

НАУЧНО • ТЕХНИЧЕСКАЯ ИНФОРМАЦИЯ

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

Издается с 1961 г.

№ 3

Москва 2016

ОБЩИЙ РАЗДЕЛ

УДК [004 : 001.102] : 001.892

А.Б. Антопольский, В.Ф. Евстафьев, С.В. Моздор

Инфосфера больших научно-производственных комплексов как форма организации: методика анализа

Рассматривается методика анализа организации информационного пространства (инфосферы) больших научно-промышленных комплексов на примере оборонно-промышленного комплекса России.

Впервые по отношению к инфосфере больших научно-производственных комплексов применена совокупность различных аспектов рассмотрения (философский, экономический, технологический, лингвистический, информационный и управленческий).

Ключевые слова: *инфосфера, организация, оборонно-промышленный комплекс, методика*

ПОСТАНОВКА ПРОБЛЕМЫ

В последнее время на государственном уровне принят ряд решений и обнародованы как общероссийские программы, посвященные поддержке инноваций [1-8], так и ряд региональных программ (по Алтайскому краю, по Ростовской области и т.д.), появились публикации на тему инновационного развития России [9-11] и многие другие, что свидетельствует не только о постоянном интересе научной об-

щественности к данной проблеме, но и об определенном развороте государственной политики в сторону поддержки экономики. Связано это прежде всего с внешними вызовами, экономическими и политическими рисками (падение цен на нефть, санкции в связи с событиями на Украине и т.д.).

Проникновение информационных технологий во все сферы жизнедеятельности значительно изменило представления о месте и роли информации в современном обществе. Все отчетливее проявляет себя но-

вая экономическая категория – **национальные информационные ресурсы**, которые являются существенной частью стратегических ресурсов общества и важным фактором развития постиндустриального мира. Иначе говоря, национальные информационные ресурсы осознаны как общественное информационное богатство страны, вопросы их воспроизводства, накопления и использования можно отнести к задачам общегосударственного значения [12].

Можно отметить три замеченные тенденции текущих процессов развития информационной инфраструктуры научно-промышленной сферы, в частности оборонно-промышленного комплекса (ОПК):

во-первых, налицо очередные попытки возврата к всеохватывающим гиперсистемам, опирающимся на концентрацию огромных информационных ресурсов физически в одном месте, на одной технологической площадке, что обычно сулит большие контракты, но никогда не решает всех проблем информационного сопровождения инновационных, текущих производственных и управленческих процессов инфосферы;

во-вторых, как встречная тенденция, давно выросший интерес к локальным автоматизированным информационным системам (АИС) и задачам сопровождения инновационных процессов на низшем и среднем уровне управления, что видно по значительному развитию и распространению корпоративных и локальных систем типа ERP (Enterprise Resource Planning – управление ресурсами предприятия – корпоративной информационной системы для автоматизации планирования, учета, контроля и анализа всех основных бизнес-процессов и решения бизнес-задач в масштабе предприятия (организации)), CRM (Customer Relationship Management – системы управления взаимоотношениями с клиентами), СЭД (системы электронного документооборота) и т.д.;

в-третьих, резко возросший интерес к коммерческой реализации продукции двойного назначения и фрагментов оборонной наукоемкой продукции.

Со времени предыдущих публикаций по вопросам обеспечения Государственного оборонного заказа [13-15] прошло много времени. Ушли в прошлое два дорогостоящих проекта, и из наблюдений отдельных управляющих решений последнего времени стало ясно, что без глубокого анализа **всех аспектов** строения инфосферы ОПК, всей динамики взглядов специалистов за последние почти 60 лет на роль научно-технической информации в общественном производстве, в новой экономической ситуации, без уточнения общей тенденции развития инфосферы, поддерживающей инновационные процессы, легко наступить на «грабли» прошлых ошибок. И велик риск остаться все в том же «колесе» технологических разработок, впуская перерабатывающему случайные потоки информации для аморфных задач в целях слабо определенного контингента малоизученных пользователей.

Некоторые исследователи считают, что информационное общество еще не находится на той стадии осмысления проблемы, чтобы говорить о полноценном процессе анализа и логического проектирования инфосферы ОПК. Выскажем также «крамольную» мысль, что это пространство надо строить иначе: неторопливо создавая информационную сеть имею-

щихся и формируемых локальных информационных ресурсов, находя взаимные интересы, и через эти интересы заставляя участвующих соблюдать нормы и законы информационного обмена.

В то же время, увлечение технологией построения больших информационных сетей, дающей возможность тиражировать слегка модернизированную действующую или разработанную информационную технологию, ведет к «перетягиванию каната» на тему чья БД мощнее, чьи ресурсы важнее и кто дольше «пашет информационное поле», и это ни на йоту не приближает нас к созданию адекватной инфосферы. Не удастся решить проблему и путем механического закачивания накопленных информационных массивов в «новые меха» информационных технологий, создатели которых не особо задумываются об адаптации опробованных разработок к новым информационным сущностям.

Выделим четыре актуальных направления развития информационной инфраструктуры и информационных процессов инфосферы ОПК:

1. Содействие развитию информационных ресурсов и формирование единого информационного пространства (поля) инфосферы ОПК.

2. Развитие инфраструктуры и технической сетевой основы инфосферы ОПК как в виде сети открытых и закрытых каналов связи, так и в виде сети в структуре Интернета.

3. Развитие лингвистики в сторону создания семантической паутины и единого справочно-поискового аппарата по информационным ресурсам инфосферы ОПК.

4. Создание нормативно-технологической основы инфосферы ОПК, особенно процессов коммерческой реализации продуктов, разработанных по Государственному оборонному заказу (ГОЗ).

В настоящей статье мы представили более глубокие, чем в [13-15], общие подходы к исследованию и прогнозу возможного развития инфосферы оборонно-промышленного комплекса.

Основой нашего анализа будут служить три составляющие:

1) рассмотрение информационной составляющей Государственного оборонного заказа (основы всего информационного сопровождения инновационных и текущих научно-производственных процессов ОПК) не просто как системы (что делалось раньше), а как инфосферы, что не одно и то же, и это будет видно при уточнении понятия инфосферы как сущности и объекта анализа;

2) представление инфосферы ОПК России как организации, поскольку теория организации рассматривает каждую конкретную организацию не только в качестве объекта исследования, но и как **методологическую основу организационных методов познания закономерностей развития природных и экономических процессов**;

3) определение сущности инфосферы ОПК России с позиций системного анализа.

Мы понимаем дискуссионность наших взглядов и будем рады, если кого-то наш труд вдохновит на продолжение и углубление исследований.

ИНФОСФЕРА КАК ОРГАНИЗАЦИЯ И ОБЪЕКТ АНАЛИЗА

Странно, но факт, что в большинстве проектов и публикаций почти не фигурирует понятие инфосферы, как сущности, объединяющей субъекты и объекты инфосферы, информатизации, информационного производства¹. Попробуем восполнить этот пробел.

Прежде всего, уточним понятие «инфосфера», особенность инфосферы ОПК как сущности, отражающей и обслуживающей большой научно-промышленный комплекс (НПК), в том числе и прежде всего ОПК.

Философский аспект анализа понятия «инфосфера»

Родоначальником термина «инфосфера» считают В.И. Вернадского, который дал свою интерпретацию концепции ноосферы на основе учения о биосфере. Ноосфера, по Вернадскому, – не просто общество, существующее в определенной среде, и не просто природная среда, подвергающаяся сильному воздействию общества, а нечто целое, в котором сливаются развивающееся общество и изменяемая природа (природная среда). При этом возникает совершенно новый объект, в котором переплетаются законы живой и неживой природы, общества и мышления. Это высшая стадия развития биосферы, характеризующаяся сохранением всех естественных закономерностей, присущих биосфере (при высоком уровне развития производительных сил научной организации воздействия общества на Природу), максимальными возможностями общества удовлетворять материальные и культурные потребности Человека [16].

Взгляд на инфосферу как на философскую категорию и на современную методологию науки развил К. Поппер, который различал три мира [17]:

первый – реальность, существующая объективно;

второй – состояние сознания и его активность;

третий – мир объективного содержания мышления, прежде всего, содержания научных идей, поэтических мыслей и произведений искусства.

Философы прошлого, как и В.И. Вернадский, уделяли большое внимание знанию в субъективном смысле, т.е. второму миру и рассмотрению проблем соотношения второго и первого миров, в то же время они мало изучали особенности жизни науки в третьем мире. А между тем, для понимания сущности науки и закономерностей ее развития, да и процесса познания вообще, по мнению К. Поппера, эта область исследований имеет важнейшее значение. Изучение продуктов научного познания является более важным, чем изучение самого процесса научного познания. Более того, в этом случае даже о самом процессе получения научных знаний мы можем узнать больше, чем при непосредственном его изучении.

Что же представляет собой этот третий мир?

«Обитателями моего третьего мира, — пишет К. Поппер, — прежде всего теоретические системы, другими важными его жителями являются проблемы и проблемные ситуации. Однако его наиболее важными обитателями... являются критические рассужде-

ния и то, что может быть названо... состоянием дискуссий или состоянием критических споров; конечно, сюда относится и содержание журналов, книг и библиотек».

Третий мир представляет собой продукт человеческой деятельности. Он постоянно растет. Вместе с тем, очень важно обратить внимание на его значительную автономность. «Мир языка, предположений, теорий и рассуждений — короче, универсум объективного знания, является одним из самых важных созданных человеком универсумов».

