МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

ВСЕРОССИЙСКИЙ ИНСТИТУТ НАУЧНОЙ И ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ
РОССИЙСКОЙ АКАДЕМИИ НАУК
(ВИНИТИ РАН)

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

Серия 2. ИНФОРМАЦИОННЫЕ ПРОЦЕССЫ И СИСТЕМЫ

ЕЖЕМЕСЯЧНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ СБОРНИК

Издается с 1961 г. **№ 5** **Москва 2019**

ОБЩИЙ РАЗДЕЛ

УДК 004.774–027.21

В.В. Таран

О развитии концепции Всемирной паутины

Рассматривается концепция Всемирной паутины с учетом ее числовых индексов. Показано влияние компьютерных наук, веб-науки и информационно-коммуникационных технологий на формирование передовых технологических отраслей Интернета. Упорядочены и перечислены основные опорные технические средства и протоколы, образующие технологическую базу, поддерживаемую и обновляемую с целью обеспечения стабильности функционирования Интернета. Приведены технико-гуманитарные характеристики концепции всемирной паутины, соответствующие установленным этапам технологического развития. Предпринята попытка систематизации процессов эволюции Интернета, опорными точками которых служит язык гипертекстовой разметки, взаимодействующий с различными технологиями и языками программирования. Приводятся авторские определения понятия всемирная паутина с технологической и общегуманитарной точек зрения. Проанализирована роль компьютерных наук в развитии интернет-технологий.

Ключевые слова: *Всемирная паутина, Интернет, Нейронет, Концепция, интернет-технологии, HTTP, HTML, CSS, WEB-1.0 – WEB-5.0, OSI, SSI, ISAPI, CGI, ASP, WAP*

Всемирная паутина сегодня – это высоконагруженный и динамически развивающийся технический проект, призванный обеспечить устойчивое функционирование Глобальной информационно-коммуникационной сети Интернет, механизм которой устроен так, что практически любая технологическая подплатформа вступает во взаимодействие с аббревиатурой WWW (World Wide WEB). Более того эта аббревиатура определяет во многом и степень общего технологического состояния и развития сети Интер-

нет. Своим появлением данный проект обязан нескольким крупным игрокам на рынке компьютерных технологий – это Массачусетский технологический институт (MIT Massachusetts Institute of Technology) и Европейский совет по ядерным исследованиям (CERN Conseil Européen pour la Recherche Nucléaire) а также входящей в совет Европейской лаборатории физики элементарных частиц (European laboratory for particle physics).

Техническая суть объекта WWW заключается в координации и всеобщей интеграции различных средств коммуникации, выполняющих конкретные прикладные операции по обработке, упорядочиванию и передачи данных как внутри локальных, так и внутри глобальной информационных сетей. Интересно, что уровень внедрения новых технологий в Интернет определяется числовыми привязками к аббревиатуре WWW[1*] и сокращенно упоминается как WEB-1.0[2], WEB-2.0[3], WEB-3.0[4]. Вследствие этих привязок мы имеем некие технологические этапы развития Глобальной сети Интернет. Например, главная технология и одновременно язык для семантической связи между различными объектами в Глобальной Сети есть технология HTML[5]. Гипертекстовые массивы, хранящиеся на различных WEB-серверах, связаны между собой информационно-распределительной технологией WWW, и взаимодействие технологий HTML с технологией-протоколом HTTP[6] фактически является неотъемлемой частью самой структуры WWW. В научно-технической периодике, посвященной изучению WWW-технологии[a)], довольно часто даются очень узкие трактовки этого понятия, но технологию WWW нельзя рассматривать только как технологию гиперсвязей. В связи с этим важно взглянуть на эту технологию с точки зрения информационных процессов, которые аккумулируются именно во Всемирной паутине. Ускоренное развитие компьютерных наук и производных от них информационно-коммуникационных технологий позволяют говорить о WWW как о нечто более широком и многогранном, нежели просто как о прикладной технологии, выполняющей рутинные функции по распределению гиперсвязей в Сети. Следует подчеркнуть и правовую сторону вопроса, поскольку гипертекстовые связи осуществляют не только транспортную функцию, но и способствуют более качественной идентификации источников, взаимосвязанных между собой, роль WWW в учете персональных данных только усиливается, [1, с. 4-7; 2].

С недавнего времени Всемирная паутина благодаря компьютерным наукам[b)] претерпевает различные технологические инновации и в соответствии с этим трансформируется и первоначальная концепция развития Интернета. В последнее время мы наблюдаем все большую степень децентрализации интернет-ресурсов и технологий [3]. Здесь часто и возникают некоторые вопросы относительно трактовок концепций технологического развития Интернета и Всемирной паутины. В оригинале инфраструктура сети Интернет включает в себя Глобальную (Всемирную) паутину и соответствующие технические протоколы канального взаимодействия, а в качестве технологической инфраструктурной подсистемы используется сетевая модель OSI (Open Systems Interconnection, basic reference model) – это эталонная модель взаимодействия открытых систем, реализующая условия, при которых сетевые устройства могут адекватно взаимодействовать между собой, имеющая несколько уровней: физический, канальный, сетевой, транспортный, сеансовый, представление данных, прикладной. Но время не стоит на месте и возникает необходимость учета новых, более совершенных моделей технического взаимодействия внутри Интернета. И на этом этапе появляется взаимозависимость прикладных технологий, обеспечивающих сетевое взаимодействие по обмену данными между локальными сетями (на канальном уровне) и на линиях оптоволоконной связи (на физическом уровне). К тому же способы управления информационным содержимым все больше начинают зависеть от грамотно спроектированного интерфейса, на базе концепции WWW – WEB-интерфейса [4]. Поскольку физический уровень совершенствуется медленнее, чем канальный, то последний является определяющим в смысле внедрения интеллектуальных технологий в Интернет. Первый уровень базовой эталонной модели взаимодействия открытых систем (OSI), который представляет собой опорную физическую основу для WWW, часто стыкует[c)] различные физические каналы распространения данных [5]. В результате чего снижается скорость соединения [см. 6] с серверами даже в зоне ближайшей географической локации. Одним из принципиальных моментов концепции WWW, по нашему мнению, должен стать аспект выработки оптимальной программной политики (усовершенствование WEB-протоколов, алгоритмов временного хранения информации) для борьбы с падениями сетевой нагрузки в пиковое время, возникающей в результате несовместимости технологий на физическом канале[d)]. Прежде всего

---

[*] Все пояснения приведены в *Приложении*

[a)] В различных источниках научно-технической литературы, посвященных вопросам исследования и функционирования сети Интернет, можно встретить понятие WWW, определяемое как технологию либо совокупность технологий, основанную на гиперсвязях.

[b)] Компьютерные науки (в России – наука «Информатика») оказывают огромное влияние на развитие Интернета в целом, и особенно на его технологическую составляющую. Благодаря новым средствам телекоммуникации информационно-коммуникационные технологии на базе Интернет открывают нам все новые и новые возможности в сфере гуманитарной коммуникации.

[c)] Под стыковкой физических каналов передачи данных понимается процесс объединения различных типов кабеля (оптоволоконный кабель, медный кабель) между собой.

[d)] Стыковочные модули физического уровня бывают разного качества. Практика показывает, что даже самый качественный стыковочный модуль в пиковые нагрузки может ещё больше усугублять ситуацию, все это сказывается на качественном обмене пакетными данными внутри сети. Многие локальные сети, имеющие выход в Интернет, на физическом уровне связаны между собой устаревшими технологиями (к примеру, медный кабель подсоединяется к оптоволокну). Рост таких сетей в силу разных причин (прежде всего экономических) не способствует улучшению качества передачи данных в Интернете и, соответст-

это важно по экономическим причинам. Если предположить, что транспортный уровень будет усовершенствован, то проблема низких скоростей (при условии маршрутизации) локальных сетей стыкующихся между собой для образования регионального сегмента Глобальной сети в регионах может быть успешно решена. По крайней мере, это будет значительный шаг вперед. С рейтинговой точки зрения рассматриваемый нами аспект может существенно улучшить состояние общей сетевой готовности страны, в соответствии с критерием готовности NRI [7]. Рассмотрим технологическую цепочку зависимости числовых привязок применительно к WWW. Технология WEB-1.0 (как аппаратная основа) появилась благодаря соревнованию между перспективными разработками NORAD[7] и технологией MILNET[8]. Прообразом этой технологии можно считать разработанную Тимом Бернерсом-Ли[e)] в Европейском совете по ядерным исследованиям программу Enquire[9], принципом действия которой стала случайная ассоциация данных. Примерно в это же время появился на свет язык гипертекстовой разметки первой версии. Интернет и его структура та, которую мы видим сегодня, начинается в 90-х гг. XX в., когда за основу Глобальной информационно-коммуникационной сети была взята система гипертекстовой коммуникации, которая взаимодействовала с Глобальной паутиной.

Но по мере развития данной концепции на различных исторических этапах требовалась технологическая модернизация, прежде всего, связанная с внедрением интеллектуальных технологий. Поскольку Всемирная паутина понятие слишком широкое, необходимо рассмотреть ее взаимодействие с ключевыми технологиями реализации контента в сети Интернет. Для более удобного понимания технологических процессов предлагаем разделение определенных технологических этапов развития сети Интернет на характеристики.

Характеристика WEB-1.0. Отправной точкой в развитии Глобальной информационно-коммуникационной сети является WEB-1.0. Ее разработка началась в 1986 году и завершилась к 1991 г. Понятие Глобальная паутина (World Wide WEB) тесно связано с изобретением технологии гипертекста и введением вспомогательного протокола передачи гипертекста HTTP (HyperText Transfer Protocol). Интернет, согласно общей технологической структуре, представляет собой объединение локальных ведомственных и иных региональных сетей посредством технических протоколов, обеспечивающих его стабильное функционирование. Чтобы не упоминать каждый раз массивную привязку World Wide WEB специалисты в области информационных технологий ограничились окончанием данной технологии WEB[f)] (Паутина). С этого момента именно WEB стал определять основные технологические вехи развития Интернета и заслужил звание неофициального индикатора, показывающего уровень внедрения информационно-коммуникационных технологий в интернет-инфраструктуру. WEB-1.0 опорной точкой которого выступает HTML-2.0 по существу – это технология, которая только объединяла WEB-странички, но не делала их динамическими.

Для эпохи WEB-1.0 характерным является статичность различных WEB-конструкций, размещаемых в Интернете. Примерно в это же время официально был доработан язык гипертекстовой разметки первого поколения (спецификация RFC-1867/HTML-2.0). Важным звеном здесь является тот факт, что язык разработан таким образом, чтобы в случае переноса актуальных гипертекстовых связей они сохранялись бы независимо от аппаратного обеспечения, т.е. копия гипертекстовых связей могла быть перенесена на любой компьютер в независимости от его мощности и аппаратной конфигурации. Возможности гипертекстовой разметки включали следующие функции:

- отображение статичных страниц на английском языке;
- установление межтекстовой связи между текстовыми блоками;
- обеспечение обратной связи по прообразу нынешних блогов и чатов;
- агрегацию примитивной линейной информации о различных событиях: курсы валют, краткие новости (в виде пейджинга), погода;
- хранение и упорядочивание информации тексто-графического содержимого.

Яркой характерной чертой технологической эпохи WEB-1.0 может служить возможность использования компьютеров как вспомогательной базы для установления примитивной мультимедийной коммуникации. Как и любая инновация, внедренная в народнохозяйственную деятельность, технология WEB-1.0 стала фундаментом для последующих модификаций структуры Интернета. Поскольку общий научно-технический прогресс диктует быстрые изменения, то уже в 1995 г. выходит версия HTML-2.0[10], которая начинает внедряться как некий официальный стандарт технического оформления документов в Интернете. До появления HTML-2.0 фактически не было единого официального стандарта, применяемого при проектировании сайтов, порталов и WEB-приложений. Существовали технологии и модификации прообраза модели HTML, наличие тегов в разных спецификациях отличалось, причем функционально они могли различаться, что приводило к путанице среди тогдашних разработчиков и снижало уровень интеграции WEB-технологий, которые внедрялись тогда на платформу Интернета. Более того, несмотря на введение официальных спецификаций, и обновляемых инструкций, которые мы наблюдаем, сегодня большинство ресурсов остаётся дезинтегрированным вследствие ещё использующихся различных разработок, уходящих

венно, противоречит положениям о качестве передачи информации в соответствии с WEB-концепцией.

e) Тим Бернерс-Ли (Timothy John Berners-Lee) – британский исследователь в области компьютерных наук, WEB-наук, наук об информации. Основатель проекта глобальной всемирной паутины совместно с бельгийским исследователем Робертом Кайо (Robert Cailliau). Внес значительный вклад в развитие сопутствующих технологий, которые в дальнейшем стали именоваться как WEB-технологии, среди них – HTML, WWW, URL.

f) В настоящей статье привязка WEB упоминается с большой буквы, WEB-технологии, WEB-индустрия и т.д. Для ассоциации именно с WEB-концепцией.

корнями в прошлое эпохи WEB 1.0. Со временем, возможности языка гипертекстовой разметки подверглись ещё некоторым технологическим инновациям:

1) введены специальные инструкции для более корректного взаимодействия с протоколом HTTP;

2) разработаны функции прикладного взаимодействия между терминальным управлением и интерфейсом, включая специальные WEB-формы;

3) в соответствии с изменениями – обновлен состав гипертекстовой разметки.

Хорошим примером взаимодействия технологий в общей концепции WEB 1.0 может служить более адаптированная под высокие скорости соединения в Сети технологическая подплатформа основанная на взаимодействии сетевых служб FTP (File Transfer Protocol) и HTTP (HyperText Transfer Protocol) технологической инфраструктурой взаимодействия которых служит IP/TCP (Internet Protocol/Transmission Control Protocol). После публикация проектной документации по гипертекстовой разметке второй версии в январе 1997 г. выходит версия HTML-3.2 – значительно улучшенная и с введением поддержки математических формул. Спустя четыре месяца появляется долгожданная первая спецификация, рекомендованная WEB-консорциумом – HTML-4.0[g)] – существенно улучшенная версия языка в сравнении с HTML-3.2. В ней модернизированы и добавлены следующие функции: возможность работы с кадрами; элемент <object> для работы с мультимедиа; элементы <th> и <td> для более эффективной работы с таблицами; улучшена работа с WEB-формами. Следующий базовый этап в развитии WWW – это выход версии HTML-4.1. И в мае 2000 г. появляется стандарт ISO HTML[h)], основанный на HTML-4.1.

Наконец, в октябре 2014 г. выходит всеми долгожданный HTML-5.0[i)]. Конечно, по нашему мнению, Глобальная паутина не ограничивается набором спецификаций языка гипертекстовой разметки. Однако HTML – это главный элемент в системе взаимосвязей Всемирной паутины. Почему мы уделили особое внимание именно этой технологии? Ответ кроется в том, что именно она диктует правила, по которым развивается и модернизируется Интернет. Поскольку WWW является технологической концепцией, необходимо указать на взаимосвязи с другими технологиями. Так, по значимости, второй технологией здесь выступает CSS[11], которая дополняет технологию HTML и в целом развивает концепцию WWW. Изначально интернет-страницы создавались, управлялись и верстались с помощью HTML, но наступили времена, когда возникла необходимость сделать страницы более динамичными и здесь понадобилась смежная технология, работающая в паре с HTML. Результатом этого и стала разработка CSS, которая, в первую очередь, является дизайн-технологией и выполняет функции дизайн-разметки, в частности:

- универсализация отображения на различных устройствах, сверстанных на базе HTML страниц;
- управление параметрами и характеристиками цвета и оттенков;
- управление атрибутами свойств различных графических блоков;
- управление и автоматизация текстовых параметров;
- поддержка верстки блоками;
- генерация стилей и их адаптация под мобильные, и стационарные компьютеры;
- управление селекторами;
- поддержка динамических указателей.

CSS-технология делится на несколько версий в соответствии с уровнями. Первый уровень введен в действие в 1996 г. и имеет название CSS1; второй уровень – в 1998 г. и его модификация CSS2[12]. В 2011 году CSS2 приобретает более удобный вид и ему присваивается модификация CSS2.1[13]. Развязка данной технологии иллюстрирует появление CSS3, модификация которого претерпевает существенные изменения, а кульминацией служит появление модификации CSS4.

Из вышеизложенного следует, что применение двух основных технологий (HTML/CSS) в сочетании с протоколом HTTP, по нашему мнению, и есть та самая база концепции World Wide WEB. Наряду с базовыми технологиями, на которые делает ставку WWW, существуют и так называемые дополняющие технологии, суть которых в том, что их линейка может время от времени меняться, модернизироваться и интегрироваться между собой. Дополняющими технологиями, прежде всего, выступают языки программирования. Приведем основные языки, по нашему мнению, способствующие революционным изменениям программной архитектуры WEB. В первую очередь, это языки, которые дополняют своими возможностями многопрофильную реализацию технически значимых функций: JavaScript[14], Python[15], Go[16], SQL[17], Ruby[18], PHP[19]. Также интерес представляют и такие технологии как: SSI[20], ISAPI[21], CGI[22], ISP[23] ASP[24], WAP[25]. Каждые из этих языков и технологий имеют довольно разнообразные функции, которые описаны в *Приложении*. Дополнительные языки это уже влияние фундаментального фактора развития компьютерных наук. Если базовые технологии, включенные нами в общую концепцию WWW, создавались профессионалами и были максимально упрощены с целью общего усвоения правил распределения и опубликования информации в Глобальной сети, то дополняющие технологии созданы профессионалами для профессионалов и упрощаются они с целью облегчить рутинные действия программистам. Это важный момент. Изначально Интернет был запланирован как свободная технология, именно поэтому у истока его технологического развития существовало так много дезинтегрирующих друг друга технических решений. Концепция Глобальной пау-

---

[g)] Техническое описание доступно по адресу: www.w3.org/TR/REC-html40-971218/ (архивная копия, дата обращения к ресурсу: 22.01.2019).

[h)] Специальная версия гипертекстовой разметки официально ISO/IEC 15445:2000 разработана и введена в действие Международной организацией по стандартизации. В свободном доступе описание находится по адресу: www.scss.tcd.ie/misc/15445/15445.HTML (дата обращения к ресурсу: 22.01.2019).

[i)] HTML (версия 5.0) сопровождалась модификациями, вышедшими в ноябре 2016 г. (HTML-5.1) и в декабре 2017 г. (HTML-5.2).

тины строится на открытости данных, и каждая спецификация, будь-то HTML или CSS, публикуется в открытом доступе в самой же Паутине.

С дополняющими технологиями все сложнее, поскольку они могут обладать коммерческими соглашениями, которые делают их недоступными для большинства пользователей. Здесь *гуманитарная концепция* Всемирной паутины пересекается с *технологической концепцией*. Поскольку в результате исследования мы приходим к выводу, что WWW условно подразделяется на две концепции: *технологическую* и *гуманитарную* считаем возможным – представить их авторские определения. С технологической точки зрения концепцию Всемирной паутины можно трактовать следующим образом: «Глобальная паутина это технологическая концепция, базирующаяся на протоколах и открытых технологиях взаимосвязи локальных и региональных сетей, одобряемых и рекомендуемых Консорциумом Всемирной паутины». С гуманитарной точки зрения определение может звучать так: «Глобальная паутина это концепция открытого доступа к семантической структуре гипертекстовых данных, функционирующих на базе технических решений, введенных в эксплуатацию Консорциумом Всемирной паутины и объединяющая различные виды сетей в зависимости от их типа, и географической распространенности». Здесь особым звеном выступают информационно-коммуникационные технологии, включающие в себя телекоммуникационные услуги.

Уровень развития информационно-коммуникационных технологий сегодня определяется темпами внедрения новых технологий, прежде всего на интернет-основе. Этот показатель сильно влияет на общее инновационно-ориентированное развитие [8]. Соответственно темпы внедрения напрямую зависят от технологической концепции Интернета. В связи с этим, необходимо обозначить горизонты концепции Всемирной паутины. Поскольку развитие Интернета очень хаотичное, то, по нашему мнению, необходимо привязываться к основным базисным его функциям, которые продиктованы введением базовой программной архитектуры. Поэтому программную архитектуру первого поколения можно представить с помощью технологической цепочки:

WEB-1.0=HTML-2.0, HTML-3.0, HTML-3.2 .

Отметим, что в этот технологический период структура сайтов и их дизайн сильно отличались от современных концепций сайтостроения. Дизайн был менее выразительным, функциональные возможности WEB-интерфейса были ограниченными, а WEB-странички были в основном статичными. Стоит также сказать, что и интернет-культура была иной, нежели чем сегодня.

Интернет-культура или в более широком понятии информационная культура во многом зависит от интернет-концепций, поскольку, по нашему мнению, на ее формирование и развитие влияет коммуникационная основа Интернета. В эпоху WEB-1.0 культура построения WEB-страниц была совершенно другой. Кстати, по одному из исследований[j] технология HTML-2.0 вызывала у пользователей больший интерес, чем аналогичная технология, которую мы наблюдаем сегодня (HTML, версия 5.0). В силу объективных причин, на тот момент было очень мало объектно-ориентированных редакторов верстки HTML-страниц, и поэтому пользователям приходилось вникать в основы построения таких страниц. Сегодня знание HTML-5.0 уже излишество, поскольку все основные действия по созданию HTML-страниц предпринимаются пользователями в объектно-ориентированных, часто работающих в режиме (онлайн) редакторах.

Эпоху второго поколения WEB-2.0 можно вывести в следующую цепочку:

WEB-2.0=HTML-4.0, HTML-4.01, ISOHTML.

WEB-2.0 поднимает развитие интернет-технологий на новый уровень. Начинают внедряться технологии, связанные с проектированием WEB-страниц, а также WEB-элементов, внедряются новые методы AJAX[26], разрабатываются более совершенные способы вещания потоковой информации на базе сетевых протоколов [6], вводится система индексации шрифтов, что помогает разработчикам более эффективно управлять шрифтами на WEB-сайтах. Кстати, именно в эту эпоху становится возможным изменять объем текста, его ширину и высоту, что позитивно сказывается на старшем поколении, людях с ограничениями по зрению и т.д.

С точки зрения влияния на интернет-культуру пользователей можно выделить процессы социализации, позволяющие пользователям объединяться в электронные социальные структуры. Использование социальных и профессионально-ориентированных сетей развивает корпоративную культуру. Важно отметить, что помимо социализации развивается и индивидуализация. Здесь обращает на себя внимание растущее количество настроек, делающих сайт индивидуальным для каждого из пользователей, чего нельзя было делать в эпоху WEB-1.0. Внедряется технология WIKI[27] и RSS[28]. Начинаются процессы адаптации сайтов под мобильные коммуникационные устройства.

