doi:https://doi.org/10.15407/emodel.41.05.035

УДК 004.932

Г.А. Кравцов[1], канд. техн. наук, **В.В. Левитин**[2],
В.И. Кошель[1], аспирант, **В.В. Никитченко**[1], аспирант, **А.Н. Примушко**[3]
[1] Институт проблем моделирования
в энергетике им. Г.Е. Пухова НАН Украины
(Украина, 03164, Киев, ул. Генерала Наумова, 15,
e-mail: hryhoriy.kravtsov@gmail.com; vlad.koshell@gmail.com;
V.nikitchenko.ua@gmail.com),
[2] Компания NaomiAI
(USA, 11238, New York, Brooklyn, 58 Prospect Place,
e-mail: levitin.vladimir@gmail.com),
[3] Национальный технический университет Украины
«Киевский политехнический институт им. Игоря Сикорского»
(Украина, 03056, Киев, пр-т Победы, 37,
e-mail: arsentiy.prymushko@gmail.com)

Сильный искусственный интеллект: предпосылки

Представлен обзор фундаментальных основ, позволяющих построить сильный искусственный интеллект (ИИ). Показана справедливость выдвинутых гипотез, доказанных методом натурных экспериментов. Утверждается, что для построения математической теории сильного ИИ следует перейти к системе аксиом Неймана— Бернайса—Геделя, и тогда появляется возможность использовать семантические структуры, представляемые компьютерными онтологиями, как алгебраические структуры. Для корректного использования онтологий в системах ИИ необходимо, чтобы онтологии по плоскостям деления представляли собой метрические пространства.

Ключевые слова: искусственный интеллект, система аксиом, аксиоматака, мозг человека, семантика, мера отличия, внутренняя модель мира, адаптируемость.

Базовая терминология. Одно из наиболее используемых определений термина «искусственный интеллект» дано в работе [1]. Н. Бостром выстраивает логическую цепочку от субъективного толкования термина «интеллект» до толкования термина «искусственный суперинтеллект», в которой интеллект представляет собой реализуемый на физическом уровне алгоритм. Он формулирует толкование термина «интеллект человеческого уровня» как интеллект, который способен решать задачи, доступные человечеству, обладает умом, интуицией, пониманием, способен к познанию, мышлению, воображению. Понятие интеллект человеческого уровня необходимо Н. Бострому, чтобы определить уровень интеллектуальных способностей человека, позволяющий установить минимум требований для определения сверхинтеллекта. В понимании Н. Бострома искусственный интеллект (ИИ) — это интеллект, созданный не в естественной биологической среде, который потенциально может превышать интеллект человеческого уровня, становясь суперинтеллектом или сверхинтеллектом (superintelligence).

Н. Бостром выделяет три формы суперинтеллекта: быстрый ИИ, коллективный ИИ и качественный ИИ.

Быстрый ИИ (Speed superintelligence) — система, которая может сделать все, что может сделать человеческий интеллект, но значительно быстрее.

Коллективный суперинтеллект (Collective superintelligence) — система, состоящая из большого числа меньших интеллектов, при этом общая производительность системы во многих очень общих областях значительно превосходит эффективность любой современной когнитивной системы.

Качественный суперинтеллект (Quality superintelligence) — система, по крайней мере, такая же быстрая, как человеческий разум, но значительно умнее.

Формулировки Н. Бострома, естественно, поддаются критике, однако, ему удалось дать наипростейшее толкование принципиального отличия ИИ от человеческого интеллекта.

Согласно теории сильного ИИ предполагается, что компьютеры, являющиеся небиологической средой, могут получить способность мыслить и осознавать себя как отдельную личность (в частности, понимать собственные мысли), хотя их мыслительный процесс не обязательно будет подобен человеческому. В то же время, теория слабого ИИ отвергает такую возможность. В 1980 г. Джон Сирл [2] впервые охарактеризовал сильный ИИ так: «Соответствующим образом запрограммированный компьютер с нужными входами и выходами и будет разумом в том смысле, в котором человеческий разум — это разум». Как видим, данное Н. Бостромом [1] определение интеллекта, в котором интеллект представляет собой реализуемый на физическом уровне алгоритм, созвучно позиции Д.Сирла, опубликованной еще в 1980 г.

Аксиоматика Д. Сирла и «китайская комната». Как известно, Д. Сирл подверг свою концепцию сильного ИИ жесткой критике, сформулировал систему аксиом, которая получила впоследствии название «аксиоматика Сирла».

Аксиома 1 Сирла утверждает, что любая компьютерная программа сводиться к манипулированию формальными или синтаксическими объектами. Иными словами, компьютерная программа не способна оперировать семантическими объектами. Данная аксиома очень важна, так как она постулирует неспособность ИИ выполнять основанные на семантике вычисления (meaningful calculations). Эта идея получает развитие во второй аксиоме.

Аксиома 2 указывает на основное противоречие (по мнению Д. Сирла) между интеллектом человека и ИИ, утверждая, что человеческий разум оперирует семантическими или смысловыми содержаниями в то время, как компьютерная программа не обладает смысловым содержанием (семантикой).

В аксиоме 3 Джон Сирл постулирует, что синтаксис сам по себе не составляет семантику и его недостаточно для существования семантики. Разделяя утверждение Д. Сирла относительно синтаксиса, далее покажем, как сопоставимы синтаксис и семантика. В тоже время, заметим, что считаем ложным следствие из аксиомы 3. В первом следствии из аксиомы 3 утверждается, что программы не являются частью разума и их наличия недостаточно для наличия разума, а следовательно, утверждение о возможности существования сильного ИИ ложно.

В аксиоме 4 утверждается, что мозг порождает разум. Эта аксиома имеет несколько следствий.

В первом следствии из аксиомы 4 речь идет о том, что любая другая система, способная порождать разум, должна обладать каузальными (взаимообусловливающими) свойствами, эквивалентными свойствам мозга.

Во втором следствии утверждается, что любой артефакт, порождающий ментальные явления, любой искусственный мозг должен быть способным воспроизводить специфические каузальные свойства мозга, и добиться наличия этих свойств невозможно, выполненяя только формальную программу.

В итоге Д. Сирл констатирует, что способ, посредством которого человеческий мозг порождает ментальные явления, не может сводиться лишь к выполнению компьютерной программы. Ментальные явления есть фиксируемое проявление ментальных моделей [3] . Ментальные модели (ММ) являются, по сути, идеями (стратегиями, пониманиями), основанными на предыдущем опыте и существуют в уме человека и направляют его действия.

Ментальные модели формируются и поддерживаются с помощью четырех инструментов: вычеркивание, конструирование, искажение и обобщение.

Вычеркивание — способность отбирать определенную информацию из всей доступной на основании интересов, настроения, озабоченности и др.

Конструирование — способ видеть то, чего нет, способность в ситуации неясности находить правдоподобные объяснения, принимая их впоследствии за реальность.

Искажение — механизм изменения пережитого в реальности с преуменьшением одних деталей и преувеличением других.

Обобщение — способ построения ментальных моделей на основании единичных случаев из опыта, который преподносится или воспринимается как типичное явление.

Для того чтобы избежать казуистических трактовок, Д. Сирл поясняет свою позицию так: «Во-первых, я не пытался доказывать, что компьютер не может мыслить. Поскольку все, что поддается моделированию вычислениями, может быть описано как компьютер, и поскольку наш мозг на некоторых уровнях поддается моделированию, то отсюда тривиально следует, что наш мозг — это компьютер, и он, разумеется, способен мыслить. Однако из того факта, что систему можно моделировать посредством манипулирования символами и что она способна мыслить, вовсе не следует, что способность к мышлению эквивалентна способности к манипулированию формальными символами.

Во-вторых, я не пытался доказывать, что только системы биологической природы, подобные нашему мозгу, способны мыслить. В настоящее время это единственные известные нам системы, обладающие такой способностью, однако мы можем встретить во Вселенной и другие способные к осознанным мыслям системы, а может быть, мы даже сумеем искусственно создать мыслящие системы. Я считаю этот вопрос открытым для споров.

