более детальное и общее рассмотрение которых можно найти в [2, 3].

Задача линейного оценивания неизвестного вектора x состоит в построении такого линейного отображения $R: \mathcal{R} \rightarrow \mathcal{D}$, что оценка $\hat{x} = Ry$ максимально близка к x . Формально, рассмотрим погрешность оценки $E\|Ry - x\|^2$.

¹ Здесь и далее $\langle \cdot, \cdot \rangle$ обозначает скалярное произведение.

$$\begin{aligned} \mathbb{E}\|Ry - x\|^2 &= \mathbb{E}\|R(Ax + v) - x\|^2 = \\ &= \|(RA - I)x\|^2 + 2\mathbb{E}\langle (RA - I)x, Rv \rangle + \mathbb{E}\|Rv\|^2 = \\ &= \|(RA - I)x\|^2 + \text{tr}RSR^*. \end{aligned}$$

В последнем равенстве мы воспользовались тем, что $\mathbb{E}v = 0$ и следующими свойствами: $\mathbf{D}(Rv) = S$ и $\mathbb{E}\|\mu\|^2 = \text{tr}\mathbf{D}\mu$ для случайного вектора μ с нулевым средним.

Поскольку в выражении для $\mathbb{E}\|Ry - x\|^2$ присутствует неизвестный вектор x , определим погрешность оценивания, обеспечиваемую оператором R , как

$$H(R) = \sup_{x \in \mathcal{D}} \mathbb{E}\|Ry - x\|^2.$$

Легко видеть, что если $RA \neq I$, то $\|(RA - I)x\|^2$ может принимать сколь угодно большие значения и, следовательно,

$$H(R) = \begin{cases} +\infty & \text{если } RA \neq I, \\ \text{tr}RSR^*, & \text{если } RA = I. \end{cases}$$

Таким образом, линейное отображение R обеспечивает конечную погрешность оценивания $H(R)$ тогда и только тогда, когда $RA = I$, что, как легко убедиться, равносильно несмещённости оценки $\hat{x} = Ry$, т.е. $\mathbb{E}Ry = x$. Однако существование R , для которого $RA = I$, равносильно тому, что ядро линейного отображения A тривиально: $\mathcal{N}(A) = \{0\}$ ².

Итак, задача линейного оценивания это задача условной минимизации:

$$\min_R \left\{ \text{tr}RSR^* \mid RA = I \right\},$$

разрешимая лишь если $\mathcal{N}(A) = \{0\}$.

Пусть $\mathcal{S}_{\mathcal{D}}$ – пространство всех самосопряженных операторов на \mathcal{D} . Определим частичный порядок на $\mathcal{S}_{\mathcal{D}}$ следующим образом:

$$P \geq Q \Leftrightarrow P - Q \geq 0.$$

Заметим, что tr является строго монотонным отображением из пространства самосопряженных операторов в вещественную прямую, а именно, если $P \geq Q$, то $\text{tr}P \geq \text{tr}Q$, а если, кроме того $P \neq Q$ то $\text{tr}P > \text{tr}Q$. Поэтому, рассмотрим задачу минимизации самого оператора RSR^* при условии $RA = I$. Докажем, что если $\mathcal{N}(A) = \{0\}$, то существует единственное линейное отображение R , доставляющее минимум оператору RSR^* .

² Будем обозначать ядро и образ линейного отображения $A: \mathcal{D} \rightarrow \mathcal{R}$, соответственно $\mathcal{N}(A) \subseteq \mathcal{D}$ и $\mathcal{R}(A) \subseteq \mathcal{R}$.

Условие $\mathcal{N}(A) = \{0\}$ равносильно тому, что оператор $A^*S^{-1}A$ – обратим. Очевидно, что

$$R = (A^*S^{-1}A)^{-1} A^*S^{-1}$$

удовлетворяет условию $RA = I$. Докажем, что среди всех линейных отображений \tilde{R} , таких, что $\tilde{R}A = I$, оператор $\tilde{R}S\tilde{R}^*$ достигает минимума при $\tilde{R} = R$. Пусть $C = \tilde{R} - R$, тогда $CA = (\tilde{R} - R)A = 0$. Отсюда имеем:

$$\begin{aligned} \tilde{R}S\tilde{R}^* &= \\ &= \left[(A^*S^{-1}A)^{-1} A^*S^{-1} + C \right] S \left[S^{-1}A(A^*S^{-1}A)^{-1} + C^* \right] = \\ &= (A^*S^{-1}A)^{-1} + (A^*S^{-1}A)^{-1} (AC)^* + AC(A^*S^{-1}A)^{-1} + \\ &+ CSC^* = (A^*S^{-1}A)^{-1} + CSC^*. \end{aligned}$$

Оператор $CSC^* \geq 0$, а из невырожденности S сразу следует, что если $C = 0$, то и $CSC^* = 0$. Следовательно, $\tilde{R}S\tilde{R}^*$ и $\mathbb{E}\|\tilde{R}y - x\|^2 = \text{tr}\tilde{R}S\tilde{R}^*$ достигают минимальных значений в единственной точке $\tilde{R} = R$.

Таким образом, задача оптимального оценивания вида

$$\min_{R: \mathcal{R} \rightarrow \mathcal{D}} \sup_{x \in \mathcal{D}} \mathbb{E}\|Ry - x\|^2$$

имеет решение тогда, и только тогда, когда $\mathcal{N}(A) = \{0\}$. При этом оценка

$$\hat{x} = Ry = (A^*S^{-1}A)^{-1} A^*S^{-1}y$$

является несмещенной оценкой вектора x , обладающей наименьшим ковариационным оператором

$$\mathbf{D}\hat{x} = RSR^* = (A^*S^{-1}A)^{-1}.$$

Отсюда, в частности, следует, что оценка $\hat{x} = Ry$ обладает минимальными дисперсиями координат \hat{x}_j в некотором ортонормированном базисе:

$$\mathbf{D}\hat{x}_j = \left((A^*S^{-1}A)^{-1} \right)_{jj} \quad \text{и} \quad \mathbb{E}\|\hat{x} - x\|^2 \text{ достигает минимального значения } \mathbb{E}\|\hat{x} - x\|^2 = \text{tr}(A^*S^{-1}A)^{-1}.$$

Итак, пусть заданы результат измерения y и модель измерения (A, S) . Тогда исходные данные для линейного оценивания представляются тройкой (y, A, S) и процедура обработки P состоит в преобразовании исходных данных в результат оценивания – оптимальную оценку \hat{x} вектора x :

$$(y, A, S) \xrightarrow{P} \hat{x} = (A^*S^{-1}A)^{-1} A^*S^{-1}y.$$

При этом отображение P определено не всюду, а лишь тогда, когда оператор $A^*S^{-1}A$ обратим.

ЛИНЕЙНОЕ ОЦЕНИВАНИЕ В СЛУЧАЕ МНОГИХ НЕЗАВИСИМЫХ ИЗМЕРЕНИЙ

Пусть теперь имеется много независимых измерений одного и того же неизвестного вектора $x \in \mathcal{D}$:

$$y_i = A_i x + v_i, \quad i = 1, \dots, n \quad (2)$$

где $y_i \in \mathcal{R}_i$ – результаты измерений; $A_i: \mathcal{D} \rightarrow \mathcal{R}_i$ – линейные отображения; $v_i \in \mathcal{R}_i$ – независимые случайные векторы с нулевыми средними $E v_i = 0$ и ковариационными операторами $D v_i = S_i: \mathcal{R}_i \rightarrow \mathcal{R}_i$. В общем случае пространства измерений \mathcal{R}_i могут быть различными.

Чтобы воспользоваться результатом предыдущего раздела для серии измерений (2), представим эту серию в виде одного измерения вида (1). Рассмотрим сначала случай двух независимых измерений (2), т.е., $n = 2$.

Пару (y_1, y_2) векторов $y_1 \in \mathcal{R}_1$ и $y_2 \in \mathcal{R}_2$ можно рассматривать как элемент произведения евклидовых пространств $\mathcal{R}_1 \times \mathcal{R}_2$. Такую пару удобно записывать

в виде столбца $\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \in \mathcal{R}_1 \times \mathcal{R}_2$. Линейные операции на $\mathcal{R}_1 \times \mathcal{R}_2$ определяются покомпонентно через операции на \mathcal{R}_1 и \mathcal{R}_2 , а скалярное произведение как

$$\left\langle \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right\rangle = \langle y_1, z_1 \rangle + \langle y_2, z_2 \rangle \quad \text{для любых } y_1, z_1 \in \mathcal{R}_1$$

и $y_2, z_2 \in \mathcal{R}_2$.

