

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

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

Издается с 1961 г.

№ 9

Москва 2020

ОБЩИЙ РАЗДЕЛ

УДК 002:004.80

А.П. Любимов

Основные подходы к определению понятия «искусственный интеллект»

Рассматриваются актуальные проблемы, связанные с определением понятия «искусственный интеллект»; используются различные отраслевые подходы, а также смежные термины, имеющие отношение к этому понятию. Показано, что за ИТ-технологиями и искусственным интеллектом – большое будущее, при условии объединения усилий и создания технологии совершенствования биологического человека с неограниченными геномными возможностями. Предполагается сотрудничество ученых и специалистов из различных областей – от гуманитариев до технарей, от медиков до селекционеров. Воплощенные в простых решениях технологии самой большой сложности будут управляться человеческим голосом и, в дальнейшем, – мозгом.

Ключевые слова: искусственный интеллект, информация, понятие, термин, определение искусственного интеллекта, естественное информационное поле, искусственное информационное поле, информационные модели, технологии, мозг

DOI: 10.36535/0548-0027-2020-09-1

ВВЕДЕНИЕ

Вопросы создания цифровых технологий и искусственного интеллекта сегодня приобретают новый, актуальный смысл и это характерно для всех развитых стран мира, в том числе и для России¹. Искусственный интеллект стал проникать во все сферы жизни человека и общества в целом.

Появление новых возможностей и выгод при расширении сферы использования искусственного интеллекта связано со многими сферами деятельности человека и его потребностями:

- 1) бытовая техника: от технологического искусственного охранника до умного дома;
- 2) компьютерная и телефонная техника. Компьютер и телефон как современный мобильный офис;
- 3) ИТ-технологии на транспорте и в связи;
- 4) ИТ-технологии в науке и образовании;
- 5) коллективный общественный интеллект, воплощенный искусственный интеллект в глобальных информационных и социальных сетях.
- 6) ИТ-технологии в экспертной деятельности, судопроизводстве, управлении и многое другое.

По мнению В.В. Рубанова, применение искусственного интеллекта в экспертной области сосредоточено на распознавании объектов и реагировании на изменения их состояний по некоторым шаблонам. Подобные технологии могут оказывать информационную поддержку экспертам и управленцам, но не заменять их [5]. Нам представляется, что невозможное сегодня, завтра будет обыденным явлением искусственного интеллекта. Это подтверждает деятельность нашего Нобелевского лауреата Ж.И. Алферова, который в начале работал с простейшими проводниками, а затем довел эти технологии до искусственного интеллекта [6].

ПРОБЛЕМЫ ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Если с ИТ-технологиями и искусственным интеллектом в бытовой сфере, а равно в промышленности для бытовых нужд, многое понятно, то в более рискованных и опасных сферах не все обстоит так просто. Об этом свидетельствуют примеры в области разработки вооружения, что приводит к опасному соперничеству и, соответственно, правовым ограничениям.

Различные проблемы уровня могут возникнуть во многих сферах деятельности человека: от экономической и военной, имплантации киберфизических систем в тело человека – до социальной сферы (например, при составлении социальных рейтингов) и создания до-

полненной реальности. В связи с этим необходимо определить само понятие «искусственный интеллект» как основы для создания цифрового рынка.

В настоящее время идет разработка Европейской стратегии развития цифровой экономики, в основе которой лежат новейшие информационно-коммуникационные технологии с использованием искусственного интеллекта [7].

Понятия, определяющие виды, сущность и опасности, сопровождающие всеобъемлющее распространение искусственного интеллекта, зависят от сферы его применения.

Так, применение искусственного интеллекта в сфере деятельности вооруженных сил опасно возможностью нарушения самого основного и неотъемлемого права человека – права на жизнь. Известно, что во всем мире новейшие достижения науки всегда брались в разработку и внедрение в первую очередь военными ведомствами. Однако предоставление возможности какой-либо военной структуре с искусственным интеллектом принимать решение о начале военных действий, об уничтожении людей или объектов создает проблему определения субъекта ответственности в случае ошибки при принятии таких решений, что формирует обширное поле для преступной безнаказанности. Из этого следует одно из требований, которое, на наш взгляд, должно быть применено в области действия искусственного интеллекта и закреплено в определении его понятия – действие структур с искусственным интеллектом должно происходить исключительно в рамках опознавательных факторов, к которым можно отнести: регистрацию, информацию о владельце и управлении, о фактах перехода контроля, в том числе путем удаленного доступа и т.п.

По мнению С.Ю. Кашкина и А.В. Покровского, другая перспективная сфера применения искусственного интеллекта – это развитие киберфизических систем, в том числе их имплантация в тело человека [8].

Киберфизические системы, состоящие из различных природных объектов, искусственных подсистем и управляющих контроллеров, призваны обеспечивать тесную связь между физическими процессами и компьютероинтегрированными встроенными устройствами или системами. Киберфизические системы возникают на стыке интернета людей, вещей и сервисов. Они распространяются практически на все виды человеческой деятельности, включая промышленные, транспортные, энергетические, военные и экономические системы, а также на все механизмы жизнеобеспечения: от медицины до создания умных домов и городов. Уже сейчас применение искусственного интеллекта в киберфизических системах усиливает аналитические способности человека посредством создания интерактивных систем нового уровня при сохранении человека в контуре управления.

Искусственный интеллект постепенно завоевывает все сферы нашей жизни. Переход человеческого разума в полностью виртуальное пространство уже не кажется мечтами фантастов. На наших глазах происходит слияние человеческого разума и компьютера. На базе развития идеологии философии трансгуманизма

¹ Указ Президента РФ от 10.10.2019 № 490 «О развитии искусственного интеллекта в Российской Федерации» (вместе с "Национальной стратегией развития искусственного интеллекта на период до 2030 года"); Указ Президента РФ от 01.12.2016 № 642 "О Стратегии научно-технологического развития Российской Федерации" // Собрание законодательства РФ, 05.12.2016, № 49, ст. 6887; Указ Президента РФ от 09.05.2017 № 203 «О Стратегии развития информационного общества в Российской Федерации на 2017–2030 годы» // Собрание законодательства РФ, 15.05.2017, № 20, ст. 2901.

идёт формирование нового, улучшенного постчеловека с искусственным интеллектом.

Однако переход к новой ступени эволюции путем модификации тела с использованием технологий искусственного интеллекта, предполагаемый трансгуманистами, оставляют без внимания вопросы духовности и возникает опасность утраты людьми тех качеств и свойств, которые формируют человечество как биологический вид. Чипы, встроенные в тело человека, уже сейчас позволяют идентифицировать людей при совершении ими покупок, при управлении автомобилем или системой «умный дом». Таким путем запущен процесс расчеловечивания личности, что неминуемо ведёт к расслоению общества, так как объективное отсутствие по разным причинам равномерного доступа к новейшим технологиям модификации тела создаёт социальное неравенство.

В 1997 г. Ник Бостом (шведский физик и философ) вместе с британским предпринимателем и философом Дэвидом Пирсом основали в Великобритании Всемирную Трансгуманистическую Ассоциацию, объединившую более 3 тыс. членов (в 2008 г. Ассоциация поменяла свое название на Humanity+). Эти трансгуманисты позиционируют себя приемниками античного гуманизма и утилитаризма, построенного на продвижение идеологии оправдания применения высоких технологий для продления жизни и достижения бессмертия финансовой олигархии. Акцент делается на расширении применения искусственного интеллекта, оцифровки человеческого сознания, крионике, а также на внедрении нанороботехники в медицине и т.п. В 2000 г. и в нашей стране (в Москве) было образовано Российское трансгуманистическое движение, представители которого позиционируют себя как «продолжатели дела Н.Ф. Фёдорова, К.Э. Циолковского, В.И. Вернадского и других представителей русского космизма» и пропагандируют в России «идеи научного иммортализма и трансгуманизма, а также новые исследования и технологии, которые позволяют каждому превзойти «естественные» пределы возможностей своего тела, своего разума, своей жизни» [9]. В США сформирована и действует Трансгуманистическая партия, и её представитель, Золтон Иштван, в 2016 г. даже балатировался на пост президента США. Эта партия является частью международного движения, в основе которого лежит мировоззрение, что биологический человек настоящего – это не окончательная ступень эволюции, а перспектива развития сферы применения технологий искусственного интеллекта ведет к неизбежности и целесообразности полной замены органов человека и отмены физического тела [10].

Однако ускоряющийся рост числа устройств со встроенными процессорами и средствами хранения большого объема данных приводит к лавинообразному увеличению объема информации, необходимой для принятия решений. Это вступает в противоречие с ограниченными когнитивными способностями человека, эволюционирующими медленнее, чем кибернетические возможности, в результате этого нарастает опасность вытеснения человека из процесса управления.

Необходимость обработки большого объема информации и возникновение глобального оборота данных обусловили разработку и внедрение новейших технологий искусственного интеллекта (ИИ). Поэтому при формулировании понятийного аппарата ИИ ключевой позицией является описание функции сбора и анализа огромного массива фактов для принятия на их основе необходимых оптимальных решений.

Для осуществления своей стратегии по искусственному интеллекту, принятой в апреле 2018 г., Европейская комиссия 7 декабря представила согласованный план, подготовленный вместе с государствами-членами для содействия развитию и использованию ИИ в Европе. В этом плане декларируется, что «... осмысливая огромные объемы данных для предложения эффективных решений, искусственный интеллект совершенствует продукты, процессы и бизнес-модели во всех секторах экономики» [11].

Европейский парламент ещё в 2017 г. обратился с рекомендацией разработать для Европейской комиссии по гражданско-правовому регулированию робототехники меры для «ограничения рисков, связанных с возможностью взлома или отключения оперативных систем программирования, встроенных в организм человека, либо стирания их памяти, что может поставить под угрозу здоровье человека, а в особых случаях даже его жизнь» [12].

ПРАВОВЫЕ ПРОБЛЕМЫ

Для защиты интеллектуальной собственности понятийный аппарат искусственного интеллекта должен включать и правовые основы, что особенно важно, так как неизбежное расширение сферы применения искусственного интеллекта предопределяет перспективы развития цивилизации и формирует главные ценности человечества. На материальный объект или услугу, создаваемые с использованием искусственного интеллекта, должны распространяться определенные права на интеллектуальную собственность, которые в перспективе станут основным источником современного синергетического и мультипликационного эффекта интеллектуального и материального производства.

Сложности в определении понятия искусственного интеллекта заключаются в том, что мы сталкиваемся с совершенно новым для человечества явлением, объединяющим человеческие характеристики (что предопределяет необходимость поднять вопрос о защите прав человека) и механические свойства искусственного интеллекта. Правосубъектность единицы искусственного интеллекта может быть определена только при четком разграничении ответственности между юридическим и физическим лицом. Поэтому возникает необходимость введения новой формы правосубъектности – электронное лицо [8].

Таким образом, становится очевидным, что для введения юридически значимого определения понятия «искусственный интеллект» (как для применения в процессе проектирования, производства и эксплуатации технических средств с искусственным интеллектом, так и для использования в процессе судебной деятельности) необходимо проработать не только существенные характеристики ИИ, обозначить воз-

возможные сферы применения, легальные объекты воздействия и механизмы защиты интеллектуальных прав на создаваемый продукт, но и ввести юридически значимые обязательства по соблюдению определенных стандартов безопасности с обеспечением должного уровня контроля за действием устройств с участием искусственного интеллекта.

При определении понятия искусственного интеллекта следует учитывать, что существующий класс ИТ-продуктов (к которым относятся программно-аппаратные системы и технологии автоматического и автоматизированного решения многообразных задач приема, передачи, накопления, обработки и управления данными и знаниями, в том числе с имитацией некоторых интеллектуальных функций человека, таких как распознавание видеообразов, анализ трехмерных изображений, обработка и перевод на разные языки текстов и речи, машинное обучение, выявление закономерностей, визуализация и анализ больших объемов данных, облачные вычисления и т.п.), по большому счету, не является в полной мере искусственным интеллектом, так как пока не обладает способностью человека «самоозадачиваться», т. е. не умеет сам себе ставить задачи, осознавая ситуацию и себя в ней. Для них применяются такие определения, как «слабый искусственный интеллект» или «интеллектуальные системы и технологии», поскольку они используют нарабатываемое и уже созданное в области искусственного интеллекта.

Специалисты в области теории управления, математической кибернетики и системного анализа, а также такие ученые, как С.Н. Васильев, Е.Н. Ведута, В.В. Рубанов, при определении ИИ выделяют в основном два понятия: «сильный» и «слабый» искусственный интеллект. К «сильному» ИИ они относят будущие системы и технологии со способностями не только автоматического достижения цели, но и самоозадачивания, то есть обладающие помимо реактивных и других регуляторных механизмов поведения также механизмами целеполагания [13, 14]. К «слабому» ИИ относят интеллектуальные человеко-машинные системы и технологии, в которых функция самоозадачивания реализуется человеком. По мнению В.В. Рубанова, те, кто непосредственно связан с предметными практиками, не спешат называть подобные программы и оснащенные ими устройства искусственным интеллектом, предпочитая более скромный термин – «высокоавтоматизированные системы». Термином «искусственный интеллект» чаще пользуются «институты развития» и бенефициары программ «цифровизации» – для обоснования затрат на проекты по соответствующей графе бюджета [5].

СТАНДАРТИЗАЦИЯ ПОНЯТИЙНОГО АППАРАТА

Возникает много вопросов по стандартизации понятийного аппарата искусственного интеллекта в различных отраслях экономики. В связи с этим, а также в целях реализации Федерального закона «О стандартизации в Российской Федерации»² и по-

вышения эффективности работ по стандартизации на национальном, межгосударственном и международном уровнях и по согласованию с заинтересованными организациями, приказом Росстандарта создан технический комитет по стандартизации «Искусственный интеллект»³. В соответствии с этим приказом за техническим комитетом закреплены объекты стандартизации с кодами ОКПД2, имеющие смежную область деятельности:

- оборудование коммуникационное (в области искусственного интеллекта);
- средства транспортные и оборудование, прочие (в области искусственного интеллекта);
- продукты программные и услуги по разработке программного обеспечения; консультационные и аналогичные услуги в области информационных технологий (в области искусственного интеллекта); а также многочисленные услуги:
- по изданию прочего программного обеспечения (в области искусственного интеллекта);
- в области информационных технологий (в области искусственного интеллекта);
- юридические (в области искусственного интеллекта);
- в области геологических, геофизических и взаимосвязанных изыскательных работ прочие, не включенные в другие группировки (в области искусственного интеллекта);
- в области технического регулирования, стандартизации, метрологии, аккредитации, каталогизации продукции (в области искусственного интеллекта);
- в области картографии (в области искусственного интеллекта);
- работы, связанные с научными исследованиями и экспериментальными разработками (в области искусственного интеллекта);
- по обеспечению безопасности и проведению расследований (в области искусственного интеллекта);
- в области образования (в области искусственного интеллекта);
- в области медицины прочие (в области искусственного интеллекта).

Закрепленные объекты стандартизации с кодами ОКПД2 по смежной области деятельности имеют свою отраслевую операциональную специфику в определениях. Термин «операциональные определения» ввел известный ученый по теории управления Эдвардс Деминг.

Приведенные выше примеры свидетельствуют о том, что предстоит большая и серьезная работа по системному выстраиванию сущностных характеристик и систематизации понятийного аппарата в области искусственного интеллекта для различных сфер и областей общественной и экономической деятельности. В этой работе возможны ошибки и неудачные решения. В.Я. Цветков писал об этом: «При

² Федеральный закон от 29 июня 2015 г. № 162-ФЗ "О стандартизации в Российской Федерации".

³ Приказ Росстандарта от 1 ноября 2019 г. № 2612 «Об утверждении Программы национальной стандартизации на 2020 год»; Приказ Росстандарта от 25 июля 2019 года № 1732 "О создании технического комитета по стандартизации "Искусственный интеллект".

получении информации человек может вносить погрешности в измерения за счет несовершенства приборов и методик измерения. Все это приводит к тому, что в результате познания человек создает свое собственное информационное поле, которое следует считать искусственным» [15].

ОРГАНИЗАЦИОННЫЕ И ГУМАНИТАРНЫЕ ПРОБЛЕМЫ

Возникает много вопросов и по другим проблемам искусственного интеллекта. Примерами могут служить:

- обозначение пределов внедрения технологий искусственного интеллекта в деятельность государственных органов власти;
- влияние искусственного интеллекта на стратегическую стабильность, меры повышения прозрачности в киберпространстве, современные и будущие технические разработки в сфере нераспространения и связанные с этим ядерные угрозы;
- создание в России научных центров мирового уровня в рамках национального проекта «Наука» предусматривает организацию до конца 2023 г. не менее девяти научных центров, выполняющих исследования и разработки по приоритетным научно-техническим областям.

Именно под таким углом зрения выступали члены РАН на осенней сессии Общего собрания РАН в 2018 г. [16].

Совет по государственной поддержке создания и развития научных центров утвердил перечень из семи приоритетных направлений развития науки⁴. В первую очередь это цифровые технологии и искусственный интеллект с использованием роботизированных системы на базе создания и применения материалов нового поколения.

По мнению специалистов есть опасность, что по мере того, как искусственный интеллект будет вытесняться, естественный интеллект (в своей массе) будет опускаться. У ученых возникает вопрос, зачем тогда прогнозная аналитика про будущее, когда массовый спрос будет формироваться рефлексом про настоящее?

Успехи в движении к искусственному интеллекту напрямую связаны с успехами в формировании языка человеко-машинных коммуникаций. Однако современные информационные системы пользуются инструментами времен «доисторического материализма»: классификаторами, словарями, тезаурусами и т.п. Их стали даже величаво называть онтологиями. Но на такой основе невозможно создание «умных» агентов и «размышляющего» искусственного интеллекта. Причина в том, что понятия словарей (семантика), не объединяемые причинно-следственными связями (синтаксис), не позволяют строить из них комбинации, несущие смысл (прагматика).

Возможно ли создание нового языка для цифровой реальности? В.В. Рубанов полагает, что эта актуальная задача хотя и суперсложна, но разрешима.

⁴ Постановление Правительства РФ от 30.04.2019 № 538 «О мерах государственной поддержки создания научных центров мирового уровня».

Следует только подчеркнуть, что для ее решения нужна не традиционная лингвистика, а относительно новая наука семиотика [5]. Считается, что подобный язык понятен человеку, а также легко переводим в алгоритмы для компьютерных вычислений.

ЗАКЛЮЧЕНИЕ

За ИТ-технологиями и искусственным интеллектом большое будущее. Ещё более могущественными они станут, если человечество и его лучшие представители объединят усилия и предложат технологии совершенствования биологического человека с большими геномными возможностями. Такое научное направление предполагает сотрудничество ученых и специалистов из различных областей: от гуманитариев – до технарей; от медиков – до селекционеров и т.д. [17].

Технологии самой значительной сложности должны быть воплощены в очень простых решениях, которые будут управляться человеческим голосом и мозгом: электронные очки, электронная авторучка, голосовой компьютер, планшет и т.д. В недалеком будущем наши технологии станут технически живыми!

