

Прикладная
герменевтика
информационного
пространства:
*картины мира,
теоретические
онтологии и веерные
матрицы*

Симон Кордонский,
Валерий Бардин

Симон Кордонский. Кандидат философских наук, профессор, заведующий кафедрой местного самоуправления департамента государственного и муниципального управления факультета социальных наук Национального исследовательского университета «Высшая школа экономики».

Адрес: 101990, Москва, ул. Мясницкая, 20.

E-mail: kordonsky@gmail.com.

Валерий Бардин. Генеральный директор компании «РелТим».

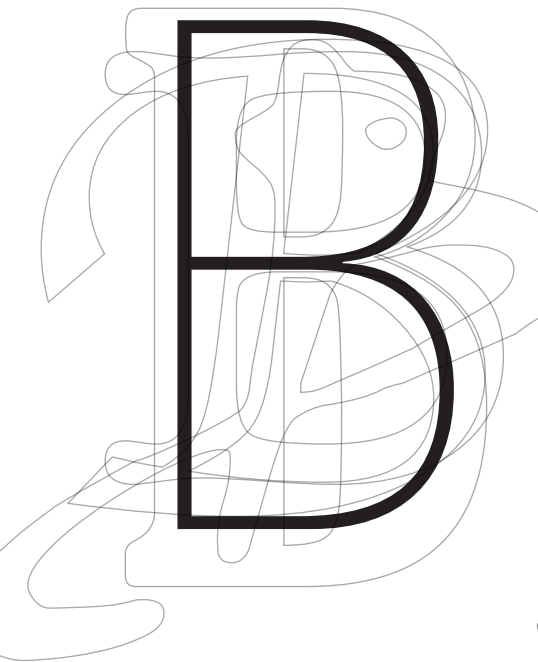
Адрес: 123308, Москва, ул. Куусинена, 6, корп. 8.

E-mail: foxgrep@gmail.com.

Ключевые слова: онтология; информатика; герменевтика; контексты; веерные матрицы.

Постановка задачи

В философской традиции анализ текстов считается предметом герменевтики. С появлением информационного пространства он стал одним из основных методов информатики, в которой сегодня доминирует альтернативный философскому эмпирический подход, основанный на статистике и лингвистике. Ниже представлен очерк третьего, собственно теоретического подхода к анализу массива текстов, образующих информационное пространство. В нем описаны методика построения теоретических онтологий с использованием логики веерных матриц



Термин «онтология» используется как в современной философии, так и в информатике, однако контексты значений этого термина практически не пересекаются. Профессиональным философам, за редким исключением, не интересны реалии, конституирующие информационное пространство, а специалистам по информатике, как показывает опыт, чужды результаты философского осмысления проблем существования. Они предпочитают избрывать собственные философии и плодотворно упорядоченным произвольным образом списки имен сущностей, наивно считая их онтологиями.

В информатике само наличие от-рефлексируемой онтологии решает множество проблем, в частности проблему элиминации информационного шума, возникающую из различий контекстов и синонимии терминов в предметных областях знания и практики. Именно поэтому специалисты по информатике направляют свои филосо-

фические усилия на построение «единой онтологии» и на унификацию способов представления знания. При этом они опираются на методы лингвистического и статистического анализа текстов, считая, по-видимому, что содержание контекстов, выделение смыслов у них возникнет само по себе, по мере усовершенствования статистических и лингвистических методов исследования и представления их результатов.

С точки зрения авторов статьи, анализ содержания контекстов и выделение смыслов составляет новый предмет, интегрирующий философские и информационные методы в область знания, которую они назвали прикладной герменевтикой. Эта работа герменевтическая по предмету, в то время как ее метод не имеет аналогов ни в философской традиции, ни в информатике. Методика построения теоретических онтологий с использованием логики веерных матриц реализована в системе анализа текстов *Gitika*.

и система анализа текстов *Gitika*¹, элементом которой являются эти онтологии.

Мы рассматриваем эту работу как герменевтическую по предмету, в то время как ее метод не имеет аналогов ни в философской традиции, ни в информатике. Она имеет прикладной характер, поскольку ее результаты непосредственно используются для анализа текстов. Именно поэтому мы назвали прикладной герменевтикой то сопряжение предмета и метода, на котором основана работа.

Картины мира и понимание текстов

Любой текст содержит некую информацию, в нем можно найти смысл. Совокупность текстов, представленная в информационном пространстве, может рассматриваться как единый большой

1. *Gitika* — исследовательский проект компании *RelTeam*, связанный с созданием систем, использующих онтологии для извлечения информа-

текст. Этот текст в целом, как принято считать, хаотичен, в нем скрыто множество смыслов, однако выделение их проблематично из-за слабой развитости методологии и техники его анализа.

Для упорядочения хаоса и извлечения смыслов создаются информационные системы, базы данных и знаний, поисковые алгоритмы и другие механизмы. Их задача состоит в том, чтобы облегчить поиск осмысленной с точки зрения заинтересованных лиц информации. При этом каждый участник пополняет информационное пространство новыми текстами и смыслами, усложняя и без того непрозрачную среду.

Люди проецируют себя в информационное пространство прямыми или косвенными описаниями личных картин мира — текстами. Под картинами мира понимается индивидуальный набор представлений о том, что существует и как оно функционирует, чаще всего смутных и лишь частично поддающихся вербализации.

Помимо индивидуальных картин мира существуют и коллективные, общие для социальных групп. Это представления о нормах и правилах, а также институтах, в которых явно или неявно задаются над- и сверхличностные сущности. Последние проявляются в информационном пространстве в той мере, в которой они описываются людьми. Описания изначально акцентированы, так как люди при фиксировании представлений исходят из собственных картин мира.

Любая картина мира имеет ядро и периферию. В ядре находятся неререфлексируемые представления о том, что существует, то есть эмпирические онтологии. На периферии — представления о том, как сущее функционирует и почему так происходит, то есть разного рода теории и философии. В силу неотрефлексированности эмпирических онтологий в текстах представлены по большей части периферии картин мира: обыденные и профессиональные теории и философии, из которых ядерные компоненты нужно каким-то образом извлекать, если в этом возникает потребность.

Диалог как инструмент взаимопонимания представляет собой согласование периферий картин мира. Люди умеют договариваться и понимать друг друга, соглашаясь (или не соглашаясь) с теориями и философиями, которые стали предметами обсуждения. Но сближение и взаимопонимание в теориях и филосо-

ции, в частности для концептуального поиска и разработки соответствующих моделей взаимодействия с клиентом.