Пара линейных отображений $A_1: \mathcal{D} \rightarrow \mathcal{R}_1$ и $A_2: \mathcal{D} \rightarrow \mathcal{R}_2$, действующих из одного и того же пространства \mathcal{D} , взаимно однозначно определяются линейным отображением

$$\begin{pmatrix} A_1 \\ A_2 \end{pmatrix}: \mathcal{D} \rightarrow \mathcal{R}_1 \times \mathcal{R}_2, \quad \text{действующим по правилу } \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} x = \begin{pmatrix} A_1 x \\ A_2 x \end{pmatrix} \quad \text{для любого } x \in \mathcal{D}.$$

Аналогично, пара линейных отображений $B_1: \mathcal{R}_1 \rightarrow \mathcal{Q}$ и $B_2: \mathcal{R}_2 \rightarrow \mathcal{Q}$, действующих в одно и то же пространство \mathcal{Q} , взаимно однозначно определяются линейным отображением

$$(B_1 \ B_2): \mathcal{R}_1 \times \mathcal{R}_2 \rightarrow \mathcal{Q}, \quad \text{действующим по правилу } (B_1 \ B_2) \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = B_1 y_1 + B_2 y_2 \quad \text{для любых}$$

$y_1 \in \mathcal{R}_1$ и $y_2 \in \mathcal{R}_2$. Манипулирование такими матрицами подчиняется обычным правилам матричной алгебры. Заметим, что сопряженной к матрице из линейных преобразований будет транспонированная матрица с сопряженными компонентами. Например,

$$\text{если } \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}: \mathcal{D} \rightarrow \mathcal{R}_1 \times \mathcal{R}_2,$$

$$\text{то } \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}^* = (A_1^* \ A_2^*): \mathcal{R}_1 \times \mathcal{R}_2 \rightarrow \mathcal{D}.$$

Наконец, для ковариационного оператора вектора $\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \in \mathcal{R}_1 \times \mathcal{R}_2$ с независимыми компонентами v_1 и v_2 имеем

$$\begin{aligned} \left(D \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \right) \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} &= E \left\langle \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}, \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \right\rangle \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \\ &= E \left(\langle v_1, z_1 \rangle + \langle v_2, z_2 \rangle \right) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \\ &= E \left(\langle v_1, z_1 \rangle v_1 + \langle v_2, z_2 \rangle v_1 \right) = \\ &= E \left(\langle v_1, z_1 \rangle v_2 + \langle v_2, z_2 \rangle v_2 \right) = \\ &= \begin{pmatrix} (D v_1) z_1 + 0 \\ 0 + (D v_2) z_2 \end{pmatrix} = \begin{pmatrix} D v_1 & 0 \\ 0 & D v_2 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \end{aligned}$$

для любого $\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \in \mathcal{R}_1 \times \mathcal{R}_2$.

$$\text{Следовательно, } D \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} D v_1 & 0 \\ 0 & D v_2 \end{pmatrix} = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix}.$$

Здесь мы воспользовались тем, что, в силу независимости v_1 и v_2 $E \langle v_1, z_1 \rangle v_2 = \langle E v_1, z_1 \rangle E v_2 = 0$ и $E \langle v_1, z_1 \rangle v_1 = 0$.

Таким образом, объединение исходных данных, отвечающих двум независимым измерениям, описывается операцией

$$(y_1, A_1, S_1) \cup (y_2, A_2, S_2) = \left(\begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}, \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix} \right) \quad (3)$$

При наличии n независимых измерений (2) требуется собрать соответствующие данные в одном месте, реорганизовать их в виде блочных матриц, возможно, очень больших размерностей и применить к объединенным данным отображение P (рис. 1).

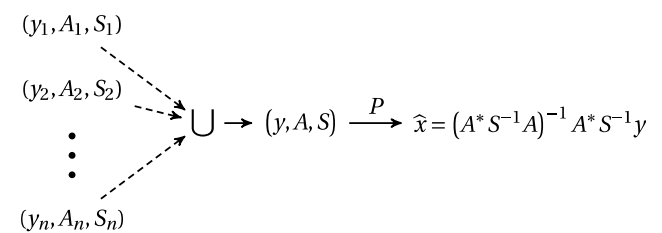


Рис. 1. Стандартная схема линейного оценивания для большого числа измерений

Здесь

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \in \mathcal{R}, \quad A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix}: \mathcal{D} \rightarrow \mathcal{R},$$

$$S = \begin{pmatrix} S_1 & 0 & \dots & 0 \\ 0 & S_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & S_n \end{pmatrix} : \mathcal{R} \rightarrow \mathcal{R},$$

$$\mathcal{R} = \mathcal{R}_1 \times \mathcal{R}_2 \times \dots \times \mathcal{R}_n, \quad \dim \mathcal{R} = \sum_{i=1}^n \dim \mathcal{R}_i.$$

При большом числе измерений размерность объединенных данных может стать крайне большой, в результате чего описываемый подход может оказаться практически нереализуем. Кроме того, добавление новых данных будет приводить к увеличению размерностей объединенных данных, что, в свою очередь, будет требовать все больше ресурсов для их хранения и обработки (применения преобразования P).

РАСПАРаллеливание обработки за счет выделения промежуточной информации

Покажем, что обработку данных в задаче линейного оценивания можно разбить на две фазы $P = P_2 \circ P_1$ первая фаза P_1 – выделение некоторой компактной промежуточной информации из исходных данных, вторая P_2 – вычисление результата оценивания на основании этой промежуточной информации. При этом, наша цель – найти такую факторизацию, что применение преобразования P_1 к объединенному набору данных может быть заменено параллельным применением P_1 к отдельным данным и последующему «сложению» полученных фрагментов информации.

Пусть (y_1, A_1, S_1) и (y_2, A_2, S_2) – два набора данных. Рассмотрим результат оценивания, отвечающий объединенному набору данных (y, A, S) , где $y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$, $A = \begin{pmatrix} A_1 \\ A_2 \end{pmatrix}$, $S = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix}$. Заметим, что для вычисления оптимальной оценки вектора x , а именно, $\hat{x} = (A^* S^{-1} A)^{-1} A^* S^{-1} y$ требуется вычислять выражения вида $A^* S^{-1} A$ и $A^* S^{-1} y$. Рассмотрим их для случая двух объединенных измерений.

Используя очевидное равенство

$$\begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix}^{-1} = \begin{pmatrix} S_1^{-1} & 0 \\ 0 & S_2^{-1} \end{pmatrix},$$

получаем

$$\begin{aligned} A^* S^{-1} A &= \begin{pmatrix} A_1^* & A_2^* \end{pmatrix} \begin{pmatrix} S_1^{-1} & 0 \\ 0 & S_2^{-1} \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = \\ &= \begin{pmatrix} A_1^* S_1^{-1} & A_2^* S_2^{-1} \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \end{pmatrix} = A_1^* S_1^{-1} A_1 + A_2^* S_2^{-1} A_2, \end{aligned}$$

и аналогично, $A^* S^{-1} y = A_1^* S_1^{-1} y_1 + A_2^* S_2^{-1} y_2$,

где $v_i = A_i^* S_i^{-1} y_i \in \mathcal{D}$, $T_i = A_i^* S_i^{-1} A_i : \mathcal{D} \rightarrow \mathcal{D}$ и T_i – неотрицательно определенный оператор. При этом объединенным данным будет отвечать пара (v, T) , в которой $v = v_1 + v_2$ и $T = T_1 + T_2$.

Это означает, что вся необходимая для дальнейшей обработки информация, относящаяся к i -му измерению, может быть представлена парой (v_i, T_i) ,

Будем называть пару $(v, T) = (A^* S^{-1} y, A^* S^{-1} A)$ **канонической информацией** для данных (y, A, S) , а множество \mathfrak{S} всех таких пар каноническим **информационным пространством** для задачи линейного оценивания вектора из пространства \mathcal{D} . Заметим, что $\mathcal{R}(A^* S^{-1}) = \mathcal{R}(A^* S^{-1} A) = \mathcal{N}^\perp(A)$ [2,3]. Следовательно, измерениям вида (y, A, S) могут отвечать лишь такие пары (v, T) , в которых $v \in \mathcal{R}(T)$. Таким образом,

$$\mathfrak{S} = \{(v, T) \mid T \in \mathbb{S}_\mathcal{D}^+, v \in \mathcal{R}(T)\},$$

где $\mathbb{S}_\mathcal{D}^+$ – множество неотрицательно определенных операторов на \mathcal{D} – выпуклый конус в линейном пространстве $\mathbb{S}_\mathcal{D}$ самосопряженных операторов на пространстве \mathcal{D} . Если $\dim \mathcal{D} = m$, то $\dim \mathbb{S}_\mathcal{D} = \frac{m(m+1)}{2}$. Тогда $\mathfrak{S} \subset \mathcal{D} \times \mathbb{S}_\mathcal{D}^+$ представляет

собой выпуклый конус в $\frac{m(m+3)}{2}$ -мерном линейном пространстве $\mathcal{D} \times \mathbb{S}_\mathcal{D}$. Отсюда, в частности, следует, что любой элемент информационного пространства \mathfrak{S} может быть задан набором из $\frac{m(m+3)}{2}$ вещественных чисел.

Очевидно, процесс линейного оценивания можно разбить на две фазы $P = P_2 \circ P_1$, где первая фаза P_1 состоит в построении канонической информации:

$$(v, T) = P_1(y, A, S) = (A^* S^{-1} y, A^* S^{-1} A),$$

а вторая фаза P_2 вычисляет результат оценивания на основании этой информации (рис. 2):

$$\hat{x} = P_2(v, T) = T^{-1} v.$$

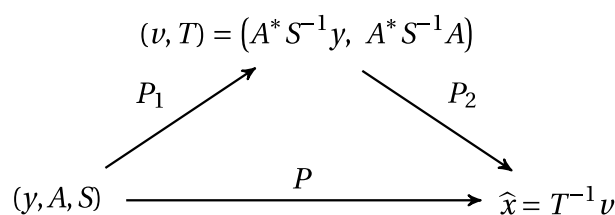


Рис. 2. Разбиение процесса обработки данных на две фазы

Как было показано, объединению исходных данных (y_1, A_1, S_1) и (y_2, A_2, S_2) отвечает композиция соответствующих элементов канонической информации (v_1, T_1) и (v_2, T_2) , определенная как $(v_1, T_1) \oplus (v_2, T_2) = (v_1 + v_2, T_1 + T_2)$.