Эпоха WEB-3.0 ознаменована следующей цепочкой:

WEB-3.0=HTML-5.0, HTML-5.1, HTML-5.2.

WEB-3.0 – это более революционная форма интернет-технологий, сочетающая в себе все наработки предыдущих концепций Глобальной сети, берущая за основу принципы функционирования, основанные на WEB-1.0 и WEB-2.0, но привносящая принципиально новый подход к распределению и управлению информацией, содержащейся на WEB-ресурсах. Преемственность и инновации – вот главный ключ этой концепции. Как правило, революции, происходящие в научно-технологической сфере, часто отвергают опыт, приобретенный до них ранее. В случае с ин-

[j] HyperText Markup Language Specification 2.0. – URL: www.marc.merlins.org/htmlearn/html2.html (дата обращения к ресурсу: 22.01.2019).

тернет-инфраструктурой это не совсем так. Дело в том, что в предыдущие интернет-эпохи, способы размещёния и создания WEB-сайтов базировались на использовании смешанных технологий, например, когда в HTML-4.0 встраивался дополнительный код, причем это могли быть совершенно разные технические решения, т. е. не существовало единых стандартов по созданию WEB-порталов. Разумеется, это шло не на пользу развитию и внедрению новых технологий.

Концепция WEB-3.0 строится на универсализации и одновременно систематизации процессов сайтостроения и хранения информации, вводится понятие семантическая паутина. Наряду с внедрением новейших технологий происходит поэтапное увеличение количества информации, аккумулируемой в Сети, и разумеется, наработки компьютерных наук приходят на помощь с точки зрения прикладной обработки информации. Идентифицировать информацию в Сети становится все труднее, и по этой причине делается ставка на семантические связи информации, методом которых является идентификация источников с помощью метаданных. Метаданные – это отдельная субстанция, которая в семантической обработке должна предоставлять максимум информации о запрашиваемом локатором ресурсе. Для решения этих задач введена поддержка метаязыка, основанного на технологиях DAML[29], OWL[30], RDF[31], OIL[32]. Однако даже с введением этих технологий остается главный пробел – (интеллектуальное) предсказание пользовательского запроса[k)]. Перечисленные нами технологии решили большинство задач, связанных с унификацией использования программных стандартизаций в Глобальной сети, но проблема логической взаимосвязи между гиперссылками остается на поверхности, и возможно она будет существовать в концепции WEB-3.0 ещё долго. Но уже сейчас не за горами концепция нового поколения WEB-4.0, на разработку которой и WEB-5.0 будут оказывать влияние несколько факторов.

Фактор первый. **Искусственный интеллект.** Сегодня среди технических специалистов, культурологов и специалистов по когнитивным направлениям в науке ведутся дискуссии, суть которых сводится к одному – существует ли сегодня искусственный интеллект и, вообще, что же это такое. По мнению автора, в настоящее время искусственный интеллект не существует. Вместо него имеются некие программные алгоритмы, способные ограниченно предсказывать последовательность действий пользователя по заданному шаблону. Но по каким-то причинам в обществе и в некоторой научно-технической литературе утверждается, что искусственный интеллект уже есть и мы им приличное время пользуемся. Если смотреть на проблему в таком аспекте, то получается, что концепция WEB-3.0 должна полностью состояться, однако как следует из нашей статьи – это не так.

Фактор второй. **Самосовершенствующиеся системы.** Это скорее промежуточный этап технологического развития, где мы частично находимся уже сегодня [9]. Самосовершенствующиеся системы – это смоделированные программным путем действия, направленные на поддержание стабильности глобальной информационно-коммуникационной сети Интернет (по определению автора статьи, на основе технических решений). Иными словами, сегодня при выходе из строя любого WEB-ресурса на экране мы видим ошибку, и данную ошибку устраняют технические специалисты. А при модели взаимодействия различных сетевых протоколов с самосовершенствующимися системами – все действия по наладке WEB-ресурса будут устраняться машинным путем. Почему это промежуточная стадия, предшествующая искусственному интеллекту? Потому что это вершина процесса автоматизации, но **не интеллектуальных действий.** Проблема решается машинным путем только по заданному шаблону. Не секрет, что у каждой ошибки есть свой код, соответственно по базе данных таких кодов нетрудно автоматизировать процессы «починки» системы. Сейчас при отказе операционной системы MS Windows[l)] пользователи получают уведомление об отправке специальных сообщений в корпорацию Microsoft, с целью улучшить качество их обслуживания, так вот при использовании самосовершенствующихся систем данные процессы будут протекать автоматически.

Фактор третий. **Дизайн и эргономика.** Когда мы говорим об искусственном интеллекте, то имеем ввиду, прежде всего, интеллектуальные действия, но сегодня доподлинно неизвестно, как в реальности обрабатывает информацию наш мозг. В соответствии с этим рассуждение о реальном искусственном интеллекте, способном полностью эмитировать человеческую деятельность и мыслительные процессы, пока преждевременны. Это все равно, что увидеть пятое измерение. Точно также нам трудно представить, что возможно какое-то другое представление информации на WEB-ресурсе, нежели то, что мы имеем сегодня. Поэтому ресурсы дизайна ограничены. Очень трудно придумать что-то лучше старого доброго колеса, по причине его геометрической формы.

Фактор четвертый. **Энергетические проблемы и Интернет.** Поскольку наблюдается колоссальный рост информации в Глобальной сети, то на аппаратную основу Интернета, состоящую преимущественно из специализированных дата-центров, тратится огромное количество электроэнергии. На этом этапе должны внедряться альтернативные источники энергии различных типов. Причем интеграция электроэнергии, её учет и управление должны лечь в основу переходной концепции WEB-4.0 в WEB-5.0. В некоторых научных публикациях, когда речь идет о WEB-4.0, упоминается революционная технология Нейронет (Neuronet), ко-

[k)] Интеллектуальное предсказание пользовательского запроса – большинство WEB-ресурсов сегодня индексируются различными поисковыми системами, однако приоритетность и релевантность контента предоставляемого машинным путем на стадии автоматизации ещё далека от совершенства. Аналогии по методам сравнения, которые используются сегодня хорошо автоматизированы, но недостаточно интеллектуальны.

[l)] Microsoft Windows – операционная система корпорации Microsoft.

торая ещё больше нацелена на интеграцию и техническую коммуникацию пользователей сети Интернет. Суть технологии основана на механизме преобразований сигналов (нейронов) человеческого (биологического) мозга в электрические сигналы, понятные аппаратно-программному интерфейсу (кремниевый мозг). Вообще проблема налаживания технической коммуникации между человеческим мозгом и компьютером не нова. В разное время над ней трудились известные исследователи: У. Эшби, Д. Ликлайдер, Б. Джонс Кристофер, Д. Энгельбартом, Фрэнсис Хейлигэн, Жак Видаль [10–15]. Однако, по мнению автора, говорить о принципиальных сдвигах в этой области пока рано, поскольку в настоящее время ещё не создан полноценный кремниевый искусственный интеллект, способный хотя бы на 80% эмитировать человеческую деятельность. Процессы же, которые мы сейчас наблюдаем, это попытка восполнить пробелы в тех сферах, куда ещё не проникла интеллектуальная автоматизация. Поэтому говорить об эпохе WEB-4.0 ещё рано, скорее, по нашему мнению, эти процессы будут происходить на переходном этапе от WEB-5.0 к WEB-6.0. Именно поэтому одним из самых важных моментов является развитие компьютерных наук и параллельно развивающихся с ними на условиях взаимодействия WEB-наук, способных дать новые решения всем остальным отраслям научного знания. Таким образом, технологическая концепция развития Всемирной глобальной паутины должна опираться на базовые технические стандарты, поддерживающие стабильное функционирование её программно-аппаратной основы, научные достижения в области компьютерных наук и дизайна и базироваться на открытой технологической основе. Понимание технологических этапов концепции развития сети Интернет должно строиться на комплексном и многофакторном анализе технологической базы, соответствующей периоду внедрения новых технологий либо обновлению уже существующей программно-аппаратной среды. От того по какому пути будет развиваться самый масштабный коммуникационный проект в истории человечества будет зависеть и дальнейшее технологическое развитие всей цивилизации.

СПИСОК ЛИТЕРАТУРЫ

1. Алексеева А.В. Человек и машина в эпоху развития информационных технологий: Защита персональных данных при использовании сети Интернет как основное право гражданина Франции. Нормативно-правовой аспект. Секция «Построение социально-психологических моделей взаимодействия людей и машин». Психофизические и социально-психологические аспекты взаимодействия в системе «человек-машина» // Материалы Всероссийской научно-практической конференции молодых ученых / отв. ред. А.В. Моров. – Ижевск, 2014. – 120 с. 4-7.
2. Бастрикова А.А. Роль интернета в современной жизни М.В. Булгакова // Вестник молодых ученых и специалистов Челябинской области. – 2015, №3(10). – С. 74-76.
3. Патаракин Е.Д. Социальные взаимодействия и сетевое обучение 2.0 – М.: НП «Современные технологии в образовании и культуре», 2009. – 176 с.
4. Скотт Б., Нейл Т. Проектирование веб-интерфейсов : пер. с англ – СПб: Символ-Плюс, 2010. – 352 с.
5. Таненбаум Э. Компьютерные сети. 4-е изд. – СПб: Питер, 2003. – 992 с. (Серия «Классика компьютерных наук»)
6. Таран В.В. К актуальности проблемы выбора стековых протоколов и их применение при IP-трансляции // Материалы XIV международной научно-практической конференции «Наука в современном информационном обществе» [Технические науки]. Т. 3. – North Charleston, USA, 2018. – С. 70-79.
7. Таран В.В. Современные подходы к оценке развития информационно-коммуникационных технологий и основные направления их совершенствования // Научно-техническая информация. Сер. 1. – 2014. – № 9. – С. 9-14; Taran V.V. Modern Approaches to the Assessment of Information and Communication Technologies and the Main Areas of Improvement // Scientific and Technical Information Processing. – 2014. – Vol. 41, №3. – P. 201-205.
8. Таран В.В. Информационно-коммуникационные технологии и их социально-экономическое и культурологическое влияние на инновационно-ориентированное развитие // Информационные технологии. – 2015. – Т. 21, №3. – С. 236-240.
9. Таран В.В. Интернет – как самосовершенствующаяся система (промежуточный этап на пути к искусственному интеллекту) // Вестник Университета РАО. – 2015. – №5. – С. 56-58.
10. Ashby W.R. An Introduction to Cybernetics. – London: Chapman and Hall (U.K.), 1956. – ix, 295 p.
11. Licklider J.C.R. IRE Transactions on Human Factors in Electronics. – 1960. – Vol. HFE-1. – P. 4-11. – URL:www.groups.csail.mit.edu/medg/people/psz/Licklider.html (дата обращения: 19. 02. 2019)
12. Christopher B. Jones Frail and Feeble Mind: Challenges to Emerging Global Consciousness / Journal of Futures Studies, May 2006. – №10 (4). – P. 5-14.
13. Engelbart D.C., 1962, Augmenting Human Intellect: A Conceptual Framework. Summary Report AFOSR-3223, under Contract AF 49 (638) – 1024, SRI Project 3578, for Air Force Office of Scientific Research, Stanford Research Institute, Menlo Park, Ca., October 1962. (URL: www.princeton.edu/~hos/h598/augmenting.pdf) (дата обращения к ресурсу: 22.01.2019).
14. Heylighen F.P. Conceptions of a Global Brain: An Historical Review Evolution: Cosmic, Biological, and Social / Technological Forecasting and Social Change, 2011 – P. 274-289
15. Vidal J.J. Toward direct brain-computer communication // Annual Review of Biophysics and Bioengineering Annual Reviews. – 1973. – Vol. 2. – P. 157-180.

ПРИЛОЖЕНИЕ

1) **WWW** (World Wide WEB) – Всемирная паутина характеризуется как система, основанная на программных решениях, позволяющих связывать различные архивы, файлы, электронные страницы и другие ресурсы, хранящиеся на серверах сети Интернет. Может упоминаться как WEB (Паутина).

2) **WEB-1.0** – технологический временной этап развития компьютерных технологий, соответствующий общей концепции World Wide WEB.

3) **WEB-2.0** – технологический временной этап развития компьютерных технологий, соответствующий общей концепции World Wide WEB.

4) **WEB-3.0** – технологический временной этап развития компьютерных технологий, соответствующий общей концепции World Wide WEB.

5) **HTML** (**H**yper **T**ext **M**arkup **L**anguage) Язык гипертекстовой разметки. Технический стандарт, осуществляющий гипертекстовые связи между различными данными в Интернете.

6) **HTTP** (**H**yper **T**ext **T**ransfer **P**rotocol) – широко распространенный в Интернете протокол прикладного уровня, обеспечивающий циркуляцию и передачу гипертекстовых связей по технологии клиент-сервер.

7) **NORAD** (North American Aerospace Defense Command) – система аэрокосмической обороны Соединенных Штатов Америки и Канады в задачи которой входит контроль воздушного пространства Северной Америки от несанкционированного проникновения противника.

8) **MILNET** (Military Network) (в историческом смысле) – Военная компьютерная сеть входящая в состав ARPANET, и в последствии разделенная с ARPANET служащая коммутационным звеном системы передачи информации военного назначения Министерства обороны Соединенных Штатов Америки.

9) **Enquire** – компьютерная программа, основанная на принципе семантического распределения данных. Принцип данной программы во многом схож с WWW.

10) **HTML** (версия 2.0) – спецификация языка гипертекстовой разметки. Вышла в свет 25 ноября 1995 г. Согласно предписанию (Request for Comments) RFC-1867 имела ограниченный набор тегов. Фактически первая официальная рабочая версия языка. В мае 1996 г. модернизирована согласно предписанию RFC-1942. В августе 1996 г. представлена версия RFC-1980. В январе 1997 г. версия интернационализирована согласно RFC-2070. Все необходимые предписания получили название HTML-3.0.

11) **CSS** (Cascading Style Sheets) – технический язык, служащий для описания и создания динамического дизайна различных WEB-проектов. Применяется как дополнение к HTML.

12) **CSS2** – модификация технологии каскадных таблиц стилей. Техническое описание доступно по адресу https://www.w3.org/TR/2008/REC-CSS2-20080411/ (дата обращения к ресурсу: 22.01.2019).

13) **CSS2.1** – модификация технологии каскадных таблиц стилей. Техническое описание доступно по адресу https://www.w3.org/TR/CSS21/ (дата обращения к ресурсу: 22.01.2019).

14) **JavaScript** – мультипарадигмальный язык программирования широкого профиля, часто используемый в Интернете. В интернет-среде может использоваться для наладки различных WEB-приложений, взаимодействующих с WEB-браузерами, либо как сценарный язык, описывающий конкретные действия на WEB-странице. Принадлежит корпорации Oracle Corporation USA.

15) **Python** – широкопрофильный аспектно-ориентированный язык программирования высокого уровня. Благодаря своей библиотеке может быть использован для различных технических нужд в сети Интернет, часто применяется для оптимизации приложений. Принадлежит некоммерческой организации Python Software Foundation.

16) **GO** (Golang) – язык программирования, разработанный в стенах корпорации Google, поддерживающий параллельные вычисления. Применяется для проектирования приложений, а также для упорядочивания данных в сети Интернет.

17) **SQL** (Structured Query Language) – язык структурированных запросов. Мультипарадигмальный язык – применяется для проектирования, контроля и управления данными в базах данных. В Интернете на данный момент широко распространен диалект этого языка SQL:2008.

18) **Ruby** (Rubinus) – язык программирования назван в честь минерала рубин. Мультипарадигмальный язык программирования высокого уровня. Как и GO, поддерживает параллельные вычисления, что делает этот язык незаменимым при проектировании приложений. В некоторых аспектах имеет схожесть с языком Python.

19) **PHP/HP** (Hypertext Preprocessor) – мультипарадигмальный сценарный язык. Изначально создавался как язык для создания персональных WEB-страниц **P**ersonal **H**ome **P**age Tools. Популярность этого языка объясняется тем, что его мобильность позволила различным технологическим площадкам и провайдерам на серверной основе в сжатые сроки создавать WEB-программы. Обладает внушительным набором технических средств, позволяющих эффективно проектировать WEB-сайты.

20) **SSI** (Server Side Includes) – включения на стороне сервера. Язык, применяющийся для стыкования отдельных фрагментов WEB-страниц для более быстрого их отображения на стороне клиента. Отображается как единый документ, имеющий объединенный код HTML. Может взаимодействовать с общим интерфейсом шлюза (CGI) для вызова внешних ресурсов.

21) **ISAPI** (**I**nternet **S**erver **A**pplication **P**rogramming **I**nterface) – интерфейс прикладного программирования интернет-сервера, сервер разработанный для программирования приложений. Приложения ISAPI, как правило, являются библиотеками DLL (**D**ynamic **L**ink **L**ibrary).

22) **CGI** (Common Gateway Interface) – общий интерфейс шлюза используется для связи с внешними объектами, взаимодействуя с WEB-сервером и обслуживающим его программным обеспечением.

23) **ISP** (Internet Service Provider) – организация, занимающаяся осуществлением предоставления доступа к сети Интернет.

24) **ASP, MS/ASP** (Active Server Pages) – техническое решение, разработанное корпорацией Microsoft, предназначенное для управления с помощью внедряемых операторов структурой WEB-страниц. В целях адаптации к технической структуре WEB-страниц может использоваться сценарный языки MS Visual Basic Script Edition.

25) **WAP** (Wireless Application Protocol) – протокол беспроводной передачи данных. Протокол предназначен для установления беспроводной связи мобильных периферийных устройств с сетью Интернет. Интегрируется с GSM-технологией (Global System Mobile), что позволяет мобильным устройствам беспрепятственно и оперативно получать доступ к интернет-ресурсам. WAP разрабатывался в соответствии с технологической концепцией WWW, поэтому хорошо адаптирован под реалии Интернета.

26) **AJAX** (**A**synchronous **Ja**vaScript and **X**ML) метод проектирования пользовательских WEB-интерфейсов базирующийся на технологии скрытого обмена данных браузера с сервером. Важным аспектом является то, что при загрузке дополнительных данных на WEB-страницу, структура WEB-страницы изменяется, а структура разметки нет. Данный аспект обеспечивает более быструю загрузку страницы (даже если физически скорость Интернета – низкая) без перегрузки дополнительными данными пропускного канала.

27) **WIKI** – разновидность технологии гипертекстовой разметки, применяемая на сайтах библиотечного и энциклопедического типа (сайты, на которых необходимо обеспечить систематизацию текстовых блоков). Отличительная особенность – это возможность редактирования и изменения текстовых массивов при помощи элементов управления сайтом.

28) **RSS** (Rich Site Summary) – формат расширяемого языка разметки (Extensible Markup Language) для описания информации с различных WEB-ресурсов и представления их в единой WEB-оболочке.

29) **DAML** (**D**ARPA **A**gent **M**arkup **L**anguage) – язык разметки, использующийся управлением перспективных исследовательских проектов Министерства обороны Соединенных Штатов Америки (**D**efense **A**dvanced **R**esearch **P**rojects **A**gency).

30) **OWL** (Ontology WEB Language) – язык, предназначенный для описания онтологий семантической паутины.

31) **RDF** (Resource Description Framework) – среда описания ресурса. Технологическая модель для управления и представления метаданных, предложенная консорциумом глобальной Всемирной паутины.

32) **OIL** (Ontology Inference Layer или Ontology Interchange Language) – специализированная инфраструктура, предназначенная для обмена онтологиями и используемая в семантической сети, может быть совместима с RDFS. Принадлежит проекту IST OntoKnowledge.

Материал поступил в редакцию 27.02.19.

Сведения об авторе

ТАРАН Василий Васильевич – кандидат культурологии, соискатель ученой степени доктора технических наук ВИНИТИ РАН, Москва
e-mail: allscience@lenta.ru

ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ

УДК 002.2–047.44:004.85

В.О. Толчеев

Исследование и анализ предметной области «глубокое обучение»

Анализируется активно развивающееся научное направление – глубокое обучение (Deep Learning), – сформировавшееся в рамках работы по системам искусственного интеллекта. С помощью наукометрических методов оценены темпы роста публикаций в ведущих странах и уровень международного сотрудничества между государствами. Исследована терминологическая структура предметной области, выявлены наиболее перспективные тематики исследований. Проведено сопоставление наукометрических показателей глубокого обучения и еще одной стремительно растущей области – квантовые технологии. Сделан вывод о более быстром росте публикационной активности по Deep Learning. Отмечено, что в обеих областях по числу статей сформировалось устойчивое лидерство США и КНР. Наукометрический анализ показал достаточно низкий уровень публикационной активности российских ученых по тематикам глубокого обучения и их слабую вовлеченность в международное сотрудничество.

Ключевые слова: *глубокое обучение, глубокие нейросети, фактографические и мультимедийные данные, наукометрический анализ, библиографическая база данных Web of Science, публикационная активность, анализ дескрипторов*

ВВЕДЕНИЕ

В системах искусственного интеллекта сформировалось новое перспективное направление, позволяющее в ряде предметных областей существенно повысить качество обработки и анализа данных (в частности, увеличить точность и быстродействие), а наилучшие результаты достигнуты при распознавании образов и в задачах обработки естественного языка и речи, а также в задачах анализа биомедицинских данных. Это новое направление имеет несколько синонимичных названий: *глубокое обучение* (*глубинное обучение, Deep Learning – DL*) или *глубокие искусственные нейронные сети* (*глубокие нейросети, Deep Networks – DN*) [1–5]. Остальные названия (например, *Learning Deep Memories* или *Hierarchical Learning*) употребляются в специализированной литературе значительно реже. В настоящей работе для обозначения данной предметной области мы будем использовать термин глубокое обучение (*Deep Learning* или *DL*) .

Глубокое обучение представляет собой раздел машинного обучения (*Machine Learning – ML*). Характерной особенностью, отличающей *DL* от *ML*, является возможность самостоятельного выделения информативных признаков для решения конкретных задач. Автоматизация процедуры отбора переменных позволяет исключить исследователя из этого слабоформализованного процесса. В случае больших объемов информации составление признакового пространства в ходе неконтролируемого обучения (обучение без учителя, *Unsupervised Learning*) или обучения с частичным привлечением учителя (*Semi-supervised Learning*) способно существенно снизить трудоемкость подготовки данных и повысить результирующую точность по сравнению с классическими (экспертно-ориентированными) методами построения системы признаков. Это значительно увеличивает универсальность глубокого обучения, позволяя создавать «типовые» архитектуры и адаптировать их к похожим задачам в пределах одной предметной области (*Transfer Learning*). Платой за подобную автоматизацию является слабая интерпретируемость получаемых переменных-предикторов.