фиях не означает формирования общих представлений об эмпирическом существовании. Понимание в диалоге, как правило, поверхностно, оно лишь косвенно затрагивает сами эмпирические онтологии — то, что существует для каждого из беседующих людей в качестве данности, и поэтому с трудом поддается рефлексии. Но даже если эмпирические онтологии отрефлексированы, содержательное понимание все равно остается проблематичным, так как дискуссии, в которых один оппонент говорит, что нечто существует, а другой утверждает обратное, явно непродуктивны.

Содержательное понимание невозможно без экспликации эмпирических онтологий и последующего формирования теоретических представлений о том, какие эмпирические сущности лежат в основе предъявляемых в диалоге философий и теорий и как именно они друг с другом соотносятся. Содержательное понимание можно рассматривать как создание теоретических онтологий и последующую с их помощью проверку на адекватность предъявляемых в диалоге теорий, философий и их эмпирических онтологий.

Такова ситуация в обыденном общении. В нем чаще всего достаточно ощущения взаимопонимания, которое формируется взаимной риторикой в силу эмпатии и других психологических механизмов. Если проблемы непонимания не снимаются обыденной риторикой, то они становятся предметом философии и решаются в той мере, в которой интересующихся устраивают философские решения. Но когда взаимодействие между людьми обезличивается и переходит в информационное пространство, проблема понимания текстов становится острой и совсем не философской.

Эмпирические и теоретические онтологии

Прямой вопрос «Что существует?», заданный конкретному человеку, введет его в ступор. В ответе он начнет перечислять то, что считает существующим, задавая попутно отношения между обозначаемыми сущностями. Но сущности, как правило, оказываются составными, связанными отношениями происхождения или вообще логически несовместимыми. Человеку трудно, если не невозможно, отрефлексировать свои эмпирические онтологии.

Здесь обычно возникает апелляция к законам природы, теориям эволюции, учениям о творении и другим всеобъемлющим конструкциям, из которых, по мнению опрашиваемого, следует объяснение того, *что* существует, *как* оно функционирует и *по-*

чему так происходит. То есть философии и теории — элементы периферии картины мира — замещают сами представления о существовании, эмпирические онтологии, которых у каждого человека заведомо больше одной.

Принуждение к рефлексии над картиной мира делает явными противоречия между разными эмпирическими онтологиями, механически склеенными друг с другом в бытовых теориях и философиях сообразно культуре и институциональной среде, в которых человек воспитывался, получал образование или работал. Оно вызывает у него раздражение, ведь личностная картина мира существует как целостность, лишь пока ее не трогают, и начинает разламываться на противоречивые фрагменты при попытке ее отрефлексировать. Линии разлома проходят по границам эмпирических онтологий в силу того, что существования, данные в опыте, не могут быть совмещены в рамках только одной известной человеку философии или теории.

Противоречия между онтологиями эмпирическими могут разрешаться в теоретических, в которых понятия, соответствующие данным в опыте сущностям, тем или иным образом интегрированы в общее представление о мире. Теоретические онтологии определяются как части картин мира, в которых непротиворечиво объединяются понятия, соответствующие элементам эмпирических онтологий. Теоретические онтологии должны быть по меньшей мере непротиворечивыми и прояснять отношения между эмпирическими онтологиями.

Построение теоретических онтологий является целью профессиональной философии. Однако если первая задача (достижение непротиворечивости) как-то решается философами в каждой конкретной философии, то вторая (объяснение отношений), видимо, неразрешима при нынешнем философском подходе к проблеме. По традиции философы фиксируются на феноменах, имеющих весьма слабые референты в многообразных эмпирических онтологиях. Они создают авторские картины мира со специфическими философскими сущностями и сами же их воспроизводят.

Информационное пространство и инструменты его структурирования

Люди, входя в информационное пространство, имеют дело не с другими людьми, а с хаотичной совокупностью картин мира, облеченных в тексты. Там не с кем договариваться. Вернее, для

договаривания и сближения картин мира отведен один сегмент пространства — социальные сети, электронные почты и другие инструменты межличностной коммуникации, порождающие тексты особого рода. Все остальное безлично, хотя и может быть персонифицировано.

В информационное пространство, в сеть в широком смысле этого слова, вывалены и вываливаются сведения о всех возможных картинах мира. Для понимания смыслов необходимо эксплицировать онтологии, которые лежат в основе теорий и философий, представленных словами и фразами анализируемых текстов.

Анализ текстов на выходе имеет некий объем слов и их сочетаний (иногда гигантский), имеющих какое-либо отношение, обычно косвенное, к тому, что действительно интересует человека. «Взаимопонимание» в общении с информационной средой возникает в том случае, если человек получает удовлетворяющий его ответ на запрос, что означает установление соответствия между эмпирической онтологией фрагмента индивидуальной картины мира и онтологией информационного массива, который получился в ходе анализа. Периферийные элементы картин мира, теории и философии в этой форме понимания, специфичной для информационного пространства, играют существенно меньшую роль, чем в межличностном диалоге.

Для достижения «взаимопонимания» специалисты по информатике разработали способы работы с информацией, имитирующие диалог. Они предполагают, что, разбивая тексты, в которых изложены философии и теории, на разного рода лингвистические элементы и анализируя статистические связи между ними, можно в конечном счете реконструировать эмпирические онтологии, то есть ядра картин мира. Проблема находится в области теоретической онтологии, но специалисты по информатике пытаются ее решать как эмпирическую задачу. По сути они исходят из той же обыденной методологии, которой пользуется человек, пытающийся ответить на вопрос о том, что существует в его картине мира.

Специалисты по информатике строят «семантические сети» и «базы знаний», описывающие все сущее в виде отношений, понятий, их свойств и прочего. Но без указания картины мира эти «понятия» превращаются просто в слова, не несущие смысловой нагрузки. Ведь интерпретация результатов работы программы, использующей сведения о мире, записанных в виде семантической сети, требует знания о соображениях, из которых эта сеть строилась, знания о том, какие отношения рассматрива-

лись и какие не учитывались, в каких онтологиях и какими картинами мира должен обладать читатель, чтобы правильно понять полученные результаты.