Это можно записать как

$$P_1(y_1, A_1, S_1) \oplus P_1(y_2, A_2, S_2) = P_1((y_1, A_1, S_1) \cup (y_2, A_2, S_2)),$$

где под $(y_1, A_1, S_1) \cup (y_2, A_2, S_2)$ понимается определяемое выражением (3) – объединение двух наборов данных в один (рис. 3).

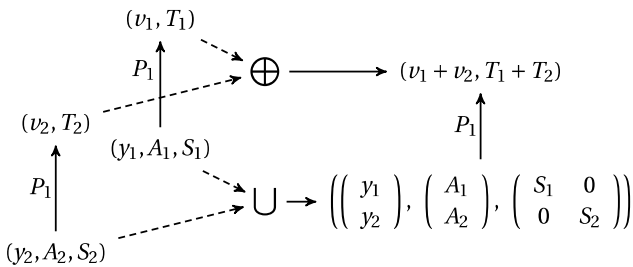


Рис. 3. Соответствие композиции фрагментов канонической информации и объединения наборов исходных данных

В результате факторизации алгоритма P на две фазы и введения канонической информации, схема обработки распределенных данных (см. рис. 1) может быть трансформирована следующим образом: из каждого отдельного фрагмента (y_i, A_i, S_i) данных выделяется каноническая информация (v_i, T_i) , которая впоследствии объединяется и используется для вычисления результата оценивания (рис. 4).

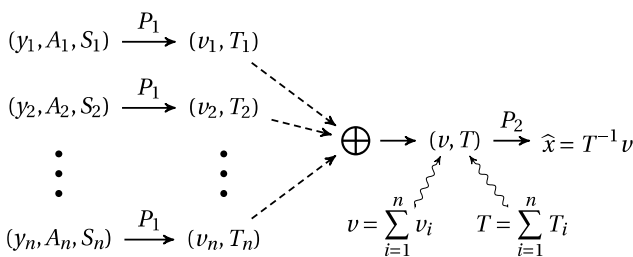


Рис. 4. Модифицированная схема обработки распределенных данных

Основные особенности такой модифицированной схемы.

1. Объем памяти, требуемый для хранения информации в каноническом виде не зависит от объема представляемых исходных данных и составляет $\frac{m(m+3)}{2}$ чисел (m -мерный вектор и самосопряженный оператор в m -мерном пространстве).

2. Выделение канонической информации (v_i, T_i) из i -го набора данных (преобразование P_1) может проводиться на тех компьютерах, где эти данные находятся, причем, параллельно и независимо.

3. Передаются лишь компактные фрагменты выделенной канонической информации одинакового объема.

4. Сложение частей канонической информации максимально упрощено и определяется покомпонентным сложением пар (v_i, T_i) .

5. Ресурсоемкость второй фазы P_2 , состоящей в построении результата по компактной накопленной информации (v, T) , определяется только размерностью m пространства неизвестных и не зависит от объема исходных данных.

6. По мере поступления новых данных, требуется лишь выделять из них каноническую информацию и «добавлять» ее к накопленной, при этом окончательную обработку P_2 будет необходимо снова применять к компактной информации фиксированного объема.

В результате, распределенность исходных данных способствует повышению эффективности обработки за счет естественного распараллеливания алгоритма.

В рассмотренной задаче линейного оценивания нашей целью было построение оценки \hat{x} , т.е. $P(y, A, S) = \hat{x}$. При построении оценки \hat{x} важно также охарактеризовать ее точность, исчерпывающую информацию о которой содержит ковариационный оператор $Q = D\hat{x} = (A^* S^{-1} A)^{-1} = T^{-1}$. В частности, он позволяет определить погрешности оценивания отдельных компонент вектора x (в произвольном ортонормированном базисе) $E(\hat{x}_j - x_j)^2 = D\hat{x}_j = Q_{jj}$ и полную погрешность оценивания $E\|\hat{x} - x\|^2 = \text{tr}Q = \sum_{j=1}^m Q_{jj}$. Таким образом, каноническая информация вида (v, T) подходит также для построения результатов оценивания $P(y, A, S)$ в виде $(\hat{x}, D\hat{x})$, $(\hat{x}, \{D\hat{x}_j\}_{j=1, \dots, m})$ или $(\hat{x}, E\|\hat{x} - x\|^2)$. В этих случаях отображение P_1 остается прежним, а отображение P_2 заменяется, соответственно, на

$$P_2(v, T) = (T^{-1}v, T^{-1}),$$

$$P_2(v, T) = \left\{ T^{-1}v, \left\{ (T^{-1})_j \right\}_{j=1, \dots, m} \right\}$$

или

$$P_2(v, T) = (T^{-1}v, \text{tr}T^{-1})$$

КАЧЕСТВО ИНФОРМАЦИИ И ИНФОРМАТИВНОСТЬ ИСТОЧНИКОВ ИНФОРМАЦИИ

Как было отмечено, ковариационный оператор $Q = D\hat{x} = (A^*S^{-1}A)^{-1} = T^{-1}$ полностью характеризует точность оценивания. При этом, чем меньше ковариационный оператор оценки \hat{x} , тем меньше погрешность оценивания: т.е. если $Q \leq \tilde{Q}$, то $Q_{ij} \leq \tilde{Q}_{ij}$ и $\text{tr}Q \leq \text{tr}\tilde{Q}$.

Будем говорить, что информация (v, T) не хуже (не менее точна), чем (\tilde{v}, \tilde{T}) , если $T \geq \tilde{T}$ и писать $(v, T) \succcurlyeq (\tilde{v}, \tilde{T})$. Если $(v, T) \succcurlyeq (\tilde{v}, \tilde{T})$ и $(\tilde{v}, \tilde{T}) \succcurlyeq (v, T)$, то будем говорить, что (v, T) и (\tilde{v}, \tilde{T}) имеют одинаковую точность и будем обозначать это $(v, T) \approx (\tilde{v}, \tilde{T})$. Очевидно, это равносильно условию $T = \tilde{T}$. Нетрудно заметить, что более точная информация обеспечивает более точное оценивание. Действительно, пусть $T \geq \tilde{T}$ и (v, T) и (\tilde{v}, \tilde{T}) позволяют построить соответствующие оценки, т.е. T и \tilde{T} – обратимы.

Согласно [2] отсюда следует, что $T^{-1} \leq \tilde{T}^{-1}$ и значит $Q \leq \tilde{Q}$, где Q и \tilde{Q} – ковариационные операторы соответствующих оценок.

Построение оценки вектора x возможно лишь если $\mathcal{N}(A) = \{0\}$. Если $\mathcal{N}(A) \neq \{0\}$, то часть вектора x , принадлежащая $\mathcal{N}(A)$, обнуляется и не может быть оценена. Однако, как показано в [2, 3], можно оценить проекцию Px вектора x на подпространство $\mathcal{N}^\perp(A)$. Здесь $P = A^-A : \mathcal{D} \rightarrow \mathcal{D}$ – ортогональный проектор на $\mathcal{N}^\perp(A)$, а $A^- : \mathcal{R} \rightarrow \mathcal{D}$ – линейное отображение, псевдообратное к $A : \mathcal{D} \rightarrow \mathcal{R}$ [2]. При этом оптимальная оценка вектора Px и ее ковариационный оператор определяются выражениями

$$\widehat{Px} = (A^*S^{-1}A)^- A^*S^{-1}y, D(\widehat{Px}) = (A^*S^{-1}A)^-$$

или, в терминах канонической информации,

$$\widehat{Px} = T^-v, D(\widehat{Px}) = T^-.$$

Более того, поскольку

$$\mathcal{N}(T) = \mathcal{N}(A^*S^{-1}A) = \mathcal{N}(A),$$

проектор P также может быть выражен через T , а именно, $P = T^-T$.

Таким образом, рассматриваемая нами каноническая информация содержит всю информацию, необходимую для построения оптимальной оценки и в этом более широком контексте.

Пусть информация (v, T) не хуже, чем (\tilde{v}, \tilde{T}) , т.е. $T \geq \tilde{T} \geq 0$. Тогда $\mathcal{N}(T) \subseteq \mathcal{N}(\tilde{T})$ и $\tilde{T}^- \geq \tilde{P}T^-\tilde{P}$ [2],

где $\tilde{P} = \tilde{T}^-\tilde{T}$ – ортогональный проектор на $\mathcal{N}^\perp(\tilde{T})$. Включение $\mathcal{N}(T) \subseteq \mathcal{N}(\tilde{T})$ влечет $\mathcal{N}^\perp(\tilde{T}) \subseteq \mathcal{N}^\perp(T)$, т.е. информация (v, T) позволяет оценить большую часть вектора x , чем (\tilde{v}, \tilde{T}) , а неравенство $\tilde{T}^- \geq \tilde{P}T^-\tilde{P}$ означает, что информация (v, T) обеспечивает более точное оценивание вектора $\tilde{P}x$. Иными словами, более точная информация обеспечивает более широкие возможности оценивания и при прочих равных условиях – менее интенсивный шум оценки.

Отметим, что рассмотренное нами понятие точности информации приводит к такому же упорядочению на множестве моделей линейного измерения, как и понятие качества моделей измерений в [3, 4] или информативности преобразователей информации в [5].

СВОЙСТВА КАНОНИЧЕСКОЙ ИНФОРМАЦИИ В ЗАДАЧЕ ЛИНЕЙНОГО ОЦЕНИВАНИЯ

Рассмотрим свойства канонического информационного пространства \mathfrak{S} , определенного выше. Эти свойства не только представляют самостоятельный интерес, но и могут выступать в качестве примера общих свойств информационных пространств, возникающих в задачах обработки больших объемов распределенных данных.