Анализ имеющихся тематических публикаций (научных трудов, отчетов и прогнозов консалтинговых агентств, учебных курсов, научно-популярных материалов) показывает, что в области глубокого обучения в настоящее время происходит исследовательский бум [6–9]. Об этом свидетельствует быстрый рост публикационной активности, создание многочисленных открытых библиотек, которые предоставляют широкому кругу специалистов доступ к уже разработанным и апробированным программно-алгоритмическим средствам, а также выпуск успешных коммерческих продуктов, использующих технологии *Deep*

Learning. Анализ финансирования, получаемого высокотехнологичными стартапами и профильными подразделениями крупнейших хайтек-корпораций, подтверждает активное формирование направления *DL*.

Несмотря на разноаспектные зарубежные и отечественные обзоры, отражающие становление и развитие *DL* [2, 10–12], среди специалистов нет единого мнения, когда эти технологии перешли в стадию ускоренного развития и стали научным направлением. В качестве «знакового» события часто рассматривается работа, опубликованная в 2006 г. Дж. Хинтоном и Р. Салахутдиновым, где авторы разработали новую модификацию глубоких нейронных сетей, в которой использовалась комбинация обучения с учителем и без учителя [13]. Однако на сайте компании *Oracle* «эра *DL*» начинается после 2010 г. [14]. В пользу такой хронологии имеются веские аргументы: в 2011 г. прошел конкурс *ImageNet* по распознаванию объектов на фотографиях, и группой профессора Дж.Хинтона были показаны точностные преимущества глубокого обучения над другими методами [4, 12, 15].

При всех очевидных успехах глубокого обучения и его важной роли при разработке передовых решений и новых бизнес-приложений остается открытым вопрос: не является ли *Deep Learning* еще одним «пузырем», которых так много появлялось за время развития искусственного интеллекта? Скептики, считающие тематику *DL* раздутой, указывают на необходимость при анализе конкретных выборок корректно определять целесообразность использования тех или иных подходов [16] и отмечают, что *DL* является, несомненно, лидирующей технологией при наличии колоссальных массивов (датасетов) фактографических и мультимедийных данных (*Big Data*) и доступности мощных вычислительных ресурсов. Вместе с тем в задачах с небольшими объемами данных преимущества глубокого обучения нивелируются и подходы интеллектуального анализа данных (*Data and Text Mining*) становятся вполне конкурентоспособными (в частности, построение моделей с помощью метода опорных векторов или комитетов решающих правил на основе деревьев решений). При этом подчеркивается весьма важное преимущество классических алгоритмов – они в своем большинстве дают хорошо интерпретируемые результаты, которые понимаются экспертами, разработчиками, заказчиками и пользователями. Применение же технологий *Deep Learning*, включающих многочисленные нелинейные преобразования, приводит к построению «черного ящика», из которого каким-то «неведомым для всех образом» извлекаются (чаще всего) правильные решения. За «кадром» остаются теоретические принципы и логические обоснования, обеспечивающие итоговые оценки. Такая ситуация в *DL* видимо сохранится достаточно долго, пока в смежной предметной области нейронаук не будет разработана комплексная теория функционирования человеческого мозга и построены нейроморфные компьютеры [17].

Все эти суждения о наличии исследовательского бума в области глубокого обучения и сроках его возникновения, опираются прежде всего на экспертные выводы и заключения. Имеющиеся оценки достаточно противоречивы и не всегда хорошо согласуются между собой. В связи с этим представляется важным провести анализ исследуемой предметной области с «другой точки зрения», используя более объективные методы, основанные на наукометрическом подходе. Несмотря на наличие большого числа обзоров, прогнозов, профильных статей и докладов, нам не удалось обнаружить какие-либо опубликованные зарубежные и отечественные исследования развития технологий *DL* с помощью инструментов наукометрии. Так по запросу "*Deep Learning Scientometric*" в БД *Google Scholar* и БД *Microsoft Academic Search* не нашлось релевантных статей (нулевая выдача также была получена в русскоязычной электронной библиотеке *eLibrary* (eLibrary.ru) по запросу *«Глубокое обучение Наукометрия»*). Вместе с тем за последние десятилетия изысканий в данной предметной области накоплен достаточно большой массив тематических публикаций, доступных для анализа и систематизации, что делает актуальным проведение комплексного изучения имеющихся документов средствами наукометрии.

МЕТОДИКА ИССЛЕДОВАНИЯ

Наукометрический анализ позволяет решать широкий круг вопросов, в частности, «измерять» активность исследовательской деятельности в предметной области и выявлять формирующиеся тенденции. В настоящей работые с помощью наукометрического подхода решаются следующие основные задачи:

1) оценка интенсивности научных исследований по направлению *DL*;

2) ранжирование стран по числу публикаций, выявление лидеров (среди стран, научных журналов и конференций);

3) выявление наиболее активно развивающихся тематик (внутри глубокого обучения);

4) анализ международного сотрудничества в области *Deep Learning*.

В любом наукометрическом исследовании имеются ключевые моменты, способные существенно повлиять на получаемые результаты:

- формализация (определение границ) предметной области и построение поискового запроса,
- выбор и обоснование базы данных для исследований,
- период времени, за который проводится анализ.

Рассмотрим эти моменты более подробно.

В настоящее время к наиболее авторитетным международным реферативным (библиографическим) базам данных относятся *Web of Science, Scopus, Google Scholar*. В работе [18] нами был обоснован выбор БД *Web of Science* для проведения наукометрических исследований в области квантовых технологий. В частности, отмечалось, что БД *Web of Science* ориентирована на индексацию статей из журналов и материалов конференций, издаваемых в США. Именно американцы, начиная с самого зарождения искусственного интеллекта, лидируют по числу публикаций и разработок по большинству перспективных направлений данной предметной области. Несмотря на то, что в развитие технологий *DL* большой вклад внес интернациональный коллектив ученых, тем не менее в последнее время основные прорывные исследования проводятся преимущественно в крупнейших американских компаниях (*Google*, *Facebook*, *Microsoft*, *IBM*) и университетах.

При выборе базы данных необходимо учитывать «китайский фактор». Существенные государственные инвестиции Китая, значительный кадровый потенциал, планомерная реализация долгосрочной национальной стратегии по достижению доминирующих позиций в искусственном интеллекте к 2030 г. позволяют рассматривать КНР в качестве второго «финалиста» в борьбе за первое место в области искусственного интеллекта [8]. Как представляется, БД *Web of Science* достаточно объективно оценивает китайские публикации, проводя разумную фильтрацию и отсекая многочисленные второстепенные работы, в которых дублируются известные результаты.

При выборе БД *Web of Science* вряд ли стоит опасаться какой-либо дискриминации российских журналов по тематике *DL*. Предварительный анализ профильных российских публикаций, проведенный нами с помощью научной электронной библиотеки *eLibrary*, показал наличие крайне незначительного числа научных трудов по запросу «*Глубокое обучение*» (было получено 157 статей, поиск осуществлялся по всем полям библиографических описаний, обращение к *eLibrary* проводилось 10 января 2019 г.).

Еще один аргумент в пользу БД *Web of Science* заключается в том, что появляется возможность сравнить результаты наукометрического анализа по глубокому обучению с ранее проведенным исследованием по квантовым технологиям. Сопоставление выявленных закономерностей может быть полезно для изучения других быстроразвивающихся предметных областей.

Следует отметить, что БД *Web of Science* не является полнотекстовой и содержит только библиографические описания научных работ, представляющие собой краткие рефераты исходных документов и включающие название, аннотацию, ключевые слова, фамилию и имя автора (авторов), место работы, издание, опубликовавшее материал, и некоторые другие служебные поля, которые достаточно редко используются в наукометрическом анализе.

При составлении запроса к БД *Web of Science* недостаточно использовать только одно ключевое слово «*Deep**». Такой подход приводит к появлению нерелевантных статей по другим (чаще всего очень далеким) тематикам. Предварительное изучение выдачи по различным запросам, проведенное в широкодоступной БД *Google Scholar*, показало, что наиболее предпочтительным вариантом является объединение в запросе двух понятий: «*Deep Learning*» и «*Deep Networks*». Изучение выдачи в БД *Google Scholar* позволило сделать еще одно важное наблюдение – в ряде публикаций термины, использованные в запросе («*Deep learning*» и «*Deep Networks*»), не следуют непосредственно друг за другом, а образуют устойчивые словосочетания из 3-5 слов (например, «Deep transfer *L*earning», «*Deep* convolutional *Networks*», «*D*eep recurrent neural Networks», «*D*eep sparse rectifier neural Networks»).

В связи с этим нами было принято решение сформировать сложный запрос к БД *Web of Science* в виде: «*Deep* Near/3 (*Learning* OR *Networks*)». Согласно этому запросу в выдачу будут включаться все документы БД *Web of Science*, которые в своих библиографических полях содержат термин «*Deep*» и термины («*Learning*» или «*Networks*), отстоящие от первого термина не более чем на три других слова, не вошедших в запрос (обращение к БД *Web of Science* проводилось 21 января 2019 г.).

Таким образом, проводимое исследование охватывает девятнадцатилетний период с 2000 по 2018 гг., выборка содержит англоязычные библиографические описания, в качестве полей для анализа использованы название, аннотация и ключевые слова; место публикации (название журнала, конференции и т.п.); авторы, их место работы и страна (страны), которые они указали в статье; год публикации.

КОЛИЧЕСТВЕННЫЙ АНАЛИЗ ПУБЛИКАЦИОННОЙ АКТИВНОСТИ

Исходные данные, полученные по сформированному запросу за 2000-2018 гг., составили 27699 англоязычных библиографический описаний (ограничения на тип документа не вводились, т.е. включались научные статьи, обзоры, прогнозы, доклады, монографии, учебники и т.п.). Все дальнейшие расчеты и визуализация результатов выполнены с помощью программы *Statistica*.

Количество публикаций по годам для первых десяти стран, ученые которых издали наибольшее число профильных научных трудов, приведено в табл. 1, а Россия сюда включена для сравнительного анализа и определения значимости российских исследований в области *Deep Learning* в общемировом массиве тематической информации.

Страны в табл. 1 упорядочены по общему числу публикаций за 2000-2018 гг. Отметим, что данные за 2018 г., полученные на момент обращения к БД *Web of Science* в январе 2019 г., являются неполными: практика показывает, что индексация статей закончится в мае-июне. Эту особенность выборки далее будем учитывать при формулировании выводов.

В наукометрии для определения скорости развития предметной области часто используется показатель «удвоение объема информации», т.е. за какой период времени годовой объем профильных документов увеличивается вдвое [19]. Скорость развития технологий *Deep Learning* чрезвычайно высока и, начиная с 2014 г., массив публикаций по проблематике ежегодно практически удваивается. Как представляется, именно этот год будет правильно принять за точку стремительного роста рассматриваемого направления. Учитывая время прохождения статей через редколлегии журналов, можно с высокой степенью уверенности предположить, что конкурсы *ImageNet* (2011-2015 гг., http://image-net.org/about-stats) стали «спусковым крючком», который запустил лавинообразный поток публикаций и вывел *DL* в число самых быстрорастущих научных направлений. Этому также способствовали и другие значимые события: в 2014 г. в открытый доступ была выложена библиотека *Caffe*, включающая большое число предварительно обученных моделей [20], а в 2015 г. – библиотека *TensorFlow* компании *Google* [21].

Приведенные на рис. 1 ретроспективные зависимости показывают существенные различия в показателях публикационной активности в рассматриваемых 11 странах, начиная с 2012 года, а с 2014 г. наблюдается переход к нелинейному росту. Этот ка-

чественный скачок хорошо иллюстрируется диаграммами размаха Тьюки («ящик с усами»), показывающими по годам медианное значение и размах (рис. 2). К 2014 г. во многом благодаря ресурсу *Kaggle* (www.kaggle.com) в области *Deep Learning* было создано исследовательское сообщество и получила распространение практика краудсорсинга, позволившая сформировать и разметить огромные выборки усилиями заинтересованных пользователей Интернета.

В 2018 г. число публикаций двух лидеров – КНР и США – практически в 10 раз превосходило научные труды, изданные в Индии или Франции, находящихся на 9 и 10 месте. При этом Россия по числу профильных документов, проиндексированных в БД *Web of Science*, отстает от Франции более чем в 2 раза. Таким образом, необходимо констатировать существенные различия публикационной активности в странах, представленных в табл. 1, и сделать вывод о значительном отрыве КНР и США от остальных государств.

Таблица 1

Количество публикаций в области *Deep Learning* в ведущих странах

Страна	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
США	32	34	32	39	40	40	45	51	62	58	72	79	116	194	287	679	1158	2255	2567
КНР	1	0	3	7	4	5	4	14	20	22	23	24	42	100	300	767	1390	2715	3189
Великобритания	10	15	7	7	20	7	12	16	21	19	29	19	29	54	60	158	276	523	599
Япония	2	0	2	6	2	0	0	3	4	3	2	3	10	22	57	129	250	433	410
Ю.Корея	0	1	0	0	0	2	1	3	1	0	1	6	2	3	18	75	179	404	620
ФРГ	1	2	2	2	7	7	13	4	7	7	7	9	14	26	41	89	194	387	418
Канада	2	2	3	6	4	3	5	4	12	9	13	14	25	39	53	99	149	322	320
Австралия	2	1	7	5	9	6	13	4	13	17	16	22	23	30	41	114	148	302	292
Индия	0	0	1	0	0	1	2	1	3	3	3	0	9	6	25	70	171	404	327
Франция	3	3	7	2	3	2	2	4	2	7	9	13	5	15	21	76	114	231	230
Россия	0	0	1	0	1	3	0	0	0	1	1	0	0	0	6	12	42	81	102

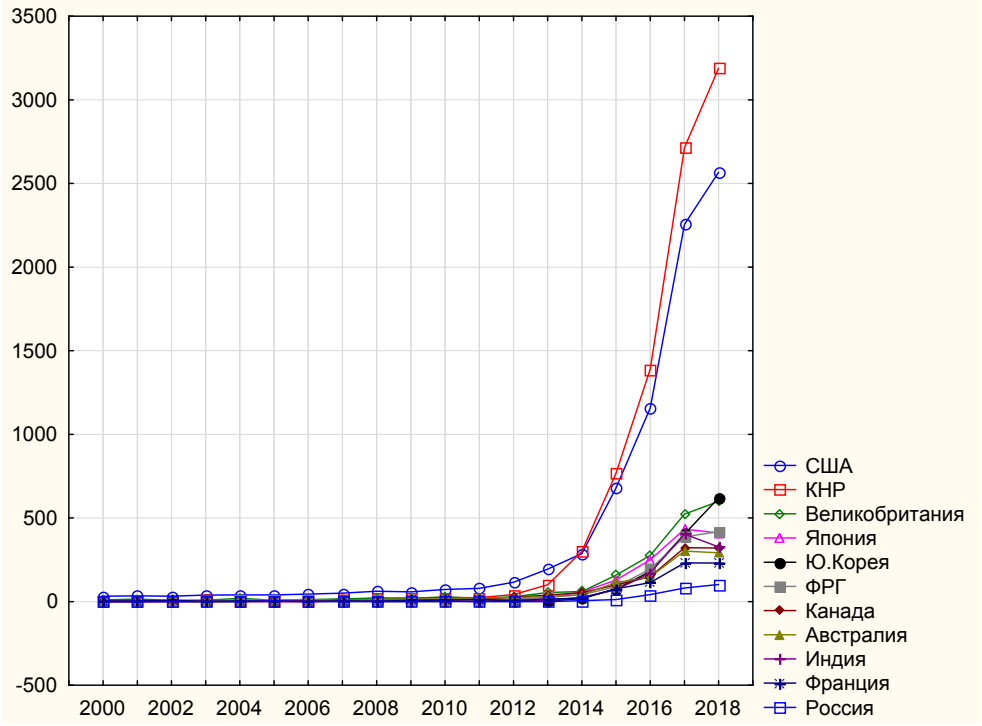


Рис.1. Количество публикаций в ведущих странах за 2000-2018 гг.

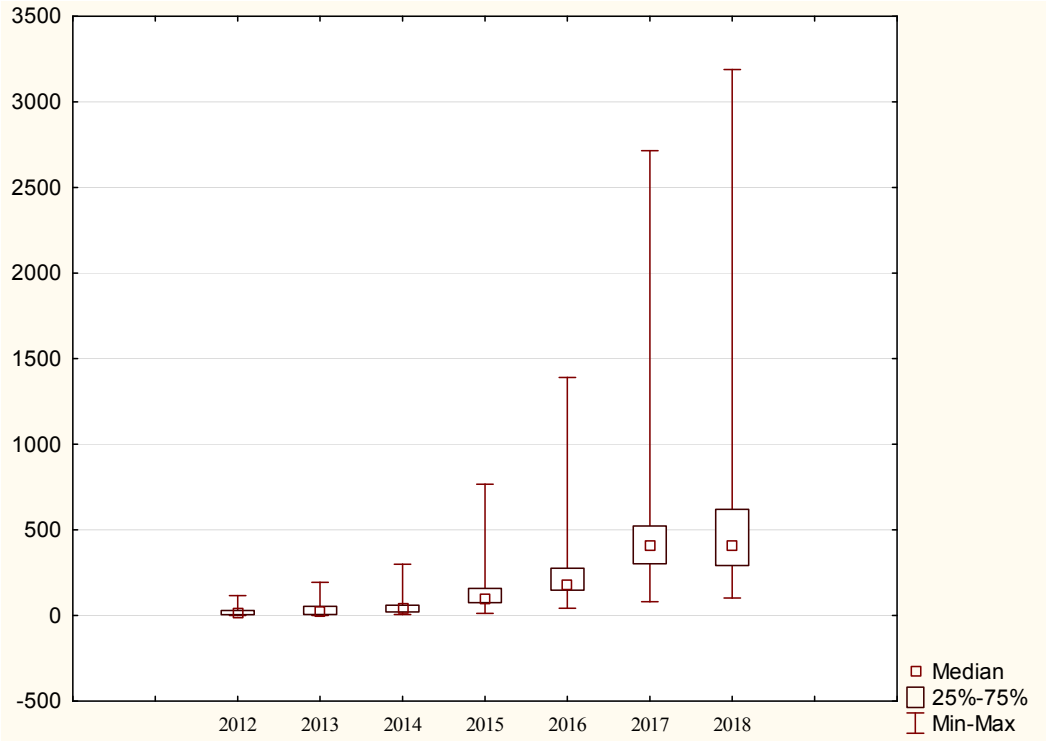


Рис.2. Диаграмма размаха по годам (2012-2018 гг.)

Представляет интерес математическая аппроксимация зависимостей, приведенных на рис. 1 для США и КНР, с помощью полиномов. Аппроксимация первоначально строится на интервале 2012-2018 гг. Затем, с учетом неполноты данных за 2018 год, зависимость определяется на интервале 2012-2017 гг., после чего делается прогноз на 2018 год. С помощью расчетов в программе *Statistica* получены следующие результаты (в приводимых выражениях x – временные отчеты).

1. Аппроксимация полиномом второго порядка за период 2012-2018 гг.

$$США = 182\text{-}165{,}1667x+75{,}7619x^2. \quad (1)$$

$$КНР = 108{,}4286\text{-}200{,}631x+95{,}4405x^2 \quad (2)$$

2. Аппроксимация полиномом второго порядка за период 2012-2017 гг.

$$США = 490{,}1\text{-}430{,}475x+118{,}5536x^2. \quad (3)$$

Рассчитываем прогноз на 2018 г. и получаем 3285 документов (на 718 больше, чем указано в табл. 1 и имеется на текущий момент в БД *Web of Science*).

$$КНР = 453{,}3\text{-}497{,}6036x+143{,}3393x^2. \quad (4)$$

Прогноз на 2018 г. – 3993 документа (на 804 больше, чем указано в табл. 1).

Таким образом можно спрогнозировать, что по окончанию периода индексирования статей в БД *Web of Science* общее количество публикаций КНР достигнет четырех тысяч и отрыв от США по числу печатных трудов продолжит увеличиваться. Отметим также, что полиноминальный рост числа статей – важный индикатор стадии подъема в предметной области и свидетельствует об увеличении количества ученых, специализирующихся в *Deep Learning*, что (косвенным образом) подтверждает значительное финансирование предметной области (т.к. большинство опубликованных результатов получено при выполнении грантов).

Ранее отмечалось соперничество за лидирующие позиции в технологиях *DL* между США и КНР. На это можно посмотреть более широко – как на борьбу двух центров научно-технологического развития: «западного», ориентированного на США, и «азиатского», в котором каждая страна развивает *Deep Learning*, опираясь на собственные силы. На основе данных табл. 1 можно увидеть, что в 2018 г. эти два центра имели практически одинаковое количество публикаций (США, Великобритания, Канада, Австралия, ФРГ, Франция – 4426; Китай, Южная Корея, Япония, Индия – 4546). Вместе с тем, учитывая стремительный рост киличества китайских и южнокорейских статей, можно сделать вывод о перемещении публикационной активности по глубокому обучению в азиатский регион (см. табл. 1 и рис. 1).

Проанализируем число научных трудов российских ученых по тематике *DL*, проиндексированных за рассматриваемый период в БД *Web of Science*. Несмотря на большое число статей с фамилиями, принадлежащими, как представляется, выходцам из стран СНГ, было выявлено очень немного работ, у авторов которых в библиографических описаниях в позиции «место работы» указывалась *Russia*. Отметим три документа (в том числе написанные россия-

нами в соавторстве с иностранными специалистами), существенно опережающие другие по показателю цитирования:

1) *Neural Codes for Image Retrieval*, **авторы:** Babenko A., Slesarev A., Chigorin A., Lempitsky V. (https://arxiv.org/pdf/1404.1777.pdf).

2) *Domain-Adversarial Training of Neural Networks*, **авторы:** Ganin Y., Ustinova E., Ajakan H., Germain P., Larochelle H., Laviolette F., Marchand M., Lempitsky V. (https://arxiv.org/abs/1505.07818).