СПИСОК ЛИТЕРАТУРЫ

1. Любимов А.П., Пономарева Д.В., Барабашев А.Г. К вопросу о понятии искусственного интеллекта в российском праве // Актуальные вопросы экономики, управления и права: сборник научных трудов (ежегодник). – 2019. – № 2-3. – С. 16-34.
2. Любимов А.П., Пономарева Д.В., Барабашев А.Г. Искусственный интеллект в европейской правовой доктрине // Вестник Дипломатической академии МИД России. Международное право. – 2019. – №1(10). – С. 32-47.
3. Любимов А.П. Достоинства и недочеты двух важных законопроектов: мнения экспертов. Круглый стол в Государственной Думе // Журнал российского права. – 2000. – № 4. – С. 26-27.
4. Любимов А.П., Пономарева Д.В., Барабашев А.Г. Основные понятия искусственного интеллекта. – Москва: ООО «Сам Полиграфист», 2019. – 116 с.
5. Рубанов В.В. Не подражать живым системам // Независимая газета 23.03.2020. – URL: http://www.ng.ru/stsenarii/2020-03-23/9_7824_consciousness.html (дата обращения: 23.04.2020).
6. Любимов А.П. Жорес Алферов - легенда мировой науки: от полупроводников до искусственного интеллекта // Представительная власть – XXI век. – 2019. – № 1-2. – С. 38.
7. Парфенова Л.Б. Европейская стратегия развития цифровой экономики: региональная дифференциация // Вестник Тверского государственного университета. Серия: Экономика и управление. – 2018. – № 3. – С. 30-38.
8. Кашкин С.Ю., Покровский А.В. Искусственный интеллект, робототехника и защита прав человека в Европейском союзе // Вестник Университета им. Кутафина (МГЮА) – 2019. – № 4(56). – С. 78.

9. Всемирная Трансгуманистическая Ассоциация / Российское трансгуманистическое движение. – URL: <http://transhumanism-russia.ru> (дата обращения: 20.05.2020).
10. Соснов А.Я. Роли для работа // Поиск. № 46. – 2019, 15 ноября. – С. 13.
11. Communication from the European Commission of 07.12.2018 // Coordinated Plan on Artificial Intelligence. – 2018. – COM 795. – P. 2.
12. European Parliament. Reports A8-0005/2017 of 27.01.2017 with recommendation to the Commission on Civil Law Rules on Robotics (2015/ 2103(INL)). – P. 15.
13. Васильев С.Н. Задача – озадачиваться. Искусственный интеллект научат ставить цели // Поиск. №50. – 2019, 13 декабря. – С. 10.
14. Ведута Е.Н., Любимов А.П., Джакубова Т.Н., Ряскова Е.С. Концепция национальной программы создания проактивного искусственного интеллекта // Представительная власть – XXI век. – 2019. – № 4. – С. 22-29.
15. Цветков В.Я. Естественное и искусственное информационное поле // Международный журнал прикладных и фундаментальных исследований. – 2014. – № 5, (часть 2). – С. 178
16. Любимов А.П., Щитов А.Н. Современные научно-технологические приоритеты РАН // Представительная власть – XXI век. – 2018. – №7-8. – С. 26-33.
17. Бабкин В.В., Промоненков В.К., Овчаренко М.М., Любимов А.П. Инновационная концепция средств защиты растений в Российской Федерации // Химическая промышленность сегодня. – 2017 – № 8. – С. 50-54.

Материал поступил в редакцию 16.06.20

Сведения об авторе

ЛЮБИМОВ Алексей Павлович – доктор юридических наук, профессор, заместитель Главного ученого секретаря Президиума РАН, Москва
e-mail: aplyubimov@presidium.ras.ru

Методология построения распределенной информационной системы поиска научно-технической информации на основе объектной модели данных

Рассматривается методология организации поиска научно-технической информации и создания соответствующей информационно-поисковой системы. Методическую основу составляет трехуровневая иерархическая структура онтологий. Первый уровень формирует понятийный аппарат предметной области, второй – детализирует терминологию конкретной области научных знаний. Онтологии третьего уровня ориентированы на определенное научное направление, позволяют структурировать объекты и методы исследования, а также научный результат и практическую ценность научной публикации. Используется объектная модель данных. Предложены основы языка информационных запросов.

Ключевые слова: онтологии, поиск научно-технической информации, архитектура информационной системы, объектная модель данных, язык информационных запросов

DOI: 10.36535/0548-0027-2020-09-2

ВВЕДЕНИЕ

Основу методического обеспечения поиска информации составляет ГОСТ 7.74-96 СИБИД. Информационно-поисковые языки. Термины и определения¹, созданный в 90-е гг. XX в., который позволил сформировать информационно-поисковый язык для индексирования документов, информационных запросов посредством понятий, рубрик, ключевых слов, классификационных признаков. С тех пор каких-либо существенных изменений в области стандартизации поиска научно-технической информации практически не проводилось. На процедуру поиска научно-технической информации существуют две точки зрения: поиск по кодам, например, УДК, ББК; поиск по ключевым словам. Если коды имеют жесткую структуру, десятки лет не меняются и достаточно сложно внести новую классификационную группировку, то ключевые слова формируются самими авторами, либо выбираются из известного списка,

содержащего сотни или тысячи различных слов. Например, раздел Информационная безопасность в системе УДК отсутствует, что приводит к необходимости классификации работ этой тематике по другим классификационным признакам. Система управления публикациями *ScholarOne* использует фиксированный список ключевых слов, из которого автор может выбрать наиболее подходящие.

Поскольку ни система классификации с применением кодов, ни ключевые слова не обеспечивают качественный поиск, существуют такие методы как тематические подборки материалов, группировка журналов и разделов в них по соответствующей тематике. В качестве примера можно привести аннотируемые издания ВИНТИ, или системы индексации научных работ, такие как *Scopus*, *Web of Science*, *Google Scholar* и др.

Таким образом, информационные запросы, построенные на классификационных признаках, могут выдавать достаточно широкий спектр публикаций, а ключевые слова могут выступать ограничителями для этой выборки. Однако в запросах иногда требуется найти информацию, которая не присутствует в ключевых словах. Информационно-поисковые системы в этом случае становятся малоприспособными. Современные ресурсы дают возможность в электронном виде получить доступ к тексту научно-технической информации, поэтому одним из направлений поиска стал

¹ Дата введения 1997-07-01. Разработан Всероссийским институтом научной и технической информации с участием Российской государственной библиотеки и ТК 191 "Научно-техническая информация, библиотечное и издательское дело", внесен Госстандартом России, принят Межгосударственным Советом по стандартизации, метрологии и сертификации (протокол № 9 от 12 апреля 1996 г.)

семантический анализ текста научной публикации. Эту работу выполняют программные боты – пауки или черви. Как правило, в тексте научной статьи разбираются семантические конструкции, оценивается совместная встречаемость слов и т.д. Такой подход основан на интеллектуальных методах обработки информации.

Современные методы информационного поиска и сравнения текстов учитывают следующие критерии при сопоставлении текстов друг с другом [1]:

1) совместную встречаемость, порядок слов запроса и расстояние между словами в текстах документов [2];

2) синтаксические и семантические связи между словами запроса [3, 4];

3) предметную область и тематику текстов [5-7];

4) отношения между словами текста в тезаурусе или онтологии предметной области (синонимические, родовидовые и др.) [8, 9].

На сегодняшний день функции полнотекстового поиска реализованы в большинстве поисковых систем общего назначения, таких как Яндекс, Гугл [10]. Подобные поисковые системы для повышения качества результатов используют большой набор критериев ранжирования документов, таких как учет географического положения пользователя, его предыдущие запросы и предпочтения, классификацию запроса по типу и т.д.. Однако не все эти критерии применимы для поиска научно-технической информации [11].

Перспективными методами поиска научно-технической информации считается построение семантических сетей и онтологий [12-18] Релевантность результатов семантического поиска информации определяется смысловым содержанием, а не синтаксисом и частотой встречаемости ключевых слов в документе.

Основополагающие принципы семантического поиска – представление данных и метаданных в модели *RDF*, использование методов *Linked Data* для представления информации в виде объектов и классов.

Онтологии используются для детальной формализации некоторой области знаний [15] в виде концептуальной схемы и позволяют определить отношения между понятиями, провести интеграцию экземпляров данных и сопоставление их с другими схемами данных.

Однако, как показано в работах [11, 14], на данный момент не существует единой методологии по созданию онтологий, отсутствуют установленные стандарты разработки онтологических моделей, что усложняет проблему повышения качества поиска научно-технической информации. Поэтому дальнейшее развитие теории и практики поиска научно-технической информации является актуальной задачей. В нашей работе предлагается использование объектной модели данных для описания структуры научной публикации и совершенствование на её основе методов поиска.

ОБЪКТНАЯ МОДЕЛЬ НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ

Объектную модель научно-технической информации будем представлять в виде древовидной структуры, где корневой вершиной выступает *DOI* статьи, а ветвями и листьями структуры являются основные параметры.

Будем различать две базовые структуры научно-технической информации – общие сведения и сведения о содержательной части. Общие сведения наследуются из стандарта СИБИД: УДК, название, авторы, аннотация, ключевые слова, организация, в которой выполнена работа, тематика, научная специальность, список использованных источников. Сведения о содержательной части включают элементы: объект исследования, методы исследования, научный результат, практическая полезность.

В общем случае структура научной статьи может быть представлена следующим кортежем:

$$(DOI, U, Type, N, A, D, KW, Org, S, Ref, B, Url, O, M, RS, P),$$

где: *DOI* – цифровой идентификатор статьи; *U* – УДК; *Type* – тип научной статьи; *N* – название статьи; *A* – авторы; *D* – аннотация; *KW* – ключевые слова; *Org* – организация, в которой выполнена работа; *S* – тематика или научная специальность; *Ref* – список использованной литературы; *B* – библиографическая ссылка; *Url* – ссылка на электронный ресурс публикации; *O* – объект исследования; *M* – методы исследования; *RS* – научный результат; *P* – практическая полезность; *T* – полный текст научной статьи.

Кортеж (*DOI, U, Type, N, A, D, KW, Org, S, Ref, B, Url*) содержит стандартные сведения о научной статье, которые могут быть использованы для классического синтаксического поиска по совпадению слов в запросе, либо по системе классификации.

Кортеж (*O, M, RS, P*) возможно применять как для синтаксического, так и для семантического поиска информации.

Семантический поиск требует реализации семантических технологий – онтологий, которые в объектном подходе реализуются через принцип наследования, множественность форм описания терминов и понятий реализуется через полиморфизм. [19-21]

Онтология представляется ориентированным графом:

$$G_E = (E, L), \text{ где } E = \{E^0, E^1, \dots, E^{O-1}\},$$

где: *E* – понятие предметной области; *L* – множество связей между понятиями. Каждая связь определяет отношения понятий. При определении отношений необходимо избегать зацикливания.

Понятие предметной области *E* представляет собой кортеж

$$(ID, Name, NV),$$

где: *ID* – идентификатор понятия; *Name* – наименование понятия; *NV* – характеристика понятия.

При полнотекстовом поиске используется само понятие, его характеристика и отношения между ними.

Для организации эффективного поиска научно-технической информации предлагается создание базы данных, в которую собираются сведения о содержательной части научной публикации. Поскольку каждое тематическое направление, область научных знаний характеризуется своим набором способов на-

учного исследования, а также терминологией, предлагается создавать специализированные методы работы с научным текстом для формирования проекции научной статьи в виде её содержания в соответствии с объектной моделью данных. Предполагается, что онтологии будут формироваться компетентным специалистом в конкретной области знаний, это позволит называть методы авторским пониманием научной сути статьи.

ПРИМЕНЕНИЕ ОНТОЛОГИЙ ДЛЯ ПОИСКА НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ

Онтологии могут стать эффективным инструментом индексации пространства, научно-технической информации [22–27]. Математически онтологии определяются как множество объектов и отношения между ними, а также процессы их изменения, которые определяются конкретной областью научных знаний:

$$D = O \cup R \cup M,$$

где: D – пространство научно технической информации; O – универсальное множество объектов; R – универсальное множество отношений; M – универсальное множество процессов изменения объектов и отношений.

При этом терминология предметной области T строится на основе нотационного бинарного отношения:

$$r = ((d : t : d \in D, t \in T),$$

где: t – термин, обозначающий элемент d предметной области.

Онтология имеет иерархическую структуру и согласно объектной парадигме представляется в виде множества классов, причем классы более нижнего уровня иерархии более детально описывают объекты пространства научно-технической информации. При этом классы могут содержать атрибуты, определяющие свойства, и внутреннюю структуру понятий, лежащих в основе классов.

Возможно, что онтология будет включать и экземпляры классов, т. е. такие классы, в которых установлены значения всех их атрибутов. Логические правила вывода при работе с онтологиями дают возможность манипулировать понятиями и данными гораздо эффективнее, позволяя проводить индексацию научно-технической информации.

По критерию ориентированности на специализацию и специфику терминологии пространства научно-технической информации выделяются: предметная, тематическая, авторская онтологии.

Предметные онтологии содержат данные об общепринятой терминологии предметной области. Онтологии могут формироваться в автоматическом и автоматизированном режимах, например по подборке научных статей, посвященных определенной тематике, либо использовать разработанные ранее онтологии.

Тематическая онтология – это часть предметной области с расширенной терминологией, ограниченной конкретной областью научных знаний.

Авторская онтология является узкоспециализированной, ориентирована на конкретное научное направление и должна обеспечивать функции предметной и тематической онтологий, иметь возможность описывать методы научного исследования, а также простоту научного результата и его практическую значимость. Авторские онтологии позволяют решать проблему поиска научно-технической информации по национальным ресурсам, представленным на различных языках [28–30]. Для отслеживания востребованности авторских онтологий предлагается применять технологию блок-чейн, которая в дальнейшем может быть использована для оценивания и монетизации авторского права ученого на онтологию.

По функциональному назначению выделяются онтологии:

- описывающие объекты и понятия области научных знаний и их видовые отношения;
- описывающие методы научных исследований;
- определяющие принадлежность научного результата;
- определяющие возможности применения научного результата в практической деятельности;
- определяющие семантические отношения между элементами других онтологий.

В соответствии с общепринятым подходом при разработке онтологии необходимо иметь глоссарий базовых понятий предметной области и их таксономические отношения, на которых строится иерархия классов и объектов пространства научно-технической информации, правила логических выводов, позволяющие оперировать данными, представленными в онтологии, и извлекать из онтологии объекты, методы и результаты научного исследования.

В основе создания онтологий лежат принципы: таким принципам как: эксплицитность, согласованность, расширяемость, минимальность.

Принцип эксплицитности предполагает явное объявление аксиом онтологии. Все понятия в онтологии должны иметь определение, а выбор того или иного термина для обозначения какого-либо понятия следует подкреплять объяснением. Любые неоднозначности, связанные с использованием естественного языка для описания терминов, подлежат выделению. Многозначность понятий минимизируется. Онтологии необходимо иметь внутреннюю связность и последовательность. Все аксиомы, содержащиеся в онтологии, а также логически следующие из них утверждения, должны быть непротиворечивыми. Для обеспечения принципа расширяемости следует убедиться, что добавление в онтологию новых терминов из этой же предметной области не ведёт к необходимости пересматривать имеющуюся схему понятий. Онтология разрабатывается по принципу необходимости и достаточности: в неё закладываются только те неотъемлемые термины предметной области, а также дополнительные термины, которые необходимы для полноценного описания ключевых понятий.

Весь процесс создания онтологии может быть разбит на четыре последовательных шага:

- 1) идентификация цели и охвата онтологии;
- 2) построение онтологии;
- 3) оценка онтологии;
- 4) создание документации.

На этапе идентификации цели и охвата онтологии специалисту предметной области необходимо ответить на вопросы о причине создания новой онтологии, выделить задачи и подзадачи, а также максимально определить группу пользователей онтологии.

Этап построения онтологии состоит из нескольких процессов. В первую очередь, перед разработчиком возникает задача идентификации ключевых концептов, выработки для них текстовые определения, выбора терминов для обозначения концептов. Также необходимо удостовериться в наличии соглашения о принятых решениях среди экспертов.

Этап оценки онтологии выполняется на основе трёх критериев: непротиворечивость, полнота и точность. Под первым понимается невозможность получить одновременно противоречащие друг другу заключения при условии корректных исходных данных. Критерий полноты подразумевает достаточность информации онтологии для разграничения значений всех её терминов. Под точностью понимается смысловое соответствие вводимых понятий коммуникативному замыслу автора, реальности, а так же их организация в рамках научной теории.

АРХИТЕКТУРА ИНФОРМАЦИОННО-ПОИСКОВОЙ СИСТЕМЫ

Информационно-поисковая система научно-технической информации основана на сервис-ориентированной архитектуре [31, 32], т.е. строится как набор независимых сервисов, обеспечивающих получение структурированных данных из подсистемы создания онтологий, в том числе распределенной индексированной базы данных, интеграцию с внешними источниками структурированных данных и цифрового пространства научно-технической информации.

На рисунке показана предлагаемая для создания информационно-поисковой системы архитектура: модуль создания онтологий; модуль интеграции с

внешними источниками структурированных данных; модуль интеграции с источниками исходных текстов научных публикаций; поисковый модуль; модуль формирования индексированной базы данных; распределенная индексированная база научных публикаций; база данных адресов и ссылок архивов научно-технической информации; модуль реализации запросов на поиск научно-технической информации; шина данных.

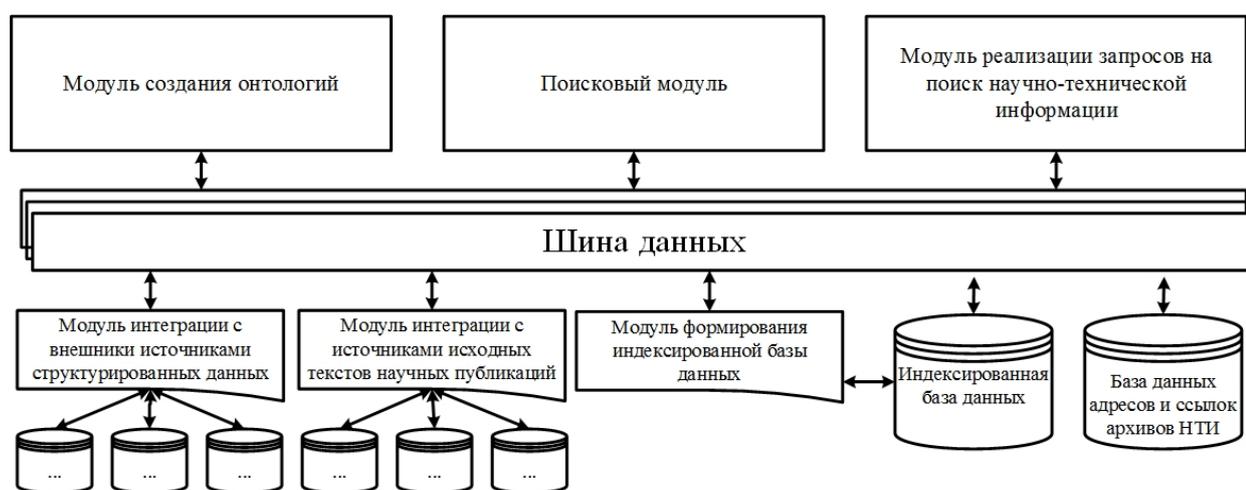
Модуль создания онтологий позволяет формировать новые онтологии понятий специалистам конкретных предметных областей, для их дальнейшего применения при решении задач поиска научно-технической информации. Поисковый модуль обрабатывает исходные тексты научно-технической информации на основе онтологий и формирует индексированную базу данных в соответствии с объектной моделью научной статьи. Составной частью модуля являются программы-пауки, осуществляющие непрерывный поиск.

Модуль запросов на поиск научно-технической информации осуществляет поиск по индексированной базе данных согласно критериям общих сведений о статье и содержательной части.

Шина данных выполняет функции интегрирующего элемента.

Модуль интеграции с внешними источниками структурированных данных позволяет использовать ранее созданные онтологии, хранящиеся на удаленных ресурсах. Модуль интеграции с источниками исходных текстов научных публикаций формирует базу данных электронных библиотек научных статей, обеспечивает доступ к исходным текстам научных статей.