Существование для любого исходного набора данных. Информации, содержащейся в данных (y, A, S) , может быть недостаточно для построения результата оценивания. А именно, если ядро отображения A нетривиально, т.е. $\mathcal{N}(A) \neq \{0\}$, то оценка неизвестного вектора не может быть построена. Тем не менее, каноническая информация (v, T) может быть построена для любых исходных данных. Отметим, что даже полное отсутствие измерений (несущее нулевую информацию) может быть представлено в каноническом виде. Формально, любое измерение (y, A, S) , в котором $A = 0 : \mathcal{D} \rightarrow \mathcal{R}$ – нулевое отображение, не несет никакой информации об измеряемом векторе. Любому такому измерению отвечает каноническая информация $\mathbf{0} = (0, 0)$, т.е. $v = 0 \in \mathcal{D}$ и $T = 0 : \mathcal{D} \rightarrow \mathcal{D}$.

Полнота (или достаточность). Каноническая форма содержит всю информацию, содержащуюся в исходных данных, а именно, она приводит к тому же результату, что и исходные данные, из которых она получена. Формально это означает, что $P(y, A, S) = P_2(P_1(y, A, S))$ для всех данных (y, A, S) из области определения преобразования P . Это свойство напоминает понятие достаточности в математической статистике.

Операция композиции \oplus . На каноническом информационном пространстве \mathfrak{S} определена операция композиции \oplus , описывающая сложение фрагментов информации, отвечающих данным. При этом $(\mathfrak{S}, \oplus, \mathbf{0})$, является коммутативным моноидом, т.е. выполнены следующие свойства для любых $a, b, c \in \mathfrak{S}$:

$$a) \quad a \oplus b = b \oplus a;$$

$$b) (a \oplus b) \oplus c = a \oplus (b \oplus c);$$

$$c) a \oplus 0 = a.$$

Отметим, что моноид $(\mathfrak{S}, \oplus, 0)$ обладает также свойством сокращения:

$$d) a \oplus b = a \oplus c \Rightarrow b = c,$$

но не имеет обратимых элементов, отличных от 0 , т.е. не существует «отрицательной» информации.

Предпорядок \succcurlyeq , отражающий понятие точности информации. На каноническом информационном пространстве \mathfrak{S} определено отношение \succcurlyeq , обладающее следующими свойствами:

$$a) a \succcurlyeq a \text{ (рефлексивность);}$$

$$b) a \succcurlyeq b \& b \succcurlyeq c \Rightarrow a \succcurlyeq c \text{ (транзитивность).}$$

Заметим, что отношение \succcurlyeq не обладает свойством антисимметричности, т.е. не является частичным порядком. Действительно, из $(v, T) \succcurlyeq (\tilde{v}, \tilde{T})$ и $(\tilde{v}, \tilde{T}) \succcurlyeq (v, T)$ следует лишь, что $T = \tilde{T}$, но не обязательно $v = \tilde{v}$. Однако на классах эквивалентной точности это отношение антисимметрично, $a \succcurlyeq b \& b \succcurlyeq a \Rightarrow a \approx b$ и, следовательно, является отношением частичного порядка.

Кроме того, алгебраическая структура информационного пространства согласована со структурой порядка, а именно:

c) $a \succcurlyeq 0$. Любая информация точнее, чем отсутствие информации;

d) $a \oplus b = a, b$. Композиция двух фрагментов информации точнее, чем каждый из них по отдельности;

e) $a \succcurlyeq b \& c \succcurlyeq e \Rightarrow a \oplus c \succcurlyeq b \oplus e$. Чем точнее фрагменты информации, тем точнее результат композиции.

Единственность представления данных в каноническом виде. Несложно убедиться, что различные данные (y, A, S) могут описываться одной и той же канонической информацией (v, T) и, как следствие, приводить к одному и тому же результату оценивания. Может показаться, что данные (y, A, S) могут быть представлены разными парами вида (v, T) . Действительно, например, $(2v, 2T)$ приведет к тому же самому результату $\hat{x} = T^{-1}v$, что и пара (v, T) . Однако, пары (v, T) и $(2v, 2T)$ при композиции с другими элементами информационного пространства будут вести себя по-разному и, в конечном итоге, будут приводить к разным результатам. Таким образом, для каждого элемента исходных данных (y, A, S) существует единственное представление элементом пространства \mathfrak{S} , согласованное с операцией композиции и обеспечивающее соответствующий результат оценивания.

В частности, поскольку результат оценивания не зависит от порядка данных в исходном наборе, каноническая информация не должна зависеть от порядка

данных. Это имеет место для информационного пространства \mathfrak{S} .

Отметим два «практических» свойства этого способа представления промежуточной информации, которые носят скорее технический характер, связанный с особенностями реализации соответствующих алгоритмов.

Компактность. Информация, представленная в канонической форме, занимает фиксированный объем (а именно, $\frac{m(m+3)}{2}$ чисел), не зависящий от объема представленных данных.

Эффективность. Представление промежуточной информации в канонической форме обеспечивает эффективное выполнение всех стадий обработки данных:

a) извлечение канонической информации из исходных данных требует несколько матричных умножений для матриц, определяемых отдельными фрагментами данных. При этом извлечение канонической информации из отдельных фрагментов может производиться параллельно;

b) комбинирование и накопление канонической информации сводится к сложению векторов и матриц фиксированной размерности и требует незначительных вычислений;

c) вычисление результата на основании накопленной канонической информации требует решения системы линейных уравнений фиксированного размера $m \times m$ (или обращения соответствующей матрицы). Даже при постоянном поступлении новых данных обновление оценки может осуществляться лишь время от времени.

ЗАКЛЮЧЕНИЕ

Несмотря на то, что задача линейного оценивания имеет самостоятельную ценность и часто встречается в приложениях, мы использовали ее, в первую очередь, в качестве иллюстрации, а также показали, что, как и в [1], проблема адаптации алгоритма к работе в системах Больших Данных приводит к построению специального вида представления информации, обладающего естественными алгебраическими свойствами.

Во многих практических задачах преобразование обработки P , трансформирующее исходные данные в окончательный результат обработки, имеет специфическое «происхождение», а именно, является решением некоторой оптимизационной задачи. В нашем случае рассматривалась задача построения оценки минимальной погрешности. Оптимизационная постановка исходной задачи, фактически, привела к тому, что понятие качества решения (точности оценки) индуцировало на информационном пространстве упорядочение, отражающее «качество» информации. Как показано в [5-9], подобные естественные упорядочение и алгебраическая структура всегда возникают при исследовании информативности различных классов источников информации, включая, например, многозначные [10] и нечеткие [11]. Можно ожидать, что подобное упорядоче-

ние, согласованное с алгебраической структурой информационного пространства, всегда будет возникать в контексте задач оптимального принятия решений в распределенных системах.

СПИСОК ЛИТЕРАТУРЫ

1. Голубцов П.В. Понятие информации в контексте задач обработки Больших Данных // Научно-техническая информация. Сер. 2. – 2018. – № 1. – С. 31-36.
2. Пытьев Ю.П. Псевдообратный оператор. Свойства и применения // Математический сборник. – 1982. – Т. 118, № 5, – С. 19–49.
3. Пытьев Ю.П. Математические методы интерпретации эксперимента. – М.: Высшая школа, 1989. 351 стр.
4. Пытьев Ю.П. Задачи редукции в экспериментальных исследованиях // Мат. сб. – 1983. – Т. 120, № 2. – С. 240–272.
5. Голубцов П.В. Информативность в категории линейных измерительных систем // Проблемы передачи информации. – 1992. – Т. 28, № 2. – С. 30–46.
6. Голубцов П.В. Информативность в категории многозначных преобразователей информации // Пробл. передачи информ. – 1998. – Т. 34, № 3. – С. 60–80.
7. Голубцов П.В. Аксиоматическое описание категорий преобразователей информации // Пробл. передачи информ. – 1999. – Т. 35, № 3. – С. 109-127.
8. Golubtsov P.V. Monoidal Kleisli Category as a Background for Information Transformers Theory // Information Processes. – 2002. – Vol. 2, № 1. – P. 62–84.
9. Golubtsov P.V. Information Transformers: Category-Theoretical Structure, Informativeness, Decision-Making Problems // Hadronic Journal Supplement. – 2004. – Vol. 19, № 4, P. 375-424.
10. Голубцов П.В., Филатова С.А. Многозначные измерительно-вычислительные системы // Мат. моделирование. – 1992. – Т. 4, № 7. – С. 79–94.
11. Голубцов П.В. Теория нечетких множеств, как теория неопределенности и задачи принятия решений в нечетком эксперименте // Пробл. передачи информ. – 1994. – Т. 30, № 3. – С. 47–67.

Материал поступил в редакцию 31.10.17.

Сведения об авторе

ГОЛУБЦОВ Петр Викторович – доктор физико-математических наук, профессор Московского государственного университета им. М.В. Ломоносова; профессор Национального исследовательского университета «Высшая школа экономики», Москва.
e-mail: golubtsov@physics.msu.ru

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

УДК [004.65:001.4](100)

С.М. Гоннова, В.И. Салохин, И.Я. Пивинский

О необходимости создания международного атласа терминов фундаментальной науки

Проанализировано современное состояние терминологической работы в области фундаментальной науки. Выявлена необходимость систематизации терминов в соответствии с современным уровнем развития науки и техники. Обсуждается создание электронных терминологических словарей различных отраслей фундаментальной науки, терминологических баз данных и в перспективе разработка электронной системы «Международный атлас терминов фундаментальной науки».