3) *Aggregating Deep Convolutional Feature for Image Retrieval*, **авторы**: Babenko A., Lempitsky V. (https://arxiv.org/pdf/1510.07493.pdf).

Как место работы у российских авторов наиболее часто указывались Сколковский институт науки и технологий – Сколтех (*Skolkovo Institute of Science and Technology – Skoltech*), Яндекс (*Yandex*), МФТИ (*Moscow Institute of Physics and Technology*). В. Лемпицкий (Сколтех) обладает самым высоким индексом цитирования среди специалистов из России – 12.

Обобщим выводы, которые следуют из анализа данных табл. 1 и зависимостей на рис. 1 и 2.

1. На рубеже 2012-2013 гг. Китай вышел на второе место по количеству публикаций в рассматриваемой области *Deep Learning*, а в период 2014-2015 гг. уверенно занял первое место, которое удерживает и в настоящее время. Можно констатировать «невероятный рывок» КНР, совершенный в последнее десятилетие (в проводимом анализе не рассматриваются вопросы «качества» публикаций, в частности, их цитируемость).

2. США занимает второе место с огромным отрывом от остальных стран. Однако согласно построенным аппроксимациям в 2018 г. отставание от КНР увеличится и скорее всего эта тенденция сохранится в ближайшей перспективе.

3. За третье место по числу публикаций в области *DL* соперничают Великобритания и Южная Корея. В 2018 г. (по количеству уже проиндексированных статей) корейские ученые впервые вышли на третью позицию и пока опережают британских специалистов на 21 документ (эти результаты могут измениться после обработки всего массива статей в БД *Web of Science*). Вместе с тем за 2014-2018 гг. Южная Корея показала устойчивое поступательное развитие.

4. В 2018 г. наметилось некоторое замедление роста числа работ, издаваемых в Японии, которая конкурировала в период 2014-2016 гг. с Великобританией за третье место.

5. После пика публикационной активности в 2017 г. некоторый спад отмечен в Индии и Австралии. Практически без изменений остались показатели Канады и Франции. Незначительное увеличение печатных работ по *Deep Learning* наблюдалось в ФРГ.

6. Россия слабо представлена в документальном потоке по тематикам *DL*. Вместе с тем, начиная с 2014 г., увеличивается количество российских научных трудов, индексируемых в БД *Web of Science*.

РАСПРЕДЕЛЕНИЕ ПУБЛИКАЦИЙ ПО ПОДТЕМАМ РАЗДЕЛА ИНФОРМАТИКА

Рассмотрим распределение документального потока по более узким тематикам (согласно рубрикатору БД *Web of Science*). Анализ этих подтем в документах выборки показывает междисциплинарность исследуемого направления. Публикации разнесены по таким рубрикам как «Информатика», «Телекоммуникации», «Роботы», «Биомедицина», «Вычислительная биология», «Нейронауки», «Оптика», «Электротехника и электроника», что и объясняет существенное рассеивание документального потока по журналам разной тематической направленности. В 2018 г. наибольшее число статей по рассматриваемой проблематике выпустили следующие журналы: «IEEE Access» (332 публикации), «Neurocomputing» (178), «Sensors» (149), «Medical Physics» (113), «IEEE Transactions on Image Processing» (85), «Pattern Recognition» (81). Таким образом, в области глубокого обучения, как и в большинстве других современных технологий, невозможно выделить узкоспециализированные издания, которые можно назвать *профильными* или *ядерными* журналам, содержащими соответственно не менее 30% или 70% статей по выбранной тематике [22].

Для более детального анализа были выбраны подтемы, попадающие в рубрику «Информатика» (*Computer Science*). В первой двадцатке рубрик, по которым в БД *Web of Science* проиндексировано наибольшее число статей, имеется шесть подтем ***Computer Science*** (табл. 2):

1) *Artificial Intelligence* – Искусственный интеллект (ИИ);

2) *Information Systems* – Информационные системы (ИнфоСис);

3) *Theory & Methods* – Теория и методы (ТиМ);

4) *Interdisciplinary Application* – Междисциплинарные приложения (МДП);

5) *Software Engineering* – Программная инженерия (ПИ);

6) *Hardware & Architecture* – Аппаратные средства и архитектура (АСиА).

Таблица 2

Количество публикаций по подтемам в разделе *Computer Science**

№	ИИ	ИнфоСис	ТиМ	МДП	ПИ	АСиА
1	7934	3363	4153	1928	1349	1143
2	1621	1073	786	517	474	270
3	1287	915	591	409	395	208

* *Примечание:* первая строка содержит число публикаций по тематикам во всех странах за период 2000-2018 гг., вторая – общее число публикаций в мире в 2018 г., третья – число публикаций первых десяти стран в 2018 г. (см. табл. 1).

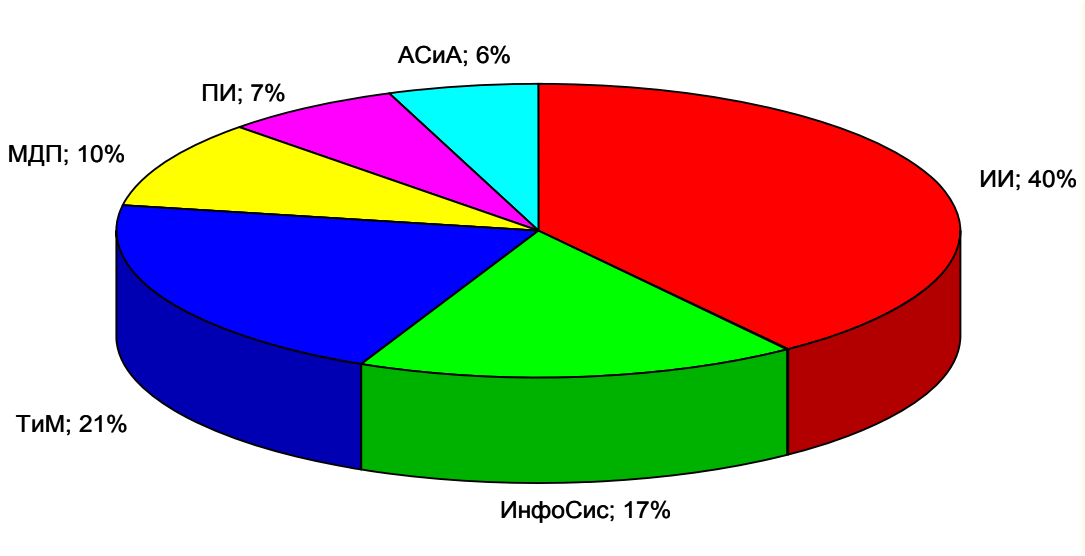


Рис.3. Процентное распределение статей по подтемам в области *Computer Science* (2000-2018 гг.)

В области информатики по тематикам *DL* в странах-лидерах издается от 75% до 85% всех статей. Важно отметить, что во всех подтемах, указанных в табл. 2, по общему количеству публикаций (в 2018 г.) лидирует КНР: по искусственному интеллекту – 673 публикации (у США – 333), по информационным системам – 554 (206), по теории и методам – 281 (154), по программной инженерии – 210 (109), по междисциплинарным приложениям – 163 (160), по аппаратным средствам и архитектуре – 89 (77). В четырех подтемах 3-ю позицию занимает Южная Корея, в двух – Великобритания.

На рис. 3 проведена визуализация для первой строки табл. 2 и показано процентное распределение публикаций за 2000-2018 гг. по указанным подтемам.

АНАЛИЗ КЛЮЧЕВЫХ СЛОВ

Для выявления наиболее активно развивающихся направлений в области *Deep Learning* проведем анализ ключевых слов. Чтобы отсечь «шумовые» термины и еще не полностью сформировавшиеся понятия, проанализируем дескрипторы, которые встречаются в изучаемой выборке не реже 70 раз (учитывая размер выборки, это означает, что слово-маркер появляется примерно в каждых 400 статьях).

Википедия дает такое определение *Deep Learning*: «в первую очередь к глубинному обучению относятся следующие методы и их вариации: определённые системы обучения без учителя, такие как ограниченная машина Больцмана предварительного обучения, автокодировщик, глубокая сеть доверия, генеративно-состязательная сеть; определённые системы обучения с учителем, такие как сверточная нейронная сеть, рекуррентные нейронные сети, позволяющие обучаться на процессах во времени, рекурсивные нейронные сети, позволяющие включать обратную связь между элементами схемы и цепочками» [23].

Проверим, насколько ключевые слова, полученные из выборки, подтверждают определение глубокого обучения из Википедии. Дескрипторы, выделенные из поля библиографического описания "*Key words*", могут быть разделены на три основные тематические группы: *нейросетевые методы*, *способы обучения* и *области интенсивного прикладного применения глубокого обучения*. Проанализируем наиболее частотные слова-маркеры по группам.

К самым рейтинговым терминам, которые характеризуют ***нейросетевые методы***, относятся:

1. Сверточные нейросети – 4444 (*(Deep) Convolutional (Convolution) Neural Network (Networks, Neural-Network), (CNN) CNNS*).
2. Глубокая нейросеть доверия – 656 (*Deep Belief Network (Networks), DBN*).
3. Рекуррентные нейросети – 377 (*Recurrent Neural Network (Networks)*).
4. Автоэнкодер (автокодировщик) – 371 (*Autoencoder, Autoencoders, Auto-Encoder*).
5. Сеть с долговременной и кратковременной памятью – 235 (*Long short term memory, LSTM)*
6. Ограниченная машина Больцмана – 223 (*Restricted Boltzmann Machine (Machines)*).

Примечание: в скобках приведены различные варианты написания терминов, результирующий численный показатель получен путем суммирования частот встречаемости всех указанных дескрипторов

Представленные результаты оказались достаточно предсказуемыми. Большинство из выявленных дескрипторов совпадает с ключевыми словами, которые используются для определения предметной области в Википедии. Доминирующее положение занимают сверточные нейросети, ставшие основным инструментом решения широкого круга проблем по компьютерному зрению и распознаванию образов. Вместе с тем получено (и это, на взгляд автора, достаточно неожиданно) значительное отставание других под-

ходов от сверточной архитектуры. Прежде всего это касается рекуррентных нейросетей, успешно применяемых в задачах обработки речи и естественного языка (разрыв в численных показателях более чем на порядок).

Среди рассмотренных высокочастотных ключевых слов отсутствуют два понятия, указанные в определении Википедии: генеративно-состязательная сеть (*Generative adversarial network – GAN*) и рекурсивная нейронная сеть (*Recursive neural network – RvNN*). В случае *GAN*–сетей возможным объяснением может стать «молодой возраст» этой архитектуры [24]. С 2014 г., несмотря на высокий исследовательский интерес, соответствующие тематические статьи еще не успели проиндексироваться в БД *Web of Science* в достаточном количестве. Что касается рекурсивных сетей, то отсутствие этого понятия значительно сложнее объяснить, учитывая высокую распространенность подхода, в частности, в задачах обработки текстовых документов [25]. Вместе с тем приведенный далее анализ «областей интенсивного прикладного применения глубокого обучения» указывает на сравнительно небольшое «представительство» публикаций по проблематике интеллектуального анализа текста (*Text Mining*) в анализируемой выборке.

Отметим также, что в группу «нейросетевых методов» не вошли и другие «ожидаемые» технологии, в частности глубокие остаточные сети (*Deep residual networks – DRN*) и нейронные машины Тьюринга (*Neural Turing machines – NMT*) [26, 27].

В группу ***способы обучения*** входят следующие ключевые слова:

1. Машинное обучение – 1153 (*Machine Learning*).
2. «Передаваемое обучение» – 399 (*Transfer Learning* – автор не нашел устоявшегося перевода этого термина, поэтому дадим пояснение: данный вид обучения предусматривает передачу ранее накопленных знаний для решения новой (аналогичной) задачи).
3. Автоматическое выявление информативных признаков в ходе обучения – 270 (*Feature Learning*).
4. (Глубокое) обучение с подкреплением– 231 (*(Deep) Reinforcement Learning*).
5. Обучение без учителя – 163 (*Unsupervised Learning*).
6. Многозадачное обучение – 119 (*Multi-Task Learning* – несколько задач обучения решаются одновременно).
7. Обучение с учителем – 80 (*Supervised Learning*).
8. Активное обучение – 79 (*Active Learning*).
9. Коллективное обучение – 73 (*Ensemble Learning*).

Группа ключевых слов, отражающих ***области интенсивного прикладного применения глубокого обучения***, включает:

1. Обработка и анализ изображений – 3276 (*Images, Image Classification, Face Recognition, Computer Vision, Image Segmentation, Image, Image Retrieval, Image Processing, Pattern Recognition, Image Recognition, Video, Scene Classification, Face Detection, Object Detection, Object Recognition, Object Tracing*).
2. Распознавание и синтез речи – 1128 *((Robust) Speech Recognition, Speech Synthesis, Speech, Speech Enhancement, Speaker Verification*).
3. Прогнозирование – 675 (*Prediction, Regression*).
4. Обработка и анализ медицинских данных – 628 (*Diagnosis, Cancer, Breast Cancer, Disease*).
5. Обработка и анализ текстовых данных – 427 (*Semantic Segmentation, Sentiment Analysis, Context, Text*).
6. Выявление аномалии – 81 (*Anomaly Detection*).
7. Диагностика неисправностей – 122 (*Fault Diagnosis*).

Примечание: в скобках приводятся ключевые слова, отнесенные к конкретному направлению. Указанные численные значения являются «оптимистичными», так как нельзя исключать, что в ряде статей некоторые ключевые слова присутствуют совместно (например, *Object Recognition* и *Object Tracing*) и суммирование частоты встречаемости ключевых слов приводит к завышенным оценкам.

МЕЖДУНАРОДНОЕ СОТРУДНИЧЕСТВО

Результаты международного сотрудничества в области *DL* за 2000-2018 гг. приведены в табл. 3. На главной диагонали стоит общее количество национальных публикаций, представляющих сумму по строке табл. 1. Остальные ячейки содержат число печатных трудов, подготовленных, как минимум, учеными из двух государств, названия которых указаны в заголовках строки и столбца. Если соавторы проиндексированной работы принадлежат к нескольким государствам, то каждой стране засчитывается по одной публикации. К российским статьям отнесены научные труды, в которых в поле «место работы» хотя бы одного автора указано *Russia*.

Международное сотрудничество в области глубокого обучения концентрируется не столько вокруг лидеров, сколько между ними. Американские и китайские ученые в соавторстве подготовили больше публикаций, чем суммарное число их совместных публикаций с исследователями из других стран (количество американо-китайских статей равно 1164, число публикаций с другими странами, представленными в табл. 3, у США – 917 (сумма по 3–9 столбцам), КНР – 646 (сумма по 3–9 столбцам)). Интерес американских специалистов связан с огромными массивами данных для обучения, доступных в Китае, а также с большим кадровым потенциалом в области искусственного интеллекта. Именно по этим причинам корпорации *Microsoft*, *Google*, *Facebook* открывают и расширяют свои исследовательские центры в КНР. Китайцы, в свою очередь, активно стажируются в ведущих американских университетах, участвуют в НИОКР, а также организуют в США инновационные стартапы по разработке ИИ.

Данные табл. 3 проиллюстрированы на рис. 4. В качестве достаточно неожиданного результата отметим активное сотрудничество Китая с Великобританией и Японией – число совместных китайско-британских и китайско-японских статей несколько больше чем публикаций, выпущенных британскими и японскими исследователями в сотрудничестве с учеными из США. Для остальных стран США остается основным партнером в области глубокого обучения.

Таблица 3

Количество работ в области *Deep learning*, выполненных в рамках международного сотрудничества (2000-2018 гг.)

Страна	КНР	США	Велико-британия	Япония	Ю.Корея	ФРГ	Индия	Франция	Россия
	1	2	3	4	5	6	7	8	9
КНР	8630	1164	298	108	56	81	23	71	9
США	1164	7840	253	103	137	196	88	105	35
Великобритания	298	253	1881	34	25	123	24	44	17
Япония	108	103	34	1338	13	31	6	17	4
Ю.Корея	56	137	25	13	1316	25	13	5	6
ФРГ	81	196	123	31	23	1237	13	56	11
Индия	23	88	24	6	13	13	1026	12	2
Франция	71	105	44	17	5	56	12	749	9
Россия	9	35	17	4	6	11	2	9	250

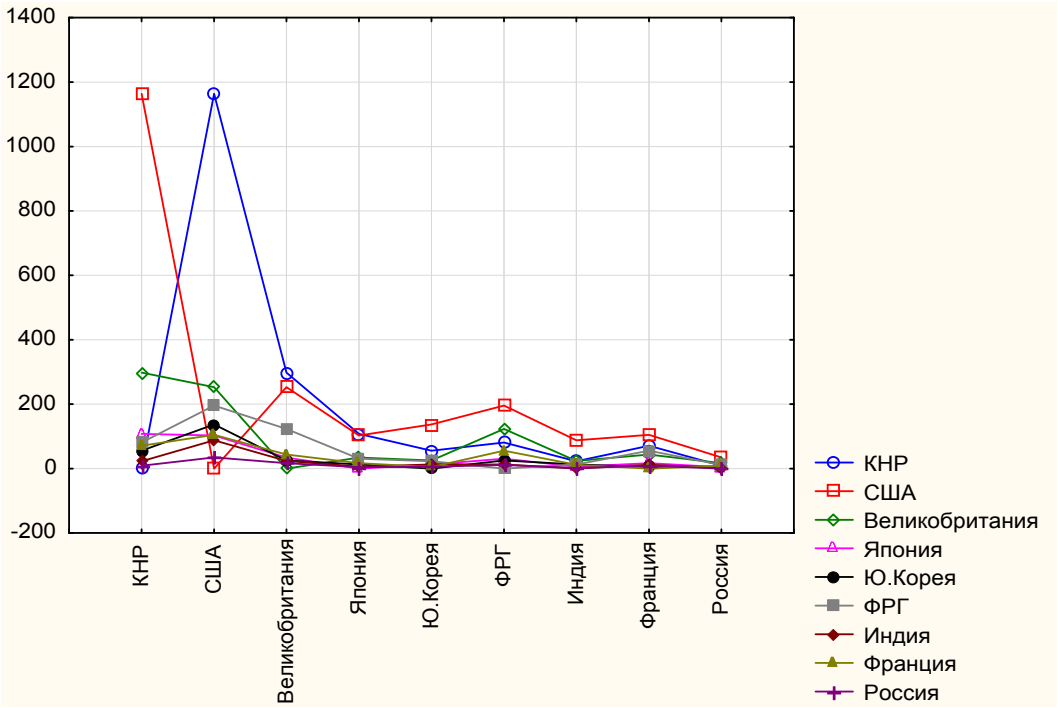


Рис.4. Количество публикаций ведущих стран, выполненных в рамках международного сотрудничества

СРАВНИТЕЛЬНЫЙ АНАЛИЗ БЫСТРОРАЗВИВАЮЩИХСЯ ПРЕДМЕТНЫХ ОБЛАСТЕЙ

Проведём сопоставление двух быстроразвивающихся областей, в которых в настоящее время наблюдается исследовательский бум: квантовые технологии (*Quantum Technologies – QT*) и глубокое обучение (*Deep Learning – DL*).

Наукометрические исследования для *QT* были выполнены ранее и представлены в [18]. Сравнительный анализ публикационной активности в *QT* и *DL* проводился на примере двух стран-лидеров (США и КНР). Численные значения представлены в табл. 4. В качестве временного интервала выбран шестнадцатилетний период – 2000-2016 гг.

Сильные различия показателей не должны вводить в заблуждение – сопоставляются не совсем равнозначные тематики. Если в *QT* запрос «Quantum*» позволяет сделать «широкий» захват тематической информации и обеспечить высокую полноту, то в случае *DL* (по запросу «*Deep* Near/3 (*Learning* OR *Networks*)») в выборку включались только узкоспециализированные публикации. При этом из рассмотрения целенаправленно исключались междисциплинарные работы на стыке тематик (нейроморфные вычисления, построение математической модели мозга, «классические» методы обработки и анализа данных и т.п.).

Таблица 4

Количество статей в области глубокого обучения и квантовых технологий, опубликованных в США и КНР за 2000-2016 гг.

	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
США- *QT*	5069	4826	5166	5604	5779	6290	6552	6634	6759	7214	7282	7313	7824	7997	7927	7934	8215
КНР- *QT*	1233	1371	1455	1604	1940	2377	3040	3483	4071	4476	4879	5611	6457	7156	7990	8972	9463
США-*DL*	32	34	32	39	40	40	45	51	62	58	72	79	116	194	287	679	1158
КНР-*DL*	1	0	3	7	4	5	4	14	20	22	23	24	42	100	300	767	1390

Таблица 5

Количество статей в области глубокого обучения и квантовых технологий в ведущих странах в 2016 г.

Предметные области	США	КНР	ФРГ	Япония	Великобритания	Россия	Франция	Индия	Ю.Корея
QT	8215	9463	4068	2559	2546	2340	2265	2532	1387
DL	1158	1390	194	250	276	42	114	171	179

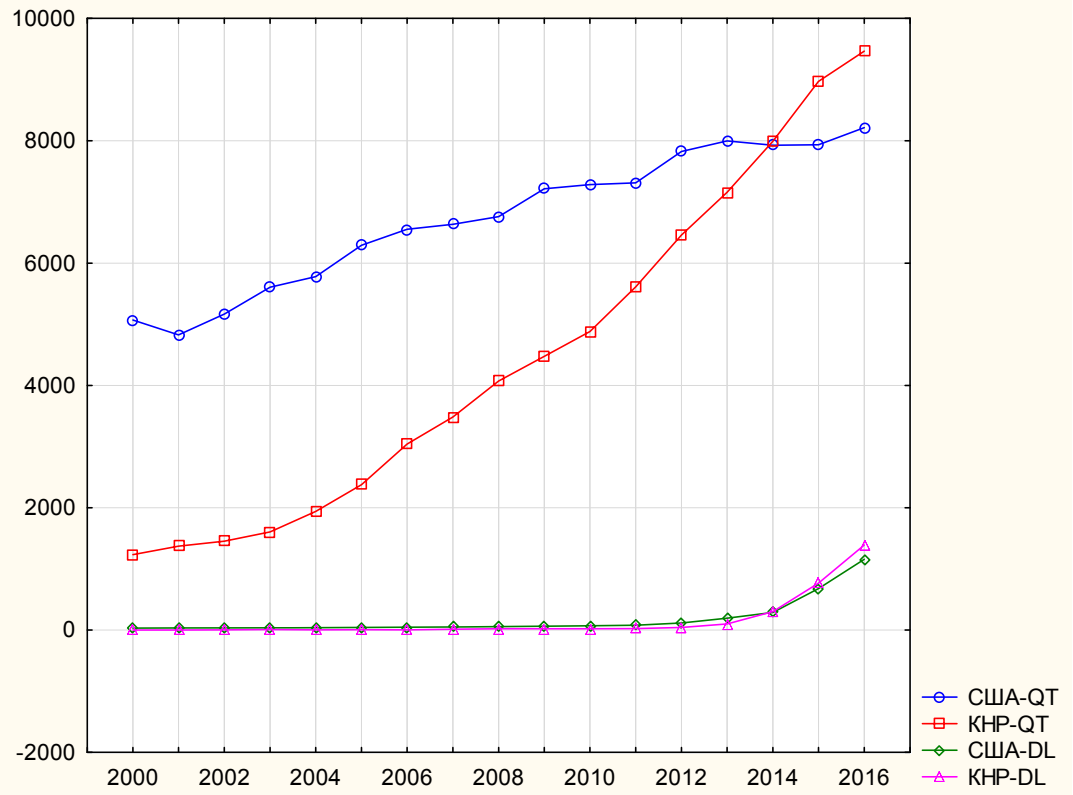


Рис.5. Рост публикаций в области глубокого обучения (*DL*) и квантовых технологий (*QT*) за 2000-2016 гг.