Модуль формирования индексированной базы данных создает структуру распределенной объектной базы данных, которая может располагаться на ресурсах электронной библиотеки, журнала, издательства или автора. База данных адресов и ссылок архивов научно-технической информации необходима для функционирования информационно-поисковой системы.



Архитектура информационно-поисковой системы

ЯЗЫК ЗАПРОСОВ ДЛЯ ПОИСКА НАУЧНО-ТЕХНИЧЕСКОЙ ИНФОРМАЦИИ

Для поиска научно-технической информации необходим специализированный язык манипулирования данными, который строится на формальном математическом описании базовых моделей объектов на основе исчисления предикатов. Для формирования алгебраической системы использованы введенные ранее математические структуры.

Элементы множеств задают основное множество (Θ) алгебраической системы A и сигнатуры (Σ). Описание представленных множеств происходит в соответствии с объявленными элементами в структуре научной публикации.

Ключевые предикаты в предлагаемой алгебраической системе описывают объект, методы и результаты исследования, его практическую полезность.

Предикат, описывающий объект исследования, имеет вид:

$$obj_res(U, N, D, KW, S, T, E),$$

предикат определен на множестве:

$$O = U \times N \times D \times KW \times S \times T \times \{(u, n, d, kw, s, t) \mid u_j \in U, n_j \in N, kw \in KW, s_j \in S, t \in T\}$$

Предикат, описывающий метод исследования имеет следующий вид:

$$Met_res(U, N, D, KW, S, T, O, E),$$

предикат определен на множестве

$$M = U \times N \times D \times KW \times S \times T \times \{(u, n, d, kw, s, t, o) \mid u_j \in U, n_j \in N, kw_j \in KW, s_j \in S, t_j \in T, o_j \in O\}$$

Предикат, описывающий результат исследования имеет вид:

$$RS_res(U, N, D, KW, S, T, M),$$

предикат определен на множестве:

$$RS = U \times N \times D \times KW \times S \times T \times \{(u, n, d, kw, s, t, o) \mid u_j \in U, n_j \in N, kw_j \in KW, s_j \in S, t_j \in T, m_j \in M\}$$

Научный результат определяется отношением элементов двух множеств – объектов и методов исследования. Возможны четыре вида отношений:

- 1) OR_1 известный объект исследуется известными методами;
- 2) OR_2 новый объект исследуется известным методом;
- 3) OR_3 известный объект исследуется новыми методами;
- 4) OR_4 новый объект исследуется новыми методами.

Новые объекты и методы исследования описываются в авторских онтологиях.

Предикат, определяющий практическую полезность исследования имеет вид:

$$P_res(U, N, D, KW, S, T, RS, E),$$

предикат определен на множестве:

$$P = U \times N \times D \times KW \times S \times T \times RS \supset \{(u, n, d, kw, s, t, rs) \mid u_j \in U, n_j \in N, kw_j \in KW, s_j \in S, t_j \in T, rs_j \in RS\}$$

Множества O, M, RS, P функционально связаны между собой.

Выделение практической полезности из научной статьи проводится с помощью онтологий.

Таким образом, полученная алгебраическая система позволяет построить модель индексированной базы, для организации хранения структурированной научно-технической информации и перейти к специализированному языку поиска.

Язык запросов для поиска научно-технической информации состоит из двух компонентов: язык описания структур данных и формирования распределенной индексированной базы данных, язык поиска научно-технической информации на основе распределенной индексированной базы данных.

Первый компонент, включая, функции создания объектов индексированной базы данных, позволяет определять структуры для сущностей: предметная онтология, тематическая онтология, авторская онтология, структура научной статьи.

Второй компонент языка содержит описание следующих функций манипулирования данными:

- создание оригинала научной статьи – заполнение базовых сведений о научной статье;
- сематического анализа текста статьи. Входными параметрами является текст статьи, авторская онтология, выходными – объект исследования, методы исследования, научные результаты, практическая полезность.

Таблица 1

Базовые функции языка описания структур данных

Описание функции	Предикат
Создание структуры публикации	$Struct_art(O, K)$
Создание онтологии	$prop_type(H, T)$
Добавление в структуру онтологии нового понятия	$obj_name(O, B)$
Назначение имени и характеристики понятия	$prop_name(H, A)$
Добавление в структуру онтологии другой онтологии	$inserted_sd(O, N, \bar{O}, L)$

Базовые функции языка манипулирования данными

Название функции	Аргументы	Значение функции	Обозначение
Добавление новой статьи	Общие сведения о статье	Идентификатор экземпляра статьи либо код ошибки	InstanceArt (DOI, PropKeyInTree, Value)
Запуск процесса анализа статьи для извлечения объекта исследования, методов исследования, научных результатов, практической полезности	Идентификатор статьи, набор онтологий, используемых при анализе	Идентификатор объекта исследования, набор идентификаторов методов исследования, идентификатор научного результата, идентификатор практической полезности, фактическое время получения данных либо код ошибки	StartAnalysis (DOI, Ont, O, M, RS, P)
Поиск научных публикаций по семантическим характеристикам: объект исследования, методы исследования, научный результат, практическая полезность	Объект исследования, набор методов исследования, научный результат, практическая полезность	Набор публикаций, соответствующих условиям поиска, либо код ошибки	Search (O, M, RS, P)

В табл. 1 представлены предикаты и соответствующие им функции языка запросов.

В табл. 2 представлены базовые функции добавления, извлечения данных.

ЗАКЛЮЧЕНИЕ

В настоящей статье представлена методология построения информационно-поисковой системы научнотехнической информации, основанная на трех видах онтологий: предметная, тематическая, авторская. Особенность последней состоит в том, что она может начинаться с любого места дерева предметной или тематической онтологии понятий и к одной и той же ветви родительской онтологии могут быть привязаны несколько авторских. Авторская онтология позволяет определять объект, методы, результаты исследования и практическую полезность научной публикации.

Система создается на основе объектной модели данных, которая включает общие сведения и содержательную часть, имеет открытую архитектуру и не ограничивает участников проекта по созданию авторских онтологий.

Предложена структура и базовые функции языка информационных запросов.

Рассмотренная информационная система позволяет формировать индексированную распределенную базу данных и организовывать взаимодействие с пространством научно-технической информации.

СПИСОК ЛИТЕРАТУРЫ

- Manning C., Raghavan P., Schütze H. Introduction to information retrieval. – New York: Cambridge University Press, 2008.
- Chi T., Tezuka T., Oyama, S., Tajima K., Tanaka K. Web search improvement based on proximity and density of multiple keywords // Data Engineering Workshops. – Edinburgh: IEEE Press, 2006. DOI: 10.1109/ICDEW.2006.164.
- Shen D, Lapata M. Using semantic roles to improve question answering // Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL). – 2007. – URL: <http://www.aclweb.org/anthology/D/D07/D07-1002> (дата обращения 23.03.2020).
- Stenichikova, S., Hakkani-Tur D., Tur G. QASR: spoken question answering using semantic role labeling // ASRU-2005, 9th biannual IEEE workshop on Automatic Speech Recognition and Understanding. – Edinburgh: IEEE Press, 2005.
- Hwang M., Kong H., Baek S., Kim P. TSM: topic selection method of web documents // Modeling & Simulation, 2007. AMS '07. First Asia International Conference. – Edinburgh: IEEE Press, 2007. – P. 369-374. DOI: 10.1109/AMS.2007.108.
- Lewandowski D. Query types and search topics of German Web search engine users // Information Services & Use. – 2006. – Vol. 26, № 4. – P. 261-269.
- Wang Y., Li H., Wang H., Zhu K.Q. Toward topic search on the Web // Proceedings of ER. – Beijing: Microsoft Research, 2012.
- Cheng T., Lauw H. W., Pappas S. Entity synonyms for structured Web search // IEEE Transactions on Knowledge and Data Engineering. – 2012. – Vol. 24, № 10. – P. 1862-1875. DOI: 10.1109/TKDE.2011.168.

9. Wei X., Peng F., Tseng H., Lu Y., Wang X., Dumoulin B. Search with synonyms: problems and solutions // *International Conference on Computational Linguistics*. – Springer Nature Singapore Pte Ltd, 2010. – P.1318-1326
10. Paananen A. Comparative analysis of yandex and google search engines // *Metropolia Ammatikorkeakoulu* // Master's thesis. 2012. – URL: <http://publications.theseus.fi/handle/10024/46483> (дата обращения 23.03.2013).
11. Синкевич Е.А. Процесс обработки информации поисковыми программами в сети Internet // *Информационно-экономические аспекты стандартизации и технического регулирования*. – 2018. – № 6(46). – С. 30-36.
12. Суворов Р.Е., Соченков И.В. Определение связанности научно-технических документов на основе характеристики тематической значимости // *Искусственный интеллект и принятие решений*. – 2013. – №1. – С. 33-40.
13. Тихомиров И.А. Вопросно-ответный поиск в интеллектуальной поисковой системе Eхactus // *Труды четвертого российского семинара по оценке методов информационного поиска РОМИП'2006*. – Санкт-Петербург: НУ ЦСИ, 2006. – С. 80-85.
14. Комаров А.Ю. Онтологии на графах // *Эпистемы : сб. научных статей. Вып. 12: Интеллектуальные системы: онтология, эпистемология и праксеология*. – Екатеринбург: Изд-во «Деловая книга»; Издательский дом «Юника», 2017. – С. 16-22.
15. Щекочихин О.В., Шведенко В.В., Шведенко П.В. Онтология понятий информационной системы с поведением // *Научно-технический вестник Поволжья*. – 2016. – № 5. – С. 223-226.
16. Maedche A., Staab S. Semi-automatic engineering of ontologies from text // *Proceedings of the 12th international conference on software engineering and knowledge engineering*. – Berlin, Heidelberg: Springer-Verlag, 2000. – P. 231-239.
17. Maedche A., Staab S. Ontology learning for the semantic web // *IEEE Intelligent Systems*. – 2001. – № 16(2). – P. 72-79. DOI: <https://doi.org/10.1109/5254.920602>.
18. Navigli R., Velardi P., Gangemi A. Ontology learning and its application to automated terminology translation // *IEEE Intelligent Systems*. – 2003. – № 18(1). – P. 22-31. – DOI: <https://doi.org/10.1109/MIS.2003.1179190>
19. Шведенко В.Н., Шведенко В.В., Щекочихин О.В. Применение структурного полиморфизма при создании информационных систем процессного управления // *Научно-техническая информация. Сер. 2*. – 2018. – №11. – С. 9-15; Shvedenko V.N., Shvedenko V.V., Shchekochikhin O.V. Using Structural Polymorphism in Creating Process-Based Management Information Systems // *Automatic Documentation and Mathematical Linguistics*. – 2018. – Vol. 52, № 6. – P. 81-84.
20. Шведенко В.Н., Шведенко В.В., Щекочихин О.В. Использование структурного и параметрического полиморфизма при создании цифровых двойников // *Научно-техническая информация. Сер. 2*. – 2019. – № 3. – С. 21-24; Shvedenko V.N., Shvedenko V.V., Shchekochikhin O.V. Using structural and parametric polymorphism in the creation of digital twins // *Automatic Documentation and Mathematical Linguistics*. – 2019. – Vol. 53, № 2. – P. 81-84.
21. Шведенко В.Н., Щекочихин О.В., Шведенко В.В. Теоретические основы создания цифровых двойников с использованием методов полиморфизма // *Информационно-экономические аспекты стандартизации и технического регулирования: Научный интернет-журнал*. 2019. – № 1(47). – URL: [http://iea.gostinfo.ru/magazine_2019_01\(47\).html](http://iea.gostinfo.ru/magazine_2019_01(47).html).
22. Клещев А.С., Артемьева И.Л. Математические модели онтологий предметных областей. Ч.1. Существующие подходы к определению понятия «онтология» // *Научно-техническая информация. Сер. 2*. – 2001. – № 2. – С. 20–27.
23. Антонов И. В., Воронов М. В. Метод автоматизированного построения онтологии предметной области // *Моделирование и анализ данных*. – 2011. – № 1. – С. 116-130.
24. Lee C.S., Kao Y.F., Kuo Y.H., Wang M.H. Automated ontology construction for unstructured text documents // *Data and Knowledge Engineering*. – 2007. – № 60(3). – P. 547-566. DOI: <https://doi.org/10.1016/j.datak.2006.04.001>
25. Ashino T. Materials ontology: an infrastructure for exchanging materials information and knowledge // *DataScience Journal*. – 2010. – № 9(1). – P. 54–61. DOI: <https://doi.org/10.2481/dsj.008-041>
26. Weng S., Tsai H., Liu S., Hsu C. Ontology construction for information classification // *Expert Systems with Applications*. – 2006. – № 31(1). – P. 1–12. DOI: <https://doi.org/10.1016/j.eswa.2005.09.007>.
27. Choi S.P. Myaeng S.H. Terminological paraphrase extraction from scientific literature based on predicate argument tuples // *Journal of Information Science*. – 2012. – № 38(6). – 2012. – P. 593–611. DOI: <https://doi.org/10.1177/0165551512459920>.
28. Zhang C., Wu D. Bilingual terminology extraction using multi-level termhood // *The Electronic Library*. – 2012. – № 30(2). – P. 295–309. DOI: <https://doi.org/10.1108/02640471211221395>.
29. Marciniak M. Mykowiecka A. Terminology extraction from medical texts in Polish // *Journals of Biomed Semantics*. – 2014. – № 5. – P. 1–14. DOI: <https://doi.org/10.1186/2041-1480-5-24>
30. Macken L., Lefever E., Hoste V. Taxis: bilingual terminology extraction from parallel corpora using chunk-based alignment // *Terminology. International Journal of Theoretical and Applied Issues in Specialized Communication*. – 2013. – Vol. 19, Iss. 1, P. 1–30. DOI: <https://doi.org/10.1075/term.19.1.01mac>

31. Sellami Z., Camps V. DYNAMO-MAS: A Multi-Agent System for Building and Evolving Ontologies from Text. // Advances on Practical Applications of Agents and Multi-Agent Systems. – Berlin, Heidelberg: Springer, 2012. DOI: https://doi.org/10.1007/978-3-642-28786-2_38
32. Шведенко В.Н., Щекочихин О.В., Черкасова Н.В. Поиск архитектурного решения информационного обеспечения цифрового двойника сложной системы // Научно-техническая информация. Сер. 2. – 2020. – № 4. – С. 18-21. DOI: 10.36535/0548-0027-2020-04-3.

Материал поступил в редакцию 01.06.2020

Сведения об авторах

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

ЩЕКОЧИХИН Олег Владимирович – кандидат технических наук, доцент, инженер информационной безопасности ООО "ММТР", г. Кострома
e-mail: slim700@yandex.ru

СИНКЕВИЧ Евгений Александрович – аспирант ВИНТИ РАН
e-mail: 503-s@mail.ru

Аналитические модели парирования негативных внешних воздействий на сетевую информационную систему

Для повышения безопасности сетевых информационных систем введены три группы средств парирования негативных внешних воздействий: однородные, разнородные единичного парирования, разнородные двойного парирования. Для каждой из этих групп поставлены задачи по обеспечению максимальной вероятности сохранения сетевой информационной системы, разработаны и описаны аналитические модели их решения, определены процедуры решений, содержащие последовательность конкретных шагов. Представлена обобщённая процедурная модель сетевой информационной системы, взаимодействующей с системой поддержки принятия решений.

Ключевые слова: сетевая информационная система, негативные внешние воздействия, средства парирования, аналитические модели

DOI: 10.36535/0548-0027-2020-09-3

ВВЕДЕНИЕ

Современное общество характеризуется широким применением сетевых информационных систем для поддержки принятия решений, обеспечивающих сбор, хранение, обработку, поиск и выдачу информации лицу, принимающему решения. Однако часто на такие системы, особенно в сфере государственного и муниципального управления, военного и двойного назначения, образования и др. оказывается влияние, относящееся к классу негативных внешних воздействий, приводящее к снижению показателя устойчивости функционирования таких систем. Особенно часто это происходит в условиях неполноты априорных сведений, а также из-за воздействия на каналы информационных потоков, на работоспособность элементов системы, на передаваемую информацию [1-8].

Оценка устойчивости функционирования сетевых информационных систем является особо важной и сложной проблемой, которая не поддаётся точному количественному выражению. В качестве противодействия негативным внешним воздействиям используются средства парирования [1-9]. На устойчивость функционирования систем под действием негативных внешних воздействий влияет множество факторов, которые требуют соблюдения серии условий для средств парирования. Во-первых, набор этих средств в идеальном случае (обеспечение максимальной устойчивости функционирования) должен быть таким,

чтобы перекрыть все пути и возможности влияния негативных внешних воздействий на систему. Во-вторых, особую значимость имеет показатель сопротивляемости средств парирования попыткам их преодоления или обхода [10, 11]. В-третьих, в случае успешного негативного внешнего воздействия на систему необходимо учитывать величину ущерба, наносимого её пользователю. Так как эти три приведённых фактора трудно формализуемы, процесс выбора средств парирования, позволяющих повысить устойчивость функционирования системы, является сложной нечёткой задачей для соответствующего лица, принимающего решения, ответственного за риски нанесения значительного материального и репутационного ущерба пользователю системы.

Для решения задач повышения безопасности функционирования сетевых информационных систем для поддержки принятия решения введём три группы средств парирования негативных внешних воздействий: 1) однородные средства парирования; 2) разнородные средства единичного парирования; 3) разнородные средства двойного парирования. Для каждой из этих групп нами разработаны аналитические модели процедур парирования [12-14], позволяющие максимизировать устойчивость функционирования сетевых информационных систем с помощью системы поддержки принятия решений [15]. Процедурная модель этих систем показана на рис. 1.