Ключевые слова: терминология, фундаментальная наука, термины, систематизация, обработка текстов, терминологические базы данных, электронные словари, публикационная активность

СОВРЕМЕННОЕ СОСТОЯНИЕ ТЕРМИНОЛОГИЧЕСКОЙ РАБОТЫ В ОБЛАСТИ ФУНДАМЕНТАЛЬНОЙ НАУКИ

В настоящее время роль науки в жизни общества динамично возрастает. Процесс глобализации технического прогресса ведет к тому, что наука приобретает интернациональный характер. Все больше фундаментальных исследований проводится международными научными коллективами, создаются международные научные лаборатории, развивается научное партнерство, что требует новых форм профессионального общения и взаимопонимания ученых.

У каждой науки есть свой профессиональный язык в предметной области. Термины служат инструментарием для формирования научных теорий, законов, положений, специальных обозначений, характерных для изучаемой сферы предметов, явлений, их свойств, закономерностей и отношений. Любая отрасль науки развивает своё знание вместе с соответствующими терминами, понятиями. Термины и терминология являются неотъемлемой составной частью науки и техники.

Важными становятся систематизация научных терминов и перевод научно-технических текстов (многозначность, тождественность означаемых терминами понятий, логичность построения предложений, смысловая точность и т.д.) для повышения уровня обмена научной информацией, научными статьями, общения, научной международной деятельности.

Согласно Википедии «термин» (от лат. *terminus* – предел, граница) – это слово или словосочетание, яв-

ляющееся названием некоторого понятия какой-либо области науки, техники, искусства и т.п. Термин представляет собой словесное обозначение понятия, входящего в систему понятий определенной области профессиональных знаний [1].

Термины являются смысловым ядром специального языка и передают основную содержательную информацию. Они отличаются от других слов однозначностью и точностью в конкретном научном контексте.

Развитие науки приводит к появлению новых понятий, которые, зачастую выражаются новыми словами или словосочетаниями. Так с начала XXI века в научных публикациях по различным разделам механики широко используются такие новые слова как, например, *наножидкости, функционально-градиентные материалы, графен, графан, фосфорен, препрег, конструкталы* и др. По словам Д.С. Лотте [2] в современном мире в результате роста научно-технических знаний свыше 90% новых слов, появляющихся в языках, составляют специальные слова. Рост числа терминов некоторых наук обгоняет рост числа общеупотребительных слов, а в некоторых науках число терминов превышает число неспециальных слов.

Термины каждой отрасли науки, техники, производства формируют свои терминологические системы, определяемые, в первую очередь, понятийными связями профессионального знания при стремлении выразить эти связи языковыми средствами.

Терминология также выполняет важную коммуникативную функцию между представителями различных областей знаний и научных дисциплин, как в

сфере обучения, так и в сфере научной деятельности, обеспечивая взаимопонимание.

Таким образом, научно-техническая терминология является одной из важных составляющих информационного сопровождения науки. Уровень современной науки требует соответствующего уровня современной терминологии.

Терминология постоянно находится в движении. Различными способами появляются новые термины, некоторые изменяют свое значение, другие устаревают и исчезают. Поэтому уточнение и систематизация устоявшихся и возникающих терминов, а также унификация и упорядочение терминов, определений в области фундаментальной науки является частью постоянной терминологической работы.

Современное состояние терминологии российской фундаментальной науки трудно признать удовлетворительным. В качестве примера недостаточной разработанности современной терминологии можно привести такой факт. При русскоязычной обработке статьи Ottino G.M. et al. *Combustion Modeling Including Heat Loss Using Flamelet Generated Manifolds: A Validation Study in OpenFOAM*, опубликованной в журнале «Flow, Turbulence and Combustion» (2016, Vol. 96, №3, p.773-800), возникла необходимость применить русский эквивалент термина «*Flamelet Generated Manifolds Technique*». Поиск по русскоязычной литературе не дал результатов. Решить проблему удалось путем контактов с русскоязычным специалистом по горению из Технического университета Чалмерса (Гётеборг, Швеция), который выпустил на английском языке книгу «*Andrei Lipatnikov. A Fundamentals of Premixed Turbulent Combustion*» (CRC Press, 2012, 548 p.). С его помощью удалось установить русскоязычный эквивалент указанного термина – «метод множества ламинарных пламен».

Работы по развитию терминологии фундаментальной науки, проводившиеся АН СССР, ее терминологическим центром – Комитетом по научно-технической терминологии, в настоящее время не ведутся. Комитет, переименованный в 1993 г. в Комитет научной терминологии в области фундаментальных наук РАН (далее – КНТ РАН), не действует. Официальные терминологические сборники по различным разделам фундаментальной науки не выпускаются [3]. Последний сборник, посвященный робототехнике, был выпущен в 2000 г.

Отдельные специалисты выпускают терминологические словари по некоторым отраслям знаний. Так вышли словари: Новый политехнический словарь под ред. А.Ю. Ишлинского [4], Крайнев А.Ф. *Механика машин. Фундаментальный словарь* [5], Резер С.М., Родников А.Н. *Логистика. Словарь терминов* [6], Русско-английский и англо-русский словарь терминологии кабельной техники, сост. П.О. Ганелес [7], Колесниченко Д.Н. *Англо-русский толковый словарь компьютерных терминов* [8] и ряд других. Однако в этих работах нет единого подхода к составлению словарей, во многих из них даются определения всем значениям многозначных профессионализмов, что не отвечает требованиям документов по однозначности терминов и отсутствию синонимов [9, 10].

Системные работы по развитию научно-технической терминологии ведет только Федеральное агентство по техническому регулированию и метрологии (Росстандарт), которое через свои Технические комитеты (ТК) выпускает терминологические стандарты. Так, комитет ТК-55 «Терминология, элементы данных и документация в бизнес-процессах и электронной торговле» выпустил национальный стандарт ГОСТ Р ИСО 704-2010. «Терминологическая работа. Принципы и методы» [11], который является аутентичным переводом на русский язык международного стандарта ISO 704:2009 «Terminology work – Principles and methods». В соответствии с действующей комплексной программой гармонизации российской и европейской систем технического регулирования, большинство национальных терминологических стандартов РФ представляют собой аутентичные переводы на русский язык международных стандартов системы ISO. Практика внедрения таких переводов порой приводит к затруднениям, обусловленным такими языковыми явлениями как полисемия и омонимия. Так, в терминологических стандартах появляются не эквиваленты терминов, а «кальки» (дословные переводы), а иногда и описательные характеристики без учета проработанности понятий и терминов в отечественной терминологии.

Следует отметить, что в последнее время появляются и оригинальные российские стандарты, например, ГОСТ Р 57188-2016 «Численное моделирование физических процессов. Термины и определения» [12]. Однако выпускаемые в РФ терминологические стандарты охватывают термины отдельных отраслей промышленности, носят рекомендательный характер и ограничено воздействуют на язык науки, что явно недостаточно для научно-технического развития страны в условиях бурно развивающейся мировой науки. Причем, в одних случаях в отечественных стандартах даются краткие пояснения к терминам, как например, в [12], а в других – англоязычные аналоги, как например, в ГОСТ Р ИСО 4063-2010 «Сварка и родственные процессы. Перечень и условные обозначения процессов» [13].

О необходимости возобновления системной работы по развитию научно-технической терминологии отмечается во многих публикациях. Так в статье [14] обсуждается использование термина «*хладноломкость*» в различных областях строительной отрасли и показано отсутствие четкого понимания его использования. В статье [15] обсуждается неразработанность системы понятий и терминов конструирования радиоэлектронных систем.

Представляется важным и необходимым проведение системной терминологической работы при переводе научных текстов на другие языки. В сентябрьском номере 2016 г. одного из ведущих и высокоцитируемых российских научных журналов «Успехи физических наук» (импакт-фактор 2014 – 2,606) опубликована статья Бункина Н.Ф. и Бункина Ф.В. «Бабстонная структура воды и водных растворов электролитов» (УФН, 186:9 (2016), 933–952). В аннотации указано о развитии теоретических представлений существования в чистой воде и водных растворах электролитов, находящихся в равновесии с внешней газовой средой,

стабильных газовых пузырьков – бабстонов. На основании этого можно сделать вывод, что появился новый термин – «бабстоны». И только из текста статьи становится понятно, что бабстон – это просто аббревиатура английских слов “bubble stabilized by ions”. В связи с этим возникает вопрос – не лучше ли было сохранить простые понятные русские слова «стабилизированные ионами пузырьки», а не придумывать новые термины, искажая смысл русскоязычного понятия? В последнее время в российской науке наметилась тенденция при конструировании новых терминов использовать англоязычные выражения. Видимо, российские ученые забывают высказывание М.В. Ломоносова о русском языке [16].

На основе анализа терминологических стандартов, словарей рекомендуемых терминов, публикаций в научно-технической литературе, а также оценке состояния терминологической работы в РФ можно сделать вывод, что назрела необходимость организации системных работ по совершенствованию научно-технической терминологии в соответствии с современным уровнем развития науки и техники. В научном плане эти работы должны расширить понятийную базу научных знаний, усовершенствовать ее и привести в соответствие с национальными и международными стандартами. Кроме того, выполнение терминологических работ должно обеспечить единообразное и непротиворечивое понимание специализированной научной информации учеными, студентами, другими специалистами, которые обмениваются информацией. Результатом этих работ должны стать терминологические базы данных различных отраслей современной фундаментальной науки.