Конечно, третий мир создается человеком. Однако человек во многом не ведает сам, что творит, а результаты его деятельности начинают вести свою собственную жизнь, о которой человек и не задумывался. «С нашими теориями, — пишет К. Поппер, — происходит то же, что и с нашими детьми: они имеют склонность становиться в значительной степени независимыми от своих родителей. С нашими теориями может случиться то же, что и с нашими детьми: мы можем приобрести от них большее количество знания, чем первоначально вложили в них».

Третий мир не мог бы возникнуть без языка науки, ведь это лингвистический мир.

Две важные функции языка – это дескриптивная (описательная) и аргументативная. Вторая из них предполагает наличие первой. Аргументы, конечно, всегда имеют дело с некоторыми описаниями, которые критикуются с точки зрения их правдоподобия и истинности.

Развитие общества приводит к тому, что возможности и значение дескриптивной и аргументативных функций постоянно возрастают. Вместо того, чтобы все больше развивать свою память, человек обзаводится различного рода приспособлениями. Он изобретает бумагу, создает печатные станки и книги, пишущую машинку и, наконец, современную вычислительную технику, что выводит его возможности в совершенно новое измерение.

Важнейшим источником роста третьего мира является критицизм.

Любое исследование начинается с постановки проблемы. Для ее решения ученый развивает теорию, которая критически оценивается через сопоставление с конкурирующими теориями и эмпирическими данными. В результате этой оценки возникает новая проблема.

Этот цикл может быть описан следующей схемой:

$$P \rightarrow TT \rightarrow EE \rightarrow P,$$

где P – исходная проблема, TT – теория, претендующая на решение проблемы, EE – оценка теории, ее критика и устранение ошибок, P – новая проблема.

Таким образом, процесс роста третьего мира «состоит в критике, обладающей творческим воображением». Мы выходим в ней за пределы нашего опыта. Критически относясь к очевидному или освещенному мнением авторитетов, все подвергая сомнению, апробируя самые невероятные возможности, ученый преодолевает границы доступной ему прежде реальности. «Вот каким образом, — пишет К. Поппер, — мы поднимаем себя за волосы из трясины нашего незна-

¹ Термин Г.Т. Артамонова.

ния, вот как мы бросаем веревку в воздух и затем карабкаемся по ней».

Естественнонаучный подход к определению понятия «инфосфера», субъекты и объекты инфосферы подробно рассмотрены в работах И.В. Соловьева [18-19], который предложил следующее определение понятия «инфосфера», от которого мы отталкиваемся в нашей работе:

Инфосфера – это совокупность информации, информационных объектов, информационных процессов, информационной инфраструктуры (электро- и радиосвязь, средства передачи, тиражирования, хранения, воспроизведения, визуализации и обработки информации); сущностей (субъектов), осуществляющих сбор (добывание), упаковку, транспортировку, обработку, хранение, распределение (распространение), производство и потребление информации, а также систем регулирования возникающих при этом отношений производства, использования, владения и распоряжения информацией.

Инфосферу И.В. Соловьев разделял на пять предметных областей деятельности, актуальных и в настоящее время:

- добывание (сбор) и производство информационных продуктов и изделий, исходной (первичной) и производной информации;
- предоставление информационных услуг, хранение и распространение информационных продуктов и изделий;
- создание и применение информационных систем (АИС, баз и банков данных, баз знаний), информационно-телекоммуникационных технологий;
- реализация прав субъектов и объектов на поиск, получение, обмен и использование информации;
- обеспечение информационной безопасности.

Основными элементами инфосферы у И.В. Соловьева являются:

- ♦ пользователи, владельцы и собственники информационных ресурсов (информационных объектов) и средств (систем) добывания, упаковки, транспортировки, распространения, хранения, обработки и визуализации информационных ресурсов;
- ♦ средства (системы) добывания, упаковки, транспортировки, распространения, хранения, обработки и визуализации информационных ресурсов;
- ♦ информационные процессы, между объектами и субъектами информационного взаимодействия².

В работе [19] И.В. Соловьев отмечает, что принципиально важным является понимание сущности **субъекта** и **объекта** инфосферы. В чем же их сущность?

Любой **субъект** инфосферы обладает свойством целостности и эмерджентности³, что позволяет говорить о нём как о системе. Но, чтобы система, объект, индивид, сущность могли быть источником активно-

сти (субъектом), необходимо и достаточно одновременное соблюдение четырёх условий:

способность в своих представлениях отделять себя от внешнего мира, от других субъектов (всякая система ограничена, но далеко не всякая может сама устанавливать свои границы);

наличие собственного (уникального) внутреннего мира, своих (субъективных) представлений, своего внутреннего логического пространства;

способность взаимодействовать с окружающим миром и с другими субъектами и объектами;

наличие интеллекта.

Без выполнения любого из этих условий невозможно в полной мере выполнение трёх других и само существование субъекта как источника активности. В то же время, любая система, объект, индивид, сущность в которой одновременно выполняются все четыре условия, может быть источником активности, следовательно, является **субъектом**.

Выполнение перечисленных четырёх условий приводит, прежде всего, к тому, что субъект инфосферы становится информационно обособленным, у него формируется свое **информационное поле** или **информационно-функциональное пространство**, локализирующее внутренние **информационные процессы** и предопределяющее **внешнее информационное взаимодействие**. Это приводит к образованию сложной **сетевой структуры (инфраструктуры)**, которая требует сложного **сетевого управления**.

Под **субъектом инфосферы** будем понимать объект, индивид, систему или сущность, обладающие интеллектом, способностью к информационному взаимодействию и являющиеся источниками активности, направленной на объект — носитель предметно-практической деятельности и познания⁴.

Объект инфосферы – часть инфосферы, имеющая четко определенное функциональное назначение и границы, существующая в форме предмета, явления, сущности [19] или процесса, на которые направлена предметно-практическая и познавательная деятельность **субъекта инфосферы**.

В качестве объекта инфосферы может выступать и субъект инфосферы (например, специалист в какой-либо области, если какая-либо система способствует решению кадровых вопросов). Объекту инфосферы всегда противостоит субъект инфосферы в его предметно-практической и познавательной деятельности [20]⁵.

Рассмотренные понятия **субъект инфосферы** и **объект инфосферы** позволяют раскрыть содержание информационного взаимодействия в инфосфере, как взаимодействия субъектов и объектов инфосферы, в ходе которого осуществляется потребление и производство информационных объектов. При этом в рамках информационного взаимодействия субъектов и объектов инфосферы реализуются **информационные процессы**, связанные с их текущими **информационными потребностями**.

² Здесь следовало бы добавить «информационные технологии» (ИТ) или информационно-коммуникационные технологии (ИКТ).

³ Эмерджентность (от англ. emergency — возникновение) — качества, свойства системы, которые не присущи ее элементам в отдельности, а возникают благодаря объединению этих элементов в единую, целостную систему.

⁴ Такое понимание закладывает возможность замены человека (специалиста) на отдельных операциях обученной специалистом машиной.

⁵ Это очень важное замечание для моделирования предметной и проблемной областей конкретной инфосферы.

Экономический аспект анализа понятия «инфосфера». Информатизация общества как задача прагматической информатики

Экономический аспект понятия «инфосфера» был подробно рассмотрен Г.Т. Артамоновым в цикле статей [21], где сформулировано новое направление информатики – *прагматическая информатика* и был утвержден новый объект экономики – *информационное производство*. Основным видом продукции информационного производства являются документы, содержащие текстовую и/или табличную запись каких-либо сведений, представленные на любом носителе, а с появлением Интернета, автоматизированных средств связи и телекоммуникаций – и виртуально.

Законы прагматической информатики Г.Т. Артамонова (рис. 1) актуальны и в настоящее время; во многом эти законы пересекаются с рассуждениями К. Поппера о третьем мире. «С позиций информатики, – писал Г.Т. Артамонов, человек представляет собой сложное информационное производство, перерабатывающее поступающую к нему информацию и вырабатывающее информацию, предназначенную другим людям. Развитие и само существование этого производства определяются не только объемами поступающей на переработку информации, но и объемами потребления произведенной им информации».

Итак, в расширенном общественном воспроизводстве участвуют и тесно взаимодействуют два компонента – материальное производство, или производство материальных продуктов, и информационное производство, или производство информационных продуктов, а более коротко – производство информации. Продукты обоих производств участвуют в расширенном общественном воспроизводстве либо как средства потребления, либо как средства производства (рис. 2).

Технологический аспект анализа понятия «инфосфера»

Продукты информационного производства Г.Т. Артамонов делил на две группы по форме их участия в расширенном воспроизводстве общества.

К первой группе относятся продукты, предназначенные для прямого потребления обществом и его членами с явной выгодой для производителя информационного продукта. К таким продуктам относятся произведения искусства и литературы, научные труды, справочники, лекции, учебники, проповеди, массовая информация и дезинформация, пропаганда и реклама, слухи и клевета, разные виды фольклора и пр.

Производство и потребление информационных продуктов этой группы, в принципе, подчиняется тем же законам рынка, что и аналогичные продукты материального производства.

Ко второй группе относятся информационные продукты, которые производители, несмотря на все изменения экономического уклада, предпочитают скрывать от общества, опасаясь какого-либо ущерба для себя. К этой группе информационных продуктов

относится информация о технологических особенностях производства, о материальных затратах на производство, об объемах продаж и полученных доходах, об отрицательных свойствах производимой продукции, о согласованности действий и решений людей с принятыми законами и установленными обществом нормативами и пр.