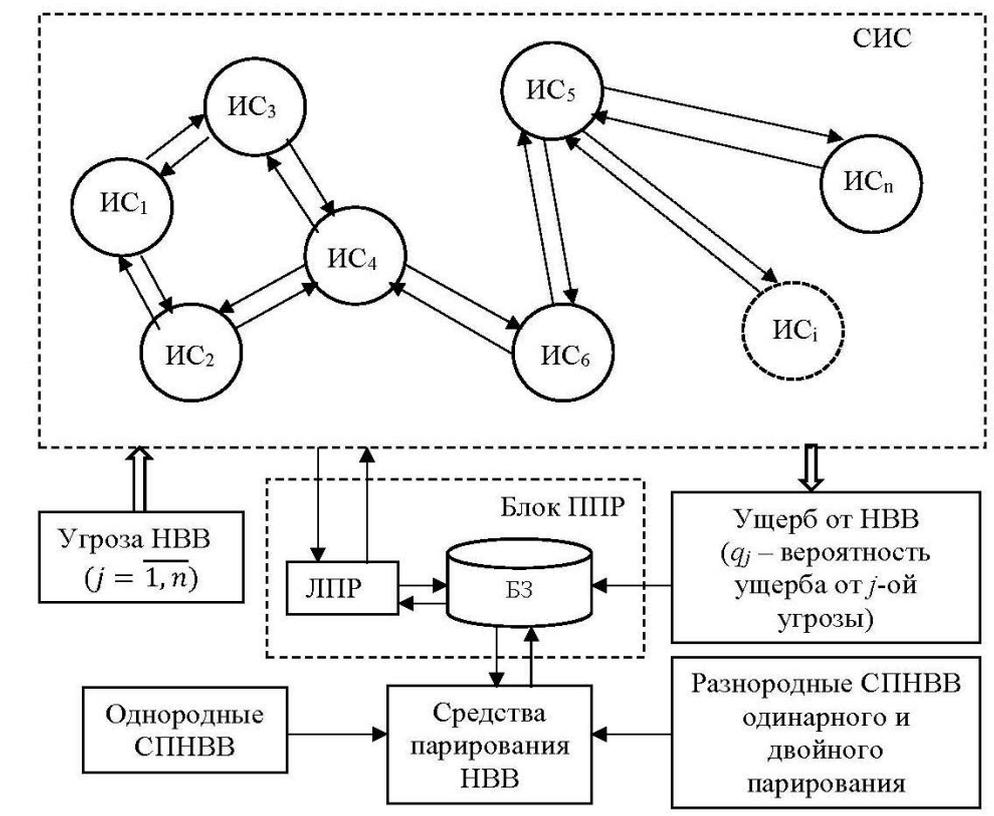


Рис. 1. Обобщённая процедурная модель сетевой информационной системы с блоком поддержки принятия решений: СИС – сетевая информационная система, ИС_i – локальные информационные системы, НВВ – негативные внешние воздействия, ППР – поддержки принятия решений, ЛПР – лицо, принимающее решения, БЗ – база знаний, СПНВВ – средства парирования негативных внешних воздействий

ПОСТАНОВКА ЗАДАЧ

Задача 1. Сетевая информационная система подвергается воздействиям j -х угроз, которые наносят ущерб с вероятностью ω_j . В системе поддержки принятия решений предусмотрена совокупность средств парирования негативных внешних воздействий. Необходимо обеспечить максимальную вероятность сохранения системы путём распределения однородных средств единичного парирования. Вероятность парирования j -й угрозы $p_j = 1 - q_j$, где q_j – вероятность отсутствия парирования. Для этого требуется найти вектор $Y = \{y_j\}_1^n$, при котором функция $f(Y)$ достигает \max :

$$f(Y) = \prod_{j=1}^n (1 - \omega_j q_j^{y_j}) \rightarrow \max, Y = \{y_j\}_1^n$$

при ограничениях:

$$\begin{aligned} \sum_{j=1}^n y_j &\leq m; y_j \in \{0, 1, 2, \dots, m\}; \\ 1 \geq q_j = (1 - p_j) &\geq 0; j = \overline{1, n} \\ 1 \geq (\omega_j = 1 - \varepsilon_j) &\geq 0 \end{aligned}$$

Задача 2. Сетевая информационная система подвергается воздействиям j -х угроз, которые причиняют ущерб с вероятностью ω_j . Необходимо обеспечить максимальную вероятность сохранения системы путём распределения m разнородных средств единичного парирования (матрица γ_{rj}). Возможности разнородных средств определяются элементами матрицы γ_{rj} . Для этого требуется найти матрицу $\gamma_0 = \|\gamma_{rj}^0\|$, обеспечивающую \max функции:

$$f(\gamma) = \prod_{j=1}^n \left(1 - \omega_j \prod_{r=1}^m q_{rj}^{\gamma_{rj}}\right) \rightarrow \max, \\ r = \overline{1, m}, j = \overline{1, n}$$

при ограничениях:

$$\begin{aligned} \sum_{j=1}^n \gamma_{rj} &= 1, \gamma_{rj} \in \{0, 1\}; 1 \geq (q_j = 1 - p_j) \geq 0, r = \overline{1, m}, \\ 1 \geq (q_\omega = 1 - \varepsilon_j) &\geq 0, j = \overline{1, n} \end{aligned}$$

Задача 3. Сетевая информационная система подвергается воздействию и угрозе, причиняющей ущерб с вероятностью ω_j . Необходимо обеспечить максимальную

ную вероятность сохранения системы из разнородных средств двойного парирования (γ_{rj}). При этом при парировании i -й угрозы с вероятностью α_{ij} парируется и j -я угроза, $j = \overline{1, n}$. Для этого необходимо найти матрицу $\gamma_0 = \|\gamma_{rj}^0\|$, обеспечивающую max целевой функции:

$$f(\gamma) = \prod_{j=1}^n \left(1 - \omega_j \prod_{i=1}^m \left[1 - \prod_{r=1}^m q_{rj}^{\gamma_{rj}} \alpha_{ij} \right] \right) \rightarrow \max,$$

при ограничениях:

$$\begin{aligned} \sum_{i=1}^n \gamma_{ri} &= 1, r = \overline{1, n}; \\ \gamma_{rj} &\in \{0, 1\}; 1 \geq (q_{ri} = 1 - p_{ri}) \geq 0, r, i = \overline{1, n}; \\ 1 \geq (\omega_j = 1 - \varepsilon_j) &\geq 0; 1 \geq \alpha_{ij} \geq 0 \end{aligned}$$

АНАЛИТИЧЕСКИЕ МОДЕЛИ РЕШЕНИЯ ПОСТАВЛЕННЫХ ЗАДАЧ

Задача 1. Представим целевую функцию в аддитивном виде:

$$R(y) = \sum_{j=1}^n \ln(1 - \omega_j q_j^{y_j}).$$

Пусть на t -м шаге процесса распределения ресурсов j -я угроза парируется с вероятностью $P_j^{(t-1)} = 1 - Q_j^{(t-1)}$, тогда значения функции R_{t-1}^+ равны следующей величине:

$$R_{t-1}^+ = \sum_{j=1}^n \ln(1 - \omega_j Q_j^{(t-1)}).$$

Далее:

$$\begin{aligned} R_t^+ &= \sum_{j=1}^n \ln(1 - \omega_j Q_j^{(t-1)}) + \ln(1 - \omega_j Q_j^{(t-1)} q_j) = \\ &= \sum_{j=1}^n \ln(1 - \omega_j Q_j^{(t-1)}) + \ln(1 - \omega_j Q_j^{(t-1)} q_j) - \\ &- \ln(1 - \omega_j Q_j^{(t-1)}). \end{aligned}$$

Приращение функции R_t^+ определим следующим образом:

$$\tilde{\Delta} t^+ = R_t^+ - R_{t-1}^+ = \ln \frac{1 - \omega_j Q_j^{(t-1)} q_j}{1 - \omega_j Q_j^{(t-1)}}, j = \overline{1, n}.$$

Полученное выражение запишем в упрощённом виде:

$$\tilde{\Delta} t^+ = \ln \frac{1 - \omega_j^{(t-1)} q_j}{1 - \omega_j^{(t-1)}}, j = \overline{1, n}; \omega_j^{(t-1)} \equiv \omega_j Q_j^{(t-1)}.$$

Это приращение соответствует полной вероятности парирования j -го негативного внешнего воздействия.

Аналогично имеем:

$$\omega_j^{(t)} = \omega_j^{(t-1)} q_j.$$

Таким образом, обеспечивается максимальная вероятность сохранения сетевой информационной системы однородными средствами единичного парирования негативных внешних воздействий.

Задача 2. Представим целевую функцию в виде:

$$R(j) = \sum_{j=1}^n \ln(1 - \omega_j \prod_{r=1}^m q_{rj}^{\gamma_{rj}}).$$

Получим формулы для приращения функций R_t^+ , R_t^- :

$$\tilde{\Delta} t_{kl}^+ = \ln \frac{1 - \omega_l Q_l^{(t-1)} q_{kl}}{1 - \omega_l Q_l^{(t-1)}} = \ln \frac{1 - \omega_l^{(t-1)} q_{kl}}{1 - \omega_l^{(t-1)}}.$$

Введём в рассмотрение матрицы:

$$\delta_{t-1}^- = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{vmatrix} \quad \delta_t^- = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{vmatrix}$$

Для R_{t-1}^- получим выражение вида:

$$\begin{aligned} R_{t-1}^- &= \sum_{j=1}^n \ln(1 - \omega_j^{(t-1)} b_j^{(t-1)}); \\ b_j^{(t-1)} &= \prod_{r \in m(t)} q_{rj} \omega_j^{(t-1)} = \omega_j Q_j^{(t-1)}, j = \overline{1, n}, \end{aligned}$$

где: $m(t)$ – массив нераспределённых к моменту t единиц парирования.

$$R_t^- = \sum_{j \neq t} \ln \left(1 - \omega_j^{(t-1)} \frac{b_j^{(t-1)}}{q_{kj}} \right) + \ln(1 - \omega_t^{(t-1)} b_t^{(t-1)}),$$

откуда следует:

$$\begin{aligned} \Delta_{kl}^- &= K_t^- - K_{t-1}^- = \sum_{j \neq l} \ln \frac{A}{B}, \\ A &\equiv 1 - \frac{\omega_l^{(t-1)} b_j^{(t-1)}}{q_{kj}}, B \equiv 1 - \omega_j^{(t-1)} b_j^{(t-1)}. \end{aligned}$$

Суммарное приращение определим следующим образом:

$$\tilde{\Delta}_{kl}^{(t)} = \Delta_{kl}^+ - \Delta_{kl}^- = \ln \frac{A}{B};$$

$$A \equiv \left(1 - \omega_i^{(t-1)} q_{kl}\right) \left(1 - \omega_i^{(t-1)} b_i^{(t-1)}\right) \prod_{j=1}^n \left(1 - \frac{\omega_j^{(t-1)} b_j^{(t-1)}}{q_{kj}}\right);$$

$$B \equiv \left(1 - \omega_i^{(t-1)}\right) \left(1 - \frac{\omega_i^{(t-1)} b_i^{(t-1)}}{q_{kl}}\right) \prod_{j=1}^n \left(1 - \omega_j^{(t-1)} b_j^{(t-1)}\right)$$

Так как распределение единиц парирования осуществляется в соответствии с *max* приращения, то будем использовать короткое выражение:

$$\Delta_{kl}^{(t)} = \frac{A}{B} \prod_{j=1}^n \left(1 - \frac{\omega_j^{(t-1)} b_j^{(t-1)}}{q_{kj}}\right);$$

$$A \equiv \left(1 - \omega_i^{(t-1)} q_{kl}\right) \left(1 - \omega_i^{(t-1)} b_i^{(t-1)}\right);$$

$$B \equiv \left(1 - \omega_i^{(t-1)}\right) \left(1 - \frac{\omega_i^{(t-1)} b_i^{(t-1)}}{q_{kl}}\right); k \in m^{(t)}, j = \overline{1, n}, l = \overline{1, n}.$$

Таким образом, обеспечивается максимальная вероятность сохранения сетевой информационной системы разнородными средствами единичного парирования негативных внешних воздействий.

Задача 3. Рассмотрим аддитивный аналог целевой функции:

$$F(\gamma) = \sum_{j=1}^n \omega_j \left\{1 - \prod_{i=1}^m \left[1 - \left(1 - \prod_{r=1}^m q_{rj}^{\gamma_r}\right) \alpha_{ij}\right]\right\};$$

$$R(j) = \sum_{j=1}^n \ln \left[1 - \omega_j \prod_{i=1}^m \left(1 - P_i^{(t-1)} \alpha_{ij}\right)\right],$$

Здесь $P_i^{(t-1)}$ – вероятность парирования i -ой угрозы на шаге $t-1$, тогда

$$R_{t-1}^+ = \sum_{j=1}^n \ln \left[1 - \omega_j \prod_{i=1}^m \left(1 - P_i^{(t-1)} \alpha_{ij}\right)\right];$$

$$R_t^+ = \sum_{j=1}^n \ln \left[1 - \omega_j \prod_{i=1}^m \left(1 - P_i^{(t)} \alpha_{ij}\right)\right],$$

где:

$$P_i^{(t)} = \begin{cases} P_i^{(t-1)}, & i \neq l \\ 1 - Q_i^{(t-1)} q_{ki}, & i = l \end{cases}; P_i^{(0)} = 0, i = \overline{1, n};$$

$$Q_i^{(t)} = 1 - P_i^{(t)}, i = \overline{1, n}$$

Приращение опишем соотношением:

$$\tilde{\Delta}_{kl}^+ = R_t^+ - R_{t-1}^+ = \sum_{j=1}^n \ln \frac{1 - \omega_j \prod_{i=1}^m \left(1 - P_i^{(t)} \alpha_{ij}\right)}{1 - \omega_j \prod_{i=1}^m \left(1 - P_i^{(t-1)} \alpha_{ij}\right)}.$$

Смежные состояния функции потерь выведем из выражений:

$$R_{t-1}^- = \sum_{j=1}^n \ln \left\{1 - \omega_j \prod_{i=1}^m \left[1 - \left(1 - Q_i^{(t-1)} b_i^{(t-1)}\right) \alpha_{ij}\right]\right\},$$

$$R_t^- = \sum_{j=1}^n \ln \left\{1 - \omega_j \prod_{i=1}^m \left[1 - \left(1 - Q_i^{(t)} b_i^{(t)}\right) \alpha_{ij}\right]\right\},$$

где $b_i^{(t)} = \frac{b_i^{(t-1)}}{q_{ki}}$; $b_i^{(0)} = \prod_{r=1}^m q_{ri}$, $i = \overline{1, n}$.

Формула, определяющая приращение, имеет вид:

$$\Delta_{kl}^{(t)} = \prod_{j=1}^n \left[1 - \omega_j \prod_{i=1}^m \left(1 - P_i^{(t)} \alpha_{ij}\right)\right] \left\{1 - \omega_j \prod_{i=1}^m \left[1 - \left(1 - Q_i^{(t)} b_i^{(t)}\right) \alpha_{ij}\right]\right\}.$$

Таким образом, обеспечивается максимальная вероятность сохранения системы разнородными средствами двойного парирования негативных внешних воздействий.

ПРОЦЕДУРЫ РЕШЕНИЯ ПОСТАВЛЕННЫХ ЗАДАЧ

Процедура решения задачи 1.

1. Вычислить компоненты текущего вектора:

$$\left\{\Delta_l^+\right\}^{(t)} l = \overline{1, n}, \omega_l^{(t)} = \begin{cases} \omega_l^{(t-1)}, & l \neq l_t, \omega_l^{(0)} = \omega_l \\ \omega_l^{(t-1)} q_l, & l = l_t, l = \overline{1, n} \end{cases}.$$

2. Закрепить средство парирования за $l = l_t$ негативным воздействием:

$$\Delta_{l_t}^+ = \max \Delta_l^+; 1 \leq l \leq n.$$

3. Вычислить значение R_t^+ :

$$R_t^+ = R_{t-1}^+ + \ln \Delta_{l_t}^+; R_0^+ = \sum_{j=1}^n \ln \varepsilon_j; \left(Q_j^{(0)} = 1, \forall j\right).$$

4. Вычислить компоненты $y^{(t)}$:

$$y_j^{(t)} = \begin{cases} y_j^{(t-1)}, & j \neq l_t \\ y_j^{(t-1)} + 1, & j = l_t \\ y_j^0 = 0, & \forall j = \overline{1, n} \end{cases}$$

5. Проверить для шага $t=t+1$ условие $t \leq m$. Если «да», то перейти к шагу 1, если «нет» – к шагу 6.

6. Сохранить результат:

$$R(y_0) = R_m, \{y_j^0\} = \{y_j^{(m)}\}.$$

Процедура решения задачи 2.

1. Вычислить элементы матрицы: $\|\Delta_{kl}\|$, $k = \overline{1, m}$, $l = \overline{1, n}$; $\omega_j^{(0)} = \omega_j$, $b_j^{(0)} = \prod_{r=1}^m q_{rj}$, $j = \overline{1, n}$.

2. Выбрать пару индексов k_t, l_t из условия: $\Delta_{k_t l_t} = \max_{k, l} \Delta_{kl}$, $l = \overline{1, n}$.

3. Вычислить $\omega_j^{(t)}, b_j^{(t)}$:

$$\omega_j^{(t)} = \begin{cases} \omega_i^{(t-1)}, j \neq l_t \\ \omega_j^{(t-1)} q_{k_t}, j = l_t \end{cases}; b_j^{(t)} = \frac{b_j^{(t-1)}}{q_{k_t}}, j = \overline{1, n}.$$

4. Проверить условие $t \leq m$. Если «да», то перейти к шагу 1, если «нет» – к шагу 5.

5. Вычислить $f(j_0)$.

6. Сохранить результаты в виде $f(j_0), \gamma_{r_j}^0$.

Процедура решения задачи 3.

1. Вычислить начальные значения:

$$P_i^{(0)}, b_i^{(0)}: P_i^{(0)} = 0; b_i^{(0)} = \prod_{r=1}^m q_{ri}; i = \overline{1, n}$$

2. Вычислить элементы матрицы: $\|\Delta_{kl}^{(t)}\|, k \in m^{(t)}, l = \overline{1, n}$

3. Выбрать пару индексов k_t, l_t из условия $\Delta_{k_t l_t}^{(t)} = \max_{k, l} \Delta_{kl}^{(t)}; k \in m^{(t)}, l = \overline{1, n}$.

4. Пересчитать вероятности:

$$P_i^{(t)} = \begin{cases} P_i^{(t-1)}, i = l_t \\ 1 - Q_i^{(t-1)} q_{k_t}, i = l_t \end{cases}; b_i^{(t)} = \frac{b_i^{(t-1)}}{q_{k_t}}, i = \overline{1, n},$$

$$t = t + 1.$$

5. Проверить соблюдение условия $t \leq m$. Если «да», перейти к шагу 2, если «нет», то к шагу 6.

6. Вычислить значение целевой функции.

7. Сохранить результаты $f(j_0), j_0$.

ЗАКЛЮЧЕНИЕ

Модели и процедуры решения поставленных задач позволяют аналитически моделировать различные ситуации парирования внешних воздействий на сетевую информационную систему. Проектировать такие системы возможно одновременно с блоком поддержки принятия решений и блоком, содержащим средства парирования негативных внешних воздействий. Надёжность функционирования таких сетевых информационных систем будет обеспечиваться с максимально возможной вероятностью.

СПИСОК ЛИТЕРАТУРЫ

1. Tyutyunnik V.M. Thermodynamics of irreversible processes and self-organization in information flows // International Journal of Advanced Research in IT and Engineering. – 2019. – Vol. 8, № 3. – P. 18-24.
2. Тютюнник В.М., Александров Е.Ю. Нарушения информационной безопасности в информационных системах с учётом распознавания объектов воздействия системами защиты информации в условиях неполноты априорных сведений // Приборы и системы. Управление, контроль, диагностика. – 2018. – № 10. – С. 10-17.

3. Александров Е.Ю., Тютюнник В.М. Моделирование оценки устойчивости сетевой информационной системы к негативным внешним воздействиям в условиях неполноты априорных сведений // Глобальный научный потенциал. – 2015. – № 9(54). – С. 102-106.
4. Александров Е.Ю., Тютюнник В.М. Методы анализа конфликтов между системами защиты информации и объектами воздействия // Наука и бизнес: пути развития. – 2019. – № 2(92). – С. 159-161.
5. Даммаг М.А.М., Тютюнник В.М. Аналитические и процедурные нечёткие модели для обработки многомерных данных // Наука и бизнес: пути развития. – 2019. – № 2(92). – С. 155-158.
6. Александров Е.Ю. Распознавание объектов системами защиты информации в условиях неполноты априорных сведений // Информационные системы и процессы: сб. науч. тр. – Тамбов; М.; СПб, Баку; Вена; Гамбург; Стокгольм: изд-во МИНЦ «Нобелистика», 2016. – Вып.15. – С. 7-12.
7. Gromov Y.Y., Karpov I.G., Didrikh V.E., Minin Y.V., Ivanova O.G. Application of linear pure birth-death processes for network-centric information systems modeling // Journal of Theoretical and Applied Information Technology. – 2016. – Vol. 85, №1. – P. 69-73.
8. Gromov Y.Y., Karpov I.G., Minin Y.V., Ivanova O.G. Generalized probabilistic description of homogeneous flows of events for solving informational security problems // Journal of Theoretical and Applied Information Technology. – 2016. – Vol. 87, № 2. – P. 250-254.
9. Gromov Y.Y., Minin, Y.V., Ivanova, O.G., Majeed H.L. Construction of continuous and discrete distribution laws for risk assessment in information systems // Journal of Engineering Science and Technology Review. – 2020. – Vol.13, № 3. – P. 1-7. DOI: 10.25103/jestr.133.01.
10. Громов Ю.Ю., Иванова О.Г., Ивановский М.А., Минин Ю.В., Александров Е.Ю., Тютюнник В.М. Применение тензорного анализа к построению информационных систем // Прикладная физика и математика. – 2018. – №5. – С. 40-44.
11. Громов Ю.Ю., Иванова О.Г., Ивановский М.А., Минин Ю.В., Александров Е.Ю., Тютюнник В.М. Применение тензорного анализа к концептуальному проектированию баз данных // Промышленные АСУ и контроллеры. – 2018. – № 11. – С. 37-43.
12. Gromov Y.Y., Karpov I.G., Minin Y.V., Ivanova O.G. Generalized probabilistic description of homogeneous flows of events for solving informational security problems // Journal of Theoretical and Applied Information Technology. – 2016. – Vol. 87, № 2 – P. 250-254.