Выполнение системной терминологической работы в области фундаментальной науки в РФ может быть осуществлено коллективом, состоящим из специалистов РАН по отдельным областям знаний, специалистов лексикологов, а также специалистов в области цифровых технологий информационных систем.

ЭЛЕКТРОННЫЕ ТЕРМИНОЛОГИЧЕСКИЕ СЛОВАРИ РАЗЛИЧНЫХ РАЗДЕЛОВ ФУНДАМЕНТАЛЬНОЙ НАУКИ

Систематизация научных терминов в соответствии с современным уровнем развития науки и техники является основной проблемой и характеризуется как комплексная, междисциплинарная.

Терминологический словарь и терминологическая система, как его основа, являются формой представления знания определенной отрасли науки.

Функциональная значимость любого терминологического словаря и терминологической системы заключается, прежде всего, в содействии развитию самого научного направления, а также в просветительской роли.

Как показывает анализ существующих терминологических словарей и научно-нормативной документации, нет единого подхода к структуре представления лексикографического материала. В сборниках научно-нормативной терминологии Комитета научной терминологии в области фундаментальных наук РАН, например, в одном из последних сборников «Робототехника», выпущенных в 2000 г. [17], приводится термин на русском языке и определение выражаемого термином понятия. В качестве справочных сведений приводятся эквиваленты терминов на немецком, английском, французском языках (рис. 1) В виде примечаний даются пояснения к некоторым определениям.

Иное представление информации в тетрадах новых терминов, выпускавшихся Всесоюзным центром переводов научно-технической литературы и документации (ВЦП). В них даются английские термины и русские эквиваленты, иногда приводятся пояснения к русским эквивалентам. Фрагмент страницы одной из тетрадей [18] представлен на рис. 2.

В терминологическом словаре [19] представлены лексические единицы с определениями на английском и русском языках (рис. 3).

<p>2. Робот</p> <p><i>D</i> Roboter <i>E</i> Robot <i>F</i> Robot</p>	<p>Многофункциональная перепрограммируемая машина для полностью или частично автоматического выполнения двигательных функций аналогично живым организмам, а также некоторых интеллектуальных функций человека.</p> <p>Примечания. 1. Под “перепрограммируемостью” понимается возможность замены, коррекции или генерации управляющей программы автоматически или при помощи человека. 2. К роботам не относятся, в частности, автооператоры, а также копирующие манипуляторы и другие машины, управляемые только человеком-оператором.</p>
--	---

Рис. 1. Фрагмент страницы сборника научно-нормативной терминологии КНТ РАН «Робототехника»

17. build-in pressure fluid system
 гидродинамический эффект смазки /за счет скоростного напора создается давление в смазке, способное разделять трущиеся поверхности/

Рис. 2. Фрагмент страницы тетради новых терминов ВЦП «Англо-русские термины по трибологии»

13	EN 1990	basic variable:	part of a specified set of variables representing physical quantities which characterize actions and environmental influences, geometrical quantities, and material properties including soil properties	базовая переменная:	Величина, выбранная из установленного набора переменных физических величин, которая характеризует воздействия и влияние окружающей среды, геометрические величины и свойства материалов, в том числе свойства грунтов.
----	---------	------------------------	--	----------------------------	--

Рис. 3. Фрагмент страницы терминологического словаря для национальных нормативных документов, реализующих Еврокоды

Планируемые электронные терминологические словари различных отраслей фундаментальной науки будут включать: 1) термины на русском языке, 2) англоязычные эквиваленты терминов, 3) определения выражаемых терминами понятий на русском языке, 4) словарные статьи к терминам. Такое представление лексикографического материала близко к методике ВИНТИ РАН, изложенной в статье [20]. Отличие состоит в том, что в методике за основу наполнения словарей предлагается брать энциклопедические словари (в первую очередь, имеющиеся в свободном доступе через Интернет) и базу данных стандартизованной терминологии Росстандарта. Если отсутствует какое-либо понятие в сетевых версиях словарей или, с точки зрения эксперта, нет достаточно четкого или полного его представления, эксперт может обращаться к печатным версиям специальных словарей. В предлагаемой модели словаря все термины отбираются из ранее сформированной терминологической базы предметной области.

Термины будут располагаться в алфавитном порядке, это облегчает поиск необходимого термина и сокращает время его поиска. В начале словаря будет приведен список устоявшихся терминов, которые являются основой для понимания.

Определение понятия, выражаемого термином, должно быть оптимально кратким и состоять из одного предложения. Краткость обеспечивается отсутствием избыточной информации – недопустимы выражения в скобках; перечисления понятий, составляющих объем определяемого понятия; сокращения, такие как «и т.п.», «и т.д.», «и пр.» [10].

Каждый термин будет сопровождаться словарной статьей. В отличие от некоторых терминологических словарей, например, [6], где многие словарные статьи имеют общепознавательный характер, в планируемых статьях пояснения к термину, будут отражать его разные значения в соответствии с контекстом и особенностями национальной трактовки, а также ука-

зывать области применения. Текст словарной статьи должен быть лаконичным, иметь четкие и убедительные формулировки при отсутствии второстепенной информации [21]. Следует употреблять синтаксические конструкции, свойственные языку научных и технических документов, избегая сложных грамматических конструкций. Объем информации, содержащийся в словарных статьях, должен быть примерно одинаковым, однако практически это сделать трудно. Главным фактором, определяющим объем словарной статьи, должна быть глубина толкования термина.

Пример:
 «Температурный глайд (temperature glide) Абсолютное значение разности температур начала и окончания процесса фазового перехода (кипения или конденсации) при постоянном давлении.

Термин *температурный глайд* применяется в холодильной и климатической технике. В качестве рабочих веществ (хладагентов) в холодильных и климатических системах широко используются смеси различных галогеносодержащих углеводородов. При заданном давлении отдельные компоненты смеси кипят или конденсируются при постоянной температуре. Температура насыщенной жидкости и температура насыщенного пара одинаковы, и температурный глайд равен нулю. В смесях при фиксированном давлении с повышением температуры сначала начинает кипеть вещество с более низкой температурой кипения. В результате содержание этого компонента в паровой фазе увеличивается, а в жидкой фазе повышается концентрация высококипящего компонента. В итоге возрастает температура кипения. Это продолжается до тех пор, пока процесс кипения не закончится. При конденсации сначала конденсируется высококипящая составляющая смеси, а затем в результате понижения температуры и низкокипящая. Таким образом, фазовые переходы (кипение или конденсация) в хладагентах при постоянном давлении происходят в некотором температурном диапазоне».

Таким образом, создание серии электронных терминологических словарей различных отраслей фундаментальной науки включает следующие работы:

- разработку концепции создания электронных специализированных словарей современной терминологии отдельных отраслей фундаментальной науки: цель, задачи, основные принципы, общие подходы к исходным данным, основные требования к разработке специализированных словарей, методику формирования электронных терминологических баз данных;
- создание пилотной версии электронного терминологического русско-английского словаря на примере одной отрасли фундаментальной науки (лексикографический анализ терминов, анализ научно-технических понятий, формирование терминологической базы данных, массив формализованных описаний, электронный макет словаря);
- апробацию методики, изложенной в концепции на примере создания и испытания пилотной версии электронного терминологического русско-английского словаря – экспертиза профессиональными организациями, оценка программных инструментов, оценка в целом создания электронных терминологических словарей различных отраслей фундаментальной науки;
- разработку на методологической основе концепции серии электронных специализированных терминологических словарей;
- формирование международного атласа терминов фундаментальной науки.

При выполнении указанных работ могут быть использованы различные методы анализа: системный, комплексный, структурный, статический, семантический, а также сочетание метода концептуального анализа текстов и методов статистической обработки, методов систематизации и упорядочения научно-технической терминологии.

ФОРМИРОВАНИЕ ТЕРМИНОЛОГИЧЕСКИХ БАЗ ДАННЫХ

Создание электронных специализированных терминологических словарей предполагает наличие автоматизированных терминологических баз данных, позволяющих получать необходимые сведения о терминах и обозначаемых ими понятиях с помощью компьютеров, без применения словарей в полиграфической форме.

Формирование терминологических баз данных различных отраслей фундаментальной науки предполагает лексикографический анализ терминов; уточнение понятий и упорядочение существующих терминов предметной области; поиск, выделение и извлечение из научных текстов новых терминов и понятий; их анализ; унификацию терминов; формирование массива формализованных описаний.

Терминологическая база данных предметной области будет включать как традиционные, устоявшиеся, так и новые термины. Пример необходимости уточнения понятий предметной области. В работе [22] отмечается, что, исходя из английской транскрипции «*electromechatronics*», в публикации [23] слова «электромеханотроника» и «электромехатроника» используются как равнозначные. Авторы ста-

ти считают первую запись более правильной, так как она сохраняет в качестве основного терминологического элемента слово «механика». Поэтому предложено понятие «электромеханотроника» использовать для обозначения отрасли науки и техники, связанной с электронизацией технических устройств, рассматриваемых в электромеханике [24].

Упорядочение существующей терминологии предметной области ранее проводилось по методике [9] и рекомендациям [10]. В свое время методика хорошо себя зарекомендовала. При этом следует отметить, что действующие в РФ рекомендации по разработке стандартов на термины и определения [10] в основном повторяют положения методики. На целесообразность ее использования в современных условиях указывает также работа [14]. Поэтому упорядочение существующей терминологии предметной области разумно проводить по методике [9] и рекомендациям [10], адаптированным к современным условиям.