Как следует из табл. 4, в 2016 г. число публикаций КНР в области глубокого обучения примерно соответствовало количеству китайских статей по квантовым технологиям в 2001 г. У США в 2000 г. публикаций по *QT* было практически в 5 раз больше, чем статей по глубокому обучению в 2016 г.

Виды зависимостей для этих двух предметных областей приведены на рис. 5. Анализ графиков показывает более высокие темпы роста в области *DL*, чем в *QT*. Если по *QT* для достижения показателя в 9463 документа (ячейка «2016 г.» в табл. 4) КНР потребовалось 15 лет, то по *Deep Learning*, согласно

построенной математической аппроксимации (см. формулу 4), для этого потребуется 10 лет. Таким образом, к 2029 г. при сохранении имеющихся темпов роста в БД *Web of Science* будет проиндексировано 9810 китайских статей.

Можно ли считать, что у остальных стран публикационная активность по *QT* и *DL* хорошо согласуется? Для этого вычислим ранговый коэффициент корреляции Спирмена по табл. 5, в которой представлено количество публикаций девяти стран за 2016 г. [28]. Расчет в программе *Statistica* показал, что значение коэффициента Спирмена равно 0,81. Это указывает на наличие весьма сильной корреляционной связи. Если из рассмотрения удалить российские публикации, выпадающие по показателю публикационной активности в *Deep Learning*, то значение коэффициента Спирмена еще более возрастает и достигает 0,88.

Резюмируем основные результаты данного раздела.

Обе рассматриваемые предметные области развиваются чрезвычайно динамично. К технологиям *QT* и *DL* предъявляются завышенные ожидания научных прорывов экстраординарного уровня (прежде всего речь идет о создании универсального квантового компьютера и сильного искусственного интеллекта), поэтому изыскания хорошо финансируются и проводятся широким фронтом. Безоговорочное лидерство в квантовых технологиях и глубоком обучении принадлежит США и Китаю, которые практически стали не достижимы для других стран по количеству публикаций и борются между собой за лидерство. Остальные государства (кроме России) планомерно развивают оба перспективных направления и находятся в первой десятке.

ЗАКЛЮЧЕНИЕ

Проведенное нами исследование позволяет утверждать, что технологии глубокого обучения находятся в стадии активного развития и к уже текущему моменту имеются «истории успеха» как в области научных результатов, так и их коммерческого использования. Кроме того, сформировалось общественное мнение о высокой значимости глубокого обучения для создания принципиально новых решений в сфере искусственного интеллекта и их практической ценности для потребителя.

Обобщим основные результаты наукометрического исследования, проведенного в настоящей работе.

1. Начиная с 2014 г., в мире наблюдается быстрый и устойчивый рост числа публикаций по *Deep Learning*. В период 2014-2018 гг. сформировался и усилился отрыв китайских и американских ученых от исследователей из других стран.

2. Основное международное сотрудничество сосредоточено между лидерами – США и КНР.

3. Исследование терминологической структуры предметной области на основе анализа ключевых слов позволило получить представление о внутренней структуре фронта научных исследований, в частности, детализировать дескрипторы в тематических группах «нейросетевые методы», «способы обучения» и «области интенсивного прикладного применения глубокого обучения».

4. Сравнительный анализ двух динамично развивающихся научных направлений – глубокого обучения и квантовых технологий – показывает более активный рост публикационной активности по первому из них.

5. Россия практически никак не проявляет себя в области глубокого обучения. Микропоток отечественных публикаций практически не виден на уровне документального массива профильных статей и докладов, издаваемых в развитых странах. Кроме того, выявлены слабая вовлеченность отечественных ученых в международное сотрудничество и низкий уровень цитирования отечественных публикаций. Осторожный оптимизм мы связываем с созданием в 2018 г. трех специализированных центров сквозных технологий в рамках национальной технологической инициативы: на базе МФТИ – центра «Искусственный интеллект»; на базе МГУ – центра «Технологии хранения и анализа больших данных» и на базе ИТМО[11] – центра «Технологии машинного обучения и когнитивные технологии». Возможно эти организационные меры позволят активизировать российские исследования в том числе в области глубокого обучения. Однако какие-то результаты от реализации данной инициативы можно будет оценивать только через 3-5 лет, когда остальные страны также продвинутся в своих исследованиях и, скорее всего, начнут активную коммерциализацию разработанных ноу-хау.

Существует также ряд проблем, которые оказались за рамками проведенных исследований. К ним относится выявление зарождающихся научных направлений в области *Deep Learning* и идентификация возможных точек роста; кадровое обеспечение исследований в различных странах и уровень финансирования (доступность грантов); вклад организаций и отдельных коллективов в развитие *DL*; частота цитирования документов.

СПИСОК ЛИТЕРАТУРЫ

1. LeCun Y., Bengio Y., Hinton G. Deep Learning // Nature. – 2015. – Vol. 521. – P. 436-444.
2. Schmidhuber J. Deep Learning in Neural Networks: an Overview // Neural Networks. – 2015. – Vol. 1. – P. 85-117.
3. Huang G., Sun Y., Liu Z., Sedra D., Weinberger K. Deep Networks with Stochastic Depth // Computer Vision. – 2016. – P. 646-661.
4. Krizhevsky A., Sutskever I., Hinton G. ImageNet classification with deep convolutional neural networks // Communications of the ACM. – 2017. – Vol. 60, Iss. 6. – P. 84-90.
5. Bengio Y., Lamblin P., Popovici D., et al. Greedy Layer-Wise Training of Deep Networks // Advances in Neural Information Processing Systems. – 2007. – P. 153-160.
6. Grace K., Salvatier J., Dafoe A., Zhang B., Evans O. When Will AI Exceed Human Perform-

[1] ИТМО – Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики

ance? Evidence from AI Experts. – URL: https://arxiv.org/abs/1705.08807/ (дата обращения 05.03.2019).
7. Preparing for the Future of Artificial Intelligence. Executive Office of the President National Science and Technology Council Committee on Technology. – USA. – 2016. – URL: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf (дата обращения 05.03.2019).
8. Ding J. Deciphering China's AI Dream. – University of Oxford, March 2018. – URL: https://www.fhi.ox.ac.uk/wp-content/uploads/ Deciphering_Chinas_AI-Dream.pdf (дата обращения 05.03.2019).
9. Coursera. – URL: https://ru.coursera.org/ specializations/deep-learning (дата обращения 05.03.2019).
10. Созыкин А.В. Обзор методов обучения глубоких нейронных сетей // Вестник ЮУрГУ. Серия: Вычислительная математика и информатика. – 2017. – Т. 6, № 3. – С. 28-59.
11. Golovko V.A. Deep learning: an overview and main paradigms // Optical Memory & Neural Networks. – 2017. – Vol. 26. – P.1-17.
12. Заметки о статьях по DeepLearning и ImageNet. – URL: http://re9ulus.github.io/2017/03/03/imagenet-paper-digest/ (дата обращения 05.03.2019).
13. Hinton G., Salakhutdinov R. Reducing the Dimensionality of Data with Neural Networks // Science. – 2006. – Vol. 313, № 5786. – P. 504-507.
14. Oracle Big Data Blog. – URL: https://blogs.oracle.com/bigdata/difference-ai-machine-learning-deep-learning (дата обращения 05.03.2019).
15. Воронцов К.В. Машинное обучение: шаг в цифровую экономику. – URL: http://www.machine-learning.ru/ wiki/images/e/e4/Voron17ai-mipt.pdf (дата обращения 05.03.2019).
16. Reese H. Understanding the differences between AI, machine learning, and deep learning. – URL: https://www.techrepublic.com/article/understanding-the-differences-between-ai-machine-learning-and-deep-learning/ (дата обращения 05.03.2019).
17. Esser S.K. et. al. Convolutional Networks for Fast, Energy-Efficient Neuromorphic Computing. – URL: https://arxiv.org/abs/1603.08270/ (дата обращения 05.03.2019).
18. Толчеев В.О. Наукометрический анализ современного состояния и перспектив развития квантовых технологий // Научно-техническая информация. Сер. 2. – 2018. – № 6. – С. 1-14; Tolcheev V.O. Scientometric Analysis of the Current State and Prospects of the Development of Quantum Technologies // Automatic Documentation and Mathematical Linguistics. – 2018. – Vol. 52, № 3. – P. 121–133.
19. Зиновьева Н.Б. Документоведение: учебно-методич. пособие. – М.: Профиздат, 2001. – 208 с.
20. Jia Y., Shelhamer E., Donahue J., et al. Caffe: Convolutional Architecture for Fast Feature Embedding // Proceedings of the 22nd ACM International Conference on Multimedia. – 2014. – P. 675-678.
21. Abadi M., Agarwal A. Barham P., et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems // Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI '16). – 2016. – P. 265-283.
22. Гиляревский Р.С., Михайлов А.И., Черный А.И. Научные коммуникации и информатика. – М.: Наука. – 1976. – 435 с.
23. Википедия. – URL: https://ru.wikipedia.org/wiki/Глубокое_обучение (дата обращения 05.03.2019).
24. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative Adversarial Networks. – URL: https://arxiv.org/pdf/1406.2661.pdf/ (дата обращения 05.03.2019).
25. Mikolov T., Karafiát M., Burget L., Cernocky J., Khudanpur S. Recurrent Neural Network Based Language Model // Interspeech. – 2010. – P.1045-1048.
26. He K., et. al. Deep residual learning for image recognition. – URL: https://arxiv.org/pdf/ 1512.03385.pdf/ (дата обращения 05.03.2019).
27. Graves A., Wayne G., Danihelka I. Neural turing machines. – URL: https://arxiv.org/abs/1410.5401/ (дата обращения 05.03.2019).
28. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL. – М.: Форум, 2004. – 463 с.

Материал поступил в редакцию 11.03.19

Сведения об авторе

ТОЛЧЕЕВ Владимир Олегович – доктор технических наук, доцент, профессор кафедры управления и информатики НИУ «Московский энергетический институт»
e-mail: tolcheevvo@mail.ru

УДК 004.94

В.Н. Шведенко, А.А. Волков

Метод создания цифрового двойника на основе агрегации информационных объектов

Развитие направления создания информационных моделей материальных систем поставило вопрос построения информационного двойника. Реализация полнофункционального цифрового двойника возможна при одновременном взаимодействии множества его проекций. Проблема взаимодействия проекций обусловлена совокупностью подходов декомпозиции объекта реального мира и агрегации информационного объекта. Предлагается организация связей проекций через выявление совместно используемых потоков данных с помощью коллинеарных структур.

Ключевые слова: *цифровые двойники, агрегация информационных объектов, проекция цифрового двойника, коллинеарные структуры, декомпозиция материальной системы, агрегация проекций цифрового двойника*

ВВЕДЕНИЕ

В настоящее время разработано много методов декомпозиции сложных технических систем, а также программных продуктов, позволяющих создавать информационные модели объектов материальных систем. Среди них можно выделить метод структурного анализа, который поддерживается такими стандартами, как *SADT, IDF0, IDEF1* и т д. Существует целая линейка программных продуктов, позволяющих строить структурные модели, например, *BPwin, Erwin.* Другой распространённый метод построения модели объектов системы основан на моделировании по физическим процессам (здесь наиболее распространёнными программами являются *ANSIS, Solidworks* и др.) Также существуют методы моделирования информационных систем, основанные на жизненном цикле [1, 2]. В этом случае используются продукты *CAD*/*CAM*/*CAE*.

Каждый из указанных методов реализуется своими способами и программными средствами, которые, в лучшем случае, могут интегрироваться через межфайловый обмен. Наибольшая степень интеграции осуществляется через *CALS* технологии, объединяющие модели на этапах жизненного цикла материальной системы.

Одним из направлений развития информационных технологий становятся теория и практика создания цифровых двойников объектов материальной системы [3].

Цифровой двойник должен позволять с высокой точностью прогнозировать поведение объектов реального мира. Это может быть достигнуто за счёт совместного использования различных методов декомпозиции материальной системы, дающих возможность получать информационную модель цифрового двойника, приближённую к реальному объекту [4].

В настоящей статье предлагается метод создания цифрового двойника, одновременно обладающего всеми преимуществами каждого из методов декомпозиции, применяемых к материальной системе на этапе выявления её свойств и функциональных особенностей.

ПРОВЕДЕНИЕ ДЕКОМПОЗИЦИИ ОБЪЕКТОВ МАТЕРИАЛЬНОЙ СИСТЕМЫ С ПРИМЕНЕНИЕМ КОЛЛИНЕАРНЫХ СТРУКТУР

Первым этапом метода создания цифрового двойника является проведение декомпозиции объектов материальной системы различными способами.

От выбранного метода декомпозиции зависит версия информационного объекта (в ней отразятся те свойства и процессы материальной системы, которые предусмотрены применённым методом декомпозиции). Каждый информационный объект имеет собственную структуру метаданных, зависящих от метода декомпозиции и определяемых на этапе агрегации версии информационного объекта.

Информационные объекты, полученные в результате применения различных методов декомпозиции, будем называть проекциями цифрового двойника материальной системы. Один метод декомпозиции может порождать несколько проекций.

Набор проекций, полученных в результате применения различных методов декомпозиции к материальной системе, стремится к полнофункциональному цифровому двойнику. Чем больше получено проекций, тем ближе цифровой двойник к полнофункциональному.

Приведем наиболее распространенные методы декомпозиции.

- Метод функциональной декомпозиции – заключается в выявлении и определении общности и различия функций, исполняемых элементами системы. Основной упор делается на детализацию функций системы; при этом не затрагивается декомпозиция структуры самого объекта.
- Метод декомпозиции по жизненному циклу – базируется на анализе процессов, протекающих в материальной системе в различные периоды времени. В результате его применения получается модель, в которой изменения свойств объектов и связей зависят от фактора времени, поэтому модель не является полнофункциональной и отражает только отдельные стороны системы.
- Метод декомпозиции по физическим процессам – осуществляется на основе присущих системе закономерностей, описываемых аналитическими или численными методами, через типовые алгоритмы. Такой подход не обеспечивает полноту функциональных связей и ограничений, накладываемых на функции системы.

Выбор одного из методов декомпозиции приводит к тому, что система может рассматриваться только с точки зрения его возможностей, при этом некоторые объекты отсутствуют в модели, а ряд связей считается незначимым. Вследствие этого модель системы в ряде случаев не удовлетворяет критериям адекватности.

Возможно совместное использование декомпозиций материальной системы для решения узкоспециализированных задач, что является частным решением и не может распространяться на другие объекты и системы.

Отличительной особенностью метода проектирования цифровых двойников объектов материальной системы является организация связей проекций через определение совместно используемых потоков данных. Для чего подходит использование коллинеарных структур [5].

Организация структуры информационной модели материальной системы, при которой информационные объекты, входящие в её состав, имеют однозначно сопоставимый между собой набор свойств или подструктур, выделенные в качестве отдельного объекта системы (далее коллинеарная структура), называется информационной моделью с коллинеарными связями [6].

Коллинеарные структуры оказывают существенное влияние на процесс декомпозиции объектов материальной системы и агрегации информационных объектов в полнофункциональную модель цифрового двойника. Коллинеарные связи позволяют объединять проекции цифровых двойников для их взаимодействия в системе. Такие модели будем называть моделями с коллинеарными связями. На рис. 1 показана схема выявления коллинеарных структур.

Под качеством декомпозиции системы будем понимать наличие признаков системы:

- отсутствие избыточных связей с минимальной сложностью алгоритмов описания системы;
- соответствие модели реальному объекту;
- учёт максимального количества факторов и их ранжирование, т.е. определение степени значимости;
- возможность управления системой через наделение её адаптивными и поведенческими функциями.

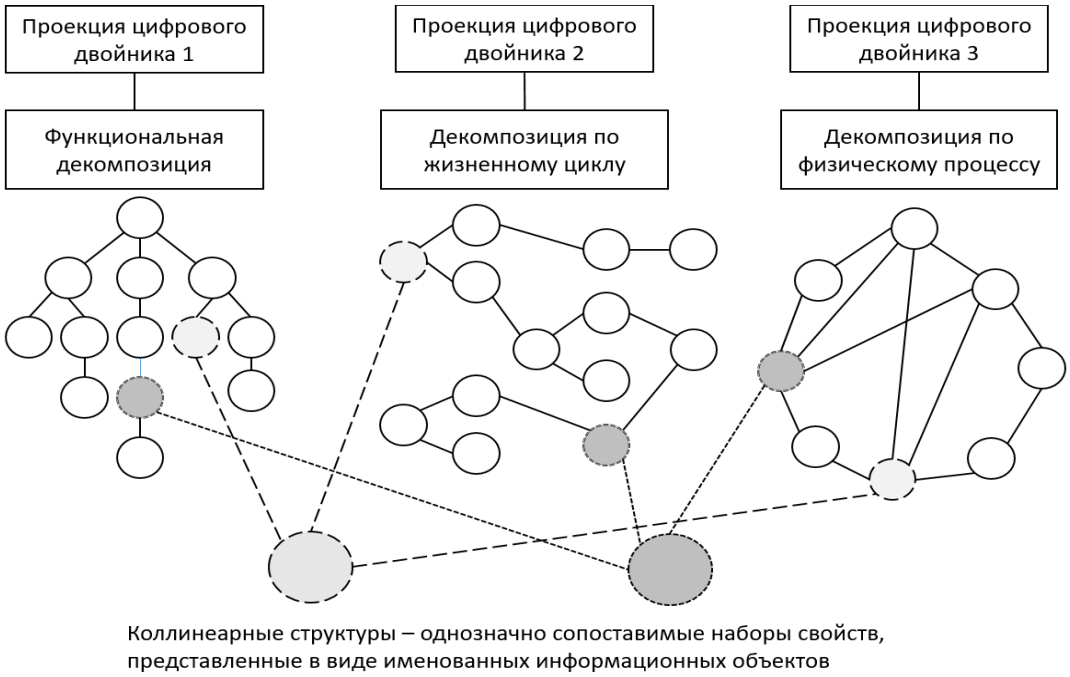


Рис. 1. Взаимодействующие проекции цифровых двойников материальной системы, полученные различными способами декомпозиции, на основе коллинеарных структур

Коллинеарные структуры обеспечивают функцию взаимодействия моделей, имеющих коллинеарные связи. Система рассматривается как объект сложной структуры, обладающий рядом признаков (атрибутов), которые определяются совокупностью свойств, входящих в него элементов (объектов) и отношений между ними, заданными в соответствии с законами композиции.

Предлагаемый нами метод проектирования информационной модели материальной системы основывается на следующих положениях:

- информационная модель материальной системы состоит из функционально независимых подсистем;
- между подсистемами устанавливаются транспортные и управляющие связи;.
- в отдельный класс выделяются подсистемы с коллинеарными структурами, которые обеспечивают однородность связей.

При построении информационной модели системы наиболее важной задачей является выбор стратегии декомпозиции.

Цель декомпозиции – разделение системы на отдельные элементы для уменьшения сложности ее описания и построения адекватной информационной модели. Под элементами системы в дальнейшем будем понимать подсистему, объект или системы. При этом подсистема – это часть сложной системы, состоящая из совокупности элементов, представляющих автономную область внутри системы, обладающую собственным системными свойствами и имеющую связи с другими элементами сложной системы. Между элементами системы существуют связи, наделенные функциями, реализующими алгоритм взаимодействия между элементами [7–11].

Декомпозиция материальной системы производится по принципу объектно-ориентированного подхода, а именно – выделение подсистемы по признаку целостного элемента и инкапсуляция (упаковывание в единый объект) с его функциями.

Следующим шагом декомпозиции материальных систем является представление их элементов в виде информационных объектов и связей между ними. При этом каждый информационный объект в зависимости от его сложности делится на некоторое количество уровней детализации. Выделяются три базовых типа уровней детализации:

1) уровни первичных данных, поступающих в информационный объект из внешнего мира;

2) уровни обработки данных. Первичные данные подвергаются преобразованиям в соответствии с уровнями абстракции информационного объекта;

3) Уровни итоговых показателей. Данные с уровней абстракции информационного объекта преобразуются в итоговые значения, характеризующие эффективность первичной информации при управлении системой в соответствии с заданными уровнями абстракции информационного объекта [4].

Функции внутри материальной системы рассматривается в трех аспектах.

1. Функции, определяющие достижение цели одного или нескольких процессов, выполняемых в рамках материальной системы на период её жизненного цикла.

2. Функции, обеспечивающие организацию связи между различными средами отдельных информационных объектов.

3. Функции, обеспечивающие регламентное выполнение связи.

АГРЕГАЦИЯ ПРОЕКЦИЙ ЦИФРОВОГО ДВОЙНИКА НА ОСНОВЕ КОЛЛИНЕАРНЫХ СВЯЗЕЙ

Процесс агрегации элементов материальной системы, полученных в результате проведенной декомпозиции, представим в виде рекурсии (выявление коллинеарной структуры происходит рекурсивно) с множеством терминальных ветвей. Каждый рекурсивный вызов увеличивает глубину масштаба информационной модели материальной системы.