Для упорядочения смысла требуются не только формализмы описания иерархий таксонов и семантических сетей, но и сами правила упорядочения. Однако теоретики и практики такого подхода считают правила построения онтологий сами собой разумеющимися, а основную проблему видят в недостатке ресурсов для создания единой непротиворечивой базы знаний.

То, что вызывает раздражение у обычного человека при попытке заставить его ответить на вопрос «Что существует?», специалисты по информатике превращают в свой хлеб — такую неисчерпаемую проблему, для решения которой необходимо беспрестанно увеличивать вычислительные мощности и разрабатывать все новые и новые алгоритмы анализа текстов. Действительно, если бы все люди были носителями единой картины мира и обладали четким представлением о том, как именно ее следует излагать, то для создания всеобъемлющего описания потребовались бы только организационные усилия и финансовые ресурсы.

Тем не менее специалисты по семантике и программисты разработали методы, которые позволяют извлекать простейшие эмпирические онтологии из текстов и формировать таксоны, то есть обобщенные представления о существовании. Эти методы являют собой имитации диалога между людьми, в которых человек, используя инструменты анализа текстов, задает вопросы информационному пространству и извлекает из него некие варианты ответов, которые оценивает как адекватные или не очень. В последнем случае он снова задает вопросы и снова получает ответы. Примером являются поисковые механизмы, такие как *Google*, которые, опираясь на внутреннюю статистику запросов, выдают списки ссылок на тексты, в которых встречаются искомые слова или их сочетания. Как правило, таких ссылок очень много, и извлечь из них нужное содержание достаточно сложно, если вообще возможно в случае нетривиального запроса.

Проблема в том, что существует не одна онтология, а целое множество, и в большей их части есть понятия, обозначаемые схожими или одинаковыми словами, но имеющими весьма различное содержание. Поэтому естественно желание построить единую непротиворечивую онтологию.

Миф о единой онтологии

Идея объединить эмпирические онтологии в одну структуру, пусть сложно устроенную, противоречит всему человеческому опыту, но весьма привлекательна для инженеров. Инженер рассматривает мир как набор деталей, элементов конструкций, кубиков, из которых он должен сделать работающий механизм. Но даже из детских кубиков, если на их гранях написаны буквы разных алфавитов, выстроить слово вряд ли возможно. Однако это не останавливает людей с инженерным подходом к миру, и они продолжают конструировать единую онтологию, искать свой философский камень. И в том случае, если они работают в узкой области, такой как товарная номенклатура, у них иногда получается нечто вполне операциональное, ведь потребительское поведение хорошо структурировано, чего не скажешь о результатах инженерных попыток расширить этот опыт на всю информационную практику.

Механическая, инженерная методология, безусловно, имеет плюсы. Она проста в своей логике, хотя машины и алгоритмы, реализующие ее, могут быть сколь угодно сложны. Инженерный подход основан на лингвистической методологии и технике обработки текстов, адаптированных к потребностям инженеров, и на неразличении эмпирических и теоретических онтологий. Слова, обозначающие сущности, при таком подходе считаются самими сущностями. Связи между словами (и иными лингвистическими феноменами) ими рассматриваются как отображение предметных (теоретико-онтологических) связей. Статистически устойчивые отношения между словами отождествляются с предметными связями между понятиями. Но связи между словами и связи между понятиями — это два различных порядка.

Если не соотносить слова с теоретическими онтологиями конкретной картины мира, то в результате поиска получается просто набор ссылок на тексты, упорядоченный внешним образом, статистически и бессодержательно, что знакомо любому человеку, искавшему что-либо в информационном пространстве. Поэтому инженерно-лингвистические ухищрения оказываются бесполезными при попытке различить, например, понятия «Путин» (онтология государственного управления и политики) и «путина» (онтология морского промысла).

Тем не менее инженерно-лингвистическая методология — за неимением других — выступает сейчас как базовая при кон-

струировании систем анализа текстов и обеспечивает понимание полученных ими результатов пускай в простейших, но самых распространенных ситуациях общения людей с информационным пространством, таких как поиск информации о товарах и услугах. Коммерческая эффективность данного подхода оправдывает инвестиции в совершенствование поисковых механизмов.

Энциклопедический подход к анализу текстов

Другие специалисты занимаются тем, что выстраивают разного рода базы знаний и данных и интерактивные энциклопедии, в которых даются определения сущностей и упорядочиваются их имена. Таковы «Википедия», многочисленные ее производные, *Wolfram|Alpha* и т.д. Эти инструменты исследования текстов по определению диалогичны, хотя формы диалога в них различны. В «Википедии» практически каждая статья становится предметом дискуссий между специалистами (и дилетантами) о той сущности, которая в статье описывается. Поэтому содержание статей «Википедии» представляет собой, как правило, компромисс между участниками дискуссии. *Wolfram|Alpha* содержит ответы (сформулированные специалистами) на конкретные вопросы и связи между этими ответами.

Связи между элементами баз данных и статьями сетевых энциклопедий (то есть совокупности гипертекстовых ссылок, идущих от статей к статьям) образуют специфические пространства, которые можно считать эмпирическими онтологиями предметных областей. Эти пространства можно определить, анализируя плотность связей между элементами баз данных теми методами, которые специфичны для инженерно-лингвистического подхода. Скажем, обработанную статистически совокупность ссылок на другие термины от статьи «физика» и на статью «физика» можно считать отображением структуры понятия «физика», то есть эмпирическим содержанием физической картины мира.

И все же базы данных и сетевые энциклопедии являются публичной демонстрацией того обстоятельства, что подрывает инженерно-лингвистическую методологию: онтологий очень много, и единственной быть не может. Однако энциклопедический подход остается ограниченным, он не позволяет работать со всем информационным пространством.

Анализ текстов и теоретические онтологии

Энциклопедический подход и выявление связей между статьями сетевых энциклопедий и баз данных позволяют эмпирически ограничить картины мира и различить их элементы. Инженерно-лингвистический подход позволяет выстраивать эмпирические онтологии так, как если бы существовала только одна картина мира, формировать таксоны и строить семантические сети. При этом основная задача анализа текстов — их понимание — не решается.

Понимание текстов возможно в том случае, если известно, к каким картинам мира они относятся. Для того чтобы соотнести текст и картину мира, вовсе недостаточно знания его эмпирической онтологии и результатов семантического анализа: необходимо также знание теоретической онтологии той картины мира, в которой текст имеет смысл.