13. Gromov Y.Y., Ishchuk I.N., Alekseev V.V., Didrikh V.E., Tyutyunnik V.M. Information support of finding a solution the problem of hidden objects // Journal of Theoretical and Applied Information Technology. – 2017. – Vol. 95, № 3. – P. 615-620.
14. Gromov Y.Y., Ishchuk I.N., Tyapkin V.N., Alekseev V.V., Tyutyunnik V.M. Information support of unmanned aerial vehicles navigation using pseudolites // Journal of Theoretical and Applied Information Technology. – 2017. – Vol. 95, № 3. – P. 570-581.
15. Сергеев А.Ю., Тютюнник В.М. Методика оценки и повышения эффективности тематико-ориентированного интернет-поиска с помощью минимизации объема поисковой выборки, обеспечивающей тематическую полноту поиска // Научно-техническая информация. Сер. 2. – 2012. – №7. – С. 18-36.

Материал поступил в редакцию 20.05.2020

Сведения об авторах

ТЮТЮННИК Вячеслав Михайлович – доктор технических наук, профессор, Заслуженный работник культуры РФ, президент и генеральный директор Международного Информационного Нобелевского Центра (МИНЦ), профессор Московского государственного института культуры, Тамбовского государственного технического университета.
e-mail: vmtutyunnik@gmail.com

ГРОМОВ Юрий Юрьевич – доктор технических наук, профессор, директор Института автоматизации и информационных технологий Тамбовского государственного технического университета.
e-mail: gromovtambov@yandex.ru

АЛЕКСАНДРОВ Евгений Юрьевич – научный сотрудник Международного Информационного Нобелевского Центра, г. Тамбов.
e-mail: evgale@gmail.com

УДК 004.89:616-006

Д.К. Чебанов, И.Н. Михайлова

О методах искусственного интеллекта для анализа онкологических данных*

Проводится краткий обзор методов искусственного интеллекта, применяемых к медицинским данным, связанным с онкологией. Перечисляются актуальные цели использования искусственного интеллекта – типы решаемых с его помощью прикладных задач. Описывается исходная информация, которая, как правило, содержит генотипические данные: о ДНК и связанных с ней молекулах, а также общеклинические параметры пациентов. Приводится описание логико-математического и программного аппарата основных решений в этой области. Работа призвана ознакомить аналитиков данных с задачами в современной онкологии для применения искусственного интеллекта, а также сориентировать биомедицинских исследователей в многообразии методов интеллектуального анализа данных и его возможностей.

Ключевые слова: искусственный интеллект, интеллектуальная система, онкология, генетические данные, мутации, иммунные данные, ДСМ-метод АПИ

DOI: 10.36535/0548-0027-2020-09-4

ВВЕДЕНИЕ

Одно из основных направлений искусственного интеллекта – интеллектуальный анализ данных, представляющий процесс поиска закономерностей в исходных данных. Его применяют в тех областях знаний, где отсутствует строгая формализация информации и нет однозначного общепринятого алгоритма её обработки. К этим областям относятся: медицина, биология, химия, социология и другие естественно-научные дисциплины.

Онкология – одно из наиболее наукоемких направлений медицины. Онкологическое заболевание характеризуется сбоем в регуляции процессов жизнедеятельности клеток и возникает в результате комбинации различных факторов: как внутренних, происходящих в организме, так и внешних. Однако процесс развития опухоли всегда происходит в результате различных нарушений в работе ДНК, которые могут являться как первопричиной заболевания, так и следствием внешних воздействий окружающей среды. В настоящий момент науке известны далеко не все биологические механизмы опухоли и функции генов, составляющих ДНК, соответственно, речь идет о плохо формализованной области знаний. В данном случае применение методов интеллектуального анализа данных является оправданным.

К инструментам его реализации относится машинное обучение, отличающееся от широко распространенных в медицине статистических методов анализа данных возможностью исследования нелинейных зависимостей для переменных, которые не являются независимыми друг от друга, что часто имеет место в биологических системах.

ТИПЫ РЕШАЕМЫХ ЗАДАЧ

Основные задачи, требующие решения в современной онкологии – это определение риска возникновения злокачественной опухоли или ее рецидива и предсказание состояния пациента (прогноз развития заболевания). В задачу определения прогноза также входит проблема выявления молекулярных механизмов (в частности, вовлеченных в них генов), которые вносят ключевой вклад в возникновение и развитие заболевания: с помощью такой информации можно подобрать персонализированную высокоточную терапию, обладающую высокой эффективностью, а также внести вклад в разработку новых препаратов.

Для пациентов, находящихся в ремиссии заболевания, крайне важно определить риски возникновения рецидива, что необходимо для индивидуального планирования схемы наблюдения или принятия решения о превентивном лечении. Решение этих задач осуществляется при помощи методов определения риска и предсказания развития опухоли, как, например, в исследовании при раке молочной железы [1].

* Работа выполнена при частичной финансовой поддержке РФФИ (проект № 18-29-03063)

Прогнозирование развития заболевания часто представляет самостоятельную – полноценную задачу [2]. Определение прогноза заболевания может быть представлено в виде задачи регрессии, когда имеет место предсказание времени выживаемости пациентов, к примеру с раком молочной железы [3] и с опухолями головы и шеи [4], либо как задача бинарной классификации в виде предсказания преодоления пациентами пятилетней выживаемости, что реализовано на примере множественной миеломы [5]. Одним из способов определения прогноза заболевания является классификация подтипов рака молочной железы по генетическим и клиническим данным, что информативно для определения прогноза развития заболевания [6].

Иммунотерапия является одним из наиболее перспективных направлений лечения онкозаболеваний. При этом до сих пор отсутствуют строгие критерии применения этого типа лечения. Поэтому важно выявлять таких пациентов, у которых ожидается положительный эффект от иммунотерапии, который превысит эффекты от других типов лечения. Исследователями решается задача предсказания иммунного ответа на опухоль в результате применения иммунотерапевтических препаратов вместе с реакцией организма на радиотерапию [7].

Важным направлением, необходимым не только для прогнозирования состояния пациента, но и для создания новых лекарственных препаратов, – моделирование протекающих в опухоли метаболических процессов [8]. Так, в работах [9, 10] решаются задачи протеин-протеиновых взаимодействий и определения статуса белков, что также важно при определении молекулярных мишеней для воздействия и перечня молекул, способных стать лекарствами.

В некоторых постановках задач исследователями выявляются факты наличия нарушений в ключевых механизмах опухоли, которые, в свою очередь, могут являться факторами прогноза заболевания или целями для назначения терапии [11]. Часто исследователи стремятся дать не прогноз состояния пациента или роста опухоли, а найти значимые биомаркеры онкологического заболевания в виде нарушений в работе ДНК и вовлеченных в них генов [12].

Сравнительно новая прикладная задача, обладающая высокой степенью практической применимости – это прогнозирование выживаемости пациента под воздействием различных способов лечения. Такой подход может быть использован для подбора оптимальной терапии онкологического заболевания [13] или для прогнозирования исхода клинических испытаний новых препаратов [14], что призвано ускорить внедрение новых способов лечения онкологических заболеваний. Этой тематике также посвящены исследования по предсказанию результата и токсичности химиолучевой терапии [15] и прогнозирование эффекта операции [16].

Значительное количество публикаций по теме искусственного интеллекта в онкологии касается проблем распознавания изображений, которые представляют либо микроскопические снимки окрашенных клеток опухоли (гистологические материалы), либо результаты лучевой диагностики. Решение таких

проблем нацелено на помощь в постановке диагноза или раннем выявлении онкозаболевания. Однако методы, исходные данные для которых не содержат признаковые описания, в нашей работе не рассматриваются.

ПРЕДСТАВЛЕНИЕ ЗНАНИЙ

Перспективным подходом интеллектуального анализа данных в онкологии является использование комбинации клинических и генетических данных. В частности, разработанные в США прикладные интеллектуальные системы *Tempus* [13] и *Flatiron* [14] используют как клинические данные, так и генотипические, которые относятся к совокупности всех генов и включают не только информацию о самой последовательности ДНК, но и о механизмах, обуславливающих ее активность, в частности: количество копий генов, их экспрессия, генетические сшивки. Такой подход повышает качество предсказания и устойчивость алгоритма [17].

Что касается диагнозов пациентов, которые преимущественно обрабатываются в задачах с применением искусственного интеллекта, то в основном к ним относятся наиболее распространенные типы опухолей: рак молочной железы, рак легкого, колоректальный рак, рак простаты и рак печени [1, 6, 12, 18]. Фокус на этих диагнозах позволяет формировать масштабные по численности массивы исходных данных. Так, в работе [12] исследуется 18100 пациентов: 3291 – с метастатическим колоректальным раком, 2409 – с метастатическим раком молочной железы, 1329 – с меланомой кожи, 11071 – с немелкоклеточным раком легкого; при этом в обучающую выборку были включены 4525 пациентов, в тестовую – 13575.

Для большинства же остальных типов опухолей классические методы искусственного интеллекта сложно применять по причине малого количества данных вследствие их редкой встречаемости. Так, в одном из наиболее востребованных источников – базе Национального института рака США [19] – представлено 53 типа опухолей по месту локализации в организме, из которых 29 имеют наборы данных объемом менее 100 пациентов.

Тем не менее, столь малочисленные выборки также активно используются в прогнозировании заболевания и поиске новых знаний. Так, в работе [3] приводятся данные 78 пациентов, которые включают экспрессию 70 генов, а также клинические данные (возраст, стадия заболевания, пол и др.). В работе [15] исследуются 12 наборов данных из 3496 пациентов, где каждый набор имеет размер от 131 до 922 пациента с числом признаков от 9 до 83.

Большое количество признаков в некоторых случаях повышает устойчивость работы алгоритма даже при небольшом количестве примеров обучающей выборки. Так, в работе [4], посвященной опухолям головы и шеи, исследовано 109 пациентов и 725 генов, при этом исследование показало приемлемую точность: показатель F1 (отношение произведения точности и полноты предсказаний к сумме этих величин) был равен 0,89.

Часто исследователями анализируются источники данных, создаваемые внутри учреждений и не нахо-

дящиеся в открытом доступе. К примеру, ученые из Великобритании провели анализ данных 668 пациентов с диагнозом рак яичников, которые отсутствуют в открытых источниках [16]. В работе [20] определяется прогноз возникновения рака поджелудочной железы на выборке из 73 пациентов с 53 признаками.

Поскольку часто различные онкологические заболевания вызываются одними и теми же нарушениями в работе ДНК, при изучении молекулярно-генетических механизмов опухоли популярностью пользуются обучающие выборки, состоящие из данных пациентов с различными диагнозами (*pan-cancer*) [2]. Так, исследование [11] проводилось по 33 типам опухолей, благодаря чему удалось достичь значительного по численности пациентов набора исходных данных: обучающая выборка насчитывала 4283 пациента, тестовая – 476, при этом признаками являлись данные размером восемь тысяч генов. В другой работе исследователи из Китая проанализировали 1216 пациентов с 12 типами опухолей с числом признаков в 311 генов [21].

Активность (экспрессия) генов, в отличие от самой структуры ДНК, является динамической характеристикой, подверженной значительным изменениям в процессе развития опухоли. В этой связи в настоящее время многими исследователями внимание уделяется не только мутациям в ДНК, но и генной экспрессии, поэтому этот тип данных учитывается во многих работах [2, 4, 6, 18].

Белок-белковые взаимодействия молекул, значимые для опухоли, организованы в особые цепочки превращений и активаций, именуемые сигнальными путями [22]. В работах американских исследователей из компании *IBM* [23], других американских коллективов [18, 24], а также корейских ученых из Сеульского национального университета [6] учитывается топология этих взаимодействий, которая включается в обучающую выборку. В работе [10] в качестве обучающей выборки используется более 30 тысяч как позитивных примеров (пар белков, взаимодействующих друг с другом), так и негативных (не вступающих в реакцию), полученных из открытых источников.

ЛОГИКО-МАТЕМАТИЧЕСКИЙ И ПРОГРАММНЫЙ ИНСТРУМЕНТАРИЙ

В области анализа онкологических данных на текущий момент отсутствует научный консенсус относительно единственного общепринятого метода, направленного на решение перечисленных выше задач. При этом существуют популярные у исследователей подходы. Авторы значительной части работ в качестве способов машинного обучения используют логистическую регрессию [11, 12, 16], метод опорных векторов (*SVM*) [1, 4], нейронные сети [10, 16]. Проводятся также сравнительные исследования, в которых различные алгоритмы применяются для одной и той же обучающей выборки. К примеру, в работе [16] сравнивались нейронные сети и логистическая регрессия, при этом нейронные сети показали более высокие результаты качества предсказания с точностью (*accuracy*), равной 0,77.

Однако этот перечень методов не является исчерпывающим. Так, в одном из исследований, проведен-

ном учеными из Нидерландов, с целью сравнения результативности различных алгоритмов применяются, помимо перечисленных алгоритмов, также деревья решений, случайный лес, и ансамблевый алгоритм *LogitBoost* на основе деревьев решений и логистической регрессии [15]. Все эти алгоритмы были реализованы при помощи среды *R*, которая используется, к примеру, и в работе [11], где логистическая регрессия реализована на языке *python* с применением *R*.

В масштабном мета-исследовании, проведенном по более чем 1800 доступным исследованиям [1], изучается точность алгоритмов машинного обучения, перечень которых составляют метод опорных векторов, нейронные сети, деревья решений, наивный байесовский алгоритм и метод *k*-ближайших соседей. Лучший результат в работе был достигнут методом опорных векторов (*SVM*), при этом площадь под кривой чувствительности-специфичности (*ROC AUC*) составила свыше 90%.

Среди динамично развивающихся направлений анализа онкологических данных в рамках подходов, связанных с нейронными сетями, можно назвать глубокое обучение на основе сверхточных нейронных сетей [6, 10]. Иногда глубокое обучение применяется в комплексе с другими методами, в частности, с регрессионным анализом, как, например, в работе исследователей из США [2]. Одной из точек приложения подобных методов является применение глубокого обучения к моделированию биологических процессов, происходящих в клетке: с этой целью используется многослойный перцептрон [25]. Авторы указанной работы признают, что существенным недостатком нейронных сетей является работа в режиме «черного ящика», что не позволяет интерпретировать полученные результаты в исходных терминах решаемой задачи и снижает, таким образом, доверие медицинских экспертов к предсказаниям. Поэтому предпринята попытка сделать нейронную сеть «видимой», при помощи выделения отдельных участков под определенные подсистемы клеток. Однако такая архитектура, созданная для определения прогноза развития и выявления значимых генов колонии дрожжей, которая по набору биологических механизмов проще злокачественной опухоли, потребовала 97 тысяч нейронов и 12 слоев, что, очевидно, обуславливает высокую вычислительную стоимость подобного алгоритма.

Представление молекулярных механизмов в качестве топологических схем делает возможным применение к их анализу методов машинного обучения на основе графовых ядер, как, например, в технологии, поддерживаемой *IBM* [23] и в другой разработке американских исследователей [9], в которых графовые ядра могут применяться для определения схожести подструктур графа. Для обоих представленных подходов авторами созданы интеллектуальные системы с помощью языка *python*.

Другой подход, связанный со схематическим представлением задачи – байесовские сети, в виде которых также представляются реакции и превращения белков [26].

Для моделирования биологической системы в некоторых подходах применяются дифференциальные

уравнения, при этом система является нестационарной [13]. В этом случае имеется возможность предсказания ее состояний в будущем, в том числе изменений в результате внешних воздействий. В работе [7] также использовались обыкновенные дифференциальные уравнения, описывающие процессы поведения иммунных клеток.

Вышеперечисленные инструменты искусственного интеллекта, судя по опубликованным описаниям, имеют в своем составе модули, отвечающие за представление исходных данных (база фактов, БФ), хранение результатов (база знаний, БЗ), а также решатель задач, реализующий исследование. В том случае, если инструмент, помимо названных компонентов, имеет комфортный для работы пользовательский интерфейс, он считается интеллектуальной системой [20, 27]. В одной из наиболее известных интеллектуальных систем для анализа онкологических данных *Flatiron* (США) используются логистическая регрессия, метод опорных векторов, деревья решений с градиентным бустингом, случайный лес [5]. Еще одна интеллектуальная система – *Tempus* (США) – не раскрывает применяемые алгоритмы, сообщая лишь о высокой точности и устойчивости исследований за счет применения комбинации исходных данных различных типов [17]. Показатель точности F1 при этом не превышает 0,62.

ЗАКЛЮЧЕНИЕ

Среди вышеперечисленных методов искусственного интеллекта, применяемых в современной науке к онкологическим данным, в большинстве случаев используются метрические алгоритмы машинного обучения, которые имеют существенный недостаток, заключающийся в сложности интерпретации результатов такого анализа в содержательных терминах исследуемой предметной области. Это вызвано тем фактом, что вводимые в таких алгоритмах метрики, чаще всего не имеют явной взаимосвязи с параметрами решаемой задачи.

Также серьезной проблемой при использовании нейронных сетей является невозможность установления порядка принятия решения этими методами, что затрудняет использование основанных на них интеллектуальных систем медицинскими специалистами в клинической сфере.

Один из инструментов интеллектуального анализа данных, используемый для решения задач исследования медицинских и онкологических данных, – ДСМ-метод автоматизированной поддержки исследований [28–31], который реализует обнаружение закономерностей в сложнотренированных эмпирических данных, содержащих причинно-следственные зависимости в неявном виде. Этот метод обеспечивает формализацию знаний предметной области средствами многозначной логики, для чего обобщает в гипотезах информацию, полученную из обучающей выборки, затем применяет эти гипотезы для предсказания исследуемого эффекта неизвестных объектов, а также имеет критерий достаточного основания правдоподобного вывода.

ДСМ-метод автоматизированной поддержки исследований был успешно применен к задачам предска-

ния возникновения онкологического заболевания [20], определения прогноза развития злокачественной меланомы кожи, а также и выявления значимых для опухоли биомаркеров (генов) [32]. В результате применения этого метода была достигнута не только высокая точность предсказания неизвестных примеров (показатель F1 равен 0,8, точность (*accuracy*) составила 0,79, что выше многих из вышеперечисленных результатов), но и обнаружены эмпирические закономерности, состоящие из комбинаций признаков и являющиеся причинами исследуемых явлений, что является актуальной задачей современной онкологии и важным подспорьем в научно-исследовательской деятельности.