В-третьих, утверждение сильного ИИ заключается не в том, что компьютеры с правильными программами могут мыслить, что они могут обладать какими-то неведомыми доселе психологическими свойствами; скорее, оно состоит в том, что компьютеры просто должны мыслить, поскольку их работа — это и есть не что иное, как мышление.

В-четвертых, я попытался опровергнуть сильный ИИ, определенный именно таким образом. Я пытался доказать, что мышление не сводится к программам, потому что программа лишь манипулирует формальными символами. А, как нам известно, самого по себе манипулирования символами недостаточно, чтобы гарантировать наличие смысла. Это тот принцип, на котором основано рассуждение о китайской комнате.»

Прежде чем резюмировать сказанное Д. Сирлом, дадим краткое представление о мысленном эксперименте под названием «китайская комната».

Д. Сирл [4] поместил в комнату человека H, ничего не понимающего в китайских символах, но знающего английский язык. В его распоряжении находятся три корзины с «данными»: в первой текст (рукопись) на китайском языке; в следующей — китайские символы и правила на английском, позволяющие сопоставить первую корзину со второй (рассказ); в последней — еще один набор правил на английском для сопоставления первых двух корзин с третьей (вопросы). Собеседник *B*, находящийся вне комнаты и свободно разговаривающий по китайски, посылает вопросы, на которые получает соответствующие ответы из комнаты. Оперируя символами трех корзин, *H* сможет выдать логически верные ответы на вопросы *B* относительно одного из рассказов. Таким образом, *H* пройдет тест на знание языка, легко убедив в этом своего собеседника, хотя и не понимает ни одного знака в наборе китайских иероглифов. Представим теперь цифровой компьютер *C*, который на основе своих программ моделирует работу. Точно так же он сможет пройти тест, оперируя системой символов, но не понимая их значения.

Аргумент Д. Сирла направлен против известного теста Тьюринга и исключает любую возможность создания сильного ИИ. Напомним, что тест Тьюринга основан на утверждении, что поведение объекта, обладающего ИИ, в конечном итоге нельзя будет отличить от поведения таких бесспорно интеллектуальных сущностей, как человеческие существа. В 2014 г. было объявлено, что тест Тюринга успешно пройден суперкомпьютером по имени Eugene, который искусственно воссоздал человеческий интеллект. Программа ИИ была разработана В. Веселовым и Е. Демченко [5].

Возвращаясь к комментарию Д. Сирла к его системе аксиом, следует заметить, что автор аксиоматики смягчает свою позицию в отношении сильного ИИ, сводя утверждение о невозможности сильного ИИ к тому, что должен и не должен делать сильный ИИ, поднимая на новый уровень вопрос этики применения технологий ИИ. Это действительно очень важный вопрос, учитывая тот факт, что сильный ИИ будет обладать способностями, которые человеку не под силу, а значит, теоретически, сильный ИИ может признать человека слабым звеном и исключить его из некоторого или даже всех процессов с целью достижения наибольшей эффективности. Вопрос этики является важнейшим с точки зрения Н. Бострома [1].

Критика аксиоматики Д. Сирла. Рассмотрим аксиому 1, которая утверждает, что любая компьютерная программа сводиться к манипулированию формальными или синтаксическими объектами, в результате чего компьютерная программа не способна оперировать семантическими объектами. Чтобы понять суть дальнейших рассуждений, необходимо обратиться к некоторым историческим событиям.

В 30-х годах 19 века перед математиками встала так называемая проблема разрешения (Entscheidungsproblem), сформулированная Д. Гильбертом. Суть ее в том, что есть некий формальный язык, на котором можно написать какое-либо утверждение. Существует ли алгоритм, который за конечное число шагов определит его истинность или ложность? Ответ был найден двумя великими учеными того времени А. Черчем и А. Тьюрингом. Они показали (первый — с помощью изобретенного им λ-исчисления, а второй — с помощью машины Тьюринга), что для арифметики такого алгоритма не существует в принципе, т.е. Entscheidungsproblem в общем случае неразрешима [6].

Так λ-исчисление впервые громко заявило о себе, но еще пару десятков лет продолжало быть достоянием математической логики, пока в середине 60-х годов того же века П. Ландин [7] не заметил, что сложный язык программирования проще изучать, сформулировав его ядро в виде небольшого базового исчисления, выражающего самые существенные механизмы языка и дополненного набором удобных производных форм, поведение которых можно выразить посредством перевода на язык базового исчисления. В качестве такой основы П. Ландин использовал λ-исчисление А. Черча [8].

Лямбда-исчисление, как формальную систему, определяющую вычислимость, разделяют на бестиповое и типизированное λ-исчисление. Само же исчисление строится на λ-термах — формальных выражениях, построенных исключительно из переменных, с применением аппликации и абстракции, не предполагающей наличия каких-либо констант. Типизированное λ-исчисление — это версия λ-исчисления, в которой λ-термам приписываются специальные синтаксические метки, называемые типами. Допустимы различные наборы правил конструирования и приписывания таких меток, они порождают различные системы типизации.

Типовые λ-исчисления являются фундаментальными примитивными языками программирования, которые обеспечивают основу типовым языкам функционального программирования — аппликативным языкам, а также типовым императивным языкам программирования. Типовое исчисление является языком декартово-замкнутой категории [9], что устанавливает прямую связь с такой моделью вычислений, как категориальная абстрактная машина. В настоящее время общеизвестно, что теория категорий, основой которой является функциональная композиция, рассматривается как формальная основа функционального программирования — подхода, помогающего при построении сложных информационных систем [10].

Особенностью как типизированного λ-исчисления, так и теории категорий является наличие семантики объектов, которыми оперируют последние. Следовательно, программы, построенные в соответствии с лучшими практиками и поддающиеся формальному описанию с помощью типизированного λ-исчисления и теории категорий, по определению используют семантические структуры, что противоречит аксиоме 1 Сирла, утверждающего, что компьютерная программа не способна оперировать семантическими объектами. Однако, теория категорий и λ-исчисление, теории о чистых вычислениях, получившие свое развитие в функциональных языках программирования, таких как Haskel, Erlang и им подобных, их мощи недостаточно, чтобы описать всю совокупность процессов, протекающих в человеческом мозге. Чтобы пояснить это утверждение о недостаточности, потребуется понимание сути некоторых процессов, протекающих в человеческом мозге.

Человеческий мозг, осознанное и неосознанное поведение. Человеческий мозг — один из самых сложных объектов исследования. Во множестве публикаций предприняты попытки пролить свет на процессы, происходящие в нашем мозге. Среди доступной литературы следует указать работу Д. Игелмана [11], которая насыщена убедительными фактами и аргументированными выводами. Изложение видения механизмов функционирования человеческого мозга будет неполным, если не обратим внимание на труд Нобелевского лауреата по экономике Д. Канемана [12].

Начнем с утверждения, что человеческий мозг обрабатывает информацию в двух режимах: осознанном и неосознанном. Установлено [11], что действия человека, доведенные до автоматизма, выполняются в неосознанном режиме.

Был проведен эксперимент с маленькими детьми на незнакомой горке для катания. Поднимаясь по лестнице на незнакомой горке, дети уделяли пристальное внимание каждому своему движению и внимание их было абсолютно сконцентрировано на подъеме по лестнице. Однако после небольшого числа повторений подъемов дети позволяли себе несколько отвлечься во время подъема по лестнице. Впервые прийдя на площадку, где были две одинаковые горки, отличающиеся только раскраской, дети осваивали первую горку за несколько итераций.

В результате наблюдения за концентрацией внимания детей на этапе подъема по лестнице и замеров времени с момента посадки для спуска и моментом начала самого спуска была установлена следующая корреляция: с каждой итерацией число моментов отвлечения на родителей увеличивалось, пока не достигало некоторого условного максимума, а время задержки перед спуском уменьшалось, достигая некоторого условного минимума. Это объясняется тем, что с каждым спуском дети видели в аттракционе все меньше опасности.