Если теоретическая онтология не эксплицирована, то запрос, включающий механизм анализа, будет содержать слова, которые что-то значат в самых разных картинах мира. В результате анализа будут смешаны эмпирические онтологии всех тех картин мира, в которых есть понятия, обозначаемые похожими или одинаковыми словами.

Чтобы минимизировать шум, следует определить в информационном пространстве сегмент, в котором размещена информация о теоретических онтологиях разных картин мира и о правилах работы с ними, позволяющих в том числе отсекать информацию о тех картинах мира, в которых есть похожие, но неинтересные в данном случае термины.

Для того чтобы сформировать такой сегмент, нужен выход за пределы концепций, доминирующих в информатике. Однако сама идея третьего подхода к анализу текстов — не инженерно-лингвистического и не энциклопедического — чужда тем людям, которые должны были бы заниматься его созданием.

Базы знаний и данных, сетевые энциклопедии и инженерно-лингвистические технологии как методы анализа текстов и упорядочения информационного пространства должны быть дополнены собственными методами теоретической онтологии. Проблема в том, в социальной практике пока нет — за рассматриваемым далее исключением — иных способов формирования теоретических онтологий, кроме философских.

Если бы теоретические онтологии были инструментальными, тогда отношения между разными подходами к анализу текстов

Табл. 1. Отношения между методами упорядочения информационного пространства

Предмет	Метод	Семантики (инженеры- лингвисты)	Специалисты по теоретическим онтологиям
	Энциклопедисты		
Энциклопедический	Базы знаний, данных и сетевые энциклопедии	Эмпирические онтологии	Картини мира, заданные в терминах теоретических онтологий
Инженерно-лингвистический	Связи между элементами баз знаний и данных и статьями энциклопедий как понятия эмпирических онтологий (семантические сети)	Сетевые поисковые системы и другие инструменты для лингвостатистического анализа текстов	Семантические сети, заданные в терминах теоретических онтологий
Теоретико-онтологический (герменевтический)	Теоретические онтологии отдельных картин мира, описанных в базах знаний, данных и энциклопедиях	Теоретические онтологии семантики	Теоретические онтологии

выглядели бы так, как это показано в табл. 1. Она демонстрирует их возможную взаимодополнительность.

Таблица выстроена из предположения о том, что есть энциклопедический, инженерно-лингвистический и теоретико-онтологический (герменевтический) предметы и соответствующие специалисты, носители методов.

Отношения между одноименными предметами и методами интерпретируются как области социальной практики и знания — базы знаний и данных, сетевые энциклопедии, поисковые системы и пока гипотетические теоретические онтологии. Это диагональные элементы таблицы.

Отношение между инженерно-лингвистическим предметом (уровнем) и энциклопедическим методом дает эмпирические онтологии, то есть связи между статьями энциклопедий, семантические (онтологические) поля.

Отношение между теоретико-онтологическим предметом и энциклопедическим методом дает теоретические онтологии отдельных картин мира, представленных в энциклопедиях и базах данных.

Табл. 2. Отношения между энциклопедическим и инженерно-лингвистическим предметами и методами

Предмет	Метод	Семантики (инженеры-лингвисты)
	Энциклопедисты	
Энциклопедический	Сетевые энциклопедии, базы знаний и данных	Эмпирические онтологии
Инженерно-лингвистический	Связи между статьями энциклопедий и элементами баз знаний и данных как понятия эмпирических онтологий (семантические сети)	Поисковые системы и другие инструменты для лингво-статистического анализа текстов

Отношение между энциклопедическим предметом и инженерно-лингвистическим методом дает собственно эмпирические онтологии, то есть сущности, выделяемые инженерно-лингвистическими методами при анализе баз данных и сетевых энциклопедий.

Отношение между теоретико-онтологическим предметом и инженерно-лингвистическим методом дает теоретические онтологии семантики, то есть теоретико-онтологические конструкции, выделяемые инженерно-лингвистическими методами.

Отношение между энциклопедическим уровнем и теоретико-онтологическим методом дает картины мира, определенные в терминах теоретических онтологий.

Отношение между инженерно-лингвистическим предметом и теоретико-онтологическим методом дает семантические сети, определенные в терминах теоретических онтологий.

Если исключить из таблицы теоретико-онтологический предмет и соответствующие методы, то таблица упростится до вполне знакомого вида (табл. 2).

В системе представлений таблицы работают все без исключения инструменты анализа текстов. Так, *Google (поисковая система)* сейчас включает в свой поисковый механизм (*основанный на построении эмпирических онтологий*) связи между элементами купленной им базы знаний *Freebase (базы знаний и сетевые энциклопедии)*, что позволяет использовать для анализа текстов формализмы *семантических сетей*.

Введение уровня теоретических онтологий существенно расширяет возможное пространство инструментов анализа тек-

стов. Другое дело, что в настоящее время неизвестен ни один метод построения теоретических онтологий, кроме веерных матриц. Собственно, примером веерной матрицы являются табл. 1 и 2, в которых структурирована одна из теоретических онтологий информатики.

Начальные условия конструирования теоретических онтологий

Формулируя подход к построению теоретических онтологий, мы исходим из следующих начальных условий:

- Реальные потребители конструируемой системы анализа текстов могут иметь дело только с актуально (или потенциально) известными теоретическими онтологиями.
- Язык, описывающий конкретные наборы онтологий, не может содержательно рассматриваться без них. То есть сказанное или написанное само по себе не несет информации в отрыве от подразумеваемой теоретической онтологии.
- Результат работы системы анализа текстов может быть понят потребителем только в рамках пересечения теоретических онтологий потребителя и информационной системы. В системе уже должны содержаться теоретические онтологии; если в тексте идет речь об одном, а потребитель считает, что о другом, то понимание невозможно.
- Теоретические онтологии могут иметь общие элементы, то есть «быть похожими», «близкими» или «соседствующими».
- «Близость», «сходство» и «соседство» двух теоретических онтологий может определяться количеством общих элементов.
- Теоретические онтологии могут быть ортогональными (не содержащими общих элементов) и взаимодополнительными (имеющими общие элементы).
- На основе одного и того же набора элементов можно построить множество теоретических онтологий, отличающихся только правилами построения.
- Теоретические онтологии, объединенные последовательным применением единых правил построения, считаются принадлежащими к одному *пространству онтологий*. Количество таких пространств внешним образом не ограничено, то есть не существует единой и всеобъемлющей теоретической онтологии.