Функциональные возможности ДСМ-метода, в отличие от большинства перечисленных подходов машинного обучения, позволяют эффективно использовать его как инструмент автоматизированной поддержки исследований, с помощью интеллектуальных систем на его основе.

СПИСОК ЛИТЕРАТУРЫ

1. Nindrea R.D., Aryandono T., Lazuardi L., Dwiprahasto I. Diagnostic accuracy of different machine learning algorithms for breast cancer risk calculation: a meta-analysis // *Asian Pacific Journal of Cancer Prevention*. – 2018 (Jul. 27). – № 19(7). – P.1747-1752.
2. Xie G., Dong C., Kong Y., Zhong J.F., Li M., Wang K. Group lasso regularized deep learning for cancer prognosis from multi-omics and clinical features // *Genes (Basel)*. – 2019. – № 10(3). – URL: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6471789/>
3. Chen H., Kodell R.L., Cheng K.F. et al. Assessment of performance of survival prediction models for cancer prognosis // *BMC Medical Research Methodology*. – 2012. – № 12. – URL: <https://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-12-102>
4. Su J., Zhang Y., Su H., Zhang C., Li W. A recurrence model for laryngeal cancer based on SVM and gene function clustering // *Acta Oto-Laryngologica*. – 2017. – № 137(5). – P. 557-562.
5. Chen R., Garapati S., Wu D., Ko S., Falk S., Dierov D., Stasiw A., Opong A.S., Carson K.R. Machine learning based predictive model of 5-year survival in multiple myeloma autologous transplant patients // *Blood*. – 2019. – URL: https://ashpublications.org/blood/article/134/Supplement_1/2156/427904/Machine-Learning-Based-Predictive-Model-of-5-Year
6. Rhee S., Seo S., Kim S. Hybrid approach of relation network and localized graph convolutional filtering for breast cancer subtype classification // *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence (IJCAI-18)*. – URL: <https://arxiv.org/abs/1711.05859>
7. Kosinsky Y., Dovedi S.J., Peskov K., Voronova V., Chu L., Tomkinson H., Al-Huniti N., Stanski D.R., Helmlinger G. Radiation and PD-(L)1 treatment combinations: immune response and dose optimization via a predic-

- tive systems model // *Journal for ImmunoTherapy of Cancer*. – 2018. – № 1 – P. 6-17.
8. Eberhard O. Voit. Modelling metabolic networks using power-laws and S-systems // *Essays in Biochemistry volume*. – 2008. – № 45. – P. 29-40.
 9. Yanardag P., Vishwanathan S.V.N. Deep graph kernels. KDD '15 // *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. August 2015. – URL: <https://dl.acm.org/doi/10.1145/2783258.2783417>
 10. Li H., Gong X., Yu H., Zhou C. Deep neural network based predictions of protein interactions using primary sequences // *Molecules*. – 2018. – № 23. – URL: https://www.researchgate.net/publication/326755315_Deep_Neural_Network_Based_Predictions_of_Protein_Interactions_Using_Primary_Sequences.
 11. Way G.P., Sanchez-Vega F., La K., Armenia J., Chatila W.K., Luna A., Sander A., Cherniack A.D., Mina M., Ciriello G., Schultz N. Machine learning detects pan-cancer ras pathway activation in the cancer genome atlas // *Cell Reports*. – 2018. – Vol. 23, Iss.1 – P.172-180.
 12. Ambwani G., Cohen A., Estévez M., Singh N., Adamson B., Nussbaum N.C., Birnbaum B. A machine learning model for cancer biomarker identification in electronic health records // *Value in Health*. – 2019. – Vol. 22, Iss. S1.
 13. Интеллектуальная система для прогнозирования онкологического заболевания Tempus (США). – URL: <https://www.tempus.com>
 14. Интеллектуальная система для прогнозирования онкологического заболевания FlatIron (США). – URL: <https://flatiron.com/>
 15. Deist T.M., Dankers F.J.W.M., Valdes G., Wijsman R., Hsu I.C., Oberije C., Lustberg T., Van Soest J., Hoebbers F., Jochems A., Naqa I. El, Wee L., Morin O., Raleigh D.R., Bots W., Kaanders J.H., Belderbos J., Kwint M., Solberg T., Monshouwer R., Bussink J., Dekker A., Lambin P. Machine learning algorithms for outcome prediction in (chemo) radiotherapy: An empirical comparison of classifiers // *Medical Physics*. – 2018. – № 45(7). – P. 3449-3459.
 16. Enshaei A., Robson C.N., Edmondson R.J. Artificial intelligence systems as prognostic and predictive tools in ovarian cancer // *Annals of Surgical*. – 2015. – № 22(12). – P. 3970 - 3975.
 17. Michuda J., Leibowitz B., Amar-Farkash S., Bevis C., Breschi A., Kapilivsky J., Igartua C., Bell J.S.K., Beauchamp K.A., White K., Stumpe M., Beaubier N., Taxter T. Multimodal prediction of diagnosis for cancers of unknown primary. AACR Annual Meeting 2020. Session PO.CL01.04 – Tumor Type-focused Translational Research Specific. Abstract 5423. – URL: <https://www.abstracksonline.com/pp8/?sf122451697=1#!/9045/presentation/3059>.
 18. Wan N., Weinberg D., Liu T.Y. et al. Machine learning enables detection of early-stage colorectal cancer by whole-genome sequencing of plasma cell-free DNA // *BMC Cancer*. – 2019. – № 19(1). – URL: <https://bmccancer.biomedcentral.com/articles/10.1186/s12885-019-6003-8>
 19. База данных The Cancer Genome Atlas Program. – URL: <https://www.cancer.gov/tcga>
 20. Шестерникова О.П., Финн В.К., Винокурова Л.В., Лесько К.А., Варварина Г.Г., Тюляева Е.Ю. Интеллектуальная система для диагностики заболеваний поджелудочной железы // *Научно-техническая информация. Сер. 2*. – 2019. – № 10. – С. 41-48; Shesternikova O.P., Finn V.K., Vinokurova L.V., Les'ko K.A., Varvanina G.G, Tyulyaeva E.Yu. An intelligent system for diagnostics of pancreatic diseases // *Automatic documentation and mathematical linguistics*. – 2019. – Vol. 53, № 5. – P. 288-291.
 21. Ding W., Chen G., Shi T. Integrative analysis identifies potential DNA methylation biomarkers for pan-cancer diagnosis and prognosis // *Epigenetics*. – 2019. – № 14(1). – P. 67 - 80.
 22. Kanehisa M., Goto S. KEGG: Kyoto encyclopedia of genes and genomes // *Nucleic Acids Research*. – 2000. – № 28. – P. 27-30.
 23. Manica M., Cadow J., Mathis R., Martinez M.R. PIMKL: pathway induced multiple kernel learning // *NPJ Systems Biology and Applications*. – 2019. – Vol. 5, № 8. – URL: <https://arxiv.org/abs/1803.11274>
 24. Sanchez-Vega F., Mina M., Armenia J., Ciriello G., Sander C., Schultz N. et al. Oncogenic Signaling Pathways in the Cancer Genome Atlas // *Cell*. – 2018. – Vol. 173, Iss. 2 – P. 321-337.
 25. Ma J., Ku Yu M., Fong S., Ono K., Sage E., Demchak B., Sharan R., Ideker T. Using deep learning to model the hierarchical structure and function of a cell // *Nature Methods*. – 2018. – Vol. 15. – P. 290–298.
 26. Friedman N., Linial M., Nachman I., Pe'er D. Using bayesian networks to analyze expression data. *Proceedings of the Fourth Annual International Conference on Computational Molecular Biology*. – 2000. – URL: <https://www.cs.huji.ac.il/~nir/Papers/FLNP1Full.pdf>
 27. Финн В.К. Об эвристиках ДСМ-исследований (дополнения к статьям) // *Научно-техническая информация. Сер. 2*. – 2019. – № 10. – С.1-34; Finn V.K. On the Heuristics of JSM Research (Additions to Articles) // *Automatic documentation and mathematical linguistics*. – 2019. – Vol. 53, № 5. – P. 250-282.
 28. Финн В.К. Дистрибутивные решетки индуктивных ДСМ-процедур // *Научно-техническая информация. Сер. 2*. – 2014. – № 11. – С. 1-36; Finn V.K. Distributive lattices of inductive JSM procedures // *Automatic documentation and mathematical linguistics*. – Vol. 53, № 5. – P. 265-296.
 29. ДСМ-метод автоматического порождения гипотез: Логические и эпистемологические основания / сост. О.М. Аншаков, Е.Ф. Фабрикантова; под. общ. ред. О.М. Аншакова. – М.: ЛИБРОКОМ, 2009. – 433 с.

30. Финн В.К. Об определении эмпирических закономерностей посредством ДСМ-метода автоматического порождения гипотез // Искусственный интеллект и принятие решений. – 2010. – № 4. – С. 41-48.
31. Финн В.К., Шестерникова О.П. О новом варианте обобщенного ДСМ-метода автоматизированной поддержки научных исследований // Искусственный интеллект и принятие решений. – 2016. – № 1. – С. 57-64.
32. Чебанов Д.К., Михайлова И.Н. Интеллектуальный анализ данных пациентов с меланомой для поиска маркеров заболевания и значимых генов // Научно-техническая информация. Сер. 2. – 2019. – № 10 – С. 35-40; Chebanov D.K. Mikhailova I.N. Intellectual mining of patient data

with melanoma for identification of disease markers and critical genes // Automatic documentation and mathematical linguistics. – Vol. 53, № 5. – P. 283-288.

Материал поступил в редакцию 16.06.2020

Сведения об авторах

ЧЕБАНОВ Дмитрий Константинович – генеральный директор ООО «ОнкоЮнайт Клиникс», Москва
e-mail: chebanov.dk@gmail.com

МИХАЙЛОВА Ирина Николаевна – доктор медицинских наук, ведущий научный сотрудник НМИЦ онкологии им. Н.Н. Блохина МЗ РФ, Москва
e-mail: irmikhaylova@gmail.com

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

УДК 81'367.335/.5

Г.И. Кустова, Н.Б. Швелидзе

Системные свойства апеллятивных вводных конструкций (по данным Национального корпуса русского языка)*

*Рассматриваются апеллятивные вводные конструкции (ВК) с ментальными глаголами 2-го лица (Я после этого, **верите ли**, всю ночь не спал) на фоне модальных вводных конструкций (Он, **я думаю**, уехал). Впервые обосновывается противопоставление вопросительного, или эмфатического (Вчера, **представляешь** [?], встретил в театре одноклассника), и парентетического, или нейтрального (А я, **знаете**, люблю посидеть с дочкой на берегу), употребления апеллятивных вводных конструкций. На материале, извлеченном из Национального корпуса русского языка, проанализированы системные свойства апеллятивных вводных конструкций **представляешь**, **веришь**, **знаешь**, **понимаешь**, **помнишь**: типы употреблений, грамматические параметры, синтаксические позиции, подгруппы, семантические различия. Результаты исследования могут быть использованы не только для анализа (в том числе типологического) разных типов ВК, но и для совершенствования программ автоматической обработки текста.*

Ключевые слова: вводные (парентетические) конструкции, ментальные глаголы, редукция, пунктуация

МОДАЛЬНЫЕ И АПЕЛЛЯТИВНЫЕ ВВОДНЫЕ КОНСТРУКЦИИ

В грамматических справочниках и учебных пособиях традиционно выделяются две группы вводных конструкций (далее – ВК), в которых фигурируют ментальные глаголы.

Одна группа – показатели достоверности / предположительности пропозиции *P* (знаком *P* мы будем обозначать вмещающее предложение, в которое включена ВК): *Он, я знаю*, уехал – ‘Он уехал’ – достоверное сообщение, знание, факт; *Он, я думаю*, уехал – ‘Он уехал’ – предположение. Эту группу будем называть модальной. Вторую группу ВК можно называть апеллятивной, ср.: *Я после этого, верите ли*, всю ночь не спал; *Вчера, представляешь*, встретил в театре одноклассника; *А я, знаете*, люблю посидеть с дочкой на берегу. В работах [1–3] считается, что функция второй группы ВК – привлечь внимание со-

беседника; в [4, 5] предлагается несколько другая формулировка функции этой группы – подчеркнуть, акцентировать, выделить то, что высказывается. Собственно, эти формулировки не противоречат друг другу, так как привлечение внимания является импликацией акцентирования: если говорящий что-то акцентирует, выделяет в своем высказывании, значит он считает это важным и хочет, чтобы адресат обратил на это особое внимание. При этом специальных исследований, посвященных тому, как ведут себя ментальные глаголы во вводных конструкциях, очень мало, и касаются они только первой группы – в нашей терминологии, модальной (ср. [6–11]). Для нас особый интерес представляют исследования, связанные с конвенциональными и неконвенциональными импликатурами (см. [12–19]), особенно – с импликатурами ментальных глаголов (ср. [20, 21]).

Вводные конструкции модальной группы (*думаю*, *знаю*, *считаю*) можно рассматривать как редукцию матричного глагола, т.е. подчиняющего глагола с пропозициональным дополнением в сложном предложении: *Я думаю, что он уехал* → *Он, я думаю*, уехал.

* Работа выполнена при поддержке РФФИ, проект № 17-29-09154 офи_м. Примеры извлечены из Национального корпуса русского языка (ruscorpora.ru)

При этом исследователей интересуют в первую очередь не глагольные конструкции, субъект которых (говорящий) очевиден и детерминирован формой глагола (*думаю*), а конструкции с другими модальными словами (ср.: *Он, видимо / очевидно / возможно, уехал*), где субъект тоже есть и определяется, например, в [22], как субъект сознания, который, в конечном счете, совпадает с говорящим или, для нарратива, с другим типом повествователя, ср. [23, 24]. Вторая группа – аппеллятивная – находится вне фокуса внимания исследователей.

В работе [25] была предпринята попытка сформулировать различия между модальной и аппеллятивной группами ВК и определить специфику аппеллятивной группы. При этом автор пришел к следующим выводам.

Во-первых, при исследовании ментальных глаголов – как в матричной, так и во вводной позиции – необходимо противопоставлять глаголы знания, или фактивные (*знать, понимать, помнить*), и глаголы мнения, или путативные (*думать, считать, полагать, предполагать*), к ним примыкают глаголы с оценочным компонентом: *бояться* – ‘думать, что плохое *P* возможно’; *надеяться* – ‘думать, что хорошее *P* возможно’. Во-вторых, необходимо противопоставлять формы 1-го и 2-го лица (говорящего и адресата), у которых достаточно разные свойства – в том числе, и во вводных конструкциях.

Сравнение ментальных глаголов модальной и аппеллятивной групп ВК показало, что для модальной группы характерны глаголы мнения; при этом в составе ВК могут употребляться формы как 1-го, так и 2-го лица глаголов мнения (фактивные глаголы, т.е. глаголы знания, для модальной группы в целом не характерны и встречаются, в основном, в особой союзной конструкции типа *как я понял, как я помню, как я узнал*, которую мы не рассматриваем). Важно, что модальные ВК могут включаться как в повествовательные, так и в вопросительные предложения. Для повествовательных предложений типичны ВК 1-го лица: *Управляющему, думаю, уже обо всем доложили*; для вопросительных предложений типичны ВК 2-го лица: *Управляющему, думаешь, уже обо всем доложили?* Но сами модальные вводные конструкции не могут быть вопросительными ни в каких предложениях, т.е. не могут произноситься с вопросительной интонацией. Они имеют редуцированный, «парентетический», нейтральный статус и произносятся с характерной для вводных конструкций «ослабленной» интонацией (пониженным тоном).

Пропозиция *P* в контексте модальной ВК с глаголами мнения является предположением, т.е. имеет не-верифицируемый, неассертивный статус (о статусе пропозиций см., например, [24, с. 95–113; 26]). При ВК 1-го лица автором предположения *P* является говорящий: *Завтра, думаю, он вернется*; ВК 2-го лица употребляется в вопросах, т.е. говорящий еще не знает, есть ли у адресата предположение *P* (‘Он вернется’), – но спрашивает он именно об этом: есть ли У АДРЕСАТА предположение *P*: *Завтра, думаешь, он вернется?*

Вводные конструкции аппеллятивной группы имеют совершенно другие – и во многом парадоксальные – свойства. Во-первых, они совместимы только с

повествовательными предложениями и никогда (!) не включаются в вопросительные предложения (т.е. невозможно: **Он, верите ли / представляешь, всю ночь не спал?*). Во-вторых, сами ВК (в большинстве своем, – об исключениях см. далее) обычно бывают вопросительными, что является уникальным свойством аппеллятивных ВК среди всех остальных групп ВК (речь идет о формах настоящего времени типа *знаешь, представляешь*. В аппеллятивной группе широко представлены также императивы, ср. *пойми, вспомни, послушай* и подобные, которых мы в настоящей работе не касаемся). Для обозначения вопросительной интонации аппеллятивных ВК мы будем использовать знак [?], как в работе [27]: *Вчера, представляешь [?], на работу опоздала*.

Для вводных слов как типа конструкций в целом характерен «редуцированный» статус, что сказывается и на их позиции (для них типична интерпозиция), и на их грамматических возможностях (вводные слова обычно имеют редуцированную парадигму и употребляются в определенных формах), и на их интонации – «вводной», «парентетической». Тем не менее в определенных случаях на ВК можно поставить логическое ударение: *Управляющий [пауза], к счастью \ [пауза], пока еще ничего не знает* (знак \ показывает понижение тона). Однако никакие ВК не могут произноситься с вопросительной интонацией (с повышением тона), кроме аппеллятивных.

В [25] было высказано предположение, что это уникальное свойство аппеллятивных ВК связано с их происхождением. Аппеллятивные ВК являются не результатом редукиции матричного глагола в исходном сложном предложении, как модальные ВК, а результатом контаминации, компрессии двух отдельных высказываний – сообщения и вопроса о реакции адресата: *Вчера у меня, представляешь [?], три раза отключался интернет* ≈ ‘Вчера у меня три раза отключался интернет + Ты представляешь [такую ситуацию]?!’ / Ты можешь себе это представить?!’. *Представляешь* здесь является не настоящим вопросом, предполагающим настоящий ответ, а риторическим, экспрессивным, эмфатическим вопросом, синонимичным риторическому же императиву: *Ты только представь!* Тем не менее формально (в первую очередь – интонационно) вопросительность сохраняется, и на такое обращение может последовать ответ: *Очень хорошо представляю, у меня у самого интернет уже сутки не работает*. Аналогично: *Это было на заправке. Мы там, помнишь [?], останавливались* ≈ ‘Мы там останавливались + Ты помнишь это?’. Здесь также ответ не требуется, но возможен (‘Помню / Не помню’). Разумеется, не во всех случаях уместен какой-либо ответ, ср.: *Этот человек был, знаешь [?], настоящим мастером* – если речь идет о незнакомом человеке, адресат не может ответить ‘знаю’ или ‘не знаю’. Нам важно лишь подчеркнуть, что аппеллятивные ВК сохраняют интонационную и семантическую связь с вопросом, что не случайно – именно вопрос (наряду с императивом) является прототипически аппеллятивным речевым актом.

Гибридной природой предложений с аппеллятивными ВК (контаминацией сообщения и вопроса) определяются и другие важные для нас свойства: пред-

ложение *P* всегда передает факт (не предположение, а знание), однако в качестве вводного слова используются не только глаголы знания, т.е. фактивные *знаешь*, *понимаешь*, *помнишь*, но и нефактивные глаголы *веришь*, *представляешь*.

