

АНАТОМИЯ ФИЛОСОФИИ. РЕПЛИКИ

© 2019 С.П. КОВАЛЁВ, А.В. РОДИН

ЗНАНИЯ И ИХ ПРЕДСТАВЛЕНИЕ В КОМПЬЮТЕРНУЮ ЭПОХУ



Ковалёв Сергей Протасович — доктор физико-математических наук, ведущий научный сотрудник. Институт проблем управления им. В.А. Трапезникова РАН. Российская Федерация, 117997 Москва, ул. Профсоюзная, д. 65. Электронная почта: kovalyov@nm.ru



Родин Андрей Вячеславович — кандидат философских наук, старший научный сотрудник сектора философских проблем естествознания. Институт философии РАН. Российская Федерация, 109240 Москва, ул. Гончарная, д. 12, стр. 1. Электронная почта: andrei@philomatica.org

Аннотация. Представление знаний (ПЗ) — это область исследований, которая лежит на стыке компьютерной науки и философской эпистемологии. В техническом смысле ПЗ можно характеризовать как раздел исследований по искусственному интеллекту (ИИ), связанный с разработкой информационных и коммуникативных систем для эффективного использования больших объемов данных. С другой стороны, знания и их представление — это одна из центральных тем философской рефлексии

Работа подготовлена при поддержке РГНФ; грант № 19-0011-00799.

и анализа как в традиционной философии, так и в аналитической философии XX века, которая во многом опирается на формальные логические подходы. Участники беседы обсудили новые эпистемические практики, использующие компьютерные технологии представления знаний.

Ключевые слова: представление знаний, искусственный интеллект, нейронные сети, онтологии, экспертное знание.

Ссылка для цитирования: Ковалёв С.П., Родин А.В. Знания и их представление в компьютерную эпоху // Человек. 2019. Т. 30, № 4. С. 94–112. DOI: 10.31857/S023620070005953-5

С.П. Ковалёв,
А.В. Родин
Знания и их
представление
в компьютерную
эпоху

Участники: С.П. Ковалёв, А.В. Родин Организатор проекта и ведущая — Ю.В. Синеокая

Ю.В. Синеокая. Дорогие друзья, сегодняшний диалог из цикла «Реплики» совместного проекта Института философии РАН и библиотеки им. Ф.М. Достоевского «Анатомия философии: как работает текст» будет посвящен проблеме современного представления знаний. Философ-эпистемолог А.В. Родин и эксперт в области компьютерного представления знаний С.П. Ковалёв обсудят важнейший феномен конца XX столетия: массовое распространение вычислительной техники и создание глобальной информационной сети. В чем заключается смысл и каковы будут последствия этого нового шага в развитии информационных технологий? Речь пойдет о существующих и перспективных технологиях компьютерного представления знаний, о новых возможностях, которые эти технологии дают для познания и для практических нужд.

А.В. Родин. Знание является предметом исследования для философской дисциплины эпистемологии. Понятие *представление знаний* указывает на формальный и материальный аспекты знания. Компьютерная революция последних десятилетий затрагивает оба этих аспекта и заставляет серьезно переосмыслить многие привычные эпистемические практики. Я начну с материального аспекта проблемы, а потом немного скажу о формальном.

По крайней мере для людей моего поколения самыми привычными материальными носителями знаний являются книга и другие виды печатных текстов. Но представление, что печатные тексты — это некие сосуды, содержащие своего рода нематериальную субстанцию, которую мы называем знанием, конечно, очень наивно. Книгами мы называем не просто материальные объекты, обладающие определенными физическими характеристиками, а объекты, которые мы определенным образом используем. Подобным образом устроены такие обиходные понятия, как «молоток», «посуда» и многие другие. Знание не «содержит-

доступ к знаниям любого рода является желательным, но я сейчас не буду начинать дискуссию по этому вопросу и приму тезис о желательности такого доступа в качестве рабочей гипотезы.