- Основным свойством теоретических онтологий должно быть единообразие в логике их построения. Причем единообразии логик построения должно реализовываться так, чтобы сохранялись предметные (содержательные) понятия, специфицирующие взятую картину мира. Это означает, что логика построения теоретических онтологий физики и психологии, например, должна быть одной. Они должны принадлежать к одному пространству, в данном случае научному.

Вероятно, может существовать множество способов формирования теоретических онтологий, однако на данный момент перечисленным начальным условиями удовлетворяет только методология, основанная на логике веерных матриц².

Прием утробения сущностей как способ формирования веерных матриц

Теоретические онтологии имеют в качестве элементов не слова, а понятия, то есть предельно идеализированные объекты, такие как идеальный газ или идеально твердое тело. Предельная идеализация достигается тем, что содержание каждого понятия полностью задается отношениями между «предметом» и «методом». Никакое иное содержание не учитывается. Под предметом понимается любая сущность или реальность, например реальность физики, химии, политики или быта. Под методом понимается специфический для конкретной реальности способ работы с ней. Реальности физики соответствует физический метод, носителями которого являются физики, а реальности быта — бытовой метод, носителями которого являются обыватели, и т. д.

Исходный материал для построения веерных матриц формируется следующим образом: если есть некая сущность, то она должна быть представлена как набор отношений, одно из кото-

2. В следующих работах продемонстрировано то, каким образом веерные матрицы позволяют формально решать проблемы существования и строить онтологии в разного рода информационных пространствах: *Кордонский С.* Циклы деятельности и идеальные объекты. М.: Пантори, 2001; *Он же.* Рынки власти. М.: ОГИ, 2000; *Он же.* Сословная структура постсоветской России. М.: Институт Фонда «Общественное мнение», 2008; *Он же.* Россия: поместная федерация. М.: Европа, 2010; *Он же.* Веерные матрицы как инструмент построения онтологий. Вашингтон: Юго-Восток, 2011; *Кордонский С., Бардин В.* О поиске информации в совокупности текстов, отображающих картины мира. Вашингтон: Русгенпроект, 2010.

рых задается названием предмета, а другое — названием метода, соответствующего (одноименного) предмету.

Методология веерных матриц позволяет формализовать процесс построения теоретических онтологий исходя из приема «утроения сущностей». Область знания (наука, социальная форма организации информации) в процедуре утроения сущностей возникает как результат формального отношения между предметом и методом или, что то же самое, между уровнем структуры (организации) и одноименным этому уровню специалистом.

К примеру, наука экономика может быть представлена отношением между предметом экономики (экономическим уровнем социальной реальности) и экономистом как специалистом, в деятельности которого воплощен соответствующий метод изучения или изменения реальности. То есть вместо одной сущности — науки экономики — появляются три: экономический уровень (или предмет экономики), специалист-экономист и собственно наука экономика как отношение специалиста-экономиста к экономическому уровню. Такая утроенная по отношению к исходной области знания сущность в дальнейшем называется элементарным агрегатом. Любая область знания может быть представлена как элементарный агрегат.

Само по себе утроение сущностей и формирование элементарных агрегатов не имеет каких-либо значимых последствий для решения задач организации и поиска информации. Но если удастся связать несколько агрегатов в одной логически прозрачной и операциональной схеме, то возникает возможность выстроить формальную структуру, объединяющую предметы и методы разных областей знания и деятельности.

Традиционно проблему объединения предметов и методов различных областей знания пытаются решить философы, методологи и классификаторы наук. Они основываются либо на предмете (философы), либо на методе (методологи), либо на науках (классификаторы).

Философы описывают связи между предметами наук. Результатом такого подхода становятся разного рода натурфилософии (или общие теории всего). Сами науки и их методы в этом случае рассматриваются как нечто внешнее по отношению к той целостности, которую пытается описать натурфилософы.

Методологи описывают методы. В результате возникают методологии науки и теории деятельности с их разделением на точные, естественные и гуманитарные методы исследования.

Методы наук связаны друг с другом. Так, математический метод сопряжен с естественно-научными методами исследования. Более того, экспансия математики в естественные области знания, по мнению методологов, во многом объясняет успехи новейшего времени в конструировании разного рода приборов и машин. А проникновение физических методов исследования в биологию стало причиной экспоненциального развития аналитических биологических наук. Науки и их предметы используются при таком подходе к интеграции знаний как иллюстрации конструкций методологов.

Классификаторы описывают многообразие социальных форм организации знания. В результате возникают классификации наук, в которых их предметы и методы связываются между собой по собственной логике классификатора. Сами науки также связаны друг с другом, что отображается в классификациях. При этом предметы и методы включаются в построения классификаторов наук только в качестве иллюстраций их логики.

Результаты философских, методологических и науковедческих попыток синтезировать структуры знания имеют много общих элементов, потому что образованы из одних и тех же элементарных агрегатов. И они различаются между собой доминированием интерпретаций, в которых определяющим для интерпретаторов выступают или предметы, или методы, или науки и области деятельности.

Некоторые результаты стремления философов, методологов и классификаторов упорядочить структуры знания имеют безусловную культурную ценность, однако в целом они совершенно не операциональны в тех случаях, когда надо решать конкретные задачи организации информации и ее поиска.

Построение элементарных агрегатов

Верные матрицы дают возможность «сцеплять» агрегаты по трем основаниям сразу. В результате возникают упорядоченные структуры, в которой науки, их предметы и методы образуют расчлененную на элементы целостность. Это происходит в ходе разбиения и собирания своеобразного многомерного пазла, в котором в качестве кусочков выступают описанные выше агрегаты.

Процесс сборки определяется тем, что кроме элементарных встречаются также и сложные агрегаты, устойчиво объединяющие объекты, которые принадлежат к разным предметам и исследуются разными методами. В предыдущих наших работах рассматривались такие агрегаты, как «организмы–клетки–молекулы», «потребности–память–эмоции» и др. Понятия объеди-

няются в последовательности на основании принадлежности их к единому предмету исследования, что позволяет выстроить многомерные агрегаты и собрать пазл, объединяющий предметы, методы и науки современного аналитического знания. Технология складывания пазлов из элементарных агрегатов, вероятно, может быть применена к любой представленной в текстах информации.

Основополагающими при построении веерных матриц выступают связи между предметами, в то время как связи между методами и областями знания формально выводятся из них. Далее на примерах будет показано, как пошагово конструируются веерные матрицы.