Когда были установлены относительно стабильные показатели отвлечений и времени задержки, родителей просили перевести детей на вторую горку. Хотя горки отличались только цветовой гаммой, первые итерации дети проходили с заметным увеличением паузы перед спуском. Однако эта ситуация очень быстро сводилась к показателям последних итераций на первой горке.

О чем свидетельствует эксперимент с горками? О том, что мозг прежде всего определяет, можно ли решить задачу неосознанно. Если обнаруживается «подходящий» прошлый опыт, мозг его использует, как предопределенную программу действий. Если задача не разрешается неосознанно, то решение приходится искать в осознанном режиме. Следует заметить, что работа мозга в неосознанном режиме требует меньше энергии, чем функционирование в осознанном режиме.

Прежде чем перейти к строгой формулировке результатов опыта, рассмотрим основы функционирования мозга человека с точки зрения биологии, стремящейся дать ответ на вопрос, как мозг выбирает программы для неосознанного исполнения. Н. Курчанов [13] пишет, что даже простой выбор из нескольких инстинктивных программ требует «инстанции», где сравниваются значения сигналов внешней и внутренней среды организма, а точность правильного выбора, определяющего поведение, имеет большое значение для благополучия вида. Он считает, что важнейшим принципом функциональной организации головного мозга является принцип иерархичности. При этом с каждым новым этапом эволюции возникают новые функциональные центры, подчиняющие себе старые, которые, однако, сами не исчезают. Внедрение все большего числа интернейронов между афферентными путями и эффекторами приводит к увеличению числа ответов на данный сенсорный стимул, увеличивая разнообразие поведения. Вершину иерархии всегда занимают ассоциативные центры.

Ассоциативные зоны [13] взаимодействуют с первичными и вторичными зонами коры мозга, куда, в свою очередь, поступает информация от рецепторов через подкорковые центры. Следовательно, ассоциативные зоны через многочисленные «переключатели» связаны почти со всеми отделами мозга, что позволяет учитывать все факторы, участвующие в программировании поведения индивида. Принято выделять три главных мозговых «регулятора» в поведении человека: теменная зона коры, нижневисочная ассоциативная зона и лобная ассоциативная зона.

Согласно Н. Курчанову, ассоциативная теменная зона коры отвечает за обработку двигательной информации и окончательное формирование пространственного зрения. Именно в ассоциативной теменной зоне сенсорная информация проприорецепторов интегрируется с информацией вестибулярной и зрительной систем, на основе чего инициируется специфическое движение, однако многие моменты этой инициации пока непонятны [13].

Нижневисочная ассоциативная зона отвечает за предметное зрение и узнавание объектов, их распределение по категориям, играя важную роль в реализации зрительной памяти. Н. Курчанов полагает, что именно в ней при определенных изменениях формируются зрительные галлюцинации. Лобные доли занимают особое место среди ассоциативных зон, поскольку им приписывают программирование и контроль наиболее сложных форм поведения человека, когда задействовано более 20% коры больших полушарий. Считается, что именно развитие лобных ассоциативных зон сыграло решающую роль в антропогенезе и становлении цивилизации.

Согласно Н. Курчанову [13] информация, идущая от рецепторов, могла бы быстро исчерпать информационные резервы мозга. Поэтому в процессе эволюции возникли специальные механизмы ограничения избыточности информации. Считается, что новая информация, как правило, более важна для человека, чем привычная, она в первую очередь и поступает в мозг, хотя в работе не дано определения «привычной» информации. Н. Курчанов пишет: «Диагностическая значимость определяется в ходе анализа ассоциативной системой мозга поступающей информации и формирования доминирующей мотивации. Однако нейрофизиологические механизмы «верхних этажей» обработки сенсорной информации во многом остаются таинственными».

Таким образом, с точки зрения биологов человеческий мозг обладает способностью устанавливать «степень подобия» между входными данными пережитого опыта и переживаемого и, в зависимости от степени подобия, «включать» программу из прошлого опыта. В данном случае под включением программы понимаем не только повторение воспроизведения некоторой программы прошлого опыта, но и отказ от ее воспроизведения в случае, если прошлый опыт имел негативные последствия для индивида.

Следовательно, модель принятия решения строится вокруг триады «возмущение — реакция — оценка реакции». И тут возникает еще один принципиальный вопрос: существует ли сознание без языка? [14]. Но прежде чем перейти к рассмотрению вопроса о языке, заметим, что для триады возмущение — реакция — оценка реакции важным является установление некоторой меры схожести между возмущениями, между реакциями и между оценками реакций. Математически предполагаемая мера есть вероятность идентичности (абсолютного сходства), где 100%-ное совпадение (или единица) свидетельствует об «однозначном равенстве» сравниваемых возмущений, реакций или оценок реакции. Понимая, что ноль — это бесконечно малая величина, которой в определенном контексте можно пренебречь, будем полагать, что если мера схожести стремится к нулю, то это лишь значит, что невозможно установить хоть какие-то свойства сравниваемых категорий (возмущение, реакция, оценка реакции). Таким образом, приходим к необходимости определения меры отличия между категориями в частности или, обобщая, меры отличия между концептами.

Мера отличия между концептами. В работе [15] предложена мера отличия на плоских классификациях, где классы являются концептами, с обобщением до меры отличия на пространственных классификациях, которые, по сути, являются онтологиями.

Пусть дана некоторая условная плоская классификация (см. рисунок). Семантически класс в классификации сам является классификацией. Например, как показано на рисунке, класс — это классификация из трех классов: $A^i_{[1,1]}$, $A^i_{[1,2]}$ и $A^i_{[1,3]}$. Очевидно, что $A^i_{[2,1]} \neq A^i_{[1,2]}$ может быть обобщено:

$$A^i_I \neq A^i_J,\ I \neq J. \tag{1}$$

Поскольку A^i представляет собой одну из плоскостей деления классификации A с единым корнем ориентированного дерева, семантически A^i и A представляют собой одно и то же.

Пусть даны классы $A^i_{[a]}, A^i_{[b]}, A^i_{[a,\,b]}$ классификации A. Ассоциативную бинарную операцию обобщения классов классификации определим так:

$$A^i_{[a]} \cdot A^i_{[b]} = A,$$

$$A^i_{[a]} \cdot A^i_{[a,\,b]} = A^i_{[a]},$$

$$A^i_{[b]} \cdot A^i_{[a,\,b]} = A,$$

$$A^i_{[a]} \cdot A = A.$$

Здесь A является идемпотентом или нулем операции обобщения классов относительно самой себя, $A \cdot A = A$, и всех уточняющих классов классификации $A^i_I \cdot A = A$.

Пусть даны классы $A^i_{[a]}, A^i_{[b]}, A^i_{[a,\,b]}$ классификации A. Определим ассоциативную бинарную операцию уточнения класса:

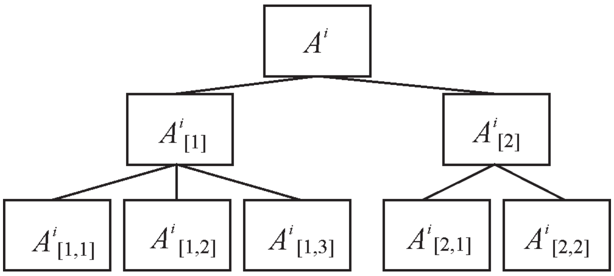
$$A^i_{[a]} + A = A^i_{[a]}, \tag{2}$$

$$A^i_{[a]} + A^i_{[b]} = A^i_{[a]} + A^i_{[b]}, \tag{3}$$

$$A^i_{[a]} + A^i_{[a,\,b]} = A^i_{[a,b]}, \tag{4}$$

$$A^i_{[b]} + A^i_{[a,\,b]} = A^i_{[b]} + A^i_{[a,\,b]}. \tag{5}$$

Семантически операция уточнения класса эквивалентна дуге дерева в случаях (2) и (4). Уравнения (3) и (5) отражают семантику уточнения одного класса другим при условии, что оба класса находятся на разных ветках дерева.



Классификация A^i

Операции обобщения и уточнения класса необходимы при построении и доказательстве существования метрики на классификациях. Наличие бинарной ассоциативной операции и идемпотенты (1) позволяет сделать вывод о том, что классификация представляет собой конечную полугруппу классов.