Для ученого-гуманитария вроде меня самая очевидная польза компьютера и Интернета состоит в том, что они позволяют делать привычные нам вещи более эффективно, то есть с меньшими усилиями. Я имею в виду создание, редактирование, публикацию, распространение и обмен текстами в электронной форме. И хотя за эти преимущества приходится платить временем и усилиями на приобретение компьютерной грамотности, эти усилия очень быстро окупаются.

На первый взгляд, в подобных случаях новые технологии не приносят ничего существенно нового и ничем не угрожают привычным институтам и практикам: мы делаем то же, что делали и в докомпьютерную эпоху, только тратим меньше времени на техническую рутину вроде переписывания и перепечатывания черновых вариантов текста, выписывания цитат из книг в читальном зале и проч. Однако это только на первый взгляд. Википедию нельзя назвать академическим проектом, но она на сегодняшний день является очень эффективным инструментом для представления и распространения знаний. Главное отличие Вики от традиционных бумажных энциклопедий в том, что у нее нет ни главного редактора, ни редакционной коллегии, ни специально отобранной группы авторов: роль авторов и редакторов выполняют активные пользователи. Администрация Вики минимальна (главная ее задача — борьба со злонамеренным вандализмом, а не контроль качества статей) и по большей части также состоит из волонтеров. Ничего подобного невозможно было бы организовать, используя традиционную бумажную технологию представления знаний.

Таким образом, Вики — это пример того, как развитие материальной базы технологий представления знаний приводит к возникновению новых социальных форм эпистемических практик. Другим таким примером является бурное развитие так называемых массовых открытых онлайн-курсов (Massive Open Online Courses — MOOC), которое, по всей видимости, обещает очень существенно видоизменить существующие сегодня образовательные институты, в первую очередь университеты, обеспечив доступ к качественному образованию миллионам и миллиардам молодых людей из бедных стран.

Разумеется, новые электронные технологии представления и распространения знаний не только открывают новые возможности, но и ставят новые проблемы. Скажу только об одной такой проблеме, которой, на мой взгляд, пока не уделяют достаточно внимания. Что будет с нашими электронными публикациями через 50, 100 и 500 лет? В наших библиотеках есть книги и рукописи,

С.П. Ковалёв,
А.В. Родин
Знания и их
представление
в компьютерную
эпоху

широкие возможности для развития новых форматов представления знаний, включая создание специально предназначенных для этой цели искусственных языков. Здесь мы касаемся темы «представления знаний» в более специальном и техническом смысле слова: в компьютерных науках так называют специальную область исследований по искусственному интеллекту (ИИ), в которой рассматривается теория программных средств для представления знаний. Такие средства уже используются в некоторых областях техники. Но и в этой специальной сфере важную роль может играть эпистемологическая экспертиза.

Базовый элемент представления знаний (ПЗ) — это спецификация *онтологии* (это тоже термин, используемый в компьютерных науках, а не только в философии!), то есть спецификация основных типов объектов данной области знаний, отношений между объектами, правил и аксиом, регулирующих объекты и отношения, и, возможно, других подобных принципов и элементов. Для задач ПЗ важно, чтобы онтология данной области знаний была сформулирована не просто в общих словах на естественном языке, но и формально, с использованием символического логико-математического формализма, который далее сможет быть реализован в программном коде.

Слушатели, знакомые хотя бы в общих чертах с аналитической философией XX века, сразу заметят, что использование одного и того же термина «онтология» в философии и компьютерных науках не случайно: начиная, по крайней мере, с «Философии логического атомизма» Рассела (1918), в рамках этой философской традиции ставились и по-разному решались задачи формального представления знаний. Сегодня формальная онтология — это та область, где философы, логики и специалисты в компьютерных науках активно сотрудничают. В числе пионеров такого сотрудничества со стороны философии можно назвать Барри Смита (Barry Smith) из Университета Буффало.