Конструирование онтологии здравого смысла

Рассмотрим следующий нарочито огрубленный пример. Известно, что существует здравый смысл как некое знание, позволяющее обывателю ориентироваться в окружающем мире. Зададимся вопросом, в каких отношениях он существует. Здравый смысл является атрибутом бытового уровня жизни, очевидным для обывателя. То есть его можно определить как отношение обывателя к бытовому уровню жизни (табл. 3).

Таким образом, из понятия «здравый смысл» сформирован агрегат, включающий в себя предмет «бытовой уровень», метод, воплощенный в «обывателе», и сам «здравый смысл» как их отношение.

На бытовом уровне может существовать множество реалий, и здравый смысл — одна из них. К примеру, известно, что существуют вещи. В данном случае они определяются как отношения потребителей к бытовому уровню. Тем самым формируется второй агрегат (табл. 4).

Табл. 3. Определение отношений, в которых существует здравый смысл

	Метод	Обыватель
Предмет		
Обывательский предмет		Здравый смысл

Табл. 4. Определение отношений, в которых существуют вещи

	Метод	Потребитель
Предмет		
Вещный уровень (предмет)		Вещи (знание о вещах)

Табл. 5. Определение отношений, в которых существуют деньги

	Метод	Покупатель
Предмет		
Покупательский уровень		Деньги (знание о деньгах)

На бытовом уровне существуют также деньги. Естественно, это не те деньги, которыми оперируют инвесторы, а те, что находятся в кошельках и на счетах обывателей. Определим их как отношение между покупательским уровнем и покупателем (табл. 5).

Сопряжение элементарных агрегатов в веерную матрицу

Эти агрегаты можно связать друг с другом, предположив, что бытовой, потребительский и покупательский уровни сопряжены между собой, то есть являются уровнями одной структуры (табл. 6).

Табл. 6. Объединение элементарных агрегатов в таблицу

Предмет	Метод	Обыватели	Потребители	Покупатели
Бытовой уровень		Здравый смысл		
Потребительский (вещный) уровень			Знание о вещах	
Покупательский (денежный) уровень				Знание о деньгах как о кэше

Незаполненные элементы таблицы означают, что в рамках развиваемого представления существуют элементарные агрегаты, которые определяются как отношения обывателей, потребителей и покупателей к неоднoименным им предметам. Например, отношение обывателей к покупательскому предмету интерпретируется как «разумные деньги», то есть деньги, которыми может располагать обыватель для совершения покупок (табл. 7).

Табл. 7. Определение отношений, в которых существуют деньги

Предмет	Метод	Обыватель
Покупательский (денежный) предмет		«Разумные деньги»

При заполнении всех отношений между предметами и методами формируется веерная матрица, в которой упорядочиваются элементарные агрегаты, относящиеся к области здравого смысла и потребительского поведения (табл. 8).

Табл. 8. Объединение элементарных агрегатов в таблицу

Предмет	Метод	Обыватели	Потребители	Покупатели
Бытовой уровень		Здравый смысл	Вещные потребности — согреться, умыться, поесть и т. д. (знание о здоровом смысле относительно вещей)	Денежные потребности — здравый смысл по отношению к деньгам (знание о здоровом смысле по отношению к деньгам)
Потребительский (вещный) уровень		«Нужные вещи», которые полагаются иметь согласно здравому смыслу (знание о вещах, которые могут иметь обыватели согласно своему здравому смыслу)	Знание о вещах	Покупаемые вещи (знание о спросе на вещи)
Покупательский (денежный) уровень		«Разумные деньги», которыми должен располагать обыватель согласно здравому смыслу (знание о деньгах, которые могут тратить обыватели)	Стоимость вещей (знание о ценах на вещи)	Знание о деньгах как о кэше

Таблица представляет собой сопряженные по предметам элементарные агрегаты, образованные отношениями между методами (персонифицируемыми обывателями, потребителями и покупателями) и предметами. Ее можно считывать как по строкам (по предметам), так и по столбцам (по методам). Так, покупательский предмет образован следующими понятиями: «разумные деньги» — стоимость вещей — знание о деньгах как кэше. А обывательский метод образован понятиями: здравый смысл — нужные вещи — «разумные деньги».

В клетках размещены два объекта: наименование феномена, соответствующего отношению между предметом и методом (между уровнем структуры — строкой таблицы — и носителем метода), и наименование области знания, соответствующей этому феномену.

Табл. 9. Вычленение элементарного агрегата второго уровня

Предмет	Метод	Потребители	Покупатели
Потребительский предмет		Здравый смысл	Вещные потребности (знание о здоровом смысле по отношению к деньгам)
Потребительский (вещный) уровень		«Нужные вещи» (знание о вещах, которые могут иметь обыватели согласно своему здравому смыслу)	Знание о вещах

Такую матрицу можно считать отражением структуры бытовых представлений о мире. Логика веерных матриц позволяет произвести их детализацию, то есть сформировать (на основе табл. 2) описание структуры второго (по отношению к табл. 6) уровня бытовой онтологии.

Формирование агрегатов второго уровня матрицы

Выберем из табл. 8 элементарные агрегаты, симметричные относительно ее диагонали (табл. 9).

Предположим, что элементы этих агрегатов могут быть объединены в новый предмет — «покупательско-потребительский». Это означает, что знание о вещах, которые могут иметь обыватели согласно своему здравому смыслу, и знание о здоровом смысле по отношению к деньгам объединяются в новую область знания об обыденном потреблении, а из методов, которые были персонифицированы покупателями и потребителями, формируется новый, который персонифицируют потребляющие обыватели (или обыденные потребители). То есть нужные вещи и вещные потребности формируют новый предмет — обыденное потребление (табл. 10).