Относительным расстоянием R между двумя классами, $A^i_{[a]}$ и $A^i_{[b]}$, классификации A назовем неотрицательное целое число, равное числу уникальных операций уточнения от ближайшего общего обобщающего класса. Получим

$$R\,(A^i_{[a]},\, A^i_{[b]}) = 2\,, \tag{6}$$

так как потребуется выполнить две операции уточнения, $A + A^i_{[a]} = A^i_{[a]}$ и $A + A^i_{[b]} = A^i_{[b]}$, для каждого из классов ближайшего обобщающего класса A. Согласно данному определению относительного расстояния

$$R\,(A^i_I\,,\ A^i_Y\,) = R\,(A^i_Y\,,\ A^i_I\,),\quad R\,(A^i_I\,,\ A^i_I\,) = 0, \tag{7}$$

где I, Y — произвольные пути уточнения (деления) классификации A в плоскости деления i. Такое расстояние $R\,(A^i_I\,,\ A^i_Y\,)$, где I, Y — произвольные пути уточнения (деления) классификации A, названо относительным, так как оно измеряется относительно выбранных классов и никак не учитывает положение этих классов в классификации.

Рассмотрим пример. Найдем относительное расстояние $R\,(A^i_{[a,b,c,d]},$ $A^i_{[a,b,c,k]})$, воспользовавшись заменой переменной $B^i = A^i_{[a,b,c]}$, и представим $R\,(A^i_{[a,b,c,d]},\, A^i_{[a,b,c,k]})$ в виде $R\,(B^i_{[d]},\, B^i_{[k]})$. По аналогии с (6) получим $R\,(B^i_{[d]},\, B^i_{[k]}) = 2$, или $R\,(A^i_{[a,b,c,d]},\ A^i_{[a,b,c,k]}) = 2$.

Следует заметить, что относительное расстояние $R\,(A^i_I\,,\, A^i_Y\,)$, где I, Y — произвольные пути уточнения (деления) классификации A, всегда единственно и конечно, так как классификация A есть ориентированное дерево. Относительное расстояние $R(A_I^i, A_Y^i)$, где I, Y — произвольные пути уточнения (деления) классификации A, инвариантно, если класс A, который уточняют значения A_I^i и A_Y^i, становится уточняющей классификацией (классом) классификации B. Поскольку структура деления класса A неизменна, то и путь между двумя классами, A_I^i и A_Y^i, в A неизменен.

Относительное расстояние $R(A_I^i, A)$, где I — произвольный путь уточнения (деления) классификации A, есть ранг. Для относительного расстояния между двумя классами, A_I^i и A_Y^i, в A верно равенство

$$R(A_I^i, A_Y^i) = R(A, A_I^i) + R(A, A_Y^i) - 2R(A, A_I^i \cdot A_Y^i), \tag{8}$$

где I, Y — произвольные пути уточнения (деления) классификации A. Так, для $A_{[a]}^i$ и $A_{[a,b]}^i$ соответственно $R(A_{[a]}^i, A) = 1$ и $R(A_{[a,b]}^i, A) = 2$. Выражение (8) характеризует связь относительного расстояния между двумя классами, A_I^i и A_Y^i, и абсолютными расстояниями до этих классов от вершины классификации A и относительного расстояния от вершины классификации до общего обобщающего класса $A_I^i \cdot A_Y^i$. Если $R(A, A_I^i \cdot A_Y^i) = 0$, то (8) принимает вид

$$R(A_I^i, A_Y^i) = R(A, A_I^i) + R(A, A_Y^i). \tag{9}$$

Если $R(A, A_I^i \cdot A_Y^i) > 0$, то для (8) будет верно неравенство

$$R(A_I^i, A_Y^i) < R(A, A_I^i) + R(A, A_Y^i). \tag{10}$$

На основании (7) равенство (9) и неравенство (10) могут быть обобщены в виде

$$R(A_I^i, A_Y^i) \le R(A_I^i, A) + R(A, A_Y^i), \tag{11}$$

что соответствует условию неравенства треугольника. Равенства (7), которые определяют свойства симметричности и рефлексивности, и неравенство треугольника (11) позволяют сделать вывод о том, что относительное расстояние $R(A_I^i, A_Y^i)$, где I, Y — произвольные пути уточнения (деления) классификации A, есть мера на классификации A. Однако относительное расстояние как мера оказывается малополезным, так как оно не помогает ответить на вопрос о том, насколько похожи или отличаются два класса классификации.

Мера сходства двух классов одной плоскости деления классификации может быть определена как мера Жаккара, если положить, что для множественной меры Жаккара

$$K_{-1,1} = \frac{n(A \bigcap B)}{n(A \bigcup B)} \tag{12}$$

определено следующее соответствие:

$$O(A_I^i, A_Y^i) = \frac{R(A, A_I^i \cdot A_Y^i) + 1}{R(A, A_I^i + A_Y^i) + 1} =$$

$$= \frac{R(A, A_I^i \cdot A_Y^i) + 1}{R(A, A_I^i \cdot A_Y^i) + R(A_I^i, A_I^i \cdot A_Y^i) + R(A_Y^i, A_I^i \cdot A_Y^i) + 1}, \tag{13}$$

где A_I^i и A_Y^i — классы в A; I, Y — произвольные пути уточнения (деления) плоскости деления классификации A^i. Добавление единицы объясняется приведением к множеству вершин графа, т.е. дерева, где для одной ветки число вершин на единицу больше числа связей (согласно свойствам дерева), для того чтобы мера (13) семантически была эквивалентна мере (12).

В выражении (13) преобразование знаменателя корректно на основании определения относительного расстояния. Для удобства меру $O(A_I^i, A_Y^i)$ обозначим $O(A^i)$ и назовем ее мерой сходства на плоскости деления классификации A^i. Покажем, что мера (13), где $I = Y$ — произвольные пути уточнения (деления) классификации A, всегда равна единице. Согласно (7) $R(A_I^i, A_I^i \cdot A_Y^i) = 0$ и $R(A_Y^i, A_I^i \cdot A_Y^i) = 0$ при $A_I^i = A_Y^i$, а следовательно,

$$O(A_I^i, A_Y^i) = \frac{R(A, A_I^i \cdot A_Y^i) + 1}{R(A, A_I^i + A_Y^i) + 1} = 1,$$

что и требовалось доказать. Строго говоря, мера $O(A_I^i, A_Y^i)$ не является математической мерой в силу невыполнения условия неравенства треугольника. Однако замечено, что выражение $\overline{O}(A_I^i, A_Y^i) = 1 - O(A_I^i, A_Y^i)$ обладает свойствами симметричности и рефлексивности:

$$\overline{O}(A_I^i, A_Y^i) = \overline{O}(A_Y^i, A_I^i), \ \overline{O}(A_I^i, A_I^i) = 0. \tag{14}$$

Пусть даны три меры: $\overline{O}(A_I^i, A_Y^i)$, $\overline{O}(A_I^i, A)$ и $\overline{O}(A_Y^i, A)$. Воспользовавшись (13), получим

$$\overline{O}(A_I^i, A_Y^i) = 1 - \frac{R(A, A_I^i \cdot A_Y^i) + 1}{R(A, A_I^i \cdot A_Y^i) + R(A_I^i, A_I^i \cdot A_Y^i) + R(A_Y^i, A_I^i \cdot A_Y^i) + 1}, \tag{15}$$

$$\overline{O}\,(A_I^i, A) = 1 - \frac{1}{R\,(A_I^i, A) + 1}, \tag{16}$$

$$\overline{O}\,(A_Y^i, A) = 1 - \frac{1}{R\,(A_Y^i, A) + 1}. \tag{17}$$

Относительное расстояние $R\,(A_I^i, A) = R\,(A, A_I^i \cdot A_Y^i) + R\,(A_I^i, A_I^i \cdot A_Y^i)$ по определению аналогично $R\,(A_Y^i, A) = R\,(A,\ A_I^i \cdot A_Y^i) + R\,(A_Y^i,\ A_I^i \cdot A_Y^i)$. Тогда (16) и (17) могут быть представлены в виде

$$\overline{O}\,(A_I^i, A) = \frac{R\,(A, A_I^i \cdot A_Y^i) + R\,(A_I^i, A_I^i \cdot A_Y^i)}{R\,(A, A_I^i \cdot A_Y^i) + R\,(A_I^i, A_I^i \cdot A_Y^i) + 1},$$

$$\overline{O}(A_Y^i, A) = \frac{R\,(A, A_I^i \cdot A_Y^i) + R\,(A_Y^i, A_I^i \cdot A_Y^i)}{R\,(A, A_I^i \cdot A_Y^i) + R\,(A_Y^i, A_I^i \cdot A_Y^i) + 1},$$