Аналогично дело обстоит с формальной эпистемологией — более новой философской дисциплиной, которая пользуется формальными логическими методами для решения традиционных и новых эпистемологических проблем. Хотя связь между формальной философской логикой и работами по искусственному интеллекту и компьютерному представлению знаний кажется очевидной, на самом деле философски мотивированные логические схемы и конструкции очень редко можно применить на практике, реализуя их напрямую в виде программного кода. Дело в том, что философски мотивированная формальная логика в XX веке, как правило, не принимала в расчет ограниченность вычислительных ресурсов и ресурсов памяти, а ставила и решала свои проблемы «в принципе», то есть с помощью сильных идеализаций. Исключением является конструктивное направле-

С.П. Ковалёв,
А.В. Родин
Знания и их
представление
в компьютерную
эпоху

который, надо признать, во многих случаях влечет за собой рост качества знаний. Но при этом отсутствуют механизмы обратной связи, которые позволяли бы контролировать этот процесс и делать его целенаправленным, делать знания полезными в широком смысле этого слова. Эту ситуацию можно сравнить с попаданием кроликов в Австралию. В Европе у кроликов есть естественные враги, такие как волки и лисы, благодаря чему европейская популяция кроликов остается ограниченной. Когда кроликов в середине XIX века завезли в Австралию, где у них не оказалось естественных врагов, они начали катастрофически размножаться: их стало намного больше, чем людей, которые могли бы их истреблять и ими питаться. Это привело к экологической катастрофе в масштабах целого континента. Есть и другие примеры подобных экологических катастроф.

Ситуация во многих областях знания сегодня схожая. Имеются огромные объемы знаний, которые мало кто способен «переварить», оценить и, тем более, применить. Приведу пример из математики. Речь идет о классификации всех простых конечных групп — результате, который был получен группой из примерно 100 математиков (часть из них работала совместно, а часть — независимо друг от друга) и опубликован в виде нескольких сотен журнальных статей в период с 1950-х до начала 2000-х годов. Если сформулировать этот результат в виде одной теоремы, то ее доказательство займет около 15 000 печатных страниц математического текста [3]. Проблема состоит в том, что это доказательство могут понять едва ли несколько человек в мире, причем ни один из них не способен проверить все шаги этого доказательства. Несмотря на такие трудности, эти несколько человек договорились о том, чтобы признать данную классификацию конечных простых групп правильной. Можно согласиться с тем, что эта классификация — значительное достижение в области чистой математики. Но давайте спросим себя: зачем оно нужно? Если бы эту теорему можно было включить в обычный университетский учебник, то студенты, изучающие эту теорему, могли бы пытаться применить ее в новых компьютерных технологиях или где-то еще. Если бы эта теорема была доступна экспертам в разных областях знаний — будь то космология, археология или какая-то другая область, они могли бы пытаться ее использовать для решения задач в своих специальных областях. Но в данном случае это, очевидно, невозможно. Как и в случае кроликов в Австралии, причина этой проблемы состоит, на мой взгляд, в отсутствии обратной связи, которая бы регулировала количество и полезность знания в соответствии с той или иной шкалой ценностей. При отсутствии такой обратной связи неконтролируемое развитие может иметь гораздо больше отрицательных и вредных последствий, чем положительных и полезных.

С.П. Ковалёв,
А.В. Родин
Знания и их
представление
в компьютерную
эпоху

ных решений действительно играет важную роль в развитии любой области знания: каждая научная теория и каждая инженерная идея проходят проверку временем, конкурируя с альтернативными теориями и идеями; в этом отношении аналогия с дарвиновским «выживанием самых приспособленных» представляется уместной. Однако Сергей использует эту аналогию иначе, говоря о том, что процесс получения новых знаний (то есть познание) должен быть подчинен каким-то высшим целям социального и этического характера. Знания тут играют роль кроликов, а социальные цели — роль хищников, которые питаются этими кроликами. Я допускаю, что социальные интересы и этические принципы могут входить в конфликт с эпистемическими интересами в определенных ситуациях, но, на мой взгляд, отсюда не следует, что вторые должны быть подчинены первым. И я не знаю примеров из истории науки, когда конфликты такого рода были продуктивными. Их, на мой взгляд, лучше по возможности избегать.