Табл. 10. Логика формирования агрегата «обыденное потребление»

Предмет	Метод	Покупающий потребитель (потребляющий покупатель)
Уровень (предмет) нужных вещей и вещных потребностей		Обыденное потребление (знание об обыденном потреблении)

Табл. 11. Структура теоретической онтологии быта (второй уровень)

Уровни структуры	Специалисты	Потребляющие обыватели (обыденные потребители)	Покупающие обыватели (обыденные покупатели)	Покупающие потребители (потребляющие покупатели)
Потребительско-обывательский уровень: <i>нужные вещи + + вещные потребности</i>		Знание об обыденном потреблении	Знание о том, какие вещи нужны: <i>[нужные вещи + + вещные потребности] / обыденные покупатели</i>	Знание о цене нужных вещей: <i>[нужные вещи + + вещные потребности] / потребляющие обыватели</i>
Покупательно-обывательский уровень: <i>разумные деньги + + денежные потребности</i>		Знание о том, сколько можно потратить: <i>[разумные траты + денежные потребности] / бытовые потребители</i>	Знание об обыденных покупках	Знание о том, сколько денег можно потратить: <i>[«разумные деньги» + + денежные потребности] / потребляющие покупатели</i>
Потребительско-покупательский уровень: <i>стоимость вещей + + покупаемые вещи</i>		Знание о том, что нужно купить: <i>[стоимость вещей + + приобретенные вещи] / бытовые потребители</i>	Знание о том, что можно купить: <i>[стоимость вещей + + покупаемые вещи] / обыденные покупатели</i>	Статистика покупок

В обыденном потреблении действуют те же самые отношения, в которых формировались первоначальные агрегаты. То есть оно рассматривается как отношения между предметом — уровнем обыденного потребления — и методом, носителем которого является обыденный потребитель.

Та же процедура, проделанная над другими элементами таблицы, позволяет трансформировать структуру табл. 9, в которой вычлененные из табл. 6 элементарные агрегаты образуют предметы (уровни) новой таблицы, соответствующей бытовой картине мира второго уровня (табл. 11).

Полученная онтология может использоваться для анализа текстов, касающихся обыденного потребления.

Реконструкция онтологии товароведческо-логистическо-складской предметной картины

Элементарный агрегат в этом случае представляет собой отношение товароведа к товарному уровню. Результатом отношения является товар и соответствующая ему область знания — товароведение (табл. 12).

Параллельно с товароведческим может существовать логистический агрегат, включающий в себя предмет «перемещение товаров», логистический метод (носителем которого выступает товаровед) и область знания — собственно логистику (табл. 13).

Табл. 12. Агрегат товароведения

	Метод	Товароведы
Предмет		
Товарный предмет		Товары (товароведение)

Табл. 13. Агрегат логистики

	Метод	Логисты
Предмет		
Вещный уровень (предмет)		Перемещение (логистика)

Табл. 14. Агрегат складского хозяйства

	Метод	Специалисты по складир.
Предмет		
Покупательский уровень		Склад (теория складирования)

С товароведческим и логистическим уровнями сосуществует складской агрегат, включающий в себя предмет «складское хозяйство», соответствующий метод, носителем которого выступают специалисты по складскому хозяйству, и область практического знания — науку о хранении и складском хозяйстве (табл. 14).

Предположим, что предметы отображены в структуре верной матрицы как ее уровни, а одноименные предметам специалисты — носители методов — отображены как наименования столбцов таблицы (табл. 15).

Ранее рассмотренные агрегаты являются структурными элементами таблицы: товарный уровень соответствует предмету товароведения, метод которого представлен товароведцами. Логистический уровень соответствует предмету логистики, метод которой представлен логистиками. А складской уровень представлен предметом «складское хозяйство» и наукой о хранении, метод которой представлен специалистами по складированию.

Предположим, что незаполненные отношения между носителями методов и предметами табл. 12 могут быть описаны так, как это сделано в табл. 16.

Товароведы, обращаясь к логистическому предмету, фиксируют существование феномена перемещения товаров (матери-

Табл. 15. Предметная картина, образованная совокупность агрегатов товароведения, логистики и теории складирования

Предмет	Метод		
	Товароведы	Логистики	Специалисты по складированию
Товарный	Товар (товароведение)		
Логистический		Перемещение (логистика)	
Складской			Склады (теория складирования)

Табл. 16. Товарно-логистико-складская теоретическая онтология

Предмет	Метод		
	Товароведы	Логистики	Специалисты по складированию
Товарный	Товар (товароведение)	Перемещаемые товары (логистически-товарное знание)	Складированные товары (складское-товарное знание)
Логистический	Перемещение товаров (товарно-логистическое знание)	Перемещение (логистика)	Внутрискладское размещение и перемещения (складское-логистическое знание)
Складской	Хранение товаров (товарно-складское знание)	Логистическое устройство складов (логистически-складское знание)	Склад (теория складирования)

альных потоков), а логистики, обращаясь на товарный уровень, фиксируют существование феномена перемещаемых товаров. Товароведы, обращаясь к складскому предмету, фиксируют феномен хранения товаров, в то время как специалисты по складскому хозяйству при обращении к товарному предмету фиксируют феномен складированных товаров. Логистики, работая на складском уровне, фиксируют феномен логистического устройства складов, а специалисты по складскому хозяйству, обращаясь к логистическому предмету, фиксируют феномен внутрискладского размещения и перемещения. Тем самым у нас получается предметная картина, содержащая сопряженные предметы товароведения, логистики и науки о складировании, а также производные от них области знания.

Табл. 17. Товарно-логистическо-складская теоретическая онтология (второй уровень)

Предмет	Метод	Специалисты по перемещаемым товарам и перемещению товаров	Специалисты по хранению товаров и складированию	Специалисты по логистическому устройству складов и внутрискладскому хранению и перемещению
Уровень перемещения товаров и перемещаемых товаров (товарно-логистическое знание)		Наука о перемещениях товаров	Наука о перемещении товаров применительно к складированию и хранению	Наука о перемещениях товаров применительно к внутреннему устройству складов
Уровень хранения и складирования товаров (товарно-складское знание)		Наука о перемещении складированных товаров	Наука о хранении товаров и их складировании	Наука о складировании и хранении применительно к устройству складов
Уровень логистического устройства складов и внутрискладского хранения и перемещения (логистически-складское знание)		Наука о логистическом устройстве складов в аспекте перемещаемых товаров	Наука о логистическом устройстве складов в аспекте хранения и складирования	Наука о логистическом устройстве складов

Верная матрица позволяет создавать и более детализированные картины. Предположим, например, что феномены, размещенные симметрично относительно диагонали таблицы (областей знания), могут выстраивать новые предметы, сопровождающиеся развитием соответствующих методов и формированием новых областей знания. Применительно к рассмотренному выше случаю это означает, что на пересечении товароведения и логистики образуются новые предметы, а также методы исследования и управления: перемещение товаров и перемещаемые товары учреждают предмет новой области знания точно так же, как хранение товаров и складированные товары, как логистическое устройство складов и складская логистика. Эти предметы соответствуют предметам и методам, представленным на табл. 17, производной от табл. 16.