При этом рекурсия приводит к построению подсистемы и образованию внутри нее связей. В результате рекурсивного масштабирования модели системы образуются динамические информационные процессы, которые можно классифицировать на трех уровнях абстракции обработки информации:

- модель последовательной обработки информации;
- модель обработки информации в иерархической системе от низших уровней к высшим уровням;
- модель обработки информации в иерархической системе от высших уровней к низшим уровням.

В процессе рекурсии образуются коллинеарные структуры, а её результатом будет появление информационной модели с коллинеарными связями, полученными различными методами декомпозиции.

Агрегация элементов на основании рекурсии в конечном итоге позволяет создать модель с коллинеарными связями и объединить модели обработки информации в единую структуру.

Итоговая информационная модель взаимодействующих материальных систем будет представлять собой децентрализованную сеть узлов. На основании этого начинать процесс моделирования можно с любой системы или подсистемы, опираясь на общий элемент, а именно на коллинеарную структуру. Таким образом, процесс агрегации начинается с выявления коллинеарной структуры для нескольких систем или подсистем (рис. 2).

Процесс построения коллинеарных связей от коллинеарной структуры *CS (Collinear Structure)* назовём функцией *CR* (*Collinear Relations*). В качестве аргументов функции берутся связываемые системы [*sys1, … , sysN*]. Связи создаются в процессе проведения рекурсии. Набор коллинеарных структур формируется на уровне модели, и в дальнейшем их количество не меняется. На концептуальном уровне устанавливается правило выхода из рекурсии.

В процессе работы рекурсивного алгоритма возможно появление коллизий соединения подсистем через коллинеарные структуры, выражающихся в:

- столкновении рекурсии. Несколько рекурсивных алгоритмов могут прийти в одну точку и продолжить построение связей в направлении, противоположном уже существующим;
- зацикливании рекурсивных алгоритмов, когда алгоритм приходит в свою начальную точку и повторяет построение связей по кругу.

После выполнения процесса агрегации возможно обращение к этапу моделирования для приближения агрегируемой информационной системы к требованиям материальной.

По завершению рекурсивного формирования структуры взаимодействующих систем появляются замкнутые структуры связей (они не существуют в иерархических или других строго структурированных моделях). Их наличие во взаимодействующих системах позволяет оказывать неоднозначное влияние систем друг на друга и на самих себя.

Выделим три базовых типа влияния взаимодействующих систем:

1. Кольцевое. Связи какой-либо коллинеарной структуры имеют замыкающий характер. Таки образом, управляющий сигнал, отправленный через коллинеарную структуру, пройдёт цепочку связей и вернётся в начальную точку. В процессе перехода между подсистемами сигнал может быть модифицирован. В этом случае он вернется в исходную точку с новыми параметрами, что позволит выполнить следующую итерацию кольцевого взаимодействия систем.

2. Противоречивое. Какая-либо коллинеарная структура имеет несколько связей с другими коллинеарными структурами. Одновременное получение конфликтующих сигналов из нескольких связанных узлов приведёт к оценке значимости этих сигналов в транспортной среде и выбору направления передачи потоков данных.

3. Дублирующее. Аналогично противоречивому типу влияния, однако сигналы, получаемые из нескольких связанных узлов, имеют одинаковый или подобный набор параметров, что повышает значимость соответствующего направления передачи потоков данных. Схематично базовые типы влияния взаимодействующих систем показаны на рис. 3.

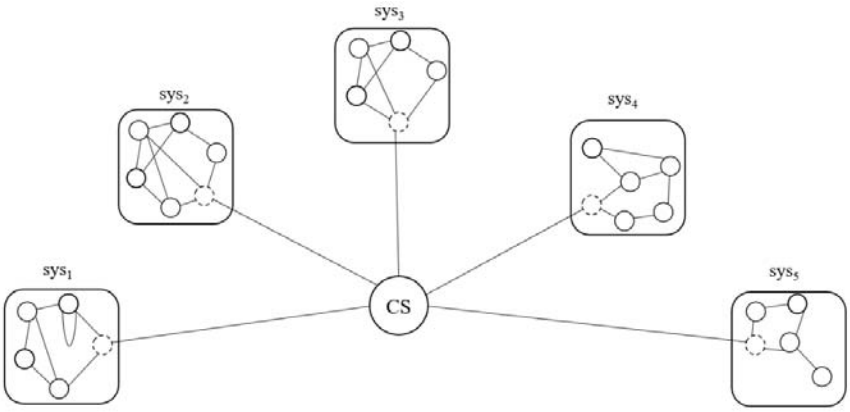


Рис. 2. Выявления коллинеарной структуры для нескольких систем

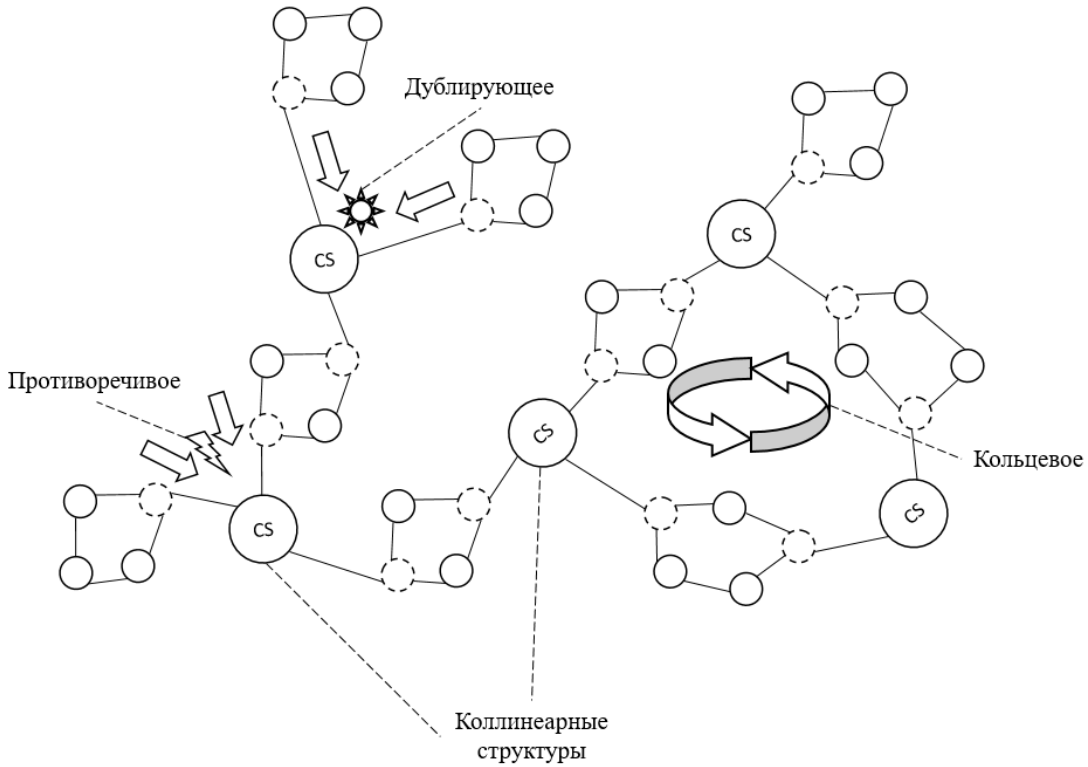


Рис. 3. Типы влияния взаимодействующих систем

ЗАКЛЮЧЕНИЕ

Предложен метод организации взаимодействия проекций цифрового двойника, полученных путем декомпозиции объектов материальных систем. Его особенность – использование коллинеарных структур, обеспечивающих совместное применение потоков данных, формируемых проекциями цифровых двойников.

Рассмотрены особенности организации связей между проекциями цифровых двойников. Установлены основные коллизии в процессе агрегации взаимодействующих проекций цифровых двойников, а также точки возможного взаимодействия цифрового двойника с объектом реального мира через его проекции.

СПИСОК ЛИТЕРАТУРЫ

1. de Bortoli M.G.D. The use of an integrated CAE/CAD/CAM system during the compressor design // International Compressor Engineering Conference. – Joinville: Research Engineer – EMBRACO, 2002.
2. Gujarathi G.P., Ma Y.-S. Parametric CAD/CAE integration using a common data model // Journal of Manufacturing Systems. – 2011. – Vol. 30. – P. 118-132.
3. Qi Q., Tao F., Zuo Y., Zhao D. Digital twin service towards smart manufacturing // 51st CIRP Conference on Manufacturing Systems CIRP 72 (2017). – URL: ScienceDirect, 2017. –P. 237–242.
4. Reeves Chris. Spotlight on the Digital Twin // ANSYS Advantage. – 2017. – № 1. – P. 1-7.
5. Шведенко В.Н., Волков А.А. Реализация параллельных структур в объектно-ориентированных СУБД // Промышленные АСУ и контроллеры. – 2011. – № 6. – С. 62-64.
6. Шведенко В.Н., Волков А.А. Модель формирования параллельных структур в объектно-ориентированных СУБД // Программные продукты и системы. – 2011. – № 3. – С. 14-17.
7. Sokolov B.V., Zelentsov V.A., Kulakov A.Yu., Pimanov I.Yu. Models and methods of reconfiguration of complex technical objects under different situation conditions // Труды международной научной конференции «Mathematical methods in engineering and technology MMET NW 2018» (10-14 сентября 2018 г., Санкт-Петербург). – СПб, 2018. – С. 361–363.
8. Shi C., Li Y., Zhang J., Sun Y., Yu P.S. A survey of heterogeneous information network analysis // IEEE Transactions on Knowledge and Data Engineering. – 2017. – Vol. 29, № 1. – P. 17–37.
9. Sokolov B., Yusupov R., Okhtilev M., Maydanovich O. Influence analysis of information technologies on progress in control systems for complex OBJECTS // New Trends in Information Technologies. Proceedings of International Conference Information-Interaction-Intellect (Bulgaria, Varna, 2010, June 23–27. – Sofia: ITHEA, 2010. – P.78-91.
10. Wang, Q.G., Zhang, Y., Chiu, M.S. Decoupling internal model control for multivariable systems with multiple time delays // Chemical Engineering Science. – 2002. – Vol.57, Iss.1. – P. 115–124.
11. Mencagli G., Vanneschi M. Towards a systematic approach to the dynamic adaptation of structured parallel computations using model predictive control // Cluster Computing. – 2014. – Vol. 17, Iss. 4. – P. 1443–1463.

Материал поступил в редакцию 11.01.19.

Сведения об авторах

ШВЕДЕНКО Владимир Николаевич – доктор технических наук, профессор, ведущий специалист ВИНИТИ РАН, Москва
e-mail: vv_shved@mail.ru

ВОЛКОВ Антон Андреевич – кандидат технических наук, инженер-программист ООО "MMTR", г. Кострома
e-mail: volkov-kostroma@yandex.ru

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

УДК 004.94:81’32:811.1/.2’01

М.А. Егорова, А.А. Егоров

Прародина народов носителей праиндоевропейского языка: математические модели для исследования лингвистической информации*

Представлены результаты теоретического анализа и компьютерного моделирования, позволяющие сделать предположение, что около 3500-4000 лет назад в рассматриваемом модельном индоевропейском языковом сообществе могли возникнуть две основные лингвистические популяции, характеризуемые сегодня как деление индоевропейских языков на так называемые языковые ареалы «сатем-кентум». Результаты компьютерного моделирования показали, что из двух основных гипотез формирования праиндоевропейцев – Анатолийской и Курганной – последняя больше соответствует полученным нами временным оценкам. Проанализированы некоторые проблемы поиска прародины народов – носителей праиндоевропейского языка.

Ключевые слова: *математическая лингвистика, математическая модель, индоевропейские языки, индоевропейская хронология, индоевропейское распространение, Анатолийская и Курганная гипотезы, компьютерное моделирование*

ВВЕДЕНИЕ

Для изучения распространения и видоизменения «лингвистической информации» в индоевропейских языковых сообществах, а также для поиска прародины народов – носителей праиндоевропейского языка – нами были исследованы две математические модели: первая – динамическая системная модель, описываемая нелинейным уравнением; вторая – модель, описываемая системой интегро-дифференциальных уравнений (см., например, [1, 2]). В рамках этих моделей было проведено численное исследование распространения и изменения лингвистической информации в некотором модельном индоевропейском (ИЕ) языковом сообществе, в том числе на начальном этапе его формирования, и обнаружены как регулярные, так и хаотические явления при передаче лингвистической информации. При этом проявились *бифуркации,* что позволило сделать вывод о возможном появлении нового качества в поведении динамической системы при малом изменении ее параметров [3, 4].

Обратимся к понятию *«информация».* Это сведения об объектах и явлениях окружающей среды (живой и неживой природы), а также об их параметрах, свойствах и состояниях. В более узком смысле под *лингвистической информацией* могут пониматься, например, семантика (определяет соотношение между словами и их значениями) и грамматика (правила, выражающие общие синтаксические свойства слов и групп слов, позволяющие производить и/или описывать правильные предложения) некоторого языка.

При построении математических моделей распространения и изменения лингвистической информации в языковых сообществах в качестве априорных данных использовались сведения, полученные из независимых исследований, как из лингвистики, так и из других научных областей, например, из истории, генетики и археологии [5–22].

Полученные результаты позволили по-новому оценить выдвинутые ранее гипотезы и позволили предложить новые перспективные подходы к исследованию явления передачи лингвистической информации в сложных социально-лингвистических системах.

ДИНАМИЧЕСКАЯ НЕЛИНЕЙНАЯ МОДЕЛЬ РАСПРОСТРАНЕНИЯ И ИЗМЕНЕНИЯ ЛИНГВИСТИЧЕСКОЙ ИНФОРМАЦИИ В СООБЩЕСТВЕ (древовидная модель)

Динамическая модель распространения и изменения «лингвистической информации» в некотором сообществе может быть описана следующим нелинейным уравнением [1]:

$$I_{m+1} = \left[a_1 I_m \left(M - I_m \right) + a_2 \left(M - I_m \right)^2 \right] \lambda , \qquad (1)$$

* Статья подготовлена при поддержке Программы Российского университета дружбы народов «5-100».

где I – величина анализируемой лингвистической информации, $m = 1, 2, ...$ ($m = 1$ соответствует первому "измерению", т.е. I_1 – начальное значение, например, в некоторый начальный момент времени t_1); a_1 – коэффициент, характеризующий распространение лингвистической информации при контактах «незнающих» со «знающими» эту информацию I_n; a_2 – коэффициент, характеризующий воздействие только на «незнающих»; M – максимальное значение лингвистической информации; λ – параметр задачи (в соответстви с теорией катастроф он может быть назван управляющим параметром). Нелинейное уравнение (1) позволяет исследовать процесс изменения распространяющейся лингвистической информации в зависимости от времени и от других параметров. Как само уравнение, так и его варианты могут использоваться, в частности, при исследовании процесса обучения, например, детей взрослыми в некотором лингвистическом сообществе.

Сделав замену переменных, представим уравнение (1) в следующем виде:

$$y = \lambda_1 x(1-x) + \lambda_2 x(1-x)^2, \qquad (2)$$

где $y = I_{m+1}/M$, $x = I_m/M$; и $\lambda_1 = a_1 M\lambda$, $\lambda_2 = a_2 M\lambda$ – новые управляющие параметры системы; $0 \le x \le 1$.

При увеличении параметра $\lambda_1 \in (0,4)$ в системе наблюдаются закономерности (см. подробнее в [1]). При $\lambda_1 < 1$ $y(x) \to 0$ (итерации I_n сходятся к $I^* = 0$ – устойчивая неподвижная точка). При дальнейшем росте управляющего параметра λ_1 происходит бифуркация: точка $I^* = 0$ теряет устойчивость, а вновь появившаяся точка становится устойчивой (итерированные значения $I_m \to const$, т.е. динамический режим является стационарным или имеет период, равный единице – наблюдается цикл S^1). В итоге в системе возникают циклы вида S^{2^p}: S^1, S^2 (здесь происходит бифуркация удвоения периода: неподвижная точка расщепляется и I_m осциллирует между двумя значениями, образующими устойчивый аттрактор с периодом 2; в этом случае отображение (2) имеет устойчивый цикл с периодом 2), S^4, S^8, S^{16}, S^{32}, S^{64} … и т.д. (где $p = 0,1,2,3,4...$) [1]. При таком подходе в качестве одной из количественных характеристик рассматриваемого процесса распространения «лингвистической информации» следует рассматривать число возникающих циклов S^{2^p} как число возникших новых языков (или диалектов) в данном языковом сообществе; так циклы S^1, S^2, S^4, S^8, S^{16}, S^{32} дают соответственно: 1, 2, 4, 8, 16 и 32 языка. При этом можно выбрать «внутренний» («встроенный») масштаб времени в данном процессе [1]: 500 лет – время расхождения двух родственных языков [10]. Отметим, что на начальных этапах развития языкового сообщества это могут быть, например, языковые семьи или языковые ареалы типа «сатем-кентум».

Дальнейший рост управляющего параметра λ_1 позволяет наблюдать переход системы в хаотический режим: непериодический, случайный процесс возникает как предел все более сложных структур (циклов вида S^{2^p}). Таким образом, хаос – предел сверхсложной организации. Наконец, при некотором значении $\lambda_1 \to \lambda_\infty$ отображение (2) дает уже непериодическую последовательность $y(x)$ типа цикла S^∞.

Используем данные, полученные при численном исследовании распространения и изменения лингвистической информации в некотором модельном индоевропейском языковом сообществе, в том числе на начальном этапе его формирования. Будем полагать, что время начала разделения (т.е. по сути «исчезновения») гипотетического праиндоевропейского языка произошло приблизительно не позднее 6500 (Курганная гипотеза) или не позднее 9500 (Анатолийская гипотеза) лет назад [5, 14].

Наиболее характерные из полученных нами результатов моделирования в рамках данной нелинейной модели приведены на рис. 1, 2. Как следует из графиков, приведенных на рис. 1*а* и рис. 2*а*, диапазон хаотического «движения» наблюдается от 0-го до 5-6-го «поколения» (итерации по m), т.е. на временном отрезке от 6500 до 2500–3000 лет назад. У графиков, показанных на рис. 1*б* и рис. 2*б*, диапазон хаотического «движения» также от 0-го до 5–6-го поколения, но в этом случае он составляет отрезок от 9500 до 6500-7000 лет назад. Сравнение приведенных на рис. 1, 2 зависимостей и полученных из них временных оценок с данными независимых исследований [5] позволяет сделать вывод, что графики на рис. 1*а*, 2*а* больше соответствует Курганной гипотезе, а графики на рис. *1б*, 2*б* – Анатолийской.

Использование рассматриваемой нелинейной модели, где есть циклы вида S^{2^p}, позволяет нам получить не только «встроенный» внутренний временной масштаб с шагом в 500 лет, но и определить временную «длину» этой лингвистической временной «линейки», вдоль которой развивается динамика исследуемой системы [1].

Далее мы рассматриваем число возникающих циклов S^{2^p} как количество появившихся новых языков в интересующем нас языковом сообществе. Тогда для циклов вида S^{2^p} получаем ряд значений для числа языков L (начиная с одного возможного индоевропейского праязыка, первоосновы всех современных индоевропейских языков): 1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024, 2048 и т.д. Исходя из этого, несложно получить простую формулу для оценки возможного числа языков в таком динамическом процессе: $L = 2^p$. Отсюда получаем формулу для численной оценки количества соответствующих данному числу языков циклов p: $p = \log_2 L$. Так для 6000 языков получаем: $p = \log_2 6000 \approx 12.55$, т.е.

примерно 13 поколений, что соответствует времени: 13·500 лет = 6500 лет. Только для индоевропейских языков имеем: $p = \log_2 500 \approx 8.97$, т.е. примерно 9 поколений или 4500 лет. Учет вероятных исчезнувших языков и всевозможных диалектов увеличит эти оценки, особенно в случае всех языков, что вполне естественно. С этой точки зрения можно рассматривать данные оценки как оценки времен, когда различные языки окончательно разошлись.

Как видно из графиков, полученные результаты компьютерного моделирования хорошо соответствуют по времени двум основным гипотезам о формировании протоиндоевропейцев: Анатолийской и Курганной [5, 14]. Напомним кратко об этих двух гипотезах.

Анатолийская гипотеза локализует индоевропейскую прародину в западной Анатолии (современная Турция). Данные, полученные Греем и Аткинсоном методами байесовского анализа, по их мнению, указывают на возраст праиндоевропейского языка в интервале от 8000 до 9500 лет и на анатолийское происхождение языка [13].

Курганная гипотеза предложена Марией Гимбутас в 1956 г. и сейчас является наиболее популярной. Для определения местонахождения прародины носителей праиндоевропейского языка были использованы данные археологических и лингвистических исследований. Согласно гипотезе протоиндоевропейские народы существовали в причерноморских степях и юго-восточной Европе примерно с V по III тысячелетие до н.э. (а возможно и ранее) [5, 6, 14, 18]. Важнейшим этапом в развитии курганной культуры было одомашнивание лошади и использование повозок, что сделало носителей культуры мобильными и существенно расширило их влияние [6, 14]. Этот факт был положен нами в основу при построении теоретических моделей. В частности, при компьютерном моделировании соотношение коэффициентов λ_1 и λ_2 выбиралось, исходя из соотношения средних скоростей движения (в км/час) всадников v_1 и пешеходов v_2 в праиндоевропейских сообществах: $\lambda_1 : \lambda_2 \propto v_1 : v_2 \approx 15 : 3$.

Сравнение полученных нами данных по обеим гипотезам с данными независимых исследователей [5] позволяет сделать однозначный вывод о предпочтительности Курганной гипотезы. Действительно, как видно из рис. 1*а* и рис. 2*а* в случае реализации сценария развития языков по этой гипотезе, период существования одного или двух языков составляет примерно 1000–1500 лет (временной интервал с 4500 до 3500 лет назад), что вполне соответствует данным [5].