Можно привести и прагматический аргумент в пользу автономии науки и против идеи регулировать развитие знаний ради каких бы то ни было внешних целей. Ситуация, когда знания, которые были получены под влиянием исключительно внутренних мотиваций (таких как стремление ответить на открытый математический вопрос), впоследствии применяются на практике, например в технике, и таким образом оказываются полезными, — это скорее правило, чем исключение. С позиции научного реализма объяснить такое положение вещей нетрудно: чтобы как-либо использовать знания в собственных интересах, нужно сначала разобраться с тем, как устроен мир, безотносительно к любым социальным и практическим интересам, включая общечеловеческие. Если бы наши знания о мире были всегда ограничены теми аспектами, которые непосредственно связаны с текущими практическими интересами самого исследователя или его социальной группы (в качестве упражнения рекомендую слушателям попробовать составить список подобных интересов для общества наших предков, живших в эпоху палеолита), то такие знания не могли бы в принципе обеспечить технический и социальный прогресс человечества в том виде, в котором он наблюдался на протяжении последних веков и тысячелетий. Я сознаю, что отказ от прогресса тоже может быть продуманной позицией, но хочу подчеркнуть, как много тут ставится на карту.

Что касается примера математического доказательства длиной в 15 000 страниц, то с ним также связана проблема, о которой Сергей не упомянул. Если это доказательство способны понять лишь несколько человек в мире и никто не способен его аккуратно проверить, то откуда у нас уверенность в том, что это доказательство не содержит ошибок? Почему мы должны считать эту теорему надежно установленным математическим результатом? Я думаю, что

С.П. Ковалёв,
А.В. Родин
Знания и их
представление
в компьютерную
эпоху

можно смоделировать на компьютере в виде простого элемента, имитирующего биологическую функцию раздражимости. На входы такого элемента поступают различные сигналы (скажем, электрические), и как только суммарный уровень поступающих сигналов превышает определенный порог, этот элемент, в свою очередь, выдает сигнал на выходе, то есть «откликается на раздражение». Далее такие элементы соединяются друг с другом таким образом, что выходные сигналы одних элементов подаются на входы других. Оказывается, что достаточно многокомпонентные и подходящим образом организованные системы подобного рода могут имитировать сложные функции мозга, такие как распознавание образов. Допустим, у нас есть картинки. На некоторых из них изображены кошки, а на других — собаки. Если человека соответствующим образом проинструктировать, он может не реагировать на изображения кошек, а при предъявлении изображения собаки, например, поднимать руку. Такое поведение можно имитировать с помощью искусственной нейронной сети, состоящей из нескольких десятков или сотен нейронов, если ее соответствующим образом обучить. На входные нейроны сети нужно подавать цифровое представление изображения, а с выходных нейронов снимать сигнал ответа. Сеть в подавляющем большинстве случаев будет выдавать выходной сигнал при предъявлении изображения собаки и не выдавать такого сигнала при предъявлении изображения кошки. Процесс обучения нейронной сети выглядит следующим образом. Сначала мы разрешаем сети работать произвольно, то есть реагировать на предъявляемые изображения случайным образом. Далее собирается статистика, которая учитывает доли «правильных» и «ошибочных» реакций, и данные этой статистики используются для корректировки удельных весов соответствующих сигналов в связях между нейронами. После этого процесс повторяется. При правильно подобранных алгоритмах корректировки весов доля ошибок при последующих итерациях обучения будет снижаться. При достаточном числе итераций сеть начинает отличать кошек от собак с высокой степенью надежности. Все это кажется похожим на то, что происходит в более обычной ситуации, когда роль обучаемого играет не машина, а человек, которого хвалят за правильную реакцию и ругают за неправильную.

Первые лабораторные эксперименты с нейронными сетями были проведены более 50 лет назад. С помощью нейронных сетей исследователи пытаются, в частности, решать инженерные задачи, такие как распознавание свойств композиционных материалов на основе их структуры. Опытный машиностроитель может предсказать «на слух», сколько прослужит та или иная конструкция, постукав по ней молотком. Конечно, явно сформулировать какие-то формальные основания для своего суждения он не мо-

*С.П. Ковалёв,
А.В. Родин*
Знания и их
представление
в компьютерную
эпоху

сти человека и естественные когнитивные механизмы, а в том, в какой степени и каким именно образом эти когнитивные возможности расширяются за счет использования нейронных сетей и других новых компьютерных технологий.