а (15) — в виде

$$\overline{O}\,(A_I^i, A_Y^i) = \frac{R\,(A_I^i, A_I^i \cdot A_Y^i) + R\,(A_Y^i, A_I^i \cdot A_Y^i)}{R\,(A, A_I^i \cdot A_Y^i) + R\,(A_I^i, A_I^i \cdot A_Y^i) + R\,(A_Y^i, A_I^i \cdot A_Y^i) + 1}.$$

Выполним замену переменных: $a = R\,(A, A_I^i \cdot A_Y^i)$, $b = R\,(A_I^i, A_I^i \cdot A_Y^i)$, $c = R\,(A_Y^i, A_I^i \cdot A_Y^i)$. Выразим $\overline{O}\,(A_I^i, A) + \overline{O}\,(A_Y^i, A) - \overline{O}\,(A_I^i, A_Y^i)$ через переменные a, b и c:

$$\overline{O}\,(A_I^i, A) + \overline{O}\,(A_Y^i, A) - \overline{O}\,(A_I^i, A_Y^i) = \frac{a+b}{a+b+1} + \frac{a+c}{a+c+1} - \frac{b+c}{a+b+c+1}. \tag{18}$$

Выполним замену переменной $\overline{O}\,(A_I^i, A) + \overline{O}\,(A_Y^i, A) - \overline{O}\,(A_I^i, A_Y^i) = f\,(a, b, c)$ и приведем правую часть (18) к общему знаменателю $W = (a+b+1)(a+c+1)(a+b+c+1)$. Тогда $f\,(a, b, c) = V / W$, где $V = 2a^2 + 3a^2 b + 3a^2 c + 4a^2 + ab^2 + 4abc + 3ab + ac^2 + 3ac + 2a + b^2 c + bc^2$.

Учитывая, что $a \geq 0$, $b \geq 0$ и $c \geq 0$ (по определению относительного расстояния), запишем $W \geq 1$ и $V \geq 0$, откуда следует

$$\overline{O}\,(A_I^i, A) + \overline{O}\,(A_Y^i, A) - \overline{O}\,(A_I^i, A_Y^i) \geq 0,$$

или

$$\overline{O}\,(A_I^i, A) + \overline{O}\,(A_Y^i,\ A) \geq \overline{O}\,(A_I^i,\ A_Y^i). \tag{19}$$

Неравенство (19) есть неравенство треугольника. Следовательно, с учетом (14) и (19) $\overline{O}\,(A_I^i, A_Y^i)$ есть мера отличия на классификации, откуда следует, что классификации с несколькими плоскостями деления являются метрическими пространствами.

Мера отличия (19), она же семантическое расстояние, обладает следующим свойством: если выбрать два произвольных класса, относительное расстояние между которыми постоянно и конечно, то мера (19) убывает при возрастании относительного расстояния от самой общей классификации до ближайшего обобщающего класса выбранных классов. Так, если $c = R(A_I^i, A_I^i \cdot A_Y^i) + R(A_Y^i, A_I^i \cdot A_Y^i)$, то

$$\overline{O}(A_I^i, A_Y^i) = \frac{R(A, A_I^i \cdot A_Y^i) + 1}{R(A, A_I^i \cdot A_Y^i) + c + 1},$$

где $c \geq 0$ по определению относительного расстояния. Следовательно,

$$\overline{O}(A_I^i, A_Y^i) = 1 - \frac{R(A, A_I^i \cdot A_Y^i) + 1}{R(A, A_I^i \cdot A_Y^i) + c + 1} = 0.$$

В то же время.

$$\overline{O}(A_I^i, A_Y^i) = 1 - \frac{R(A, A_I^i \cdot A_Y^i) + 1}{R(A, A_I^i \cdot A_Y^i) + c + 1} = \frac{1}{2}. \qquad (20)$$

Выражение (20) свидетельствует о том, что если известен только один уточняющий класс классификации, то мера отличия с обобщающей его классификацией равна 1/2 или эквивалентна вероятности, что классифицируемый объект либо попадает в уточняющий класс либо нет, с вероятностью 1/2. Уточнение классификации по одному классифицируемому признаку на каждом уровне приводит к следующей мере различия:

$$\overline{O}(A_I^i, A_Y^i) = 1 - \frac{R(A, A_I^i \cdot A_Y^i) + 1}{R(A, A_I^i \cdot A_Y^i) + c + 1} = \frac{c}{c+1}. \qquad (21)$$

Если в выражение (21) вместо константы c подставить удвоенную максимальную длину пути уточнения (максимальный ранг), то получим максимальное значение меры отличия на плоскости деления классификации A^i. Из (21) следует

$$\overline{O}(A_I^i, A_Y^i) = 1 - \frac{R(A, A_I^i \cdot A_Y^i) + 1}{R(A, A_I^i \cdot A_Y^i) + c + 1} = \frac{c}{c+1}.$$

Субъективность меры отличия между концептами. В работе [14] мера отличия на плоской классификации обобщена на случай пространственной классификации, представляющей собой компьютерную онтологию, и показана применимость предложенной меры для математического моделирования. Однако не доказано, что мозг использует такой простой механизм определения меры отличия между концептами. Более того, был проведен эксперимент, который показал, что мера отличия между концептами зависит от контекста. Эксперимент состоит в следующем.

1. На трех листках бумаги нарисовать:

1) на первом листке — линию (без линейки) длиной семь сантиметров и перевернуть бумагу, чтобы скрыть нарисованную линию;

2) на втором листке — линию 15 см и также перевернуть бумагу;

3) на третьем листке —линию длиной 10 см.

2. Измерить линейкой длины нарисованных линий и записать результаты в виде трех пар оценочной длины и реальной длины.

3. Вычислить разницу между ожидаемой и фактической длиной для каждой строки и найти среднее значение из этих отношений, которое характеризует внутреннюю метрику как статистическую оценку без контекста.

Результаты будут удивительными. Если ваш друг также выполнит это упражнение, велика вероятность того, что ваши линии и линии вашего друга будут значительно отличаться. Более того, если вы повторите алгоритм через неделю, вы, вероятно, будете удивлены новыми результатами, потому что они будут зависеть от вашего опыта и вашего нового состояния ума.

Исходя из результатов описанного эксперимента, а также на основании гипотезы о том, что используемая мера для сравнения концептов имеет более сложную природу [16], можно сделать вывод о том, что внутренняя мера мозга есть линейная взвешенная комбинация набора теоретических мер в системе аксиом NBG (Neumann — Bernays — Gödel) (табл. 1) [17—28].

Введем следующую суперпозицию мер:

$$K=\beta_0\cdot(1-K_\infty(\alpha))+\beta_1\cdot(1-K_{0,1})+\beta_2\cdot(1-K_{0,+1})+\beta_3\cdot(1-K_{0,-1})+$$
$$+\beta_4(1-K_{0,2})=1-\left|\beta_0\cdot K_\infty(\alpha)+\beta_1\cdot K_{0,1}+\beta_2\cdot K_{0,+1}+\beta_3\cdot K_{0,-1}+\beta_4\cdot K_{0,2}\right|, \quad (22)$$

где $\sum_{i=1}^{4}\beta_i=1,\ 0\le\beta_i\le 1,\ -1<\alpha<\infty$. Суперпозиция мер K является линейной комбинацией мер, в которой β_i — весовой коэффициент. Как показано в [15], все меры подобия в (11) могут быть заменены мерами отличия, которые являются строгими мерами:

$$K=\beta_0(1-K_\infty(\alpha))+\beta_1(1-K_{0,1})+\beta_2(1-K_{0,+1})+\beta_3(1-K_{0,-1})+\beta_4(1-K_{0,2}),$$

где $\sum_{i=0}^{4}\beta_i=1;\ 0\le\beta_i\le 1;\quad -1<\alpha<\infty$.