Традиционные термины могут извлекаться из существующей терминологической литературы (сборники рекомендуемых терминов, терминологические стандарты, энциклопедии, толковые терминологические словари), а новые – из публикаций, отобранных по результатам поиска в индексируемых российских и международных информационно-аналитических системах мирового потока информации.

При формировании массива терминов традиционные термины будут извлекаться из крупных существующих международных терминологических баз данных (ТБД), которых существует несколько десятков. Мы не отдаем предпочтение ни одной из них. Использование той или иной базы данных зависит, в первую очередь, от ее доступности для составителей массива терминов. Следует отметить, что научная терминология на русском языке представлена базой данных «Российская терминология» (БД РОСТЕРМ) [25], содержащей свыше 140 тысяч терминологических статей из национальных и международных стандартов, регламентов, а также терминологических приложений к ним, терминологических сборников Комитета научной терминологии в области фундаментальных наук РАН. Термины и определения даны на русском языке, а также представлены эквиваленты терминов на английском языке.

Существенную помощь при отборе традиционных терминов могут оказать сборники «Тетради новых терминов», выпускавшиеся Всесоюзным центром переводов научно-технической литературы и документации (ВЦП) в 1970–1993 гг. [25]. Они содержат русскоязычные термины и их англоязычные эквиваленты по разным разделам науки и техники.

Для отбора новых терминов, в первую очередь, будут использованы результаты анализа последних публикаций мирового потока научной и технической литературы, в том числе международных полнотекстовых и реферативных баз данных Scopus, Web of Science и других профессионально значимых мировых журналов, а также информационные ресурсы ВИНТИ РАН. Эти ресурсы представляют собой федеральную базу отечественных и зарубежных публикаций по естественным, точным и техническим нау-

кам, которая генерируется с 1981 г., включает 28 тематических фрагментов, состоящих из 200 разделов, сформированных на основе литературы, получаемой более чем из 80 стран на 40 языках, обновляющаяся ежемесячно, пополнение её составляет около 1 млн документов в год. Практика показывает, что такой подход позволит охватить до 95% новых терминов, появляющихся в научной литературе.

При формировании терминологических баз данных для выявления понятийного состава текстов будет проведен их концептуальный (семантический) анализ, на основе которого из текстов будут выделяться фразеологические единицы с учетом частоты их встречаемости. При этом практика составления терминологических словарей показывает, что частотный принцип отбора ключевых слов должен сочетаться со смысловым подходом [26]. Именно смысловой подход позволяет включать в список ключевых слов и редко встречающиеся термины, которые составляют фундамент понятийного и терминологического аппарата научных направлений. Частотный принцип отражает лишь наиболее разрабатываемые направления, по которым имеется наибольшее число публикаций, а те термины и понятия, которые отражают фундаментальные положения науки, уже прошли свой пик исследований и максимум публикаций. В этой связи они могут употребляться в теоретических работах и учебной литературе. Однако их значение для науки диктует необходимость включения таких терминов в список ключевых слов, несмотря на низкую частоту встречаемости. К подобной категории при применении только частотного принципа отбора относятся и термины, описывающие вновь открытые явления и новые технологии, например, новый метод – «*визуализация магнитных частиц - magnetic particle imaging*».

Об унификации терминов. Для чего нужна унификация терминов? В современной науке существует стремление к семантической унификации терминов одной и той же области науки в разных языках (однозначное соответствие между терминами разных языков). В то же время полисемия английского языка приводит к тому, что в переводах на русский язык появляется неоднозначная трактовка терминов.

Например, в некоторых переводах на русский язык англоязычный термин «*condensation heat transfer*» встречается как «*конденсационный теплообмен*». В то же время в оригинальных русскоязычных публикациях употребляется термин «*теплообмен при конденсации*», что предпочтительнее. Действительно, конденсация (от лат. *condensatio* – сгущение, уплотнение) – физический процесс перехода вещества из газообразного состояния в жидкое, происходит с выделением теплоты фазового перехода и сопровождается переносом теплоты и массы. Причем перенос теплоты происходит на микроуровне (теплопроводность) и на макроуровне (конвекция). Поэтому термин – «теплообмен при конденсации» полнее и правильнее отражает физическую сущность процесса.

Другой пример, в русскоязычной литературе используется термин «метод группового учета аргументов», введенный разработчиками этого метода. В англоязычной литературе применяется термин

«*group method of data handling*», дословный перевод которого на русский язык отличается от русскоязычного термина.

Из приведенных примеров ясно, что унификация нужна для создания эквивалентных терминов, а не «калек».

По результатам обработки текстов научных статей, рефератов и других документов составляется список ключевых слов и словосочетаний (словник), выражающих устойчивые современные научно-технические понятия в соответствующей предметной области.

При переводе отобранных терминов на другие языки необходимо обеспечить адекватный перевод отдельно взятых терминов определенного текста. Каждый переводимый термин должен проверяться с точки зрения терминосистем, фигурирующих в языке оригинала и языке перевода, которые служат для обозначения системы терминов какой-либо отдельной науки, техники и т.д. Также должны быть учтены различия терминов, определяемые спецификой передачи мысли на каждом из этих языков.

Развитие информационно-коммуникационных технологий, внедрение компьютеров в различные сферы человеческой деятельности ведут к возможности отказа от бумажных носителей лексикографической информации. Сейчас многие иностранные словари представляются в бумажном и электронном виде, например, Collins Online Dictionary, Macmillan English Dictionary for Advanced Learners и др., а некоторые планируют вообще отказаться от бумажного варианта, например, четвертое издание Oxford English Dictionary [27]. Поэтому, следуя общемировой тенденции, разрабатываемые терминологические словари предметных областей должны быть представлены в электронном формате. Отметим только, что этому вопросу посвящены многие исследования, например [27], где для представления словарей в электронном виде и в качестве обменного формата предлагается использовать лексикографический формат TEI.

Электронные терминологические словари различных отраслей фундаментальной науки в связи с общепризнанным английским языком предполагается создавать в виде двуязычных, русско-английских, словарей, которые будут содержать термины на русском языке и их англоязычные аналоги. Это позволит, помимо учебных целей, широко использовать словари в системах автоматической обработки текстовой информации на естественных языках как, например, КОРЕФ и в системах машинного перевода с одного языка на другой, таких как PROMT, RETRANS и др.

МЕЖДУНАРОДНЫЙ АТЛАС ТЕРМИНОВ ФУНДАМЕНТАЛЬНОЙ НАУКИ

Предлагаемый «Международный атлас терминов фундаментальной науки» представляет собой интернет-ресурс, который содержит серии электронных терминологических русско-английских словарей различных отраслей фундаментальной науки, таких как математика, механика, информатика, физика и астрономия, химия и науки о материалах, биология и медицинские науки, науки о Земле и др.

Разработка и создание электронной системы «Международный атлас терминов фундаментальной науки» направлены на систематизацию, унификацию, упорядочение терминологии на русском и английском языках, на распространение русскоязычных терминов, понятий, образующихся в новых научных направлениях за рубежом (которые возникают, в основном, на английском языке).

Создание системы «Международный атлас терминов фундаментальной науки» способствует совершенствованию систем автоматической обработки текстовой научной и научно-технической информации и систем фразеологического машинного перевода текстов с одного языка на другой.

Сформированные терминологические базы данных и серии электронных терминологических словарей различных отраслей фундаментальной науки создадут условия для более активного участия России в развитии международной скоординированной терминологической работы, обмене терминологическими данными между национальными БД, унификации методологии терминологической деятельности и стандартов, гармонизации терминологической лексики в разных языках.

Применение разрабатываемых электронных словарей с расширенным понятийным аппаратом научных знаний в системах автоматической обработки текстовой информации позволит обеспечить более глубокую переработку информации (поиск, классификацию, индексирование, реферирование).

Применение словарей в издательствах, использующих интерактивные системы русско-английского и англо-русского машинного перевода, позволит повысить качество перевода текстов.

ЗАКЛЮЧЕНИЕ

Рассматривая вопросы информационного сопровождения современной науки, следует признать, что формирование международного атласа терминов фундаментальной науки является форматом, отражающим развитие и достижения науки.

Необходимость создания электронных терминологических словарей различных отраслей фундаментальной науки в соответствии с современным уровнем развития науки и техники очевидна.

Актуальность систематизации терминов, развитие терминологии фундаментальной науки обосновывается реализацией Стратегии научно-технологического развития РФ [28], также разработкой государственной программы «Научно-технологическое развитие Российской Федерации». Государственная программа включает в себя подпрограмму по фундаментальным научным исследованиям и другие подпрограммы, которые касаются: исследований, разработок и инноваций в целях реализации приоритетов научно-технологического развития Российской Федерации, развития национального интеллектуального капитала, эффективной открытой системы организации науки, технологий и инноваций и т.д.

Из программы видно, что государственные задачи сегодняшнего дня касаются и фундаментальной науки.

В связи с этим представляется важным понимание сути и однозначное использование научных терминов не только при проведении научных исследований, но и при формировании, мониторинге, реализации государственных, международных комплексных междисциплинарных научно-технических программ. Внедрение научных достижений в производство (проекты полного инновационного цикла) требует проведения экспертизы научных исследований, их результатов, технологической и технической оценки и др.