С.П. Ковалёв. Я с этим согласен и могу привести пример ограниченности основанного на имитации подхода к нейронным сетям. Если в деталях знать алгоритмы, по которым происходит обучение нейронной сети, то натренированную сеть можно обмануть с помощью очень небольших и незаметных для человеческого глаза модификаций предъявляемых изображений, так что эта сеть будет принимать, например, кошку за страуса [2]. Это показывает, что системы, основанные на имитации нейронных связей, работают неустойчиво. Когда изображение анализирует человек, то наряду с биологическим механизмом, который, как можно предположить, функционирует подобным образом, важную роль играет понимание структуры изображения, расположения и связей между частями изображенного объекта. Так что я согласен, что попытки имитировать какую-то определенную биологическую функцию очень быстро себя исчерпывают и что самые интересные технические решения обычно ничего не имитируют, а имеют самостоятельную ценность. Вместе с тем хочу отметить, что в области представления знаний таких решений на сегодняшний день нет, так что они остаются делом будущего.

Ю.В. Синеокая. Спасибо. Теперь переходим к вопросам и репликам слушателей.

Вопрос из зала. Можно ли сказать, что в компьютерных системах представления знаний знания всегда представляются в виде текстовых нарративов?

А.В. Родин. Я думаю, что нет. Тот способ представления знаний, о котором вы говорите, связан скорее с книжной эпохой, с идеей, что книга и текст на естественном языке — это универсальные инструменты для фиксации, сохранения и передачи любого рода знаний. На практике далеко не все виды знаний передавались таким способом: многие практические знания передавались и до сих пор передаются из рук в руки от более опытных людей к менее опытным. Письменные, да и вообще словесно сформулированные инструкции могут играть при этом какую-то роль, но не всегда главную. В качестве примера можно привести знание того, как управлять автомобилем. Если же теперь посмотреть, что происходит со знаниями в электронных коммуникациях, то даже если оставить в стороне специальные вопросы, связанные с использованием искусственных программных языков для представления знаний, можно заметить существенные отличия от стандартного книжного формата. Как я уже говорил, хотя Википедия для пассивного пользователя выглядит как обычная бумажная энциклопедия, на самом деле она устро-

С.П. Ковалёв,
А.В. Родин
Знания и их
представление
в компьютерную
эпоху

Примерно в то же время, когда такой запрос был явно сформулирован, то есть в середине XX века, философия по каким-то причинам сфокусировала усилия на решении проблем совсем другого рода — с некоторыми исключениями, которые подтверждают это правило. Во всяком случае, мои попытки найти в современной философской литературе ответы на интересующие меня философские вопросы не дали никакого результата. На мой взгляд, это гораздо более серьезная проблема, чем возможные ошибки в стандартной теории информации.

Вопрос из зала. Можно провести аналогию между развитием знания и биологической эволюцией. Некоторые знания устаревают, появляются новые знания и так далее. Вопрос состоит в том, кто властен над этим процессом, как им можно управлять и как определить приоритетные направления развития. Или же этот процесс нужно предоставить самому себе?

А.В. Родин. Я согласен с этой биологической аналогией и думаю, что изучение эпистемических проблем в контексте биологии вообще и эволюционной теории в частности может быть очень интересным и продуктивным. Если опираться на эту аналогию, то можно сказать, что знания, как и жизнь на нашей планете, регулируют сами себя и не нуждаются ни в каком внешнем контроле.

С.П. Ковалёв. Мой тезис о важности экологии знаний также опирается на эту биологическую аналогию.