Таким образом, верные матрицы позволяют конструировать теоретические онтологии произвольных картин мира, упорядочивая их эмпирические онтологии.

Система анализа текстов *Gitika*

Выше описан понятийный аппарат и подходы к решению проблем, сформировавшиеся в процессе четырехлетней работы над проектом *Gitika*. Отсутствие понятийного аппарата, в рамках которого можно было бы описать возникавшие в ходе работы проблемы и методы их решения, вынудило предпринять меры по уточнению аппарата, переопределению старых и введению новых понятий, что, собственно, и привело к появлению данной статьи.

Технология для экстракции знаний из произвольно взятых текстов предполагает следующие умения:

1. Умение описывать пространство (онтологий), ограничивающее компетенции системы и позволяющее определять рассматриваемые онтологии как «близкие», «далекие», «высокого уровня» и пр. Собственно, для этого используется аппарат веерных матриц как инструмент построения пространств сравнимых онтологий.
2. Умение определять то, в каких онтологиях написан текст.
3. Умение определять то, как онтологии модифицируют данный текст.
4. Умение определять позиции наблюдателя в данном пространстве онтологий.
5. Умение выяснять, о каких объектах, связях, событиях и процессах в этих онтологиях идет речь.

Информационная система *Gitika* «умеет» совершать названные выше операции. В ней сопрягаются техники формирования эмпирических онтологий путем индексирования связей между статьями сетевых энциклопедий и баз данных и техники формирования теоретических онтологий посредством конструирования веерных матриц.

Система может позиционировать тексты, то есть определять наборы векторов, которые указывают позиции исходного текста в пространстве сравнимых онтологий. Это позволяет искать документы не только по ключевым словам (частично известному содержанию), но и по затронутым темам.

Сопоставление слов и выражений с объектами и экземплярами формализованной онтологии является отдельной задачей, связанной со спецификой языка документа. Платформа *Gitika* является языково независимой в своем принципе, однако для реализации его необходима обработка национальных сетевых энциклопедий.

При наличии нескольких наборов сравнимых онтологий позиционирование текстов и идентификация их объектов выполняются в несколько этапов:

- Сначала выдвигаются гипотезы о связи слов и выражений текста с понятиями, включенными в поддеревья известных онтологий, полученные при экстракции связей «Википедии».
- Затем определяются поддеревья онтологий, получившие наибольшие веса (*вес* — число, определяемое количеством и плотностью в поддереве сопоставленных и связанных между собой слов и выражений).
- На заключительном этапе исключаются слабые с точки зрения веса поддеревьев гипотезы.

Вектора строятся на основе идентификации *контекстов* (тематических, географических, временных), определяющих интерпретацию текста, *объектов* текста, а также описываемых в нем *событий* и *процессов*.

Логика концептуального поиска может быть представлена в следующей форме: энциклопедии и базы данных, как и другие источники информации, индексируются, и из них извлекаются статистически устойчивые связи между статьями. Эти связи получают статус эмпирических онтологий. Параллельно конструируются наборы веерных матриц.

Эмпирические онтологии сопрягаются с соответствующими веерными матрицами и получают собственные имена, то есть соотносятся с теоретическими онтологиями. Запрос к системе формулируется либо в терминах матриц, либо в терминах эмпирических онтологий. В первом случае результатом поиска будут некие эмпирические онтологии, во втором — конкретные тексты, в которых представлены понятия, а не слова обыденного или предметных языков.

Практика реализации системы такова, что она позволяет различать контексты употребления одинаковых слов. Система способна «учиться», то есть запоминать фиксированные понятия и предлагать их (или их синонимы) человеку, осуществляющему поиск.

Таким образом, систему анализа текстов *Gitika* можно рассматривать как реализацию идеи о формировании языково независимого в своем принципе инструмента, позволяющего структурировать информационную среду и осуществлять кон-

цептуальный поиск, равно как и другие операции, связанные с анализом текстов. Необходимым компонентом системы является технология конструирования теоретических онтологий, основанная на методологии верных матриц.

Если обратиться снова к табл. 1, то можно видеть, что система анализа текстов *Gitika* включает все ее элементы. В ее основе лежат связи между статьями сетевых энциклопедий и баз данных и технология выстраивания эмпирических онтологий. Она позволяет создавать теоретические онтологии картин мира, описывать и сравнивать их между собой, определять их близость/удаленность. Также она позволяет экстрагировать из семантических сетей теоретические онтологии и сравнивать их по степени сходства или различия и в прочих отношениях.

Applied Hermeneutics of the Information Space: Worldviews, Theoretical Ontologies, and Fractal Matrix Tables

Simon Kordonsky. PhD in Philosophy, Professor, Head of the Department of Local Administration at the School of Public Administration of the Faculty of Social Sciences of the National Research University—Higher School of Economics. Address: 20 Myasnitskaya str., 101990 Moscow, Russia.
Email: kordonsky@gmail.com.

Valery Bardin. RelTeam CEO. Address: Bldg 8, 6 Kuusinena str., 123308 Moscow, Russia. 8.
E-mail: foxgrep@gmail.com.

Keywords: ontology; information technology; hermeneutics; contexts; fractal matrix tables.

The term *ontology* is used both in modern philosophy and in the sphere of IT, but the contexts and the matters of its meanings do not coincide or overlap. With few exceptions, professional philosophers are not interested in addressing the information space, or in investigating and comprehending it. At the same time, IT specialists ignore philosophical approaches to existential problems. They prefer to invent their own philosophies and to produce randomly ordered terms

for phenomena which they naively regard as ontologies.

In IT, the very existence of the reflected ontology solves many problems. In particular, it solves the issue of information noise which arises from differences in contexts and from synonymous terms in the domain of knowledge and the domain of practice. Therefore IT specialists focus on producing a common ontology and thinking of ways to standardize methods of presenting knowledge. In undertaking this task, they rely mostly on linguistic and statistical text analysis methods, as they believe that improving these methods will help to automatically establish meanings,

The authors of the article argue that an integration of philosophical and IT methods and practices will lead to the establishment of a new subject called *applied hermeneutics*. This new subject will deal with hermeneutics, and its methods would be new both in philosophy and in IT. The technique of constructing theoretical ontologies by using matrix tables will be implemented using the text analysis system called *Gitika*.