Электронная система «Международный атлас терминов фундаментальной науки», содержащая терминологические русско-английские словари по различным отраслям фундаментальной науки, будет полезна при создании информационных интернет-ресурсов министерств, ведомств, научных и инновационных учреждений, предприятий, организаций. Ее создание направлено на защиту идентичности российской научной сферы в условиях интернационализации науки, на повышение эффективности российской науки в рамках международного взаимодействия, на развитие форм и механизмов научной дипломатии. Электронная система будет способствовать повышению уровня международного научно-технического сотрудничества в части развития компетенций и научной коммуникации, в первую очередь молодых исследователей, повышения уровня обмена научной информацией, научного партнерства, участия российских научных коллективов в реализации многосторонних (международных) программ развития, участия в международных системах научно-технической экспертизы и прогнозирования и т.д.

Появление специализированных русско-английских словарей, развитие и распространение русскоязычных терминов, понятий, поднимет имидж российской науки, повлияет на публикационную активность российских ученых, в том числе в мировых системах Scopus, Web of Science.

СПИСОК ЛИТЕРАТУРЫ

1. Википедия – сводная энциклопедия. Термин. – URL: <https://ru.wikipedia.org/wiki/> (дата обращения 01.03.2017).
2. Лотте Д.С. Вопросы заимствования и упорядочения иноязычных терминов и терминологических элементов. – М.: Наука, 1982. – 152 с.
3. Википедия – сводная энциклопедия. Комитет научной терминологии в области фундаментальных наук. – URL: <https://ru.wikipedia.org/wiki/> (дата обращения 01.03.2017).
4. Новый политехнический словарь / под ред. А.Ю. Ишлинского. – М.: Большая Российская энциклопедия, 2003. – 671 с.
5. Крайнева А.Ф. Механика машин. Фундаментальный словарь. 2-е изд., испр. – М.: Машиностроение, 2001. – 904 с.
6. Резер С.М., Родников А.Н. Логистика. Словарь терминов / под ред. С.М. Резера. – М.: ВИНТИ РАН, 2007. – 412 с.

7. Русско-английский и англо-русский словарь терминологии кабельной техники / сост. П.О. Ганелес. – М.: Журн. «Кабели и Провода», 2006. – 464 с.
8. Колесниченко Д.Н. Англо-русский толковый словарь компьютерных терминов. 2-е изд., перераб., доп. – СПб.: Наука и Техника, 2008. – 280 с.
9. Краткое методическое пособие по разработке и упорядочению научно-технической терминологии. – М.: Наука, 1979. – 129 с.
10. Р 50.1.075 – 2011 Рекомендации по стандартизации. Разработка стандартов на термины и определения. – М.: Стандартинформ, 2012. – 23 с.
11. ГОСТ Р ИСО 704-2010 Терминологическая работа. Принципы и методы. – М.: Стандартинформ, 2012. – 51 с.
12. ГОСТ Р 57188-2016 Численное моделирование физических процессов. Термины и определения. – М.: Стандартинформ, 2016. – 12 с.
13. ГОСТ Р ИСО 4063-2010 Сварка и родственные процессы. Перечень и условные обозначения процессов. – М.: Стандартинформ, 2011. – 16 с.
14. Мойсейчик Е.А. Основные понятия и проблемы терминологии металлостроительства // Вестник Московского государственного строительного университета (МГСУ). – 2014. – № 8. – С. 7-19.
15. Бобков Н.М. Систематизация терминологии в области конструирования радиоэлектронных систем // Труды Нижегородского государственного технического университета им. Р.Е. Алексеева (НГТУ). – 2014. – № 3. – С. 19-29.
16. Ломоносов М.В. Полн. собр. соч., т.7. Труды по филологии. 1739-1758. Российская грамматика. Под ред. В.В. Виноградова и др. – М.-Л.: АН СССР, 1952. – С. 389-578.
17. Сборники научно-нормативной терминологии. Вып. 115. Робототехника. Терминология. – М.: КНТ РАН, 2000. – 47 с.
18. Тетради новых терминов, № 181. Англо-русские термины по трибологии. – М.: ВЦП, 1991. – 64 с.
19. Терминологический словарь для национальных нормативных документов, реализующих Еврокоды. – М.: ЦНИИПСК им. Н.П. Мельникова, 2014. – 208 с.
20. Калёнов Н.Е., Белоозеров В.Н. Формирование терминологических словарей по лексике классификационных систем // Научно-техническая информация. Сер. 1. – 2015. – № 3. – С. 60 – 70.
21. ГОСТ 7.9-95 Система стандартов по информации, библиотечному и издательскому делу. Реферат и аннотация. Общие требования. – М.: ИПК Издательство стандартов, 1996. – 7 с.
22. Коськин Ю.П., Самохвалов Д.В. О терминологии и подготовке специалистов по электромеханотронике // Известия Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» им. В.И. Ульянова (Ленина) (СПбГЭТУ «ЛЭТИ»). – 2013. – Вып. 1. – С. 57-66.
23. Электротехническая энциклопедия. В 4 т. – М.: изд-во МЭИ, 2005–2010.
24. Коськин Ю.П. Введение в электромеханотронику. – СПб.: Энергоатомиздат, СПб. отд-ние, 1991. – 192 с.
25. БД РОСТЕРМ. – URL: <http://www.standards.ru/nd.aspx> (дата обращения 01.03.2017).
26. Соболева И.С., Пронина Т.А., Суворова Л.А. Русско-английский список ключевых слов по радиационной биологии. – М.: ВИНТИ РАН, 2017. – 45 с. Рук. деп. в ВИНТИ.
27. Буторова У.В., Панков ДИ, Пурицкая Е.В., Захаров В.П., Герд А.С., Хохлова М.В. Структурная типология словарных статей в словарях русского языка и способы их формального представления // Научно-техническая информация. Сер. 2. – 2016. – № 2. – С. 29–35.
28. Указ Президента Российской Федерации от 1.12.2016 г. № 642 «О Стратегии научно-технологического развития Российской Федерации» // Правительство России. – URL: <http://government.ru/docs/28270/> (дата обращения 03.07.2017)

Материал поступил в редакцию 07.07.17.

Сведения об авторах

ГОННОВА Светлана Михайловна – ст. научный сотрудник, магистр «Технологического менеджмента» РАНХиГС при Президенте РФ, нач. отдела инноваций и перспективных разработок ВИНТИ РАН, Москва
e-mail: gonnova@viniti.ru

САЛОХИН Владимир Иванович – кандидат технических наук, заведующий отделом научной информации по механике ВИНТИ РАН.
e-mail: v.salohin@yandex.ru

ПИВИНСКИЙ Игорь Ярославович – студент Российского университета дружбы народов, Москва
e-mail: ipivinskiy@gmail.com

ВНИМАНИЮ ПОДПИСЧИКОВ!

С 2018 года возобновляется издание информационного бюллетеня «Иностранная печать об экономическом, научно-техническом и военном потенциале государств-участников СНГ и технических средствах его выявления» серии «Экономический и научно-технический потенциал» (56741) взамен информационного бюллетеня «Экономика и управление»

Периодичность выхода – 12 номеров в год. Объем 48 уч.-изд. л. в год.

В бюллетене освещаются материалы иностранной печати по широкому спектру вопросов, касающихся сфер экономического и научно-технического развития России и стран СНГ: общие вопросы, финансы, промышленность, рынки, сельское хозяйство, космос, транспорт и связь, природные ресурсы, трудовые ресурсы, внешние торгово-экономические и научные связи

Оформить подписку на информационный бюллетень, начиная с любого номера, можно в ВИНТИ РАН по адресу: 125190, Россия, Москва, ул. Усиевича, 20,

Телефоны: (499) 151-78-61; (499) 155-42-85

Факс: (499) 943-00-60;

E-mail: contact@viniti.ru; sales@viniti.ru

Центр (Отдел) научно-информационного обслуживания (ЦНИО) ВИНТИ РАН

Информационные услуги, предоставляемые ЦНИО ВИНТИ РАН:

- проведение тематического поиска и консультации поисковых экспертов;
- подготовка списков научной литературы;
- подбор, копирование полнотекстовых материалов из первоисточников на бумажном носителе и в электронном виде;
- библиометрическая оценка публикационной активности исследователей и научных организаций с использованием российских и зарубежных баз данных;
- информационное обеспечение информационно-аналитической деятельности по подготовке и предоставлению аналитических обзоров и других научных материалов.

ВИНТИ РАН располагает следующими информационными ресурсами:

- фондом НТЛ, включающим более 2,5 млн. отечественных и иностранных журналов, книг, депонированных рукописей, авторефератов диссертаций и другой научной литературы, ретроспектива – с 1991 года;
- базами данных и Интернет-ресурсами: БД ВИНТИ (разработка ВИНТИ), БД SCOPUS, БД Questel (патенты) и другими реферативными ресурсами;
- полнотекстовыми электронными ресурсами (статьи, патенты, материалы конференций).

Ознакомиться с информацией о доступных полнотекстовых и реферативных ресурсах можно на сайте ВИНТИ www.viniti.ru

К услугам пользователей – **Электронный Каталог ВИНТИ** <http://catalog.viniti.ru>
и **служба электронной доставки документов.**

Осуществляется платное информационное обслуживание по разовым заказам и на договорной основе с предоставлением всех необходимых финансовых документов.

Проводится индивидуальное обслуживание пользователей в читальном зале ЦНИО ВИНТИ.

Обращаться в ЦНИО ВИНТИ:

- адрес: 125190, Россия, г. Москва, ул. Усиевича, 20;
- телефоны: 8(499) 155 -42 -43, 8(499) 155 -42 -17;
- эл. почта cnio@viniti.ru, fdk@viniti.ru;
- факс 8(499) 930 -60 -00 (для ЦНИО).