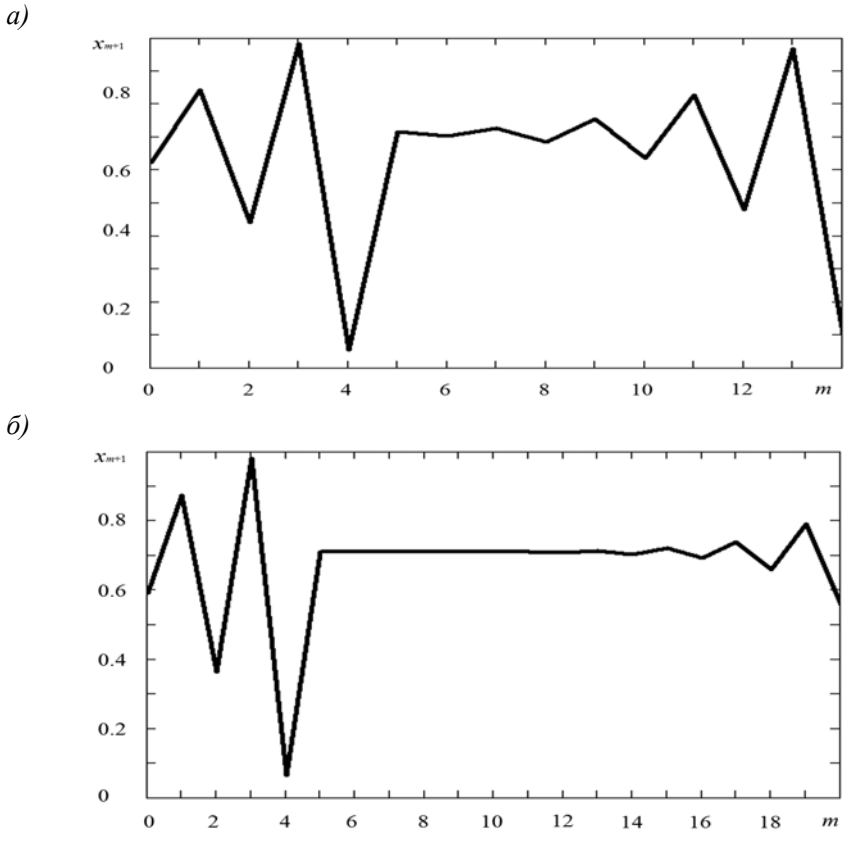
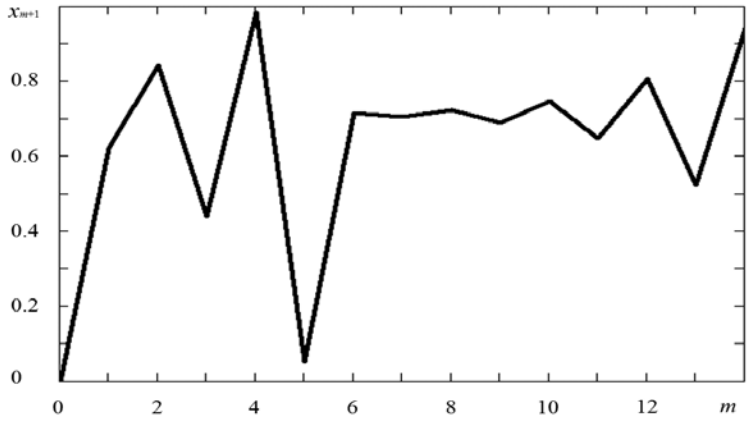


Рис. 1. Графики зависимости функции $x_{m+1} = \lambda_1 x_m (1 - x_m) + \lambda_2 x_m (1 - x_m)^2$ от m, характеризующие возникающие в системе циклы вида S^{2^p} (S^1, S^2, …) при $1 << p < +\infty$:
а – при параметрах: $\lambda_1 = 3.2$ и $\lambda_2 = 0.62$,
начальное значение $x_0 = 10^{-4}$; *б* – при параметрах: $\lambda_1 = 3.2$ и $\lambda_2 = 0.59$,
начальное значение $x_0 = 10^{-4}$.

а)



б)

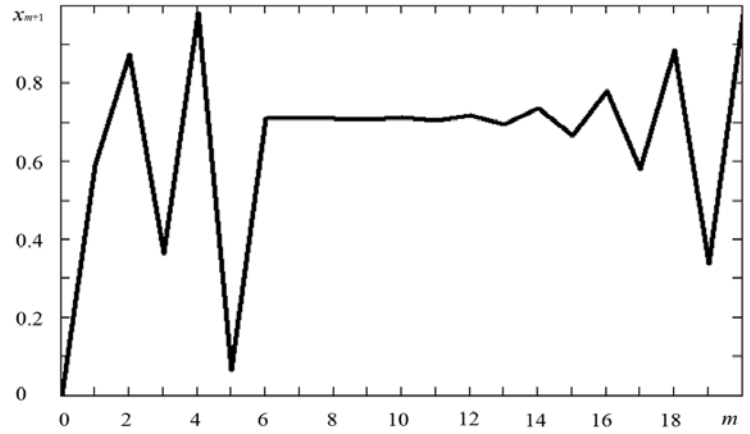


Рис. 2. Графики зависимости функции $x_{m+1} = \lambda_1 x_m (1 - x_m) + \lambda_2 x_m (1 - x_m)^2$ от m, характеризующие возникающие в системе циклы вида S^{2^p} (S^1, S^2, …) при $1 << p < +\infty$:
а – при параметрах: $\lambda_1 = 3.2$ и $\lambda_2 = 0.62$,
начальное значение $x_0 = 1$; *б* – при параметрах: $\lambda_1 = 3.2$ и $\lambda_2 = 0.59$,
начальное значение $x_0 = 1$.

ВОЛНОВАЯ МОДЕЛЬ РАСПРОСТРАНЕНИЯ И ИЗМЕНЕНИЯ ЛИНГВИСТИЧЕСКОЙ ИНФОРМАЦИИ В СООБЩЕСТВЕ

При описании языковых явлений в сообществах наиболее часто употребляются термины: передача и распространение (см., например, [16, 17]). Действительно, лингвисты давно обратили внимание на волновую природу многих языковых явлений. Однако пока предложено мало математических моделей, позволяющих это учитывать.

Для описания процессов распространения информации в языковом сообществе предлагается применить математический аппарат, который широко используется при исследовании различных волновых явлений (см., например, [23–26]). Как известно, скалярное волновое уравнение

$$c^2 \frac{\partial^2 \Psi(z;t)}{\partial z^2} - \frac{\partial^2 \Psi(z;t)}{\partial t^2} = 0$$

описывает математические модели различных реальных процессов в физических, биологических, экологических и социальных системах (см., например, [3, 23–26]).

Применительно к нашей задаче пространственная координата (ось z) расположена вдоль направления распространения волны; при этом: c – скорость волны, t – время, Ψ – волновая функция. В общем случае волновой процесс может сопровождаться процессами диффузии информации в окрестности ареалов, куда доходят волны лингвистической информации.

Интересующие нас сведения об информационных волновых процессах в рассматриваемой системе могут быть получены с помощью функции Ψ. На данном этапе исследования для нас важны две характеристики волны: амплитуда Ψ и интенсивность $|\Psi|^2$. Более того, полученное в итоге решение, как среднее по ансамблю статистически идентичных лингвистических систем (из данного языкового сообщества), описывает по сути интересующую нас лингвистическую информацию I (скалярная функция), которая

определяется с помощью величины $|\Psi|^2$. Для удобства решения все величины могут быть приведены к безразмерному виду с использованием характерных параметров системы: размера L_c (критический размер области, занятый сообществом, ниже которого начинается его вымирание), частоты $f_c = 1/T_c$ (в качестве T_c может быть взято, например, время разделения двух языков, т.е. 500 лет), и времени T_c. Для модельных расчетов можно положить начальное значение $|\Psi_0|^2 = 1$, либо использовать полученные в первой (древовидной, см. (1)) модели значения I как некоторые начальные значения.

Математическая модель процесса волнового распространения и изменения информации в рассматриваемой системе описывается системой интегро-дифференциальных уравнений [1]:

$$\frac{\partial^2 a_n}{\partial z^2} - \beta_n \frac{\partial a_n}{\partial z} + \sum_n a_m B_{mn}(z) + \\ + \int q(\rho; z) D_n(\rho; z) d\rho = 0 \quad , \qquad (3)$$

где a и q – некоторые коэффициенты (собственные значения); β_n – комплексный волновой параметр системы; B_{mn}, D_n – некоторые функции, определяемые с помощью собственных (базисных) скалярных функций системы (см. подробнее, например, в [23]); пространственная координата z расположена вдоль направления распространения волны; параметр ρ, по которому производится интегрирование, является функцией параметра β_n и в общем случае тоже комплексный: $\rho = \rho(\beta_n)$.

Приближенное искомое решение ищется с помощью спектрального метода в следующем виде [1]:

$$\Psi(X;t) = \sum_n a_n(\rho;t)\Psi_n(X_n;t) + \\ + \int q(\rho;t)\Psi_p(X_p;t) d\rho \quad , \qquad (4)$$

где первое слагаемое в правой части (как и сумма по n в (3)) определяет величину «информации» в распределенной системе, обусловленную контактами «незнающих» со «знающими» (сумма по дискретному набору значений информации у ее $N < +\infty$ носителей); второе информационное слагаемое (аналогично интегралу по ρ в (3)) определяется как интеграл (т.е. в общем случае это может быть некоторый континуум, включающий только «незнающих»); $\rho \geq 0$; $X_{n;p}$ – некоторые функции, характеризующие свойства рельефа (где находятся исследуемые лингвистические сообщества), и оказывающие влияние на процесс распространения лингвистической информации.

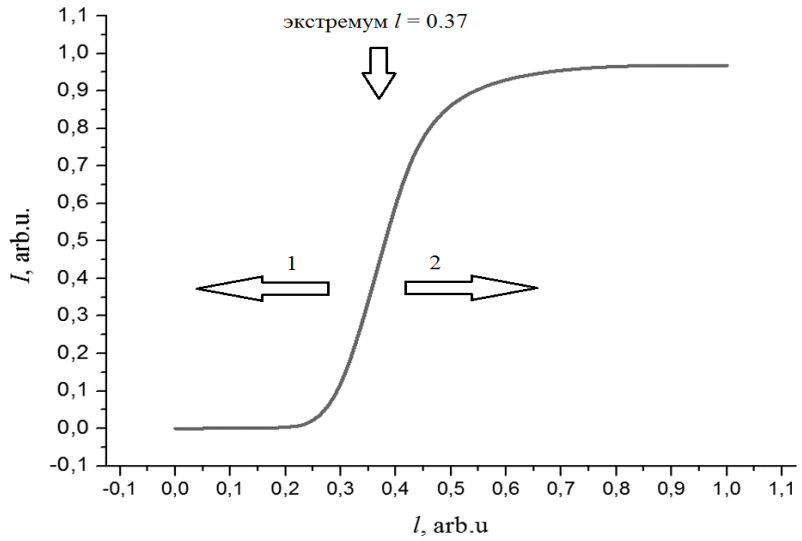
Коэффициенты a и q возможно рассматривать как некоторые эффективные амплитуды в соответствующих (дискретном и интегральном) разложениях. В общем случае коэффициенты a_n и q_i могут меняться как во времени, так и в зависимости от места (ландшафта) действия, описываемого с помощью функций X. Учитывая положительную определенность $I \geq 0$ [27], мы представляем информацию так: $I \propto |\Psi|^2$, поскольку в общем случае Ψ – волновая функция, которая по определению является комплекснозначной величиной; в частном случае это может быть действительная функция U, характеризующая некоторое возмущение (волну) в рассматриваемой системе. Заметим, что наш теоретический подход аналогичен подходам, использованным в работах [24–26, 28, 29] при построении математических моделей динамики полей различной природы.

Два типичных графика, характеризующие некоторые стационарные распределения информации (среднее по ансамблю статистически идентичных лингвистических систем), которые можно построить рамках комплексной нелинейной модели распространения лингвистической информации в системе, приведены на рис. 3.

Используя методы теории катастроф [3, 4], поясним характер поведения рассматриваемой открытой лингвистической системы в зависимости, например, от параметра системы l (нормированная длина участка, на котором распространяется исходная информация; соответствует некоторому нормированному числу поколений в данном сообществе). С этой целью исследуем график первой производной этой зависимости. Эту функцию можно рассматривать как некоторую потенциальную функцию диссипативной системы, имеющую экстремумы (локальные и глобальные). Для определения интервала значений функции, где модуль первой производной не превышает 1, проводятся горизонтальные линии +1 и -1. В этом интервале соблюдается условие устойчивости особой точки (точек), где $|dI/dl| \leq 1$.

Полученный на рис. 3*б* график первой производной соответствует известному случаю бифуркации состояния равновесия («частица» в потенциальной яме с барьером или полочкой). В этом месте возможны два состояния равновесия системы: при $l < 0.25$ и при $l > 0.75$, а между ними есть высокий потенциальный барьер. Здесь можно применить понятие фазового перехода, при котором происходит качественное изменение системы. Например, если система начинает движение из любой точки участка $0 < l < 0.35$, то итерации I_n сходятся к $I^* = 0$ – устойчивая неподвижная точка. Аналогично, если система начинает движение из любой точки участка $0.35 < l < 0.5$, то итерированные значения $I_n \to const$, т.е. динамический режим становится стационарным или имеет период, равный единице: возникает цикл S^1. Проведенные расчеты показали, что при вариации параметров в системе возможны также и другие циклы типа S^{2^p}.

а)



б)

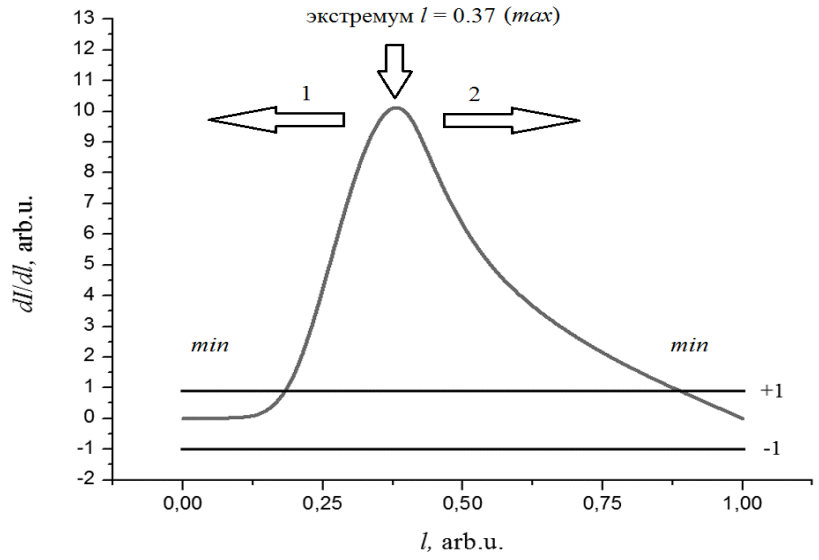


Рис. 3. Стационарное распределение информации в праиндоевроейском сообществе:
а – график зависимости функции $I(X)$ от нормированной величины l;
б – на графике первой производной зависимости, приведенной на рис. 3*а*, отмечены три экстремума: максимум в центре (l = 0.37) и два минимума по краям (слева и справа).

Для этой модели (рис. 3*а*) видно, что наиболее характерной чертой является известный в математической лингвистике логистический характер поведения моделируемой динамической системы [2, 5–8]. Полученные результаты позволяют, в частности, высказать предположение, что около 3500–4000 лет назад в рассматриваемом модельном индоевропейском языковом сообществе могли возникнуть две основные лингвистические популяции, характеризуемые сейчас как деление индоевропейских языков на так называемые языковые ареалы «сатем-кентум» [6, 8]. Действительно, если две лингвистических популяции одновременно начинают движение (последовательность итераций с шагом в 500 лет) в окрестности точки (ареала совместного проживания в рамках некоторого праиндоевропейского сообщества) экстремума l = 0.37 (см., например, рис. 3), одна слева (отмечено стрелочкой 1), а другая справа от нее (отмечено стрелочкой 2), они полностью расходятся примерно за 8–10 суммарных итераций (количество поколений в ансамбле лингвистических систем), что соответствует примерно 2000 г. до н.э. Необходимо отметить, что возможны и другие изоглоссы (линии на лингвистической карте, обозначающие границы распространения данного языкового явления), которые делят индоевропейские языки на две подгруппы иначе, чем «сатем-кентум» [8], что тоже может представлять интерес при исследовании. Заметим, что термины «сатем» и «кентум» происходят от слов, означающих числительное «сто» в репрезентативных

языках каждой группы (латинский *centum* и авестийский *satəm*). Делением сатем-кентум называется изоглосса в семье индоевропейских языков, относящаяся к эволюции трёх рядов дорсальных согласных (заднеязычные согласные; например, в русском языке к заднеязычным относятся согласные, обозначаемые буквами **г**, **к** и **х**), реконструированных для праиндоевропейского языка. Обычно к группе «сатем» относят индоиранские, балтославянские, албанский, армянский; возможно, какое-то число недокументированных мёртвых языков, таких как, например, фракийский и дакский. К группе «кентум» – италийские, кельтские, германские, греческий; возможно, небольшие и неизвестные группы мёртвых языков, таких как венетский, древний македонский, и, предположительно, иллирийские языки (см., например, [6–12]).

Приведем результаты одного сравнения данных, полученных по первой (древоподобной (1)) и второй (волновой (3)) моделям. Такие сравнения возможны, потому что в обеих моделях было задано одинаковое соотношение коэффициентов: λ_1 и λ_2 – в первой модели, и соответствующих им коэффициентов a_n и q_i – во второй модели. При компьютерном моделировании соотношение этих коэффициентов выбиралось также, исходя из соотношения средних скоростей движения всадников v_1 и пешеходов v_2 в праиндоевропейских сообществах, например: $a:q \propto v_1:v_2 \approx 15:3$.

На графиках (рис. 1*а* и рис. 2*а*) видно, что величина «лингвистической информации» в первой (ветвящейся) модели после перехода от хаотического «движения» динамической системы к периодическому (предположительно это время формирования собственно индоевропейских языков – этап их окончательного расхождения или *дивергенции*; начальный хаотический период соответствует некоторому периоду совместного развития, определенной *конвергенции* предков индоевропейских языков до их окончательного разделения) достигает первый раз максимума на 13–14-м поколениях. В волновой модели для ряда полученных графиков (характерных *S*-кривых), подобных графику на рис. 3*а*, аналогичная величина лингвистической информации, после перехода динамической системой критической точки l = 0.37, может достигать максимума примерно за 4–8 поколений в случае цикла S^2, и примерно за 8–48 итераций (т.е. не более $48 \cdot 500$ лет = 24000 лет – теоретически возможное максимальное время существования рассматриваемого языкового праиндоевропейского сообщества в рамках данной математической модели) в случае цикла S^1, т.е. в среднем ($(4+8+8+48)/4 = 17$ поколений) чуть больше, чем в первой модели (13-14 поколений). Таким образом, волновая модель по этому показателю включает в себя древовидную.

Так оценка, полученная для цикла S^1 в рамках волновой модели, для некоторого разумного возможного периода времени, когда, например, могло возникнуть исходное (архаичное, раннее) праиндоевропейское языковое сообщество, включавшее в себя также Анатолийскую и Тохарскую группу языков, а возможно и некоторые другие исчезнувшие группы праиндоевропейских языков: $26 \cdot 500$ лет = 13000 лет назад. Для расчета было взято число итераций, соответствующее выходу ряда рассчитанных для ансамбля зависимостей $I(l)$ примерно в центр правой части (на рис. 3*а* отмечено стрелочкой 2) анализируемых *S*-кривых. Аналогичная оценка, полученная для цикла S^2: $6 \cdot 500$ лет = 3000 лет назад. По этим двум циклам получаем среднюю оценку: 8000 лет назад. Эта оценка соответствует данным работ [5, 6, 14].

Приведем еще одну полезную оценку возможного времени возникновения исследуемого сообщества. С этой целью воспользуемся методикой определения времени диффузии T_{dif} системы в минимум (см. рис. 3*б*) по величине кривизны (т.е. по значению второй производной d^2I/dl^2) потенциальной функции системы в точках максимума и минимума [4]. Такая оценка использует тот факт, что исследуемая лингвистическая система рассматривается как динамическая система, возникшая в процессе самоорганизации исходных ранних праиндоевропейских сообществ, а затем, стремящаяся как любая диссипативная система в состояние исходного равновесия (см. подробнее в [4]), определяемого положением некоторого (глобального) минимума, когда развитие сообщества стабилизируется ($I \to const$, при $l \to +\infty$, см. рис. 3*а*). Для этого необходимо также определить величину «коэффициента диффузии» особей данного лингвистического сообщества, для чего достаточно оценить ее критическое значение для популяции по формуле $D_c = vL_c^2/\pi^2$, где v – линейная скорость роста популяции, L_c – критический размер области, занятый популяцией, ниже которого начинается ее вымирание, $\pi \approx 3.14$ [3, 4, 24, 25]. Тогда время возникновения исследуемого сообщества (примерно равное времени диффузии) можно определить по формуле:

$$T_{dif} \approx \left(2\pi/\sqrt{|\lambda_1\lambda_2|}\right)\exp\left(\Delta V/D_c\right),$$

где λ_1, λ_2 – кривизна d^2I/dl^2 функции в точках локального минимума и локального максимума, ΔV – скачок потенциальной функции (высота барьера) при переходе системы из локального максимума в локальный минимум.

Разумно определить L_c, исходя из размеров территории, на которой были разбросаны праиндоевропейские сообщества: $L_c \le$ 500–1000 км [20]. Например, рассмотрим так называемую «страну городов» – условное название территории на Южном Урале, в пределах которой найдены древние городища синташтинской культуры эпохи средней бронзы (около 2000 лет до н. э., т.е. сравнимо по времени с древними Египетскими пирамидами). Самым известным поселением «страны городов» является Аркаим. Городища были разбросаны по территории диаметром более 350 км, что и характеризует протяженность данной территории, которую занимали жившие там праиндоевропейские сообщества. Для шумерских

«номов» (мелкие государства), которые существовали около 4500–5000 лет назад, данные похожи: L_c не менее 63 км [30]. Скорее всего, величина L_c для подобных сообществ не превышала 50-100 км, что в принципе соответствует расстоянию, которое в светлое время суток мог проехать всадник.

В результате для заданных значений $\lambda_1 \approx 0.29$, $\lambda_2 \approx -0.22$, $\Delta V \approx 498 - 560$, L_c (см. значения выше), и $D_c \approx 21–24$ можем найти оценку для времени диффузии сообщества: $T_{dif} \approx 2600–6000$ лет. Здесь размерности всех величин были взяты в системе единиц СИ и учтено, что скорость роста человеческой популяции обратно пропорциональна длительности поколения. Такой большой разброс в определении величины T_{dif} определяется, в первую очередь, погрешностью нахождения величин D_c, λ_1, λ_2, ΔV и L_c. Подчеркнем, что имеющиеся исторические и археологические данные не позволяют определить величину L_c абсолютно точно (если это вообще возможно).