От авторов. Полемизируя с Г.Т. Артамоновым на основе сегодняшних реалий, мы можем сказать, что современное массовое информационное производство имеет свою явно выраженную специфику, связанную с привычкой общества получать многие виды информации бесплатно. Так работают телевидение, радио, многие интернет-сервисы, например, Яндекс, которые получают свою прибыль от рекламы. Такая бизнес-модель оказалась более эффективной, чем модель платного доступа к информации, например, через подписку.

Кроме того, надо учитывать, что в современном информационном пространстве присутствуют субъекты (акторы⁶), имеющие различную мотивацию (государственные, коммерческие, некоммерческие организации, физические лица, группы лиц). Для многих акторов миссия распространения информации оказывается важнее задачи получения прибыли. Так, зарубежные бюджетные организации, в том числе научные и образовательные, а также научные журналы массово переходят на бизнес-модели открытого доступа, что приводит к большей общественной эффективности, нежели рыночные модели. Но в то же время для информационных производств, работающих в бизнес-моделях открытого доступа должны существовать альтернативные источники финансирования, от государственного до краудфандинга⁷.

В целом можно утверждать, что экономические модели распространения массовой общественно полезной информации в Интернете сейчас находятся в процессе становления. Подробно эти процессы исследованы в фундаментальной монографии А.Б. Долгина [22].

Производство информационных продуктов изначально развивалось вместе с материальным производством. Его назначением было информационное обслуживание материального производства, а точнее – *информационное обслуживание управления материальным производством*. При этом под управлением мы понимаем процесс подготовки, принятия и исполнения решений.

Информационное обслуживание играло второстепенную роль, и на него при организации управления соответствующим производством старались тратить как можно меньше финансовых, материальных и кадровых ресурсов. Информационное обслуживание ориентируется на выполнение запросов на ту или иную информацию, поступающую от органов управления основным производством, и играет пассивную вспомогательную роль по отношению к основному производству.

⁶Актор – субъект деятельности в обществе, выступающий в качестве источника социального действия (термин предложен Т. Парсонсом).

⁷Краудфандинг – механизм привлечения финансирования от широких масс с различной целью.

<p><u>Закон неистребимости информации:</u> информация не исчезает при ее потреблении или использовании в качестве предмета труда и не амортизируется при ее использовании в качестве средства труда</p>	<p><u>Закон самодостаточности информационного производства:</u> информационное производство может развиваться за счет использования в качестве предмета труда собственных продуктов, а в материальном производстве информация никогда не может использоваться в качестве предмета труда</p>	<p><u>Закон информационного таянения:</u> интерес субъектов информационного производства друг к другу обратно пропорционален квадрату расстояния между ними</p>	<p><u>Закон информационной сложности:</u> интерес к документу обратно пропорционален его сложности</p>	<p><u>Закон старения информации:</u> интерес к информации обратно пропорционален времени, прошедшему с момента ее последнего проявления</p>
<p><u>1-й закон информационной динамики:</u> только полное копирование информационных продуктов сохраняет информацию</p>	<p><u>2-й закон информационной динамики:</u> никакой информационный продукт не гарантирован от случайных искажений при своем движении в общественном информационном производстве</p>	<p><u>3-й закон информационной динамики:</u> намеренные искажения информации обнаруживаются только по их последствиям</p>	<p><u>Закон минимального осмысления текста:</u> потребитель сложного текста стремится предельно ограничить тиреозахваты на выявление части его смысла, достаточной для принятия какого-либо решения</p>	
<p><u>ГЛАВНАЯ ЗАКОНОМЕРНОСТЬ РАЗВИТИЯ ВЕРБАЛЬНЫХ ЯЗЫКОВ:</u> языковые конструкции живут, пока и поскольку они используются в реальном информационном производстве</p>	<p><u>Закон относительной однозначности восприятия смысла текстов:</u> однозначность восприятия смысла текстов на каком-либо языке может быть обеспечена только строгой технологической дисциплиной в рамках единого информационного производства</p>	<p><u>Принцип разделения функций общественных отношений:</u> в каждом общественном процессе функции заказчика, подрядчика и аудитора должны выполняться тремя различными, максимально не зависимыми друг от друга субъектами</p>	<p><u>Принцип баланса между доступностью и тайной:</u> в информационном общественном производстве в отдельных его частях функции обеспечения необходимой полноты осуществленной информации, функции установления порядка сохранения тайны, производства информации и функции определения равенства между полнотой опущения и мерами защиты тайны должны выполняться различными субъектами не зависимо друг от друга субъектами</p>	
<p><u>1-й принцип управления информацией:</u> полнота осуществления общественной необходимой информации от субъектов информационного производства достигается только средствами принуждения</p>	<p><u>2-й принцип управления информацией:</u> качество согласования предлагаемого информационного продукта обеспечивается полнотой базы прецедентов, доступной производителю этого продукта, и эффективностью принуждения производителя</p>	<p><u>Принцип разделения функций общественных отношений:</u> в каждом общественном процессе функции заказчика, подрядчика и аудитора должны выполняться тремя различными, максимально не зависимыми друг от друга субъектами</p>	<p><u>Решающий принцип управления информацией:</u> управление информационным производством или информацией в организации должно быть функцией первого лица в ее руководстве</p>	

Рис. 1. Законы прагматической информатики Г.Т. Артамонова [21]

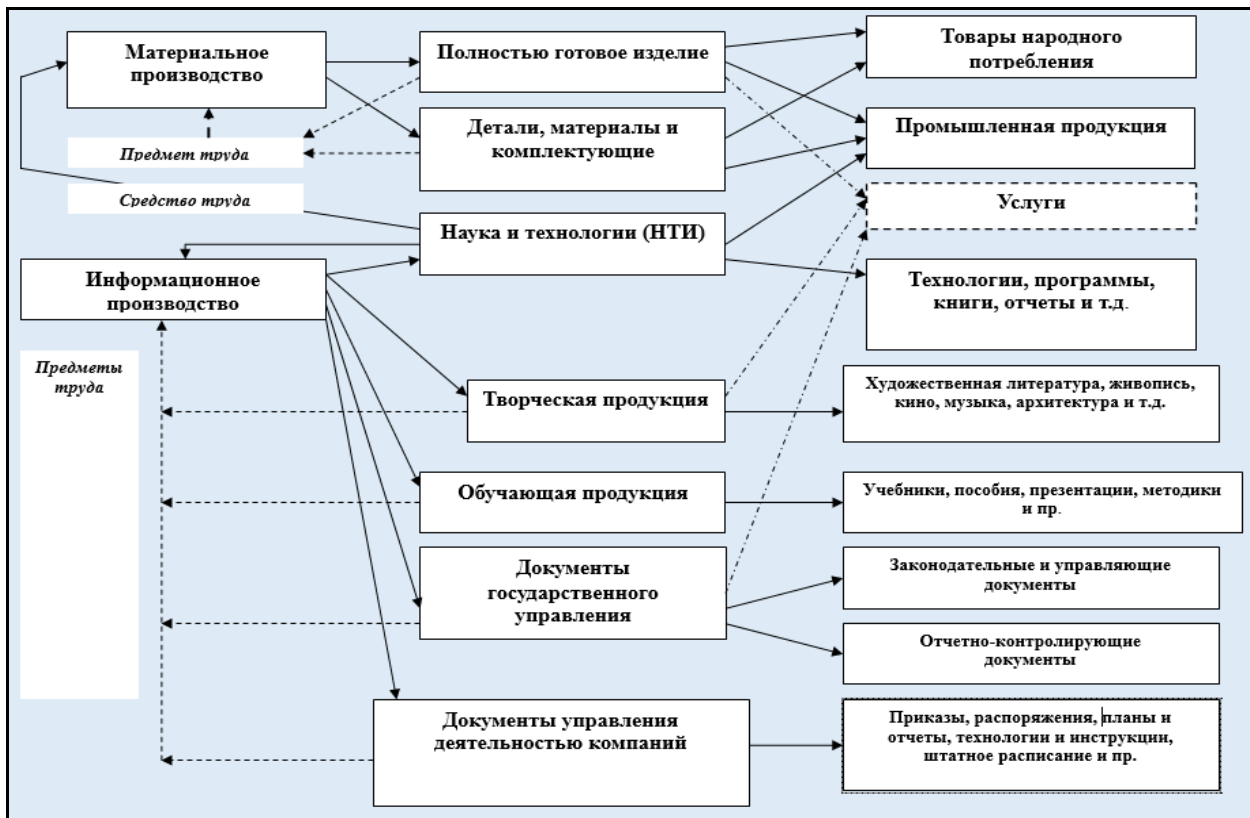


Рис. 2. Взаимоотношение материального и информационного производства

Управленческий аспект анализа понятия «инфосфера»

Под «управлением информацией» Г.Т. Артамонов понимал управление информационным производством любой организации. Он затронул сложнейшую, не разрешенную до настоящего времени проблему столкновения личных, коммерческих интересов разработчика, как участника информационного производства, и общества, как потребителя результатов информационного производства. Именно эта проблема сейчас является двигателем развития инфосферы ОПК. За время, прошедшее с момента написания статей [21], ситуация в России изменилась, пожалуй, к худшему. Практически исчезло директивное управление информационным производством на общегосударственном уровне, лишь кое-где оно реализуется на уровне отраслей и регионов. Однако за рубежом качество управления информационным производством постоянно растет [20].

В сложном процессе информационного взаимодействия разработчиков Г.Т. Артамонов выделял два момента: первый касается согласования предложения разработчика (это предложение и есть новый информационный продукт) с ранее принятыми решениями в рамках текущего или предыдущих проектов, а также со смежниками по разработке; второй – относится к оповещению всех заинтересованных разработчиков о решениях, принятых по предложениям их коллег.