А.В. Родин. Давайте попробуем немного развить эту аналогию, чтобы применить ее к вопросу о контроле и регулировании научных исследований и вообще к процессу получения новых знаний. Поскольку мы, люди, являемся биологическим видом наряду с другими биологическими видами, населяющими нашу планету, мы, разумеется, влияем на ход биологической эволюции уже постольку, поскольку в ней участвуем: боремся за выживание и процветание нашего вида, как это делают и представители любых других биологических видов, включая тех из них, которые в эволюционном отношении находятся очень далеко от человека. Наши знания о мире и основанные на этих знаниях технологии, включая, в частности, генную инженерию, ставят нас в особое положение и позволяют влиять на развитие жизни на планете более глубоко и более эффективно, чем это могли делать наши предки. Однако те же самые научные знания о нашей планете, ее биосфере и наших собственных обществах дают нам и серьезные поводы опасаться, что непродуманное использование этой технологической мощи может легко привести наш вид, а может быть, и всю жизнь на планете, к катастрофе. Изучение такого рода обстоятельств и контекстов и является предметом экологии. Понятия регулирования и контроля играют в практических аспектах экологии важную роль, но нужно отдавать себе отчет, что в таких

С.П. Ковалёв,
А.В. Родин
Знания и их
представление
в компьютерную
эпоху

ни на планете и человеческую историю под чей-то безусловный контроль, а научиться ставить практические цели в соответствии с правильно сформулированными общими человеческими интересами и с человеческими возможностями и достигать этих целей, используя все доступные технологические ресурсы. Условием для такого разумного и эффективного отношения к природе и человеку является понимание того, как устроен окружающий нас природный мир без учета специальных человеческих целей. Именно в этом состоит, на мой взгляд, фундаментальное значение естественных наук для человечества.

Ю.В. Синеокая. Благодарю наших блестящих докладчиков и до следующего четверга!

Литература

1. *Rodin A.* Axiomatic Method and Category Theory (Synthese Library, vol. 364). Cham: Springer, 2014.
2. *Szegedy Ch., Zaremba W., Sutskever I. et al.* Intriguing properties of neural networks. URL: [/arxiv.org/abs/1312.6199](https://arxiv.org/abs/1312.6199) (last revised: 19.02.2014).
3. *Weisstein E.W.* Classification Theorem of Finite Groups. URL: <http://math-world.wolfram.com/ClassificationTheoremofFiniteGroups.html>

Knowledge and its Representation in the Computer Era

Sergei P. Kovalyov

DSc in Physics and Mathematics, Leading Research Fellow.
RAS Institute of Control Sciences.
65 Profsoyuznaya Str., Moscow 117997, Russian Federation.
E-mail: kovalyov@nm.ru

Andrei V. Rodin

PhD in Philosophy, Senior Research fellow.
RAS Institute of Philosophy.
12/1 Goncharnaya Str., Moscow 109240, Russian Federation.
E-mail: andrei@philomatica.org

Abstract. Knowledge Representation (KR) is a research area at the borderline between Computer Science and Philosophical Epistemology. Technically KR can be described as a division of Artificial Intelligence focused on prospective information and communication technologies aiming at the efficient use of large data sets. On the other hand, knowledge and its representation is a central theme of philosophical reflection and analysis; this remark concerns not only the traditional philosophy but also the Analytic philosophy of the 20th century, which essentially uses formal logical methods. The participants discussed new epistemic practices made possible by new computer-based technologies of knowledge representation.

*С.П. Ковалёв,
А.В. Родин*
Знания и их
представление
в компьютерную
эпоху

Keywords: knowledge representation, artificial intelligence, neural networks, ontologies, expert knowledge.

For citation: Kovalyov S.P., Rodin A.V. Knowledge and its Representation in the Computer Era // *Chelovek*. 2019. Vol. 30, N 4. P. 94–112.

DOI: 10.31857/S023620070005953-5

References

1. Rodin A. *Axiomatic Method and Category Theory* (Synthese Library, vol. 364). Cham: Springer, 2014.
2. Szegedy Ch., Zaremba W., Sutskever I. et al. *Intriguing properties of neural networks*. URL: [/arxiv.org/abs/1312.6199](http://arxiv.org/abs/1312.6199) (last revised: 19.02.2014).
3. Weisstein E.W. *Classification Theorem of Finite Groups*. URL: <http://math-world.wolfram.com/ClassificationTheoremofFiniteGroups.html>