Ещё одна парадоксальная особенность предложений с аппеллятивными ВК: из-за того, что это контактирование разных предложений (вводный глагол берется из одного предложения, а содержание *P* из другого), *P* является не знанием адресата, а знанием говорящего: *Там, знаешь* [?], *документы перепутали – знаешь* – предикат адресата, а факт *P* ‘перепутали документы’ – знание говорящего, адресату этот факт как раз не известен и сообщается ему впервые.

Термин «вопросительные» к аппеллятивным ВК применяется до известной степени условно. Поскольку речь идет не о настоящем вопросе, на который говорящий хочет получить ответ, а об использовании вопросительной интонации в аппеллятивных целях (для воздействия на адресата, для привлечения его внимания), правильнее было бы говорить, что ВК имеют эмфатический статус, или эмфатическое употребление. Тем не менее интонация аппеллятивных ВК близка к вопросительной и отчетливо противопоставлена другому типу употребления аппеллятивных ВК – невопросительному, когда ВК произносятся с классической парентетической (вводной) интонацией. Поэтому для простоты мы будем использовать термин «вопросительный (тип употребления)» наряду с терминами «эмфатический», или «акцентированный». Невопросительный тип употребления мы будем называть редуцированным, или парентетическим, или нейтральным.

Все аппеллятивные ВК имеют эмфатическое (акцентированное, вопросительное) употребление, но только *знаешь* и *понимаешь* имеют, в дополнение к этому, парентетическое (нейтральное) употребление. К ним мы обратимся далее.

Если бы функция аппеллятивных ВК исчерпывалась привлечением внимания адресата, достаточно было бы одного вводного слова – например, перцептивного *смотри(те)* или *слушай(те)*. Однако в аппеллятивной группе используются еще и формы ментальных глаголов *представляешь*, *веришь*, *помнишь*, *знаешь*, *понимаешь*. И для этих форм противопоставлены не только два типа употребления (эмфатическое и парентетическое), но и отдельные ВК внутри этих типов. Это означает, что у разных ВК разные функции и разные семантические особенности, ради которых их и выбирают говорящие.

Задача нашей работы – на материале, извлеченном из Национального корпуса русского языка (НКРЯ), проанализировать системные свойства аппеллятивных ВК *представляешь*, *веришь*, *знаешь*, *понимаешь*, *помнишь*: типы употреблений, грамматические параметры, синтаксические позиции, подгруппы, семантические различия.

ВК аппеллятивной группы употребляются в формах единственного (*веришь*) и множественного (*верите*) числа, имеют вариант с частицей *ли* (*веришь ли*, *верите ли*) или без нее, а также вариант с местоимением (*ты представляешь*) или без него (*представляешь*). Далее мы для краткости будем обозна-

чать всю парадигму ВК основной формой ед. ч. (*веришь*, *знаешь* и т.д.), если речь не идет об особенностях какой-либо отдельной формы. В силу ограниченности объема статьи мы сосредоточимся, в основном, на «минимальной» форме ВК – без местоимения и без частицы *ли*.

ВВОДНЫЕ КОНСТРУКЦИИ ПРЕДСТАВЛЯЕШЬ И ВЕРИШЬ

Как уже было отмечено, *представляешь* и *веришь* имеют только вопросительный (эмфатический) статус. Они употребляются во всех трех позициях.

Препозиция: — *Представляешь* [?], *в вагоне только три купе, и в каждом отдельная ванна* [Виктор Пелевин. Желтая стрела (1993)];

Жор напал. Начала с печени трески, представляешь [?], *никогда не пробовала* [Александр Снегирев. Вера (2015)] – ВК относится ко второй части предложения (*никогда не пробовала*) и стоит в препозиции относительно нее;

Такие масштабы! Верешь [?], *газеты некогда читать. Новости по дороге ухватываю* [Семен Данилюк. Бизнес-класс (2003)].

Интерпозиция: *И ему, представляешь* [?], *разрешили прямо в Малом Эрмитаже жить, деду этому* [«Сибирские огни», 2012];

Месяц назад у меня случилась великая радость — сын родился. Так я, верите [?], *даже не пригубил* [Труд-7, 2006];

И так мне, веришь [?], *хорошо, так спокойно на душе, что хоть не просыпайся...* [Евгений Шкловский. Будь мужчиной, Макс (1996) // «Новый мир», 1998].

Постпозиция: *Это для двукратных олимпийских чемпионов места не хватило, представляете?* [«Совершенно секретно», 2003.07.04];

Мама величает Максима Малановича «дорогим гостем». Я здороваюсь с ним за руку, верите? [Геннадий Башкуев. Маленькая война // «Сибирские огни», 2013].

ПРИМЕЧАНИЕ. В такой же функции употребляется сленговое *прикинь*. Особенно примечательно, что *прикинь* употребляется с такой же «вопросительной» (восходящей) интонацией, хотя это форма императива: *Он уже вторую неделю на Канарах, прикинь* [?].

У *представляешь* и *веришь* есть некоторые структурные различия, на которых мы не будем подробно останавливаться, а просто упомянем их. *Представляешь* свободно присоединяет *ты* (выступает в форме *ты представляешь* [?]), но почти не употребляется с *ли* (*представляешь ли*) – в Национальном корпусе русского языка (НКРЯ) обнаружился только один пример: — *А вот я, представляешь ли, хочу*, — *сказал отец задумчиво* [Альберт Лиханов. Чистые камушки (1967)] – при том что в матричной позиции (т.е. в позиции подчиняющего глагола в главной части сложного предложения) *ли* свободно присоединяется к глаголу: — *Представляете ли вы, какую тяжесть на себя взваливаете* [Г. Я. Бакланов. Жизнь, подаренная дважды (1999)].

Верюшь, напротив, неохотно употребляется с *ты* (в НКРЯ нашлось только два примера, причем оба – в препозиции): *Ну зачем мне было воровать? И поэтому я не могу. Мне кажется, что все видят, глядят. Мне стыдно... Ты верюшь [?], братушка, мне стыдно...* [Борис Екимов. Пиночет (1999)]; *Ты верюшь [?], Параскева, пятого дня смотрю — возле этого ферта из ДОСААФ хвостом вращат* [Виль Липатов. Деревенский детектив (1967–1968)], – но свободно сочетается с *ли*: *На устах моих цвела улыбка, которая была для меня, верите ли [?], родом судороги* [Евгений Водолазкин. Лавр (2012)].

Верюшь и *представляешь* маркируют сообщаемый факт *P* как необычный, удивительный, неожиданный и т.д.: ‘*P* настолько удивительно / неожиданно / необычно, что это трудно представить / в это трудно поверить’, ср. синонимичные выражения: *Ты удивишься; Ты не поверюшь*.

Вводное *представляешь* близко к матричному употреблению *представляешь* с косвенным вопросом (по существу, риторическим) или косвенным восклицанием. Такие предложения тоже маркируют некоторый факт как удивительный или неожиданный и не требуют ответа: *Представляешь, как я испугался!* = ‘очень сильно’.

Вводное *верюшь*, в принципе, соотносится с вопросительным предложением, где матричный глагол *верюшь* присоединяет союзное придаточное, ср.: *Нашел я ее в таком расстройстве, что, верите [?], я не мог выдержать... заплакал...* [П.Д. Боборыкин. Труп (1892)] или: *Вы верите, что я не выдержал и заплакал?* Однако если вопрос *Ты верюшь, что P?* может быть буквальным, то вопрос вводной конструкции является риторическим и не требует ответа, поскольку ситуация *P* имеет статус факта. Просто ситуация *P*, с точки зрения говорящего, необычная, редкая, маловероятная. Аналогично «нормальный» вопрос *Ты в это верюшь?* часто связан с обсуждением маловероятных ситуаций и имеет импликацию ‘в это трудно поверить’, ‘это странно / маловероятно’, которая становится характеристикой ситуации *P* и наследуется вводной конструкцией. При этом сам по себе факт *P* может быть действительно редким и необычным, ср.: *Верите [?], неделю не сплю, не ем. Лишь дурнину эту глотаю, — целкнул ногтем по бутылке* [Борис Екимов. Высшая мера (1995)] – маловероятно, чтобы человек не спал и не ел целую неделю, и в это действительно трудно поверить, – а может быть вполне обычным, ср.: *Верите [?], мне даже в голову не пришло спросить, кто они такие* [Людмила Гурченко. Аплодисменты (1994–2003)], т.е. необычность события *P* – вопрос интерпретации говорящего.

Итак, *представляешь* относится к ситуациям, которые удивили самого говорящего, и он хочет, чтобы слушающий тоже разделил его чувство – восприятие ситуации *P* как удивительной или неожиданной. *Верюшь* относится к ситуациям, которые, по мнению говорящего, должны адресату казаться необычными, маловероятными, выпадающими из «нейтрального» жизненного потока, т.е. *представляешь* маркирует ситуацию как удивительную для говорящего, *верюшь* – как необычную с точки зрения любого человека, в том числе адресата.

ВК *представляешь* и *верюшь* семантически связаны с устойчивыми выражениями–восклицаниями: *Ты можешь себе представить?!; Ты можешь в это поверить?!; Кто бы мог подумать!; Ничего себе!; Вот это да!; Что творится!* и подобными. Кстати, путативный глагол *подумать* совершенного вида, в отличие от *думать* несовершенного вида, тоже употребляется в аппеллятивной группе, но только в форме императива: *Вчера, ты подумай, полтора часа ждали трамвая*.

ВВОДНАЯ КОНСТРУКЦИЯ ПОМНИШЬ

Только вопросительное употребление имеет вводная конструкция *помнишь* (-ите):

— *В сводке за вчерашний день... ну, помните [?], о гастрономе... один из налётчиков был в офицерском кителе без погон...* [Анатолий Азольский. Облдрамтеатр // «Новый Мир», 1997].

Помнишь по функции близко к императиву *вспомни*: *Кстати, вспомни, пожилой звонил кому-то по сотовому* [Петр Галицкий. Опасная коллекция (2000)].

Как отмечалось в работе [25], общий тезис о том, что ВК 1-го лица сочетаются с пропозицией говорящего (*Он, думаю, придет* – ВК – 1 лицо, *P* ‘он придет’ – предположение говорящего), а ВК 2-го лица (*думаешь*) – теоретически – должны сочетаться с пропозицией адресата, для аппеллятивной группы не верен. ВК 2-го лица *знаешь, понимаешь, представляешь, верюшь* вводят знание не адресата (как можно было бы ожидать, исходя из формы вводного глагола), а говорящего (1-го лица): *Я, знаешь, не люблю курорты, лучше на даче; Я, знаешь [?], решил остаться* – адресат этого не знает и знать не может, так как слышит эту информацию впервые: говорящий сообщает ему факт из своего личного опыта или своей внутренней информационной сферы. В этом смысле ВК *помнишь*, на первый взгляд, занимает особое место в аппеллятивной группе, потому что, как и матричное *помнишь*, вводит факт, известный обоим собеседникам, а не только говорящему, факт, относящийся к совместной истории говорящего и адресата. В противном случае *помнишь* нельзя употребить, ср.: – *Я еще, помнишь [?], документы не мог найти. – Как я могу это помнить, я был в командировке*. Однако на самом деле, используя ВК *помнишь*, говорящий не имеет цели узнать, помнит ли адресат факт *P* – в своем высказывании говорящий уже напомнил о *P*, независимо от того, помнит адресат или нет.

Сферой действия (о понятии сферы действия см. [28]) *помнишь* может быть не только целая ситуация, ср.: *Мы еще, помнишь [?], на выставку пошли*, но и отдельный ее аспект или участник: — *Откуда у тебя эти сведения? — Я свидетеля опросил, помнишь [?], в синем костюме* – здесь *помнишь [?]* относится не к факту опроса свидетеля, а к самому свидетелю и его костюму как «опознавательному» знаку.

Итак, предложения с ВК *помнишь*, как и предложения с другими вопросительными (эмфатическими) ВК, содержат факт (сообщение о факте). Отличие *помнишь* от других ВК в том, что это факт не новый для адресата, и говорящий исходит именно из из-

вестности этого факта адресату. При этом *помнишь* не столько напоминает этот факт, сколько маркирует тему или эпизод, который адресат должен актуализировать в своем информационном поле. Этим ВК *помнишь* отличается от ВК *знаешь* и *понимаешь*, которые вводят совершенно новую для адресата, неизвестную ему информацию.

ВВОДНЫЕ КОНСТРУКЦИИ ЗНАЕШЬ И ПОНИМАЕШЬ

Особое место в классе апеллятивных ВК занимают *знаешь* и *понимаешь*, так как у них есть два разных интонационных варианта – вопросительный (эмфатический) и нейтральный (редуцированный), которым соответствуют и два разных значения (две разные семантические функции).

ПРИМЕЧАНИЕ. Для *знаешь* и *понимаешь* широко используется вариант с вопросительной частицей (*знаешь ли, понимаешь ли*), – при этом *ли* возможна и при вопросительной, и при нейтральной интонации.

Знаешь – это наиболее «нейтральный» фактивный глагол по сравнению с *понимаешь* или *помнишь*: он не имеет никаких семантических «добавок». Вероятно, именно поэтому *знаешь* так широко используется в экспрессивных высказываниях уже в матричной позиции: *Да ты знаешь, сколько я с ним возился!*; *Ты не знаешь, как мы намучились с этой дачей!*; *Если бы ты знал, как мне это надоело!* Интересно, что в подобных экспрессивных контекстах *знаешь* практически синонимичен *представляешь*: *Ты не представляешь, как мне это надоело*; *Ты не можешь представить...*

В системе вводных конструкций такой синонимии нет – ВК *знаешь*, в отличие от ВК *представляешь*, не маркирует удивительные факты. Тем не менее говорящий, по каким-то причинам, считает нужным отметить сообщаемую адресату информацию *Р* с помощью ВК *знаешь*.

Интонация *знаешь / знает* зависит от позиции.

В препозиции *знаешь* (*знаешь ли / знает / знает ли*) всегда «вопросительное» (эмфатическое): **Знаешь** [?], *я никак не ожидал от него такого*.

ВК должна находиться в препозиции к содержательной части высказывания, но при этом может стоять после служебного слова (союза, частицы) или после обращения:

Голубчик, не надо мне объяснять про «творчество» — я и сам все это знаю, на себе испытал и испытываю. И, знаешь [?], *убедился, что это всегда временно: что-то сместится, щелкнет, откроется* [Юлий Даниэль. Письма из заключения (1966-1970)];

И, знаешь [?], *лизнула, на спор, металлический поручень подъемника...* [Марина Палей. Дань саламандре (2008)];

Конечно, я шучу, но, знаете ли [?], *в каждой шутке есть рациональное зерно* [«Зеркало мира», 2012];

Следователь явно все на то поддавливал, не убийство ли. Хотя, знаешь [?], *заметно было, что не по себе ему. Руки дрожали, голос...* [Роман Сенчин. Квартирантка с двумя детьми (2010)];

Таня, знаешь [?], *я думаю, дело обстояло так* [Дарья Донцова. Рождественский кролик (2011)].

В интерпозиции произнесение обычно редуцированное:

Очень, знаете ли, похоже на впечатление от выступления группы «Блестящие» [«Русская Жизнь», 2008].

Легко увидеть (вернее, услышать), что при переносе «начальной» (препозитивной) ВК в интерпозицию произнесение обычно становится редуцированным:

Знаете [?], *все это не ново* → — *Всё это, знаете, не ново*, — *сказал он* [Виктор Пелевин. Желтая стрела (1993)];

Четвёртый месяц я здесь. И, знаете ли [?], *нахожусь, что здесь очень и очень неплохо* [М. А. Булгаков. Мастер и Маргарита (1929-1940)] → *И нахожусь, знаете ли, что здесь очень и очень неплохо*.

Тем не менее в интерпозиции возможно и экспрессивно-вопросительное произнесение:

Валентина сходит с ума. Она, знаешь [?], *Паша, на седьмом месяце...* — *Гольдберг как будто извинялся* [Людмила Улицкая. Казус Кукоцкого // «Новый Мир», 2000].

В постпозиции возможен только редуцированный вариант ВК:

А то, как всегда, сначала вялаешься, а потом разберёшься, — *нехорошо, знаешь* [Андрей Волос. Недвижимость (2000) // «Новый Мир», 2001];

— *Ты ещё ходишь?* — *спросил матрос*. — *Хожу*, — *сказал Виктор*. — *Привык, знаешь* [Василий Аксенов. Пора, мой друг, пора (1963)];

Мне всё это было ничем. Я был счастлив, знаете. Молодой, способный. Возможно, даже талантливый [И. Грекова. Хозяева жизни (1960)].

Итак, в интерпозиции, в принципе, возможны оба произнесения: *А ещё, знаете, перед показом я жутко волновалась: хотелось выглядеть достойно на фоне профессиональных моделей* [«Дело» (Самара), 2002.05.26] – здесь возможна и вопросительная, и нейтральная ВК. Это означает, что интерпозиция – позиция противопоставления, а пре- и постпозиции – позиции неразличения.

Таким образом, с формальной точки зрения правило следующее: экспрессивно-вопросительное *знаешь* встречается в пре- и интерпозиции, хотя предпочтительна препозиция; редуцированное – в интер- и постпозиции. Это не означает, однако, что позиция определяет значение (а значения, как мы предположили, у вопросительного и редуцированного *знаешь* разные). Значение определяется коммуникативным намерением говорящего. Но при этом выбор говорящего ограничен: если значение «вопросительное», он выбирает препозицию, реже – интерпозицию; если «парентетическое» – помещает *знаешь* в интер- или постпозицию.

Каковы функционально-семантические различия двух *знаешь*, которые мы условно назвали вопросительным и парентетическим?

Вопросительное *знаешь* используется в тех случаях, когда говорящий считает, что информация важна, значима, интересна для адресата. Например, сообщаемое может прямо касаться адресата: *Я привезу тебе девочку, обязательно. Я, знаешь* [?], *хочу много детей. Девочек и мальчиков, штук пять* [Людмила Улицкая. Казус Кукоцкого // «Новый Мир», 2000].

Особый случай – если говорящий апеллирует к личному опыту адресата (т.е. у адресата действительно предполагается соответствующее знание):

*Там росли цветы, такие, **знаешь** [?], как фиалки, только желтые;*

*В супермаркетах, **знаете** [?], бывает, товар вывешивают, ценник один, а подходишь на кассу, и там цена уже другая [«Русская жизнь», 2012].*

Здесь вводное слово приближается к настоящему вопросу и, как и в случае *помнишь*, на него может быть получен ответ: ‘да, знаю’; ‘нет, никогда не сталкивался с *P* / не видел *P* / *P* не встречалось’. Однако говорящий употребляет вводное вопросительное *знаешь* все-таки не ради получения ответа на вопрос, в отличие от настоящего вопроса, который задается ради ответа, ср.: *Таких, кстати, мы называем толстолобики, рыба такая есть, **знаете**?*.. [Максим Милованов. Естественный отбор (2000)] – здесь говорящий действительно интересуется, известна ли адресату такая рыба.

В случае редуцированного (парентетического) *знаешь* говорящий сообщает свой личный опыт или даже свое мнение, и *знаешь* показывает, что этот факт, мнение, жизненные представления и установки важны для самого говорящего:

— *Со мной тут однажды случай был, **знаешь*** [Дина Рубина. Медная шкатулка (2011-2015)];

*Пела когда-то. И даже играла на гитаре. Со временем перестала. Некогда, **знаете ли*** [Дина Сабитова. Цирк в шкатулке (2007)];

*Если задуматься, то жить вообще вредно, **знаете ли*** [Форум (2005)].