Три вида ресурсов производства – кадровые, финансовые и материальные – составляют неразделимую триаду, определяющую эффективность деятельности любой организации, любого производства. Давно уже осознано, что к этой триаде добавился

еще один ресурс – *информационный*, ибо сложность взаимодействия при управлении ресурсной триадой потребовала существенного развития информационного производства, обеспечивающего это взаимодействие.

Таким образом, наряду с информационными ресурсами отдельных информационных производств, взаимодействующих друг с другом в процессе общественного воспроизводства в соответствии со своими потребностями, в обществе функционируют информационные производства, обязанные по своему положению формировать общедоступные государственные информационные ресурсы. Состав, а также порядок формирования и использования таких информационных ресурсов должны определяться государством с учетом интересов общества в целом. Этот порядок должен определяться законодательно с учетом права граждан на информацию, с одной стороны, и необходимостью сохранения личной, коммерческой, служебной и государственной тайн – с другой.

Лингвистический аспект анализа понятия «инфосфера»

Ситуация, складывающаяся в настоящее время в логико-лингвистическом описании и представлении метаданных, опираясь на опыт научных коммуникаций в Интернете, прежде всего в WWW, детально рассмотрена в работе [23], и здесь мы только наметим основные направления.

В информационном пространстве, значительная часть которого формируется в Интернете, в основном в Web, генерируются многочисленные информационные ресурсы, различающиеся по самым различным

параметрам – структуре данных, тематике, объектам, программному инструментарию. Эти ресурсы существенно разрознены, недостаточно структурированы и систематизированы.

При их описании недостаточное внимание уделяется вопросам интероперабельности⁸, слабо применяются соглашения по стандартизации электронного представления информационных ресурсов, средства, поддерживающие интеграцию информационных ресурсов, повышение качества поиска. В результате пользователь не может получить полную и достоверную информацию, представляющую для него интерес.

Очевидно, что каждая область науки оперирует своими специфическими данными и имеет потребность в собственных форматах их представления, что обусловлено требованиями функциональности соответствующих систем обработки информации. Этим объясняется малая степень интеграции научных данных, по сравнению с образовательными, библиотечными и другими типами ресурсов. Тем не менее, необходимость обеспечения активных научных коммуникаций, прежде всего на междисциплинарном уровне, требует решения задачи интеграции научных ресурсов, начиная с верхнего уровня.

В сложившейся в Интернете ситуации данные о ресурсах, с которыми оперируют поисковые системы, осуществляющие полнотекстовый поиск, представлены в основном в слабоструктурированной форме. Пользователь в этих условиях получает огромное количество шума, среди которого трудно выбрать релевантную информацию. Учитывая это обстоятельство, для представления сведений о ресурсах стали использовать структурные формы, выделять метаданные, описывающие содержимое ресурса в виде набора именованных значений, для связи с другими ресурсами. Метаданные используются для автоматизированного анализа содержания ресурса, построения поисковых индексов и позволяют обеспечить достаточно высокую точность и эффективность поиска разнородной информации.

Важнейшей характеристикой любой информационной системы является глубина (специфичность) структуризации метаданных. Слишком большая глубина усложняет процессы подготовки метаданных, затрудняет их интеграцию, слишком малая глубина снижает функциональность системы. Понятно, что чем уже тематика системы, чем более однородные ресурсы в ней используются, тем глубже могут быть структурированы данные. Однако во многих случаях при интеграции разнородных ресурсов высокая степень детализации не требуется.

В связи с этим необходимо определить минимальную глубину структуризации метаинформации о ресурсах, которая может наращиваться при увеличении функциональности системы. Для этого нужна методика развития схем метаданных, формирования профилей метаданных.

Для обеспечения взаимодействия существующих разнородных научных систем на информационном уровне необходимо выработать стандарты научно-информационного сообщества на интерфейсы взаимодействия и профили метаданных, что позволило бы реализовать инструментальные средства, обеспечивающие интеграцию данных в единую среду. Эти стандарты должны включать:

- типовые интерфейсы взаимодействия (форматы данных, протоколы обмена) отдельных информационных источников;
- профили метаданных, предоставляемые этими источниками;
- справочники и классификаторы ресурсов;
- требования информационной безопасности и разграничения прав доступа к ресурсам.

МЕТОДОЛОГИЧЕСКАЯ СУЩНОСТЬ ПОНЯТИЯ «ОРГАНИЗАЦИЯ»

Теория организации, вторая составляющая часть нашего исследования, впервые изложенная еще в 1925 г. в фундаментальном труде «Тектология (Всеобщая организационная наука)» А.А. Богдановым [24], отличается особым подходом к изучению каждого явления материального и духовного мира с позиций организационного опыта. Любой исследуемый объект может быть рассмотрен как отношение между частями целого или целого с окружающей его внешней средой. При этом выясняется, что законы организации едины для любых объектов, а сами разнородные явления отождествляются через аналогию связей и закономерностей.

В теории организации объектом изучения выступает организационный опыт окружающей нас действительности. При этом основные задачи познания заключаются в систематизации данного опыта, в осмыслении способов организации природы и человеческой деятельности, объяснении и обобщении этих способов, установлении тенденций и закономерностей их развития и роли в экономике мирового процесса. Общий план этой теории (по А.А. Богданову – всеобщей организационной науки) «аналогичен плану любой из естественных наук, но объект существенно иной». Теория организации «имеет дело с организационным опытом не той или иной специальной отрасли, но всех их в совокупности; другими словами, она охватывает материал всех других наук и всей той жизненной практики, из которой они возникли; но она берет его только со стороны метода, т.е. интересуется повсюду способом организации этого материала» [24].

Предмет теории организации – организационные отношения, т.е. связи и взаимодействия различного рода целостных образований и их структурных составляющих, а также процессы и действия организующей и дезорганизующей направленности. Многообразие видов организационных отношений наглядно раскрывается через введенные А.А. Богдановым регулирующие механизмы: **конъюгации** (соединение элементов и комплексов между собой); **ингрессии** («вхождение», образование связующего промежуточного звена между разнородными звеньями при формировании новой целостности); **дезингрессии** («вхождение», образование нейтрализующего, разрушающего звена в процессе дезорганизации некой целостности); **цеп-**

⁸ Интероперабельность (англ. *interoperability*) — способность к взаимодействию двух и более систем или компонентов для обмена информацией и использования этой информации.

ной связи (объединение посредством общих звеньев); **отбора и подбора**, (стихийно регулирующие меры); **бирегуляции** (обратная связь); **агрессии и депрессии** (централистский и скелетный способы формирования комплексов).

Современное определение представляет понятие «организация» в более широком смысле таким образом [25]:

1) внутренняя упорядоченность, согласованность, взаимодействие более или менее дифференцированных и автономных частей целого;

2) совокупность процессов или действий, ведущих к образованию и совершенствованию взаимосвязи частей целого;

3) объединение людей, совместно реализующих некоторую программу или цель и действующих на основе определенных процедур и правил.

Для нашего анализа наиболее близок третий пункт, имеющий более узкое, социальное приложение

применительно к человеческой организации. Таким образом, в самом общем случае под **организацией** следует понимать упорядоченное состояние элементов целого и процесс по их упорядочению в целесообразное единство.

Структурное представление понятия «организация» возможно в статике и динамике (рис. 3). В статике – это некоторое целостное образование (социальное, техническое, физическое, биологическое), имеющее вполне определенное предназначение. В динамике – это разнообразные процессы по упорядочению элементов, формированию и поддержанию целостности вновь создаваемых или функционирующих природных объектов. Эти процессы могут состоять из целенаправленных действий людей, и тогда можно говорить об организации как функции управления, либо состоять из естественных физических процессов, т.е. иметь самоорганизующееся начало.

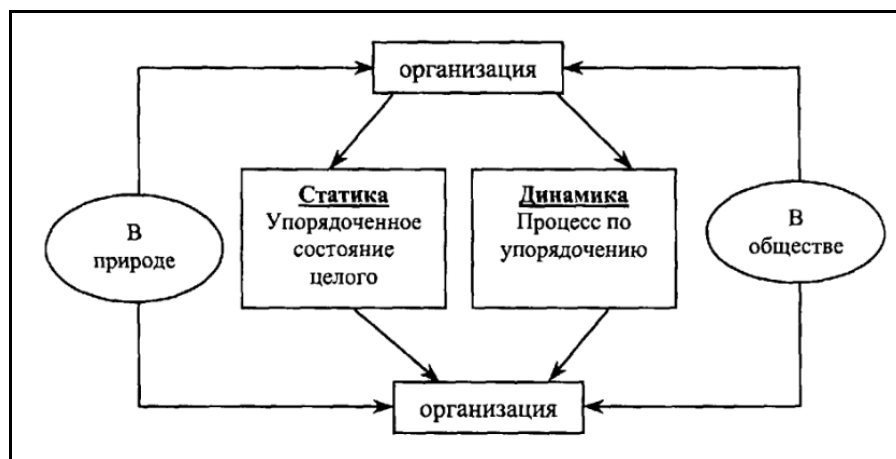


Рис. 3. Структурное представление понятия «организация» [26]



Рис. 4. Структура понятия «организация» применительно к производственным системам [27]

Р.А. Фатхутдинов [27] конкретизирует и уточняет содержание компонентов понятия «организация» и их связи применительно к *производственным системам* (в том числе инфосфере ОПК)⁹ Он организацию рассматривает с точки зрения (рис. 4):

1) как субстанцию системы, ее структуру и содержание, суть и форму;

2) как процесс по переработке входа системы в ее выход как функцию управления по достижению миссии, целей системы (т.е. в динамике).