Сравнивая меры Жаккара и Серенсена, можно сделать вывод об «оптимистичности» или «пессимистичности» одной меры по отношению к другой. Предложенная мера (22) в зависимости от значения α и β_i может вести себя пессимистично, нейтрально или оптимистично, что позволяет говорить о возможности ее «обучения».

В работе [15] предложена модель вычислений на многомерных классификациях с определенной степенью доверия:

$$O_K(F) = \frac{1}{NK} \sum_{i=1}^{N} O(F^i)\, p(F^i),$$
$$p(F^{k+1}) > (2L_{F^{k+1}} + 1) \sum_{i=1}^{k} p(F^i), \tag{23}$$

где $F^1, ..., F^i, ..., F^N$, $i = \overline{1, N}$ — совокупность плоских классификаций (плоскостей деления), формирующих многомерную классификацию $F = F^1 \times \times F^2 \times ... \times F^N$, $i = \overline{1, N}$; $p(F^i)$ — вес или коэффициент влияния плоскости деления F^i на меру отличия $O_K(F)$; L_{F^K} — максимальный ранг плоской классификации F^K.

Таблица 1. **Сводная таблица мер**

Наименование	Обозначение	Формула
Континиум эквивалентных мер	$K_\infty(\alpha)$	$\frac{2[n(A \cap B) + 1]}{(1+\alpha)[n(A) + n(B) + 2 - 2\alpha[n(A \cap B) + 1]]}$, $-1 < \alpha < \infty$
Коэффициент Кульчицкого	$K_{0,1}$	$\frac{n(A \cap B) + 1}{2}\left[\frac{1}{n(A)+1} + \frac{1}{n(B)+1}\right]$
Мера Шимкевича–Симпсона	$K_{0,+1}$	$\frac{n(A \cap B) + 1}{\min(n(A), n(B)) + 1}$
Мера Брауна–Бланка	$K_{0,-1}$	$\frac{n(A \cap B) + 1}{\max(n(A), n(B)) + 1}$
Коэффициент Отиаи	$K_{0,2}$	$K_{0,2} = \frac{n(A \cap B) + 1}{\sqrt{(n(A)+1)(n(B)+1)}}$

Таблица 2. **Система ошибок деления**

Правило деления	Ошибка деления
1. Правило соразмерности Объем делимого понятия должен быть равен сумме объемов членов деления Среди членов деления не должно быть предметов, которые не входят в объем делимого понятия. В то же время, все элементы объема делимого понятия должны быть распределены	Неполное деление Не все члены деления названы. Остаются такие предметы из объема делимого понятия, которые не распределены вообще *Пример:* *Науки бывают гуманитарные и естественные. В данном случае названы не все виды наук: технические, математические, социальные*
	Деление с лишними членами Названы члены деления, не входящие в родовое понятие *Пример*: *Книги бывают художественные, учебные, научные и рукописные свитки. Рукописные свитки вообще книгами не являются*
2. Правило единого основания. Деление должно быть выполнено по одному основанию Для классификации не может быть использовано несколько основных признаков. Классификация в таком случае становится путаной	Сбивчивое деление В основу деления положено несколько основных признаков. Часть членов деления образуется по одному признаку, а другая часть — по другому *Примеры*: *Преступления бывают умышленные, неумышленные и заказные. Первые два преступления выделены по признаку мотивации, а третье — по принципу организации* *Науки бывают гуманитарные, естественные, технические, социальные, математические, прикладные и фундаментальные. Фундаментальные и прикладные вычленены по одному признаку – отношению к практике, а оставшиеся виды наук – по другому, а именно, по предмету*
3. Запрет пересечения членов деления Пересечение членов деления означает, что один и тот же предмет является элементом нескольких членов деления. Классификация получается путанной	Пересечение членов деления Приводит к тому, что один и тот же элемент входит в объем двух или нескольких членов деления. *Примеры*: *Преступления бывают умышленные, неумышленные и неосторожные. Объемы понятий «неосторожные» и «неумышленные» преступления совпадают* *Науки бывают гуманитарные, естественные, общественные, социальные, технические, математические. Общественные и социальные науки — это одно и то же*
4. Правило непрерывности Деление должно быть непрерывным. Это означает, что в процессе деления должен быть выполнен постепенный переход от родовых понятий к ближайшим видовым понятиям	Скачок в делении Возникает в процессе деления, когда смешивают роды и виды *Примеры*: *Науки бывают естественные, физические, химические, биологические, гуманитарные, философские, математические, технические, медицинские.* *Физические, химические, биологические науки являются видами естественных наук и выделены преждевременно. Они должны быть указаны при следующем шаге классификации. Произошло смешивание на одной ступени деления родов и видов*

Таблица 3. **Система выявленя ошибок деления**

Ошибка деления	Возможность выявления ошибки
Неполное деление	Если эксперт не знает о существовании других уточняющих классов для некоторого произвольно выбранного класса, для которого уже определены уточняющие классы, то данная ошибка невыявляема
Деление с лишними членами	Лишние члены не входят в родовое понятие, т.е. не являются уточняющими классами родителя. Следовательно, среди всех комбинаторных сочетаний по три встретится комбинация, для которой треугольник взаимоудаленности не будет равносторонним. Данная ошибка выявляема
Сбивчивое деление	Поскольку часть членов деления образуется по одному признаку, а другая часть – по другому, то среди всех комбинаторных сочетаний по три встретится комбинация, для которой треугольник взаимоудаленности не будет равносторонним. Данная ошибка выявляема. Выявление сбивчивого деления требует обязательной проверки всех возможных комбинаторных сочетаний по три среди уточняющих классов одного класса-родителя
Пересечение членов деления	Поскольку один и тот же элемент входит в объем двух или нескольких членов деления, выявить данную ошибку проверкой всех комбинаторных сочетаний из трех среди уточняющих классов одного и того же родительского класса не представляется возможным, т.е. ошибка на текущий момент невыявляема. Однако она может быть выявляема, если рассмотреть гипотезу о возможности построения треугольника, в котором неподвижным классом является родительский класс
Скачок в делении	Поскольку ошибка возникает в процессе деления, когда смешивают роды и виды, то среди всех комбинаторных сочетаний по три встретится комбинация, для которой треугольник взаимоудаленности не будет равносторонним. Данная ошибка выявляема

Обучаемую модель вычислений на классификациях можно получить заменой меры отличия на плоской классификации $O_K(F^i)$ мерой отличия (23) в виде $O_K(F^i, \alpha^i, B^i)$, где $B^i = \{\beta_1^i, \beta_2^i, \beta_3^i, \beta_4^i\}$ и $\sum_{i=0}^{4} \beta_i = 1$:

$$O_K(F) = \frac{1}{NK} \sum_{i=1}^{N} O_K(F^i, \alpha^i, B^i)\, p(F^i),$$
$$p(F^{k+1}) > (2L_{F^{k+1}} + 1) \sum_{i=1}^{k} p(F^i). \tag{24}$$

Модель (24) является адаптируемой (обучаемой) параметрами α^i и B^i для каждой плоскости деления F^i. Следовательно, суть обучения модели (24) сводится к нахождению значений матрицы размером $5 \times N$, в которой α^i и $B^i = \{\beta_1^i, \beta_2^i, \beta_3^i, \beta_4^i\}$ формируют строку из пяти значений для каждой плоскости делений из N.