Итак, есть два разных *знаешь* (*знаете*) – вопросительное и редуцированное, – которые различаются не только интонационно, но и функционально (семантически). Предложения с вопросительным *знаешь* сообщают информацию *P*, которая значима, по мнению говорящего, для адресата. В частном случае это может быть информация, прямо касающаяся адресата или апеллирующая к его личному опыту. Но даже если эта информация прямо не относится к адресату, говорящий считает, что адресату ее важно или интересно будет узнать.

Редуцированное *знаешь* / *знаете* – это информация, значимая для самого говорящего, иногда – прямо касающаяся говорящего, иногда – важная для понимания событий. Это может быть не только факт, но и, например, наблюдение говорящего, которым он хочет поделиться с адресатом. Говорящий по каким-то причинам считает нужным выделить, маркировать эту информацию с помощью *знаешь* – в противном случае он не стал бы употреблять ВК. Тем самым он приобщает адресата к своему опыту, к своим ценностям, к своей логике рассуждений и размышлений – и хочет, чтобы адресат тоже обратил внимание на *P*. Весьма показательно, что редуцированное *знаешь* присоединяется не к очевидным и наглядным фактам (ср.: ²*Вон, **знаете**, трамвай идет*), а к фактам, неизвестным адресату, или к оценочным, интерпретационным, объяснительным, комментирующим высказываниям, которые являются не обязательными для развития повествования, а раскрывают позицию и воззрения говорящего: *Он, **знаете**, был со сдвигом в*

голове [Маринетти] *и потом приветствовал фашизм в Италии* [«Эксперт», 2014]; *У нас, **знаешь**, насчет равенства было очень строго* [Дина Рубина. Медная шкатулка (2015)].

Именно к парентетическому *знаешь* относится та интонация доверительности, о которой упоминается в литературе (ср. [29, с. 424]): говорящий показывает, что хотя *P* – это информация личного характера, неизвестная адресату и прямо не касающаяся его, адресат имеет право ее узнать, он может «заглянуть» во внутренний мир говорящего, приобщиться к его мыслям и переживаниям. Итак, термин «доверительное» применим не к любому употреблению *знаешь*, а только к парентетическому.

Таким образом, вопросительное и редуцированное *знаешь* маркируют две разные стратегии взаимодействия говорящего с адресатом: воздействие на личную сферу адресата и вовлечение адресата в свою собственную личную сферу.

Понимаешь. Глагол *понимать*, как отмечалось в [29, с. 413–414], отличается от глагола *знать* качеством информации в зависимой пропозиции *P*. Если знание может возникнуть в результате восприятия ситуации, получения информации о ней из внешнего источника, то понимание предполагает более сложную, неочевидную информацию и требует интеллектуальных усилий – анализа, обработки данных, сопоставления, «проникновения в глубину». В случае ВК *понимаешь* эти усилия могут быть связаны не со сложностью самой ситуации *P*, не с логическими операциями, требуемыми от адресата, а с особой коммуникативной целью говорящего. *Понимаешь* репрезентирует своего рода призыв к адресату войти в положение говорящего, принять во внимание какие-то специальные обстоятельства, сблизить свою позицию с позицией другого человека или просто вникнуть в ситуацию, ср. речевые формулы: *Ты пойми ... / постарайся меня понять ...; Ты не хочешь меня понять ...; Ну, как ты не понимаешь ...; Неужели это трудно понять?!* и подобные. В этом смысле семантика ВК *понимаешь* также связана с усилиями (как и не-вводное употребление *понимать*), но речь идет скорее о психологических усилиях, эмоциональном отклике. Говорящий предлагает адресату разобрататься в ситуации, и именно это требует определенных усилий от адресата:

— *Надоело, Миша, **понимаешь** [?], надоело, — заговорил Алексей, — что она меня всю жизнь пугает и пугает...* [Михаил Кураев. Записки беглого кинематографиста // «Новый Мир», 2001];

— *А я, **понимаешь** [?], Марин, слышу, воду кто-то льет, — радостно хохоча, приговаривала тетя Валя* [Борис Минаев. Детство Левы (2001)].

Иногда ситуация требует объяснения:

*Борька, конечно, не удержался и все **объяснил** сразу, по телефону. — Там, **понимаешь** [?], пацан потерялся, — сказал он. — Наши, московский, семнадцать лет* [Анна Берсенева. Возраст третьей любви (2005)].

Вообще, здесь речь идет о том, чтобы адресат как можно более наглядно представил себе ситуацию, поставил себя на место говорящего или на место участников ситуации. Но глагол *представляешь* здесь неуместен, поскольку у него другая функция – он за-

креплен за выражением удивления, необычности ситуации.

Для эмфатического *понимаешь* (*понимаете*) возможны все три позиции.

Препозиция:

— *Спасибо! — обрадовалась Наташа, когда Светлана вручила ей подарки, так и не подобрав слова — от кого все это. — Натусь, понимаешь* [?], *это я купила, — завела разговор Светлана* [Маша Трауб. Замочная скважина (2012)] — здесь ВК стоит после обращения, но перед Р;

— *Понимаете* [?], *я играла Офелию, а он в это время играл Эйса Вентуру* [«Экран и сцена», 2004.05.06];

Понимаете [?], *о России, то есть о Советском Союзе, как и вообще о Восточном мире, даже и речи не шло* [«Автопилот», 2002.02.15];

— *Понимаешь ли* [?], — *продолжал он, — для него, и для меня тоже, это очень важно* [Елена Халецкая. Синие стрекозы Вавилона (2004)].

Интерпозиция:

Стасик заметил мой удивлённый взгляд. Объяснил, показывая на Митрофанова: — Ему, понимаешь [?], *оса залетела в рот. — Господи, — сказала моя жена, — она и сейчас там? — Да нет. Он, понимаешь* [?], *заканчивал экскурсию в монастыре. И тут ему в рот залетела оса* [Сергей Довлатов. Заповедник (1983)].

Постпозиция:

— *Да не мог он не вырезать, понимаете?* [«Финансовая Россия», 2002.09.19];

У тебя внутри твой скелет. Настоящий скелет, понимаешь? [Андрей Геласимов. Нежный возраст (2001)].

Иногда позиция ВК допускает двоякую трактовку:

А важно то, что он заранее, понимаете ли [?], *заранее знал о подсолнечном масле!* [М. А. Булгаков. Мастер и Маргарита (1929-1940)] — здесь можно говорить о постпозиции по отношению к первому *заранее* или препозиции по отношению ко второму *заранее*.

Постпозиция для *понимаешь* — наиболее слабая позиция с точки зрения вводности. В постпозиции *понимаешь* легко превратить, при некотором изменении интонации (на письме — пунктуации), в самостоятельное высказывание, ср.: *Он не мог не прийти. Понимаешь?* Конечно, это высказывание будет иметь ту же аппеллятивную функцию (аппелляция к адресату); кроме того, такое высказывание нельзя считать самостоятельным в том смысле, что оно привязано к предыдущему высказыванию и без него не имеет смысла.

В «Русской грамматике» Н.Ю. Шведовой [5, § 2221] *знаешь* и *понимаешь* трактуются как асемантические, «пустые вставки». В действительности, это относится только к паразитическому употреблению *понимаешь* (оно бывает только интер- и постпозитивным): *Вы тут, понимаешь, пьянствуете, а я, понимаешь, работаю* [Андрей Белозеров. Чайка (2001)]. В этом случае *понимаешь* может быть заменено на другие паразитические ВК: *значит, так сказать, вообще* и подобные, а также на другие паразитические элементы речи (слова-паразиты), ср. *как бы,*

ну вот. Такое *понимаешь* может не согласовываться с параметрами высказывания: *Вы, понимаешь, всегда чем-то недовольны* (при семантическом употреблении должно было бы быть множ. ч. — *понимаете*). *Знаешь* никогда так не употребляется и всегда соответствует грамматическому числу других форм, связанных с адресатом: *Вы, знаете ли, кое-что забыли*, но: **Вы, знаешь ли, кое-что забыли / *Ты, знаете ли, кое-что забыл.* Это означает, что *знаешь* и в парентетическом статусе сохраняет свою семантику (а именно — семантику доверительности, при всей расплывчатости этой формулировки), противопоставлено другим ВК и не может быть заменено на них, т.е. *знаешь*, в отличие от *понимаешь*, никогда не становится асемантической вставкой.

Таким образом, нейтральное (редуцированное) употребление, которое есть у *знаешь*, и асемантическое, «пустое», паразитическое употребление, которое есть у *понимаешь*, — это разные вещи.

Аналогичным нейтральному, т.е. «доверительно-му», *знаешь* является *понимаешь ли* (*понимаете ли*) — с частицей *ли*:

У него с женой, понимаете ли, школьная еще любовь [Дина Рубина. Медная шкатулка (2011-2015)];

Вскоре после ареста Москвина газета перестала выходить — рекламодатели как-то, понимаете ли, притормозили [«Русская жизнь», 2012].

У *знаешь* и *знаешь ли* такого расхождения нет: обе эти ВК могут употребляться в доверительной функции, но не бывают асемантическими (паразитическими).

Наконец, есть еще один особый тип употребления *понимаешь ли / понимаете ли* — риторический (полемический), который синонимичен *видите ли* (он характерен для таких речевых актов, как упрек, осуждение, выражение возмущения и под.), ср.: *Он, видите ли, не знал! — Он, понимаете ли, не знал! А кто должен знать?!; Ему, понимаете ли, некогда!*

ЭКСКУРС В ОБЛАСТЬ ПУНКТУАЦИИ

В заключение — небольшой экскурс в область пунктуации. Как уже неоднократно отмечалось, предложение, в которое вставляется аппеллятивная вводная конструкция, — повествовательное (оно никогда не бывает вопросительным), при этом ВК, наоборот, имеет вопросительную интонацию. Если ВК находится в пре- или интерпозиции, проблемы не возникает — в конце повествовательного предложения ставится точка (или, в экспрессивном варианте, восклицательный знак). При этом вопросительность самой ВК в русской пунктуации никак не отмечается. Если же ВК оказывается в конце предложения, то ставится вопрос, — и поскольку предложение повествовательное, этот знак вопроса относится только к ВК: *Сказали, что нельзя приходить после девяти вечера, представляешь?* [«Столица», 1997.06.17]; *Где старший с младшим, неизвестно, а средний приехал учиться к нам. Сын миллионера, представляешь?* [«Столица», 1997.04.01]; — *В той газете, — продолжала она, — ну в той, старой, которую мы с тобой смотрели, помнишь?* [Вера Белоусова. Второй выстрел (2000)]; *Но её нигде нет. Нигде, понимаешь?* [Вера Белоусова. Второй выстрел (2000)].

Эта ситуация сама по себе необычна, так как ни в одном пособии по пунктуации не сказано, что знак препинания может относиться только к вводной конструкции, а не к предложению в целом, и что, вообще, интонация вводного слова может обозначаться на письме (при том что интонация всего предложения может никак не обозначаться). Между тем в данной конструкции это именно так.

Однако у ВК *помнишь* ситуация еще более необычная. Видимо, из-за того, что *помнишь* выбивается из группы, отличается от других аппеллятивных ВК наличием у адресата знания *P* и близостью к матричному употреблению (*Помнишь, как я документы не мог найти?* или *Я еще, помнишь [?], документы не мог найти; Помнишь, что говорилось в сводке о гастрономе?* или — *В сводке за вчерашний день... ну, помните [?], о гастрономе...*), у предложений с ВК *помнишь* имеется следующая уникальная пунктуационная особенность. В предложениях вида: — *Нет, не из-за поездки. — Тогда из-за тех баб? — Каких баб? — Ну, помнишь [?], летели с нами в Сочи?* [Андрей Геласимов. Ты можешь (2001)]; — *Перед отъездом он подарил мне микроскоп, маленький, медный, чудо какой хорошенький, — улыбнулась Катя, — а я его сразу же отнесла к Тане Завидоновой, помнишь [?], во втором классе со мной училась?* [Людмила Улицкая. Пиковая дама (1995-2000)], — говорящий не спрашивает: «Они летели с нами в Сочи?»; «Она со мной училась во втором классе?» — наоборот, он это сообщает (напоминает) в качестве факта. Предложение является повествовательным, а вопросительная ВК *помнишь* находится не в конце, а в середине предложения. Тем не менее в конце предложения стоит вопросительный знак, т.е. фактически знак вопроса относится к вводному слову — при этом формально он стоит в конце повествовательного (!) предложения. Очевидно, что ни в одном пособии по пунктуации такой случай не упоминается и не предусматривается.

ВЫВОДЫ

Итак, мы выделили и рассмотрели два основных типа употребления вводных конструкций аппеллятивной группы — акцентированный, или вопросительный, и редуцированный, или нейтральный.

Те формулировки, которые используются в литературе (привлечь внимание собеседника, выделить то, что высказывается), не отражают всего спектра функций и значений вводных конструкций аппеллятивной группы. Вся группа в целом служит для выделения, подчеркивания, а *знаешь, понимаешь* могут употребляться как асемантические, «пустые вставки».

В грамматиках и словарях служебных слов акцентированный и редуцированный типы не выделяются и даже не упоминаются. Не отмечается также, что доверительность, которая часто упоминается в связи с данной группой ВК, относится только к редуцированному типу и, кроме того, зависит от позиции: высказывание *Я, знаете, люблю посидеть на берегу с удочкой* может произноситься с нейтральной («доверительной») интонацией, а в высказывании *Знаете [?], мне все это надоело!* — *знаете* может быть только вопросительно-эмфатическое.

В ходе рассмотрения аппеллятивных вводных конструкций с ментальными глаголами обнаружилось, что предложения с вопросительными ВК реального вопроса не содержат и ответа не предполагают. Автор вопроса: *Ты знаешь, что Сидоров уехал?*, как уже отмечалось, хочет получить ответ («знаю» / «не знаю»). У автора предложения с вопросительной ВК другая коммуникативная цель: вопросительная интонация ВК является приемом «вовлечения» адресата, эмфатического воздействия на него, чтобы придать особое значение сообщаемой информации, акцентировать ее.

При этом, с одной стороны, специфические функции каждого вводного слова и его отличие от других вводных слов связаны, разумеется, с исходной семантикой ментального глагола, которую он имеет в текстовом (не вводном) употреблении. С другой стороны, во вводном употреблении происходит как грамматическая, так и семантическая редукция исходного глагола, т.е. вводный глагол не является механическим переносом «текстового» глагола во вводную позицию, и исходная семантика претерпевает определенные модификации.

СПИСОК ЛИТЕРАТУРЫ

1. Остроумова О.А., Фрамполь О.Д. Словарь вводных слов, сочетаний и предложений. Опыт словаря-справочника. — М.: Изд-во СГУ, 2009.
2. Скобликова Е.С. Современный русский язык. Синтаксис простого предложения. — М., 2006.
3. Правила русской орфографии и пунктуации. Полный академический справочник / Под ред. В. В. Лопатина. — М., 2006.
4. Розенталь Д.Э., Джанджакова Е.В., Кабанова Н.П. Справочник по правописанию, произношению, литературному редактированию. — М., 1994.
5. Русская грамматика. В 2-х тт. Т. 2 / Под ред. Н. Ю. Шведовой. — М., 1980.
6. Appelt D.E. Planning English sentences. — Cambridge: Cambridge University Press, 1985.
7. Banik E. Parenthetical Constructions — an Argument against Modularity // Proceedings of the Workshop on Grammar Engineering Across Frameworks, ACL and AFNLP, Suntec, Singapore, 6 August 2009. — 2009. — P. 46–53.
8. Dehé N., Kavalova Y. Parentheticals. (Linguistik Aktuell / Linguistics Today 106). — Amsterdam: John Benjamins, 2007.
9. Dehé N., Wichmann A. Sentence-initial *I think (that)* and *I believe (that)*. Prosodic evidence for use as main clause, comment clause and discourse marker // Studies in language. — 2010. — Vol. 34, № 1. — P. 36–74.
10. Urmson J. O. Parenthetical verbs // Philosophy and ordinary language / ed. Ch. E. Caton. — Urbana, Chicago, London: University of Illinois Press, 1963. — P. 220–240.
11. Зализняк А.А., Падучева Е.В. О семантике вводного употребления глаголов // Вопросы кибернетики. — М., 1987. — С. 80–96.

12. Blakemore D. Denial and Contrast: A Relevance Theoretic Analysis of 'but // *Linguistics and Philosophy*. – 1989. – № 12. – P. 15–37.
13. Blakemore D. Relevance and Linguistic Meaning: The Semantics and Pragmatics of Discourse Markers. Number 99 in Cambridge Studies in Linguistics. – Cambridge: Cambridge University Press, 2001.
14. Grice H.P. Logic and Conversation // *Syntax and Semantics*. Vol. 3: Speech Acts / eds. Peter Cole, Jerry Morgan. – New York: Academic Press, 1975. – P. 43–58.
15. Kehler A. Coherence, Reference and the Theory of Grammar – Stanford University: CSLI. – 2002.
16. Potts Ch. The Lexical Semantics of Parenthetical-As and Appositive Which // *Syntax*. – 2002. – № 5. – P. 55–88.
17. Potts Ch. The Syntax and Semantics of As-Parentheticals // *Natural Language and Linguistic Theory*. – 2002. – № 20. – P. 623–689.
18. Potts Ch. Conventional Implicatures, a Distinguished Class of Meanings // *The Oxford Handbook of Linguistic Interfaces, Studies in Theoretical Linguistics* / eds. Gillian Ramchand, Charles Reiss. – Oxford: Oxford University Press, 2003.
19. Potts Ch. The logic of conventional implicatures. – Oxford, UK: Oxford University Press, 2005.
20. Кустова Г.И. Некоторые проблемы описания ментальных предикатов // «Научно-техническая информация». Сер. 2. – 1998. – № 2. – С. 22–28.
21. Шатуновский И.Б. Эпистемические глаголы: коммуникативная перспектива, презумпции, прагматика // *Логический анализ языка. Знание и мнение*. – М., 1988.
22. Падучева Е.В. Говорящий: субъект речи и субъект сознания // *Логический анализ языка. Культурные концепты*. – М.: Наука, 1991.
23. Падучева Е.В. Семантические исследования. Семантика времени и вида в русском языке. Семантика нарратива. – М., 1996.
24. Падучева Е.В. Эгоцентрические единицы языка / 2-е изд. – М: ЯСК, 2019.
25. Кустова Г.И. О коммуникативном статусе вводных конструкций с ментальными предикатами // *Труды ИРЯ РАН*. – 2020. – № 3. (в печати).
26. Падучева Е.В. Высказывание и его соотносительность с действительностью (Референциальные аспекты семантики местоимений). – М.: Наука, 1985.
27. Кобозева И.М. О двух типах вводных конструкций с парентетическим глаголом // *Типология и теория языка: от описания к объяснению. К 60-летию А. Е. Кибрика* / ред. Е.В. Рахилина, Я.Г. Тестелец. – М., 1999. – С. 539–543.
28. Богуславский И.М. Сфера действия лексических единиц. – М., 1996.
29. Апресян Ю.Д. Проблема фактивности: *знать* и его синонимы // Ю.Д. Апресян. Избр. труды в 2-х т. Т. 2. – М., 1995. – С. 405–433.

Материал поступил в редакцию 31.05.20.

Сведения об авторах

КУСТОВА Галина Ивановна – доктор филологических наук, ведущий научный сотрудник Института русского языка им. В. В. Виноградова РАН, Москва
e-mail: galinak03@gmail.com

ШВЕЛИДЗЕ Нина Бидзиновна – доктор филологических наук, профессор кафедры языкознания, русской филологии, литературного и журналистского мастерства Пятигорского государственного лингвистического университета, г. Пятигорск
e-mail: sh-nina@yandex.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, ВИНТИ РАН

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

E-mail: feo@viniti.ru

<http://www.udcc.ru>