Субстанцию и процесс следует рассматривать в единстве. Субстанция первична, процесс вторичен.

МЕТОДИКА СИСТЕМНОГО АНАЛИЗА ИНФОСФЕРЫ

В качестве аппарата системного анализа инфосферы ОПК как организации нами выбрана методология, изложенная в работе В.Г. Овчинникова [28] и состоящая в следующем. Любой объект можно рассматривать различными способами, например, задавая относительно него вопросы и получая соответствующие ответы. Результаты такого рассмотрения (и, в частности, ответы на поставленные вопросы) обычно называют *«предметами»*.

Инфосфера научно-промышленного комплекса объективно не может быть одной автоматизированной информационной системой (АИС) а только их объединением. Однако при самом общем взгляде на нее мы можем *условно*, как допуск, который делают иногда в своих расчетах математики, назвать ее су-

персистемой, свойства которой в значительной степени повторяют характерные свойства локальных и больших систем, прежде всего АИС. Аспекты рассмотрения информационной системы определяются теми вопросами, которые интересуют проектировщика. Предпочтительно, очередность вопросов должна быть такова (см. таблицу).

Системные модели в таблице упорядочены по возрастанию их сложности (сверху вниз): каждая последующая модель включает предыдущую, добавляя к ней небольшое число новых отношений, связывающих ранее введенные предметы с предметами нового типа, такими, как моменты времени, средства реализации, элементы физического пространства и организационные элементы.

Количество дополнительных типов отношений в каждой новой модели невелико, обуславливает небольшое усложнение моделей. Каждый новый набор предметов и типов отношений образует отдельный «слой», накладываемый на ранее построенные модели, в результате чего получается иерархия вложенных и достаточно простых моделей – своеобразный «слоеный пирог», «матрешка» или «луковица» моделей (рис. 5).

Автор работы [28] представляет предмет своего исследования (информатику) в виде двух конусов¹⁰, напоминающих «песочные часы», и рассматривает его в нисходящем порядке (сверху – вниз, как показано на рис. 6).

Аспекты рассмотрения информационной системы

Модель (тип структуры)	Вопросы, на которые отвечает модель
Прикладная функциональная модель (ПФМ) инфосферы ОПК	Что делает (должна делать) система? Для чего она предназначена? Какие функция она исполняет в рассматриваемой предметной области?
Информационно-логическая модель (ИЛМ) инфосферы ОПК	Как (каким образом, с помощью каких методов) система осуществляет (должна осуществлять) свои функции? Какова схема функционирования системы? Какая информация и в какой логической последовательности обрабатывается (должна обрабатываться) в системе?
Функционально-техническая модель (ФТМ) инфосферы ОПК	Какие материальные средства применяются (должны применяться) в системе? Где (на каких носителях) записана и хранится (должна быть записана и храниться)? Какими средствами информация обрабатывается (должна обрабатываться)?
Территориально-распределенная модель (ТРМ) инфосферы ОПК	Как все эти средства распределены (должны быть распределены) в физическом пространстве (в организации, группе организаций и предприятий, в области, в России)
Организационно-управляющая модель (ОУМ) инфосферы ОПК	Какова организационная структура системы? Как организовано (должно быть организовано) взаимодействие системных компонентов? Кто и как ими управляет (должен управлять)? Какие средства управления («кнут и пряник»)? В каких условиях они применяются?

⁹ Инфосферу научно-производственных комплексов, в частности инфосферу ОПК, можно отнести к производственным организациям.

¹⁰ Некоторые специалисты (И.В. Лоханов) называют такую схему «биконусом».

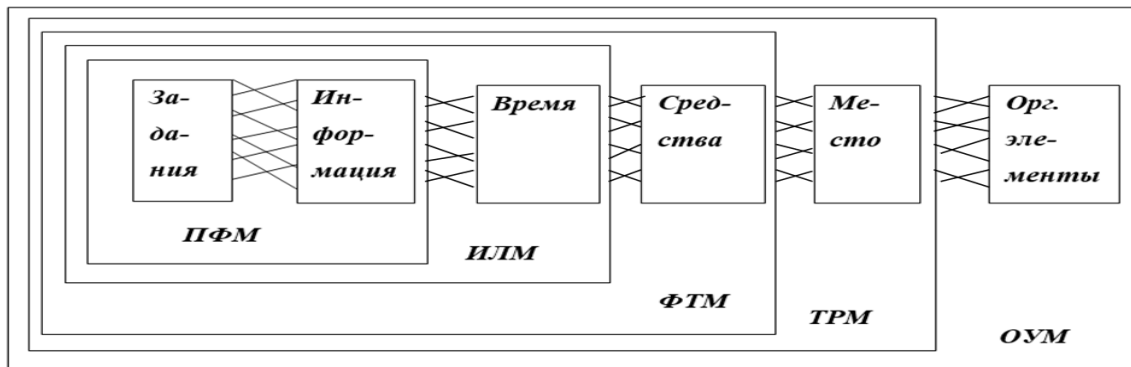


Рис. 5. Иерархия вложенных аспектных моделей информационной системы

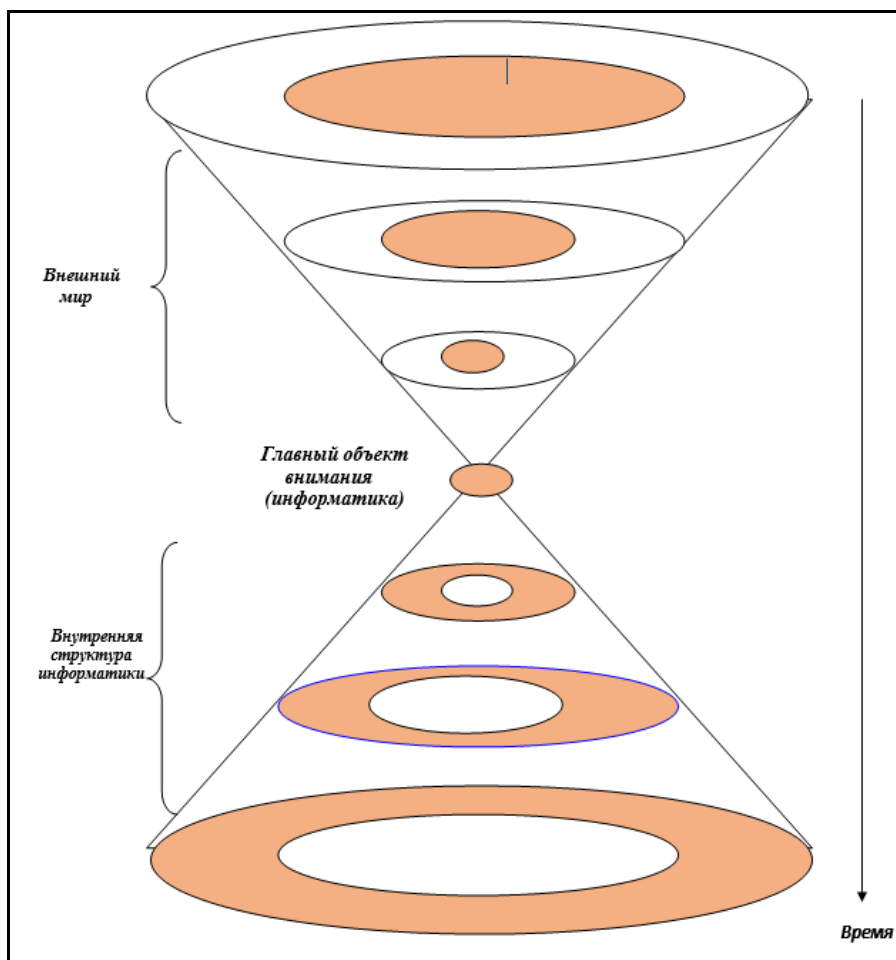


Рис. 6. Порядок рассмотрения инфосферы «Информатика» в работе [28].

Выделенные на рис. 6 области рассматриваются на соответствующем этапе познания, а не заштрихованы – те, что были объектом внимания на предыдущих этапах.

В процессе рассмотрения в верхней части «конуса» («внешний мир») идет сужение предметных областей: вначале рассматривается ряд крупных явлений, каждое из которых включает, как часть, другое, представленное ниже, после чего описывается внутренняя структура этого, «внутреннего», объекта.

Так постепенно все более и более уточняется объект внимания до основного («главного»). Нижний

конус детализирует внутреннюю структуру этого главного объекта.

Таким образом, мы представили, с каких точек зрения можно рассматривать инфосферу, в том числе инфосферу больших научно-производственных комплексов, включая инфосферу оборонно-промышленного комплекса, а также что представляет собой теория организации по отношению к интересующим нас объектам анализа. Представленная модель определяет содержание дальнейшего анализа организации инфосферы оборонно-промышленного комплекса.