Обоснованность использования компьютерных онтологий. Прежде чем постулировать обоснованность построения сильного ИИ на основании онтологий, укажем на недоступность любого иного подхода, который не основан на знаниях, представленных концептами. Теоретически иной подход может существовать, но доказать это крайне сложно ибо невозможно четко сформулировать предмет доказательства. Изучение механизмов хранения и обработки знаний неразрвно связано с изучением механизмов их передачи. Результаты исследований показывают, что передача знания о некотором более сложном концепте основана на передаче знаний о более простых или похожих концептах и всего множества отличий между известными концептами и новыми. Настоящая система образования построена на этом принципе.

В работе [29] рассмотрено построение корректных онтологий как развитие идеи классификаций. Сформулированы принципы, называемые правилами деления [30], что приводит к ошибкам, которые могут быть систематизированы, как показано в табл. 2 [31]. Составителем классификации [31] также показаны возможности самостоятельного выявления ошибок (табл. 3).

Таким образом, можно утверждать, что корректные компьютерные онтологии являются формой толковых словарей, в которых более сложные концепты или классы выражены через более простые или наиболее используемые концепты и совокупность отличий от них.

Выводы

Описанное высокоуровневое представление об основах позволяет построить теорию сильного ИИ, оперирующего не набором символов и числами, а семантическими объектами. В основу подхода положена гипотеза о том, что мозг человека оперирует концептами, о чем свидетельствует механизм передачи знаний, принятый как единственно верный в обществе, а именно объяснение более сложных концепций через более простые. Мозг человека оперирует концептами, образуя из них внутреннюю модель мира. При этом мера семантической близости между концептами изменяется в зависимости от контекста.

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. *Бостром Н.* Искусственный интеллект. Этапы. Угрозы. Стратегии. СПб.: Манн, 2016, 496 с.
2. *Minds S.J.* Brains, and programs // Behavioral and brain sciences, 1980, Vol. 3, No 3, p. 417—424.

3. *О'Коннор Д., Макдермотт И.* Искусство системного мышления: Необходимые знания о системах и творческом подходе к решению проблем. М.: Альпина Паблишер, 2014, 256 с.
4. *Harnad S.* Searle's Chinese Room Argument. Encyclopedia of Philosophy // Macmillan Reference, 2006, Vol. 2, p. 239—242.
5. *Ликопастов Е.* Тест Тьюринга пройден // Gadjets news. [Электронный ресурс] Режим доступа: http:gadgets-news.ru/test-tyuringa-projden . Дата обращения: 12.09.19.
6. *habr:* λ-исчисление. Часть первая: история и теория. [Электронный ресурс] Режим доступа: https:habr.com/ru/post/215807 . Дата обращения: 12.09.19.
7. *Landin J.P.* The mechanical evaluation of expressions // The Computer Journal. British Computer Society, 1964, Vol. 6, No. 4, p. 308—320.
8. *Барендрегт Х.* Ламбда-исчисление. Его синтаксис и семантика: Пер. с англ. М.: Мир, 1985, 606 с.
9. *Маклейн С.* Категории для работающего математика. М.: Физматлит, 2004, 349 с.
10. *Milewski B.* Category Theory for Programmers [Electronic Resource]. Source: https: bartoszmilewski.com/2014/10/28/category-theory-for-programmers-the-preface .
11. *Eagleman D.* The brain: the story of you // Canongate Book (November, 2015), 257 p.
12. *Kahneman D.* Thinking, Fast and Slow // Farrar, Straus and Giroux; 1 edition (October 25, 2011), 528 p.
13. *Курчанов Н.А.* Поведение: эволюционный подход. Учеб. пособие. С.-Пб.: СпецЛит., 2012, 232 с.
14. *Редозубов А.* Искусственный интеллект как совокупность вопросов. Habr. [Электронный ресурс] Режим доступа: https://habr.com/ru/post/151102/. Дата обращения: 12.09.19.
15. *Кравцов Г.А.* Модель вычислений на классификациях // Электрон.моделирование, 2016, **38**, № 1, p. 73—87.
16. *Кравцов Г.А., Кошель В.И., Долгоруков А.В., Цуркан В.В.* Обучаемая модель вычислений на классификациях // Там же, 2018, **39**, № 3, с. 63—76.
17. *Семкин Б.И., Горшков М.В.* Аксиоматическое введение мер сходства, различия, совместности и зависимости для компонентов биоразнообразия // Вестник Тихоокеанского государственного экономического университета, 2008, № 4, с. 31—46.
18. *Ким Дж.-О., Мьюллер Ч.У., Клекка У.Р.* Факторный, дискриминантный и кластерный анализ. Под ред. И.С. Енюкова. М. : Финансы и статистика, 1989, 215 с.
19. *Jaccard P.* Distribution de la flore alpine dans le Bassin des Dranses et dans quelques regions voisines // Bulletin de la Societe Vaudoise des Sciences Naturelles, 1901, Vol. 37, № 140, p. 241—272. DOI : 10.5169/seals-266440.
20. *Levandowsky M., Winter D.* Distance between Sets // Nature, 1971, Vol. 234, p. 34—35. DOI : 10.1038/234034a0.
21. *Sörensen T.* A method of establishing groups of equal amplitude in plant sociology based on similarity of species content // Biologiske Skrifter, 1948, Vol. 5, No. 4, p. 1—34.
22. *Kulczynski S.* Zespoly ryslin w Pieninach (Die Pflanzenassociationen der Pienenen) // Bulletin International de L'Acadremie Polonaise des Sciences et des Letters, Classe des Sciences Mathematiques et Naturelles. Serie B. Supplrement II, 1927, Vol. 2, p. 57—203.
23. *Sokal R.R., Sneath P.H.A.* Principles of numerical taxonomy. N.-Y.: W.H. Freeman & Co, 1963, 359 p.
24. *Szymkiewicz D.* Une contribution statistique a la gйographie floristique // Acta Soc. Bot. Polon, 1934, Vol. 34, № 3, p. 249—265.
25. *Simpson G.G.* Holarctic mammalian faunas and continental relationship during the Cenozoic // Bull. Geol. Sci. America, 1947, Vol. 58, № 2, p. 613-688.
26. *Braun-Blanquet J.* Pflanzensoziologie Grundzьge der Vegetationskunde. Berlin : Springer-Verlag Wien, 1951, 632 p. DOI : 10.1007/978-3-7091-4078-9.

27. *Ochiai A.* Zoogeographical studies on the soleoid fishes found Japan and its neighboring regions-II // Bull. Jap. Soc. sci. Fish, 1957, Vol. 22, No. 9, p. 526—530. DOI : 10.2331/suisan. 22.526.
28. *Сёмкин Б.И.* Эквивалентность мер близости и иерархическая классификация многомерных данных // Иерархические классификационные построения в географической экологии и систематике, 1979, p. 97—112.
29. *Кравцов Г.А., Кошель В.И.* Вычисления на классификациях. Корректность классификации // Электрон.моделирование, 2017, **39**, № 5, p. 59—69.
30. *Ивлев Ю.В.* Логика. М.: Изд-во «Проспект», 2008, 304 с.
31. *Буквы!:* Правила деления в логике и ошибки в делении [Электронный ресурс]. Режим доступа: http://bukvi.ru/pravo/logika/pravila-deleniya-v-logike-i-oshibki-v-delenii.htm. Дата обращения: 12.09.19.