О ПРАРОДИНЕ НАРОДОВ НОСИТЕЛЕЙ ПРАИНДОЕВРОПЕЙСКОГО ЯЗЫКА

Результаты проведенного теоретического анализа и компьютерного моделирования позволили нам получить данные, определяющие места, где могла располагаться прародина народов (условно будем считать их древними европейцами (*Old Europeans*)) носителей праиндоевропейского языка или так называемая *Urheimat (Urhemeit)* – гипотетическая (лингвистическая?) прародина. Это пять основных групп, отмеченных на карте (см. рис. 4) как: I – ареал, включающий территории современных Польши, Германии, Чехии, Словакии, Беларуси, Литвы, Латвии, России, Дании и Швеции; II – ареал, включающий территории современных России, Украины, Молдавии и Казахстана; III – ареал, включающий территории современных Узбекистана, Таджикистана, Киргизстана, Туркменистана, Ирана, Афганистана, Пакистана и Индии; IV – ареал, включающий территорию «Страны городов» (Южный Урал, Казахстан); V – ареал, включающий часть территории современных Армении, Грузии и Турции (территория бывшего Армянского царства Багратидов).

Как видно из [31], в центральной и западной частях современной Турции было обнаружено присутствие гаплогруппы У-ДНК R1a, которая попала туда через Кавказ в Бронзовый век (примерно 3500–2000 до н.э.). Возможно, что ареалы IV и V возникли примерно в одно время (как подгруппы исходного праиндоевропейского сообщества), они, как I и III, тоже расположены симметрично относительно ареала II (предположительно II-й ареал отражает стадию самого раннего праиндоевропейского языкового сообщества, включавшего в себя также Анатолийскую (в последующем это ареал V с прилегающими к нему западными районами современной Турции) и Тохарскую группу языков (в последующем это частично ареал III и прилегающие к нему западные районы современного Китая), до их отделения от ареала II).

Для понимания полученных результатов, представленных на карте, приведем некоторые из свойств, которыми обладает стационарное распределение, характеризующее процесс распространения лингвистической информации в рассматриваемой системе, которое получается как среднее по ансамблю статистически идентичных лингвистических систем: 1) ограниченность (есть некоторые максимумы и минимумы); 2) симметричность; 3) асимптотическое убывание от максимального значения данного распределения.

Учитывая, что более предпочтительной в нашем анализе является Курганная гипотеза, в качестве исходного выбираем II-й ареал. Принимая во внимание свойства стационарного распределения, характеризующего рассматриваемую лингвистическую систему, мы можем найти области решения, удовлетворяющие, в частности, данным, приведенным при нахождении оценки возможного времени возникновения T_{dif} исследуемых праиндоевропейских сообществ. В результате получаем несколько групп решений, которые и объединены в ареалы I–V. При этом мы также учитываем, что, по крайней мере, некоторые ареалы должны совпасть с двумя известными регионами, где есть преобладание мужчин гаплогруппы У-ДНК R1a. Очевидно, что мы делаем дополнительное предположение, что носителями праиндоевропейских языков были, в первую очередь, мужчины с этой гаплогруппой. Здесь важно отметить, что ареалы I, III и IV отсутствовали в исходной постановке проблемы, а V-й ареал представляет только малую и периферийную часть Анатолии.

Как видно, решение поставленной проблемы оказалось таково, что отсутствовавшие сначала элементы были восстановлены в силу вышеуказанных свойств решения (ограниченность, симметричность, асимптотическое убывание). Начальная постановка проблемы о местоположении прародины народов носителей праиндоевропейского языка (языков) не дает нам возможности провести подробный анализ полученного нового результата с учетом возможных ареалов I, III и IV. Кратко можно сказать следующее. С одной стороны, нам удалось математически точно определить место, где собственно возник язык – в ареале II, при этом в моделях мы сравнивали только два возможных ареала его возникновения: ареалы II и V (в целом вся Анатолия). С другой стороны, мы не могли в рамках исходной модели точно определить, где располагается прародина праиндоевропейских народов. При этом исходим из того, что носители языка первоначально могли проживать не только в ареале II, где впоследствии этот язык мог сформироваться, т.е. вероятно это могли быть два разных процесса: лингвистический и генетический. Поэтому правомерно поставить вопрос: где была прародина соответствующих народов как биологического вида до того, как они стали носителями праиндоевропейских языков? Это могли быть, например, одновременно ареалы I, II и III, где вероятно возникли зачатки близкородственных диалектов, а затем в ареале II из них развился ранний праиндоевропейский язык.

Рис. 4. Карта, иллюстрирующая, места, где могла располагаться прародина праиндоевропейских народов.

Основываясь на исследованиях, приведенных в [5, 6, 14, 18, 19, 21, 22], можно сделать вывод, что ареалы I и III почти равноправны с точки зрения примерного времени возникновения гаплогруппы У-ДНК R1a, а с учетом наших данных следует предположить, что ареалы I и III вполне могли служить генетическими резервуарами для ареала II. При этом понятно, что природный ландшафт (в основном равнинный), несомненно, ставит ареал I в приоритет перед ареалом III (в основном горный ландшафт). Действительно, для представителей курганной культуры равнинный ландшафт, несомненно, являлся более предпочтительным (использование лошадей и повозок с лошадьми) по сравнению с горным ландшафтом. Здесь уместно заметить, что самая ранняя колесница (2026 г. до н.э.) найдена недалеко от «Страны городов» (в районе ареала IV). С учетом полученных нами временных оценок можно сделать вывод, что в рамках рассматриваемых моделей сообщество ареала IV существовало до его выделения в самостоятельный ареал, скорее всего, именно в пределах ареала II. Действительно, ареал IV датируется примерно рубежом III–II тысячелетия до н.э., либо началом II тысячелетия до н.э., т.е. эти ареалы были взаимосвязаны и после их разделения, между ними существовал обмен лингвистической (и генетической) информацией, а также новыми технологиями (колесницы и т.д.).

Все ареалы надо рассматривать как взаимосвязанные элементы одной динамической системы, однако их вклад в развитие праиндоевропейских сообществ, в первую очередь с точки зрения лингвистики, был, скорее всего, неравнозначным. В качестве грубого приближения можно считать, что основную роль в развитии праиндоевропейского языка играли первые три ареала. Именно вдоль этого «информационного канала» (его роль выполнял географический маршрут между этими ареалами, своего рода «праиндоевропейский путь» по аналогии с «шёлковым путем») в течение тысячелетий осуществлялась устойчивая передача лингвистической и генетической информации между ареалами рассматриваемой системы. По-видимому, массовые миграции народов с востока (гунны, середина IV века н.э.; тюркские племена, VIII–X вв.) и с севера (готы, первые века н.э.) прервали эти потоки.

В заключение этого раздела уместно напомнить о научной гипотезе черноморского потопа, согласно которой около 5600 г. до н.э. имел место масштабный катастрофический подъём уровня Чёрного моря [32], возможно ставший исторической основой легенд о Всемирном потопе. Вероятно, вследствие потопа под водой оказалась общая прародина индоевропейских и ряда других народов. В этой связи заметим, что когда II ареал распадается (при вариации параметров) на два ареала, то тот, который расположен в районе северного Причерноморья, всегда захватывает прилегающую к нему часть Чёрного моря. Возможно это совпадение, а возможно наша модель отражает реальное историческое событие.

Оценка числа возможных «символов» алфавита праиндоевропейцев

Рассмотрим, можно ли что-то сказать о гипотетическом праиндоевропейском алфавите, а именно – о числе его возможных «символов», «знаков» (или «букв») на основании полученных нами сведений.

С этой целью мы использовали данные численных расчетов возможного количества информации примерно на 5-7-й итерациях по m у графиков, приведенных на рис. 1*а* и рис. 2*а* после хаотического «движения», т.е. примерно 4000 ± 500 лет назад [1].

Для подобной оценки были использованы формулы для полной информации, содержащейся в некотором сообщении, учитывающие общее число символов n в сообщении (непрерывный сигнал заменен дискретной последовательностью отсчетов) и число различных символов $\Im$ алфавита. Для простоты под числом символов n может пониматься определенное число условных «символов», соответствующих числу знаков в некотором списке, например, в 100-словном списке Сводеша[1] (*Swadesh list*). Установлено, что разумная оценка $\Im_n$ ближе к диапазону: $\Im_n$ = 11 – 38 символов [1], что совпадает с известными в научной литературе данными по числу знаков в алфавитах некоторых языков (финикийское письмо – 22; греческий алфавит – 24; глаголица – 41; кириллица – 43; санскрит – 36).

ПРОБЛЕМА КОРРЕКТНОСТИ ЗАДАЧ РАССЕЯНИЯ ПРИ РАСПРОСТРАНЕНИИ ИНФОРМАЦИИ В ПРАИНДОЕВРОПЕЙСКИХ СООБЩЕСТВАХ

В общем случае волновое распространение информации сопровождается процессами ее рассеяния и диффузии в окрестности ареалов, куда доходят ее волны. Поэтому необходимо кратко затронуть проблемы корректности задач рассеяния информации в лингвистическом сообществе.

Доказательство существования и единственности решения поставленных задач при наличии помех основывается на существовании регуляризирующих операторов для интегральных уравнений [23]. В отсутствие помех (при точных входных данных) существует единственное решение поставленных задач. Более подробный анализ этих проблем выходит за рамки настоящей статьи.

Эти выводы хорошо соответствуют полученным нами результатам, когда в некоторый небольшой период времени существуют только циклы вида S^1, что означает, что в лингвистическом сообществе развивался один праиндоевропейский язык, например, в случае первой (древовидной) модели это было примерно 4500 лет назад (а 3500 лет назад появляется цикл S^2: было уже два языка), а во второй (волновой) модели – это было примерно 4300 лет назад (средняя оценка). Также примем во внимание оценку для времени диффузии исследуемого сообщества: $T_{dif} \approx$ 2600–6000 лет. В результате приходим к выводу, что в рассматриваемом случае самое позднее возможное время существования одного праиндоевропейского языка – 3800 лет назад, а самое раннее – около 5000 лет назад. Теперь учтем среднюю оценку, полученную для циклов S^1 и S^2 по второй модели, позволяющую определить время, когда могло возникнуть раннее праиндоевропейское языковое сообщество – 8000 лет назад. Эта оценка получена как некоторое среднее по ансамблю статистически идентичных лингвистических систем, когда близкие праиндоевропейские языки мало отличались друг от друга (например, диалекты). С учетом этой последней оценки получаем в среднем: 4850 лет назад – самое позднее время, и 5700 лет назад – самое раннее время существования одного праиндоевропейского языка.

ЗАКЛЮЧЕНИЕ

В настоящей работе приведены результаты теоретического анализа и компьютерного моделирования, которые показывают в частности хорошее соответствие двум основным гипотезам о формировании праиндоевропейцев: Анатолийской и Курганной. Необходимо также подчеркнуть, что есть хорошая корреляция наших данных с выводами ряда генетиков о возможном времени появления (примерно 14000-20000 лет назад) гаплогруппы У-ДНК R1a (мы предполагаем, что носителями праиндоевропейских языков были в первую очередь мужчины с данной гаплогруппой) (см., например, [18, 19, 21, 22, 31]). Отметим, что наши результаты хорошо соответствуют выводам других независимых исследователей (см., например, [5, 6, 14, 31]). А совокупность данных, полученных при исследовании обеих моделей, позволяет говорить о предпочтительности Курганной гипотезы.

Изложены теоретические принципы нового метода исследования лингвистических сообществ (и процессов распространения в них лингвистической информации) как динамических диссипативных систем. Впервые установлено, что можно рассматривать процесс распространения и изменения лингвистической информации не только как регулярный, но и как типично хаотический процесс. При этом число возникающих циклов в ветвящемся древовидном процессе (вследствие бифуркации удвоения периода) рассматривается как число возникших новых языков, в данном лингвистическом сообществе.

Впервые нами было предложено использовать для анализа распространения лингвистической информации в праиндоевропейских сообществах тот факт, что праиндоевропейские народы первыми в истории человечества одомашнили лошадь. Именно это преимущество всадников перед пешеходами и было положено нами в основу при построении теоретических моделей и компьютерном моделировании.

СПИСОК ЛИТЕРАТУРЫ

1. Egorova M., Egorov A., Solovieva T., Stezhenskaya L. On two models of distribution and changes of linguistic information in Indo-European model language communities // Proceed-

[1] Список Сводеша – предложенный американским лингвистом Моррисом Сводешем метод для оценки степени родства (близости) между различными языками по такому признаку, как схожесть наиболее устойчивого базового словаря. Как правило, это стандартизированный перечень базовых лексем какого-либо языка, упорядоченный по убыванию их «базовости» (исторической устойчивости к исчезновению из употребления). В стандартный список Сводеша на любом языке включают только наиболее простое, очевидное, базовое, современное значение слова. Минимальный набор важнейшей лексики содержится в 100-словном списке Сводеша. Например, первые 10 единиц в 100-словном списке Сводеша для русского языка: я; ты, вы; мы; это; то; кто; что; не; всё, все; много.

ings of the 4th International Multidisciplinary Scientific Conference on Social Sciences and Arts SGEM'2017, 28 – 31 March, 2017, Extended Scientific Sessions. – Vienna: STEF92 Technology Ltd; Sofia, 2017. – Book 3, Vol. 1. – P. 155-168. – URL: https://doi.org/10.5593/SGEMSOCIAL2017/HB31/S10.021
2. Егоров А.А., Егорова М.А. О моделях распространения лингвистической информации в языковом сообществе // Тез. докл. XXI Всерос. конф. "Теоретические основы конструирования численных алгоритмов и решение задач математической физики" (5–11 сентября 2016 г., г. Новороссийск, Абрау–Дюрсо). – М.: ИПМ им. М.В. Келдыша РАН, 2016. – С. 82-83.
3. Компьютеры и нелинейные явления: Информатика и современное естествознание / авт. предисл. А.А. Самарский. – М.: Наука, 1988.
4. Gilmore R. Catastrophe theory for scientists and engineers. – NY: Wiley, 1981.
5. Chang W., Cathcart C., Hall D., Garrett A. Ancestry-constrained phylogenetic analysis supports the Indo-European steppe hypothesis // Language. – 2015. – Vol. 91, № 1. – P. 194-244.
6. Anthony D.W. The Horse, the wheel, and language: how bronze-age riders from the Eurasian steppes shaped the modern world. – Princeton: Princeton University Press, 2007.
7. Kornai A. Mathematical Linguistics. – London: Springer, 2008.
8. Бурлак С.А., Старостин С.А. Сравнительно-историческое языкознание. – М.: Издательский центр «Академия», 2005.
9. Атлас языков мира. Происхождение и развитие языков во всем мире. – М.: Лик пресс, 1998.
10. Яхонтов С.Е. Оценка степени близости родственных языков / Теоретические основы классификации языков мира. – М.: Наука, 1980. – С. 148-157.
11. Ethnologue: languages of the world. – URL: https://www.ethnologue.com
12. Старостин Г.С. и др. К истокам языкового разнообразия. – М.: Издательский дом «Дело» РАНХиГС, 2015.
13. Gray R.D., Atkinson Q.D. Language-tree divergence times support the Anatolian theory of Indo-European origin // Nature. – 2003. – Vol. 426. – P. 435-439.
14. Pereltsvaig A., Lewis M.W. The Indo-European controversy: facts and fallacies in historical linguistics. – Cambridge: Cambridge University Press, 2015.
15. Михайлов А.П., Петров А.П., Маревцева Н.А., Третьякова И.В. Развитие модели распространения информации в социуме // Математическое моделирование. – 2014. – Т. 26, № 3. – С. 65-74.
16. Labov W. Transmission and diffusion // Language. – 2007. – Vol. 83. – P. 344-387.
17. Heggarty P., Maguire W., McMahon A. Splits or waves? Trees or webs? How divergence measures and network analysis can unravel language histories // Philosophical Transactions of the Royal Society B. – 2010. – Vol. 365. – P. 3829-3843.
18. Haak W., Lazaridis I., Patterson N., et. al. Massive migration from the steppe is a source for Indo-European languages in Europe // Nature. – 2015. – Vol. 522. – P. 207-211.
19. Lazaridis I., Patterson N., Mittnik A., et. al. Ancient human genomes suggest three ancestral populations for present-day Europeans // Nature. – 2014. – Vol. 513. – P. 409-428.
20. Аркаим – Синташта: древнее наследие Южного Урала // Сб. науч. тр. – Челябинск: Изд-во Челябинского гос. ун-та, 2010.
21. Underhill P.A., Myres N.M., Rootsi S., Metspalu M., Zhivotovsky M.A., King R.J., et al. Separating the post-Glacial coancestry of European and Asian Y chromosomes within haplogroup R1a // European Journal of Human Genetics. – 2010. – Vol. 18. – P. 479-484.
22. Klyosov A.A., Rozhanskii I.L. Haplogroup R1a as the Proto Indo-Europeans and the legendary Aryans as witnessed by the DNA of their current descendants // Advances in Anthropology. – 2012. – Vol. 2, № 1. – P. 1-13.
23. Арсенин В.Я. Методы математической физики и специальные функции. – М.: Наука, 1974.
24. Нахушев А.М. Уравнения математической биологии. – М.: ВШ, 1995.
25. Мари Дж. Нелинейные дифференциальные уравнения в биологии. – М.: Мир, 1983.
26. Гуц А.К., Коробицын В.В., Лаптев А.А., Паутова Л.А., Фролова Ю.В. Математические модели социальных систем. – Омск: ОГУ, 2000.
27. Яглом А.М., Яглом И.М. Вероятность и информация. – М.: Наука, 1973.
28. Mathematical modeling / eds. J.G. Andrews, R.R. McLone. – London: Butterworts, 1976.
29. Ли Ц.-Д. Математические методы в физике. – М.: Мир, 1965.
30. Гуляев В.И. Шумер. Вавилон. Ассирия: 5000 лет истории. – М.: Алетейя, 2004.
31. Cinnioglu C., King R., Kivisild T., Kalfoglu E., et al. Excavating Y-chromosome haplotype strata in Anatolia // Human Genetics. – 2004. – Vol. 114. – P. 127-148.
32. Ryan W., Pitman W. Noah's Flood: The new scientific discoveries about the event that changed history. – NY: Symon & Schuster, 1999.

Материал поступил в редакцию 10.02.19.

Сведения об авторах

ЕГОРОВА Майя Александровна – кандидат педагогических наук, доцент кафедры иностранных языков факультета гуманитарных и социальных наук Российского университета дружбы народов (РУДН), Москва.
e-mail: Mey1@list.ru

ЕГОРОВ Александр Алексеевич – доктор физико-математических наук, внештатный профессор-консультант РУДН.
e-mail: alexandr_egorov@mail.ru

ВНИМАНИЮ ЧИТАТЕЛЕЙ!

ВИНИТИ РАН, как единственный в России владелец лицензии Консорциума УДК, предлагает издания УДК полного четвертого издания на русском языке в печатном и электронном виде:

1. Таблицы УДК

УДК. Том I Общая методика применения УДК. Вспомогательные таблицы. Основные таблицы. Общий отдел. Алфавитно-предметный указатель к Общему отделу

УДК. Том II **1/3** Философия. Психология. Религия. Богословие. Общественные науки (только электронное издание)

УДК. Том III **5/54** Математика. Естественные науки (только электронное издание)

УДК. Том IV **55/59** Геологические и биологические науки (только электронное издание)

УДК. Том V **6/61** Медицинские науки (только электронное издание)

УДК. Том VI (часть 1) 6/621 Прикладные науки. Технология. Инженерное дело (только электронное издание)

УДК. Том VI (часть 2) 622/629 Техника. Инженерное дело (только электронное издание)

УДК. Алфавитно-предметный указатель к т. VI (1 и 2 части) (только электронное издание)

УДК. Том VII **63/65** Сельское хозяйство. Домоводство. Управление предприятием (только электронное издание)

УДК. Том VIII **66** Химическая технология. Химическая промышленность. Пищевая промышленность. Металлургия. Родственные отрасли (только электронное издание)

УДК. Том IX **67/69** Различные отрасли промышленности и ремесел. Строительство (только электронное издание)

УДК. Том X **7/9** Искусство. Спорт. Филология. География. История.

УДК. АПУ (с в о д н ы й) к полному 4-му изданию

УДК. Изменения и дополнения. Выпуск 2 (к т.т. 1–3) (только электронное издание)

УДК. Изменения и дополнения. Выпуск 3 (к т.т. 1–6) (только электронное издание)

УДК. Изменения и дополнения. Выпуск 4 (к т.т. 1–7) (только электронное издание)

УДК. Изменения и дополнения. Выпуск 5 (к т.т. 1–10)

УДК. Изменения и дополнения. Выпуск 6 (к т.т. 1–10)

УДК. Изменения и дополнения. Выпуск 7 (к т.т. 1–10), 2017 г. (только электронное издание)

Для подписки необходимо направить заявку по адресу:

125190, Россия, Москва, ул. Усиевича, 20, ВИНИТИ РАН

Телефоны: 499-155-42-85, 499-151-78-61

E-mail: feo@viniti.ru

ВНИМАНИЮ ПОДПИСЧИКОВ!

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

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

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

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

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

Факс: (499) 943-00-60;

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

ВНИМАНИЮ ЧИТАТЕЛЕЙ!

ИЗДАНИЕ УДК

УНИВЕРСАЛЬНАЯ ДЕСЯТИЧНАЯ КЛАССИФИКАЦИЯ

АЛФАВИТНО-ПРЕДМЕТНЫЙ УКАЗАТЕЛЬ

в 2-х томах

Алфавитно-предметный указатель (АПУ) к 4-му полному изданию УДК на русском языке:

Том I содержит АПУ от буквы А до Н;

Том II содержит АПУ от буквы М до Я и указатель латинских наименований к классам УДК 56 Палеонтология, 57 Биологические науки, 58 Ботаника, 49 Зоология, 61 Медицинские науки.

АПУ содержит около 100 000 понятий, представленных в полных таблицах УДК.

При его составлении были учтены изменения, опубликованные в Выпусках № 1 – 6 «Изменения и дополнения к УДК»

Для подписки необходимо направить заявку для оформления счета по адресу:

125190, Россия, Москва, ул. Усиевича, 20, ВИНИТИ РАН

Телефоны: 499 155-42-85, 499 151-78-61

E-mail: feo@viniti.ru

http://www.udcc.ru