СПИСОК ЛИТЕРАТУРЫ

1. Концепция развития национальной информационной системы научной, научно-технической и инновационной деятельности России // Межотраслевая информационная служба. – 2006. – №3. – С. 3-21.
2. Концепция государственной инновационной политики Российской Федерации на 2002-2004 гг. // Межотраслевая информационная служба. – 2001. – №. 4 (117). – С. 5-9.
3. Концепция и основные положения проекта федерального закона «О государственно-частном партнерстве в сфере военно-технического обеспечения обороны и безопасности страны». – URL: http://www.tpprf.ru/common/upload/19.07._Kontseptsiya_zakona_GCHP_v_OPK2.ppt.
4. Послание Президента Федеральному Собранию. – URL: <http://www.kremlin.ru/news/19825>.
5. Постановление Правительства РФ от 23 мая 2015 г. N 497 «О Федеральной целевой программе развития образования на 2016-2020 годы». – URL: <http://www.garant.ru/hotlaw/federal/627520> (дата обращения: 05.07.2015).
6. Программа фундаментальных научных исследований в Российской Федерации на долгосрочный период (2013–2020 годы). – URL: http://www.consultant.ru/document/cons_doc_LAW_160303/?frame=1 (дата обращения: 05.07.2015).
7. Распоряжение Правительства Российской Федерации от 20 октября 2010 г. N 1815-р г. Москва «О государственной программе Российской Федерации «Информационное общество (2011-2020 годы)». – URL: <http://www.rg.ru/2010/11/16/infobschestvo-site-dok.html> (дата обращения: 05.07.2015).
8. Стратегия инновационного развития Российской Федерации до 2020 г. (проект). – URL: http://www.economy.gov.ru/minec/activity/sections/innovations/doc20101231_016 (дата обращения: 05.07.2015).
9. Брижань А.В. Национальная инновационная система как ключевой элемент развития экономики России. – URL: <http://www.bestreferat.ru/referat-94374.html>.
10. Голиченко О.Г. Российская инновационная система: проблемы развития. // Вопросы экономики. – 2005. – №12. – С.16-34.
11. Иванов В. В. Инновационная политика России: варианты и перспективы. – URL: http://www.issras.ru/papers/inn148_2011_Ivanov.php (дата обращения: 04.07.2015).
12. Нечипоренко В.П. Из опыта создания и обеспечения функционирования информационной инфраструктуры науки и техники // Межотраслевая информационная служба. – 2015. – №1. – С. 23-27.
13. Головатый И.Н., Евстафьев В.Ф., Моздор С.В. Информационная среда оборонного комплекса в русле программы построения информационного общества (взгляд из прошлого в ближайшее будущее) // Межотраслевая информационная служба. – 2011. – № 4 (157). – С. 3-13.
14. Головатый И.Н., Евстафьев В.Ф., Моздор С.В. Информационная среда оборонного комплекса. Часть 2. Концепция развития информационной инфраструктуры Рособоронзаказа//Межотраслевая информационная служба. – 2012. – № 2 (159). – С. 3-13.
15. Иванов М.О., Моздор С.В., Ольхов Е.Н. Информационная среда оборонного заказа и передача достижений оборонного комплекса (новый взгляд на межотраслевой обмен) // Оборонный комплекс – научно-техническому прогрессу России. – 2004. – №2 (122). – С. 8-18.
16. Вернадский В.И. Несколько слов о ноосфере // Успехи современной биологии – 1944. – № 18, Вып. 2. – С. 113–120.
17. Концепция Третьего мира К.Поппера и научные революции // Биофайл. Научно-информационный журнал. – URL: <http://biofile.ru/his/2050.html>.
18. Соловьёв И.В. О происхождении и содержании понятия «инфосфера». Инфосфера как объект исследования об информации // Фундаментальные исследования. – 2013. – № 6–1. – С. 66-71. – URL: www.rae.ru/fs/?section=content&op=show_article&article_id=10000626 (дата обращения: 30.03.2015).
19. Соловьёв И.В. О субъекте и объекте инфосферы // Научная библиотека КиберЛенинка. – URL: <http://cyberleninka.ru/article/n/o-subekte-i-obekte-infosfery> (дата обращения: 30.03.2015).
20. Чернова С.А. Оценка направлений политики ЕС и РФ в области исследований, технологических инноваций и образования. – URL: <http://riep.ru/upload/iblock/05c/05caa23d22e503c29bbce8ae276a413f.pdf>
21. Артамонов Г.Т. Информатика: теория и практика (заготовки к книге) // Научно-техническая информация. Сер 1. – 1997. – № 8. – С. 30-33; 1998. – №1. – С. 29-34; №4. – С. 31-36; № 6. – С. 31-35; №12. – С. 29-34; 1999. – №6. – С. 36-43.
22. Долгин А. Манифест новой экономики. Вторая невидимая рука рынка. – URL:http://www.koob.pro/dolgin/new_economy.
23. Бездушный А.А., Бездушный А.Н., Серебряков В.А., Филиппов В.И. Интеграция метаданных Единого Научного Информационного Пространства РАН. – М.: ВЦ РАН, 2006.
24. Богданов А.А. Тектология (Всеобщая организационная наука). Кн. 1 – 2. – М.: Экономика, 1989. – URL: http://www.pseudology.org/science/Bogdanov_Tektologia1a.pdf.

25. Юдин Б.Г. Организация // БСЭ. 3-е изд. Т. 18. – М.: Советская энциклопедия, 1974. – С. 473-475.
26. Теория организации: учебник / под ред. В.Г. Алиева, изд. 2-е, перераб. и доп. – М.: Экономика, 2003. – 432 с.
27. Фатхутдинов Р.А. Организация производства: учебник. 3-е изд., перераб. и доп. – М.: ИНФРА-М, 2007. – 544 с.
28. Овчинников В.Г. Методологическое введение в информатику. – М.: Компания Спутник+, 2004. – 217с.

Материал поступил в редакцию 22.10.15.

Сведения об авторах

АНТОПОЛЬСКИЙ Александр Борисович – доктор технических наук, профессор, главный научный сотрудник Института управления образованием РАО, Москва
E-mail: ale5695@yandex.ru

ЕВСТАФЬЕВ Владимир Федорович – доктор технических наук, профессор, начальник управления ФГУП ВИМИ, Москва
E-mail: evf@vimi.ru

МОЗДОР Сергей Владимирович – кандидат технических наук, доцент, консультант ФГУП ВИМИ
E-mail: dr44@inbox.ru

Вычисление закона распределения случайной величины

Дается классификация непрерывных случайных величин, строятся три системы непрерывных распределений, излагается суть устойчивого метода вычисления закона распределения (в том числе рангового) и оценок параметров, а также универсального метода моментов. Обосновываются универсальные законы рассеяния и старения публикаций.

Ключевые слова: непрерывные случайные величины, непрерывные распределения, вычисление закона распределения и оценок параметров, рассеяние и старение публикаций

ВВЕДЕНИЕ

В настоящее время при отыскании наилучшего аппроксимирующего распределения для описания статистического вариационного ряда используется метод выдвижения гипотез. При этом вид теоретического распределения устанавливается по форме гистограммы, что не дает однозначного результата. Например, для аппроксимации симметричной гистограммы чаще других используется нормальный закон из всего разнообразия симметричных распределений. В этом случае по выборочным значениям случайной величины вычисляются среднее их значение и среднее квадратическое отклонение. Далее эти значения принимаются в качестве оценок параметров теоретического закона (того же нормального) и осуществляется проверка степени близости теоретического и статистического распределений по критериям согласия (например, «хи-квадрат» Пирсона) при заданном уровне значимости α – обычно в пределах от 0.01 до 0.1. Величина α представляет собой вероятность отклонения верной гипотезы. Вводится также мощность критерия $1-\beta$, где β – вероятность принятия неверной гипотезы. Однако, несмотря на такие предосторожности, некоторые неверные гипотезы могут быть приняты. В итоге задача нахождения закона распределения может оказаться нерешенной даже при выдвижении нескольких гипотез подряд.

Та же ситуация возникает при аппроксимации статистических ранговых распределений, которые широко применяются в информатике, технике, математической лингвистике, библиотечной деятельности, социологии и других областях знания. Для аппроксимации таких распределений многие авторы используют законы Дж. Ципфа, С. Бредфорда (закон рассеяния публикаций), а также более общие модели

– логарифмически нормальный закон и др. Метод выдвижения гипотез в этих случаях также не дает однозначных результатов. Это происходит по нескольким причинам:

- набор используемых распределений крайне ограничен;
- отсутствует классификация случайных величин;
- рассматриваются выборки недостаточного объема (особенно это касается ранговых распределений);
- попытки описать неоднородные выборки (например, ранговые распределения слов) одним теоретическим распределением;
- неудачно выбранная форма представления ранговых распределений (в системе координат «логарифм ранга – логарифм частоты»), которая несет слишком мало информации о статистическом распределении и не имеет вероятностного смысла;
- отсутствие общего подхода, т.е. предпринимаются попытки решить одну из частных задач без предварительного решения общей задачи.

В такой ситуации любой метод оценивания будет неэффективным.

Несмотря на недостатки метода выдвижения гипотез, он все еще используется в научных исследованиях и преподается в университетах. Использование этого метода обосновывается тем, что по статистическому ряду якобы невозможно установить вид теоретического распределения. Это глубокое заблуждение, поскольку любая случайная величина, если она представляет собой статистически однородную совокупность большого числа значений, несет в себе достаточное количество информации, позволяющее отличить ее от другой случайной величины. И задача

исследователя заключается в том, чтобы эту информацию извлечь из выборочных данных путем нахождения наилучшего аппроксимирующего распределения. Если же пытаться найти его путем перебора небольшого числа теоретических распределений (как правило, не более 10–15), которые обычно заложены в известные приложения, то действительно будет слишком мала вероятность правильного решения этой задачи.

Поскольку закон распределения является наиболее полной характеристикой случайной величины, то наиболее важной задачей при обработке статистических данных является его **вычисление**, так как закон распределения позволяет осуществлять различные исследования, в том числе давать прогноз и решать множество практических задач. Приложения, которые не в состоянии решать эту задачу, перекладывают всю ответственность по отысканию закона распределения на пользователя, не предоставляя ему практически никаких гарантий точного решения этой задачи. Использование таких приложений в системах менеджмента качества может нанести огромный материальный и моральный ущерб предприятию, и в целом – экономике государства, но разработчики этих приложений не несут никакой ответственности.