Получена 06.09.19

REFERENCES

1. Bostrom, N. (2016), Artificial Intelligence. Stages. Threats. Strategies, Mann.
2. Searle, J. and Minds. (1980), “Brains, and programs”, *Behavioral and brain sciences*, Vol. 3, no. 3, pp. 417-424.
3. O’Konnor, J. and Makdermott, I. (2014), The Art of Systems Thinking: Essential knowledge of systems and a creative approach to problem solving, Alpina Publisher.
4. Harnad, S. (2006), Searle’s Chinese Room Argument, Encyclopedia of Philosophy, Vol. 2, pp. 239-242.
5. Likopastov, Y. (2014), “Turing test passed”, Gadjets news, available at: http://gadgets-news.ru/test-tyuringa-projden/ (accessed September 6, 2019).
6. habr: (2014), “λ-calculus. Part One: History and Theory”, available at: https://habr.com/ru/post/ 215807/ (accessed September 6, 2019).
7. Landin, J.P. (1964), “The mechanical evaluation of expressions”, *The Computer Journal. British Computer Society*. Vol. 6, no. 4, pp. 308-320.
8. Barendregt, X. (1985), Lambda-ischisleniye. Yego sintaksis i semantika [Lambda calculus. Its syntax and semantics], Mir.
9. MacLane, C. (2004), Kategorii dlya rabotayushchego matematika [Categories for working math], Fizmatlit, Moscow, Russia.
10. Milewski, B. (2014), “Category Theory for Programmers”, available at: https://bartoszmilewski.com/2014/10/28/category-theory-for-programmers-the-preface/ (accessed September 6, 2019).
11. Eagleman, D. (2015), The brain: the story of you, Canongate Books, 257 p.
12. Kahneman, D. (2011), Thinking, Fast and Slow, Farrar, Straus and Giroux, 528 p.
13. Kurchanov N.A. (2012), Povedeniye: evolyutsionnyy podkhod. Uchebnoye posobiye [Behavior: an evolutionary approach. Tutorial], SpetsLit, St. Petersburg, Russia.
14. Redozubov, A. (2012), “Artificial Intelligence as a Set of Issues”, available at: https://habr.com/ru/post/151102/ (accessed September 6, 2019).
15. Kravtsov, G.A. (2016), “Classification Computing Model”, Elektronnoye modelirovaniye, Vol. 38, no. 1, pp. 73-87.
16. Kravtsov, G.A., Koshel’, V.I, Dolgorukov, A.V. and Turcan, V.V. (2018), “Classification Classification Learning Model”, Elektronnoye modelirovaniye, Vol. 39, no. 3, pp. 63-76.
17. Semkin, B.I. and Gorshkov, M.V. (2008), “Axiomatic introduction of measures of similarity, difference, compatibility and dependence for components of biodiversity”, Bulletin of the Pacific State Economic University, no. 4, pp. 31-46.

18. Joe, Kim., Muller, C.W. and Kleck, W.R. (1989), Faktornyy, diskriminantnyy i klasternyy analiz [Factor, discriminant and cluster analysis], Finansy i statistika.
19. Jaccard, P. (1901), "Distribution of alpine flora in the Dranses Basin and in a few neighboring regions", Bulletin of the Vaudoise Society of Natural Sciences, Vol. 37, no. 140, pp. 241-272, DOI : 10.5169/seals-266440.
20. Levandowsky, M. and Winter, D. (1971), "Distance between Sets'', Nature, Vol. 234, pp. 34-35, DOI : 10.1038/234034a0.
21. Sцrensen, T. (1948), "A method of establishing groups of equal amplitude in plant sociology based on similarity of species content", Biologiske Skrifter, Vol. 5, no. 4, pp. 1-34.
22. Kulczynski, S. (1927), "Plant complexes in the Pieniny", International Newsletter of the Polish Academy of Sciences and Letters, Class of Mathematical and Natural Sciences, Vol. 2, pp. 57-203.
23. Sokal, R.R. and Sneath, P.H.A. (1963), Principles of numerical taxonomy, New York : W.H. Freeman & Co.
24. Szymkiewicz, D. (1934), A statistical contribution to floristic geography, Acta Soc. Bot. Polon, Vol. 34, no. 3, pp. 249-265.
25. Simpson, G. G. (1947), "Holarctic mammalian faunas and continental relationship during the Cenozoic'', Bull. Geol. Sci. America, Vol. 58, no. 2, pp. 613-688.
26. Braun-Blanquet, J. (1951), Plant Sociology Fundamentals of Vegetation Science, Springer-Verlag Wien, Berlin, Germany, DOI : 10.1007/978-3-7091-4078-9.
27. Ochiai, A. (1957), "Zoogeographical studies on the solenoid fishes found Japan and its neighboring regions-II", Bull. Jap. Soc. sci. Fish, Vol. 22, no. 9, pp. 526-530, DOI : 10.2331/suisan.22.526.
28. Sлmkin, B.I. (1979), Ekvivalentnost' mer blizosti i iyerarkhicheskaya klassifikatsiya mnogomernykh dannykh [Equivalence of proximity measures and hierarchical classification of multidimensional data], Iyerarkhicheskiye klassifikatsionnyye postroyeniya v geograficheskoy ekologii i sistematike, pp. 97-112.
29. Kravtsov G.A. and Koshel' V.I. (2017), "Classification calculations. Classification classification", Elektronnoe modelirovaniye, Vol. 39, no. 5, pp. 59-69.
30. Ivlev, YU.V. (2008), Logika [Logics], Prospekt, Moscow, Russia.
31. "Letters !: Rules of division in logic and errors in division", available at: http://bukvi.ru/pravo/logika/pravila-deleniya-v-logike-i-oshibki-v-delenii.html (accessed September 6, 2019).

Received 06.09.19

Г.О. Кравцов, В.В. Левітін,
В.І. Кошель, В.В. Нікітченко, А.Н. Примушко

СИЛЬНИЙ ШТУЧНИЙ ІНТЕЛЕКТ: ПЕРЕДУМОВИ

Подано огляд фундаментальних основ, які дозволяють побудувати сильний штучний інтелект (ШІ). Показано справедливість висунутих гіпотез, доведених методом натурних експериментів. Стверджується, що для побудови математичної теорії сильного ШІ слід перейти до системи аксіом Неймана—Бернайса—Геделя, і тоді відкривається можливість використовувати семантичні структури, які відтворюються комп'ютерними онтологіями, як алгебраїчні структури. Для коректного використання онтологій в системах ШІ необхідно, щоб онтології в площинах розподілу являли собою метричні простори.

К л ю ч о в і   с л о в а: штучний інтелект, система аксіом, аксіоматика, мозок людини, семантика, міра відмінності, внутрішня модель світу, адаптованість.

G.A. Kravtsov, V.V.Levitin,
V.I. Koshel', V.V. Nikitchenko, A.N. Primushko

STRONG ARTIFICIAL INTELLIGENCE:
BACKGROUND PRECONDITION

The article provides an overview of the fundamental foundations for building strong artificial intelligence. The validity of the hypotheses put proved by the method of field experiments is shown. It is argued that in order to construct a mathematical theory of strong artificial intelligence (AI) it is necessary to go over to the von Neumann-Bernays-Godel system of axioms and then the possibility of using semantic structures represented by computer ontologies as algebraic structures opens up. For the correct use of ontologies in artificial intelligence systems, it is necessary that ontologies along the division planes are metric spaces.

K e y w o r d s: artificial intelligence, axiom system, human brain, semantics, measure of difference, internal model of the world, adaptability.

КРАВЦОВ Григорий Алексеевич, канд. техн. наук, докторант Ин-та проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины. В 2000 г. окончил Севастопольский военно-морской ин-т им. П.С. Нахимова. Область научных исследований — кибербезопасность смарт-грид, криптография, программирование, разработка распределенных гетерогенных вычислительных систем.

ЛЕВИТИН Владимир Викторович, генеральный директор Компании NaomiAI. В 2015 г. окончил Научно-исследовательский университет информационных технологий, механики и оптики (г. Санкт-Петербург, Россия). Область научных исследований — искусственный интеллект, автоматизация процессов, распределенные системы.

КОШЕЛЬ Владимир Иванович, аспирант Ин-та проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины. В 2002 г. окончил Харьковский национальный университет им. В.Н. Каразина. Область научных исследований — искусственный интеллект, интеллектуальный анализ данных, искусственные нейронные сети, обработка естественного языка.

НИКИТЧЕНКО Владимир Владимирович, аспирант Ин-та проблем моделирования в энергетике им. Г.Е. Пухова НАН Украины. В 1991 г. окончил Полтавское высшее военное командное училище связи. Область научных исследований — информационно-коммуникационные технологии, искусственный интеллект.

ПРИМУШКО Арсентий Николаевич, магистр Национального технического университета Украины «Киевский политехнический институт им. Игоря Сикорского», бакалаврат которого окончил в 2019 г. Область научных исследований — искусственный интеллект, машинное обучение, искусственные нейронные сети, программирование, системы навигации и ориентации.