В литературе по теории вероятностей и математической статистике приводится множество методов оценивания параметров, даже по несколько для каждого отдельного распределения (все это частные задачи), но нет надежного **метода вычисления закона распределения** по статистическим данным (т.е. не решена общая задача). А без такого метода практически бесполезно вычислять оценки параметров наугад выбранного аппроксимирующего распределения. Таким образом, **любой метод оценивания параметров должен в первую очередь обеспечить вычисление закона распределения.**

Естественно, что к разработке различных методов оценивания параметров принуждают различные факторы, например, неполнота статистических данных, разная форма их представления, но и в этом случае необходимо в первую очередь вычислить закон распределения.

В настоящей статье дается обоснование и построение методов вычисления закона распределения и оценок параметров по статистическому ряду без выдвижения гипотез и проверки каждой из них по критериям согласия, строятся три системы непрерывных четырехпараметрических распределений, излагаются два метода вычисления закона распределения и оценок параметров, разработанные автором, – общий устойчивый метод, который по точности не уступает методу наибольшего правдоподобия, и универсальный метод моментов. В обоих случаях для установления типа аппроксимирующего распределения достаточно решить систему двух уравнений с двумя неизвестными параметрами формы (k , u). Два других параметра (α , β) вычисляются по простым формулам при известных значениях двух параметров формы, которые найдены на первом этапе.

Из вышесказанного следует, что закон распределения должен и может быть вычислен по статистическому распределению, причем без выдвижения гипо-

тез, но с использованием свойств случайной величины. Для решения этой важнейшей задачи необходимо иметь общий метод, включающий как минимум три составные части [1]:

- классификацию случайных величин;
- универсальные (обобщенные) законы распределения, а лучше – системы распределений, включающие как частные случаи множество распределений своего класса;
- хотя бы один **метод вычисления закона распределения и оценок его параметров, единый для всех систем распределений** (при условии полных статистических данных, однородности выборки, достаточном ее объеме).

Необходимые расчеты желательно выполнять в автоматизированном режиме с помощью компьютерной программы.

Наличие нескольких систем непрерывных распределений и общего метода оценивания гарантирует правильное вычисление закона распределения с весьма высокой вероятностью, близкой к единице. При этом отпадает необходимость использования критериев согласия, поскольку вычисленное распределение будет наилучшим для аппроксимации имеющихся статистических данных. Чем больше объем выборки и выше степень однородности случайной величины, тем точнее будет аппроксимация.

Рассмотрим каждую из трех отмеченных выше составных частей.

КЛАССИФИКАЦИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

Анализ свойств непрерывных случайных величин показывает, что можно четко выделить по крайней мере три их класса.

Отнесем к первому классу такие случайные величины, которые могут принимать как положительные, так и отрицательные значения, а в некоторых случаях и те и другие в одном распределении, например, в системах качества – разброс отклонений от среднего размера детали в большую или меньшую сторону. Распределения таких случайных величин будем описывать **первой системой непрерывных распределений**. Эти распределения обладают характерным для них свойством: с ростом среднего значения вся кривая распределения, т.е. график плотности перемещается по горизонтальной оси без изменения своей формы.

В качестве примера можно привести статистическое распределение сотрудников некоторого учреждения по возрасту. При постоянном составе сотрудников средний возраст и возраст отдельных сотрудников с каждым годом увеличивается на единицу. Это значит, что последующие значения случайной величины (возраста) образуются из предыдущих путем ежегодного прибавления постоянной величины $C=1$. Для других распределений величина C может принимать другие значения, в том числе разные на одинаковых отрезках времени. Но главным здесь остается правило прибавления ее к предыдущим значениям.

Если величина C постоянна на равных отрезках времени, то среднее значение случайной величины растет во времени по линейному закону. Это еще очень важное свойство распределений первого клас-

са. Количество свойств этих распределений может пополняться. Всем перечисленным свойствам должна удовлетворять первая система непрерывных распределений.

Отнесем ко **второму классу** такие случайные величины, которые заданы на положительной полуоси. Последующие значения таких случайных величин образуются из предыдущих путем их умножения на некоторую положительную величину C . На равных отрезках времени она может принимать различные значения. Если эта величина на равных отрезках времени постоянна, то среднее (и отдельные значения) такой случайной величины растут во времени по показательному закону, а логарифм среднего – по линейному закону.

Распределения таких случайных величин будем описывать **второй системой непрерывных распределений**.

Примером такого распределения является статистическое распределение сотрудников того же учреждения по уровню заработной платы. С учетом инфляции государственные органы корректируют заработную плату, причем, как правило, путем умножения предыдущих значений на некоторую положительную величину $C > 1$. Правда, такой метод повышения заработной платы приводит к негативным последствиям: он увеличивает расслоение общества по доходам – богатые богатеют, бедные беднеют. А отсюда и множество других проблем. Поэтому необходимо периодически использовать метод прибавления некоторой величины C , обеспечивающей приемлемый уровень минимального потребительского бюджета и в то же время – экономию бюджетных средств. Кроме того, этот метод будет сдерживать темп расслоения общества по доходам.

Другими примерами статистических распределений, которые могут быть с высокой точностью описаны второй системой непрерывных распределений, являются распределения тех же сотрудников по росту и весу [1].

Отнесем далее к **третьему классу** такие существенно положительные случайные величины, последующие значения которых образуются из предыдущих путем их возведения в некоторую степень C . Если величина C на равных отрезках времени постоянна, то среднее значение такой случайной величины растет во времени по двойному показательному закону, его логарифм – по показательному закону, а двойной логарифм – по линейному закону.

Распределения таких случайных величин будем описывать **третьей системой непрерывных распределений**. Эта система наряду со второй может быть использована в математической лингвистике, информатике, библиотечном деле для описания статистических ранговых распределений слов частотно-го словаря, ключевых слов.

Из анализа свойств случайных величин следует важный вывод: **распределение и динамика случайных величин взаимосвязаны**. По известному закону роста среднего значения безошибочно определяется система распределений. Отметим, что это не выдвижение гипотезы, а выбор системы распределений в зависимости от свойств случайной величины.

Зная закон распределения, можно легко решать различные задачи. Например, как изменится распределение работающих по уровню заработной платы при ее новом среднем уровне, как оптимизировать налоговую систему и т.д.

Все вышесказанное относилось к трем классам случайных величин, средние значения которых растут во времени по законам соответственно: линейному, показательному и двойному показательному. Для описания распределений этих величин используются три системы непрерывных распределений.

Но далее неизбежно возникает вопрос: какими распределениями описывать случайные величины, рост которых задается другими законами? На этот вопрос можно дать простой ответ: необходимо найти промежуточные (дополнительные) системы непрерывных распределений. Они должны находиться между первой и второй системами, второй и третьей. А это значит, что с учетом дополнительных систем будет достаточное количество теоретических распределений для описания практически всего многообразия статистических распределений. При этом аппроксимирующие распределения можно будет вычислять в большинстве случаев по двум системам – первой и второй.

Таким образом, выбор системы упрощается до предела и при этом обеспечивается вычисление наилучшего аппроксимирующего распределения, т.е. по сути – закона распределения.

ПОСТРОЕНИЕ СИСТЕМ НЕПРЕРЫВНЫХ РАСПРЕДЕЛЕНИЙ

Поставим общую задачу: построить три системы непрерывных распределений, удовлетворяющие требованиям трех классов случайных величин. Только решение этой общей задачи позволит с высокой точностью аппроксимировать практически любые статистические распределения однородных случайных величин, в том числе ранговые распределения. Как правило, общая задача решается значительно проще частных задач.

Из свойств случайных величин следует, что наиболее просто можно построить вторую систему непрерывных распределений, которые заданы на положительной полуоси.

Для решения этой задачи используем **метод моделирования**. Рассмотрим самые простые распределения, плотности которых заданы **уравнением прямой**. Оно позволяет записать три плотности распределения некоторой случайной величины T с одним параметром α [2]:

$$p(t) = \alpha(1 - \alpha t / 2), \quad (0 < t < 2/\alpha); \quad (1)$$

$$p(t) = \alpha, \quad (0 < t < 1/\alpha); \quad (2)$$

$$p(t) = 2\alpha t, \quad (0 < t < \sqrt{1/\alpha}). \quad (3)$$

Графики этих плотностей, т.е. кривые распределения, имеют вид прямой и заданы на положительной полуоси. Эти распределения записаны из условия, что площадь под кривой распределения равна единице.

Первая из трех приведенных формул представляет собой треугольное убывающее распределение. Вторая – равномерное распределение. Третья – треугольное возрастающее распределение.

Поскольку графики этих плотностей заданы уравнениями прямой, то они принадлежат одной, а именно второй системе непрерывных распределений, которая должна быть задана общей формулой – плотностью распределения с несколькими параметрами. Назовем это распределение обобщенным, или универсальным.

Итак, должно существовать некоторое универсальное распределение, частными случаями которого являются приведенные выше три плотности распределения. Теперь необходимо найти метод, который поможет выявить предполагаемую обобщенную плотность распределения.

Здесь уместно отметить, что закон распределения может быть задан не только плотностью, но и функцией распределения, которая представляет собой интеграл от плотности $F(t) = \int_0^t p(t)dt$.

$$F(t) = \int_0^t p(t)dt.$$

При интегрировании плотностей распределения (1)–(3) появятся **новые значения параметров**, т.е. будет выявлена новая информация. Вместо этих значений введем новые параметры.

Итак, интегрируя плотности (1)–(3), найдем три функции распределения

$$F(t) = 1 - (1 - \alpha t / 2)^2; \quad (4)$$

$$F(t) = \alpha t = 1 - (1 - \alpha t); \quad (5)$$

$$F(t) = \alpha t^2 = 1 - (1 - \alpha t^2). \quad (6)$$

Далее используем **метод обобщения**. Обобщим попарно функции распределения (4) и (5), (5) и (6) путем введения новых параметров (например, u , β) вместо показателей степени 1 и 2. В первом случае в результате обобщения функций распределения (4) и (5) будем иметь

$$F(t) = 1 - (1 - \alpha ut)^{\frac{1}{u}}. \quad (7)$$

Во втором случае при обобщении функций распределения (5) и (6) получим

$$F(t) = 1 - (1 - \alpha t^\beta). \quad (8)$$

Теперь замечаем, что в формуле (8) имеется параметр β , но его нет в формуле (7). Введем его в формулу (7). В результате получим трехпараметрическую функцию распределения

$$F(t) = 1 - (1 - \alpha ut^\beta)^{\frac{1}{u}}, \quad (9)$$

откуда дифференцированием по t найдем плотность распределения с тремя параметрами α , β , u :

$$p(t) = \alpha \beta t^{\beta-1} (1 - \alpha ut^\beta)^{\frac{1}{u}-1}. \quad (10)$$

Последняя плотность может быть еще более расширена за счет **введения нового параметра**. Параметр β в формуле (10) используется дважды в качестве

показателя степени. Пусть это будут два разных параметра: в одном случае – параметр β , в другом – произведение $k\beta$, где k – новый параметр. Тогда вместо (10) можем записать искомую четырехпараметрическую плотность в виде [2]

$$p(t) = N t^{k\beta-1} (1 - \alpha ut^\beta)^{\frac{1}{u}-1}, \quad (11)$$

где N – нормирующий множитель, зависящий от четырех параметров α , β , k , u . Он вычисляется из условия, что площадь под кривой распределения равна единице.

Итак, простейшими средствами – моделирования плотностей на базе уравнения прямой, интегрирования, обобщения функций распределения, дифференцирования трехпараметрической функции распределения и, наконец, введения четвертого параметра k , – нами получена универсальная четырехпараметрическая плотность распределения, предназначенная для аппроксимации существенно положительных случайных величин ($T > 0$), в том числе статистических ранговых распределений. Эта плотность задает **вторую систему непрерывных распределений**. Отметим, что в библиотечно-информационной деятельности она может использоваться как **универсальный закон рассеяния публикаций** [2, 3]. Приведем формулировку закона Бредфорда: «Если научные журналы расположить в порядке убывания числа помещенных в них статей по какому-либо заданному предмету, то в полученном списке можно выделить ядро журналов, посвященных непосредственно этому предмету, и несколько групп или зон, каждая из которых содержит столько же статей, что и ядро. Тогда числа журналов в ядре и последующих зонах будут относиться как 1:n:n²...». Из этой формулировки следует, что закон рассеяния Бредфорда основан на статистических ранговых распределениях, но об их свойствах Бредфорд ничего не сообщает.

Плотность (11) можно получить и другим, но более сложным методом – на базе кривых роста новых событий и взаимосвязанных с ними законов распределения вероятностей новых событий. Кривые роста позволяют также построить систему дискретных распределений [4]. Однако приведенный выше метод построения обобщенного распределения значительно проще.

Нетрудно убедиться в том, что плотность (11) включает как частные случаи нормальный закон, законы Стьюдента, Коши, Вейбулла, Максвелла, «хи-квадрат», гамма-распределение, бета-распределение и множество других, в том числе закон Ципфа. Она с успехом может описывать гистограммы с различной формой, а также статистические ранговые распределения. Исследования автора на большом числе статистических распределений, в том числе ранговых, показали, что ни в одном случае не был вычислен закон Ципфа, хотя в различных исследованиях он используется довольно часто. Нормальным законом описывается незначительная доля статистических распределений. Следовательно, при наличии обобщенной плотности метод выдвижения гипотез лишается смысла, как и частично – критерии согласия, поскольку для однородной выборки достаточно боль-

шого объема они покажут высокую точность аппроксимации статистического распределения вычисленной четырехпараметрической плотностью.

На базе плотности (11) легко получить другие плотности как распределения функций случайного аргумента. Например, при $T = e^x$ по известной формуле $p(x) = p(t)(dt/dx)$ найдем

$$p(x) = Ne^{k\beta x} (1 - \alpha ue^{\beta x})^{\frac{1}{u}-1}. \quad (12)$$

Здесь случайная величина X может быть задана на всей числовой оси, т.е. значения x могут быть как положительными, так и отрицательными. Эта плотность задает **первую систему непрерывных распределений**. Отметим, что она может использоваться в библиотечно-информационной деятельности как **универсальный закон старения публикаций** [2].

Полученные плотности содержат два параметра формы k , u . Главным из них является параметр u . От его значений зависит тип распределения [2, с. 144]. Параметр β может быть параметром формы или масштаба, а параметр α – масштаба или сдвига.

Найдем, наконец, третью систему непрерывных распределений.

Если в формуле (11) принять $T = \ln Y$, то получим

$$p(y) = \frac{N}{y} (\ln y)^{k\beta-1} [1 - \alpha u (\ln y)^\beta]^{\frac{1}{u}-1}. \quad (13)$$

Последняя плотность задает **третью систему непрерывных распределений**. Она может использоваться в математической лингвистике для описания ранговых распределений слов частотного словаря.

На базе полученных плотностей можно достаточно просто решать различные задачи.

Так, график плотности (12) при значениях параметра формы $u \leq 1/2$ имеет три характерные точки: моду x_C и две точки перегиба x_A , x_B , расположенные на равных расстояниях по обе стороны от моды. При $1/2 < u < 1$ имеются две характерные точки – x_A и x_C (для распределений с левосторонней асимметрией). При $u \geq 1$ кривая не имеет характерных точек.

График плотности (11) при определенных значениях параметров имеет вид убывающей кривой. Следовательно, эта плотность может описывать ранговые распределения, представленные в системе координат $p_r = f(r)$, где r – ранг события, p_r – его относительная частота. Но такая форма представления ранговых распределений несет слишком мало информации о них, поскольку убывающая кривая не имеет никаких характерных точек. А их можно было бы использовать для **вычисления границ ядра и зон рассеяния** публикаций. Со времени окончательной формулировки С. Бредфордом своего закона рассеяния, т.е. с 1948 г. никто не предложил метода вычисления границ ядра и зон рассеяния публикаций по статистическому ранговому распределению, несмотря на то, что было предпринято множество попыток уточнения закона С. Бредфорда. Но это частная задача, решить которую невозможно без предварительного решения общей задачи – разработки систем непрерывных распределений и исследования их свойств.

При наличии систем распределений эта задача решается весьма просто. Изложим метод ее решения.

Координаты трех характерных точек рангового распределения можно найти путем преобразования плотности (11) к форме плотности (12). Это достигается путем умножения левой и правой части выражения (11) на t и использования равенства $t^\beta = e^{\beta \ln t}$. В итоге имеем [3, 5]

$$tp(t) = Ne^{k\beta \ln t} (1 - \alpha ue^{\beta \ln t})^{\frac{1}{u}-1}. \quad (14)$$

Последнее равенство также представляет собой плотность распределения. Действительно, если ввести обозначение $x = \ln t$, то плотность $p(\ln t)$ можно получить на базе плотности $p(x)$ как распределение функции случайного аргумента:

$$p(\ln t) = p(x) \frac{dx}{d \ln t} = p(x).$$

С учетом плотности $p(x)$ и равенства $x = \ln t$ можем записать

$$p(\ln t) = Ne^{k\beta \ln t} (1 - \alpha ue^{\beta \ln t})^{\frac{1}{u}-1}. \quad (15)$$

Из формул (14) и (15) следует соотношение между плотностями $p(t)$ и $p(\ln t)$

$$tp(t) = p(\ln t). \quad (16)$$

Из тех же формул следует также равенство функций распределения

$$F(\ln t) = \int p(\ln t) d \ln t = \int tp(t) d \ln t = \int tp(t) \frac{dt}{t} = F(t),$$

т.е.

$$F(\ln t) = F(t). \quad (17)$$

При дифференцировании последнего равенства по t из него следует равенство (16). Следовательно, плотность $p(t)$, приведенная к форме $tp(t) = f(\ln t)$, представляет собой плотность $p(x)$ и обладает всеми свойствами последней, т.е. при значениях параметра $u \leq 1/2$ она имеет три характерные точки A , C , B . Отсюда вытекает правило: чтобы для убывающего рангового распределения найти характерные точки, его необходимо привести к форме плотности $p(x)$, т.е. изобразить графически в системе координат $tp(t) = f(\ln t)$ [3, 5, 6] (см. рис.).

Тогда ранговое распределение будет иметь моду $\ln t_C$ и две точки перегиба $\ln t_A$ и $\ln t_B$, которые находятся на равных расстояниях от моды:

$$\ln t_C - \ln t_A = \ln t_B - \ln t_C.$$

Абсциссы этих характерных точек **приняты автором настоящей статьи в качестве границ ядра журналов и зон рассеяния** (см. рис.). Из последнего равенства следует соотношение

$$\frac{t_C}{t_A} = \frac{t_B}{t_C} = n, \quad (18)$$

которое можно принять в качестве уточненной формулировки закона рассеяния публикаций в толковании С. Бредфорда, хотя оно представляет собой новую формулировку закона рассеяния публикаций.

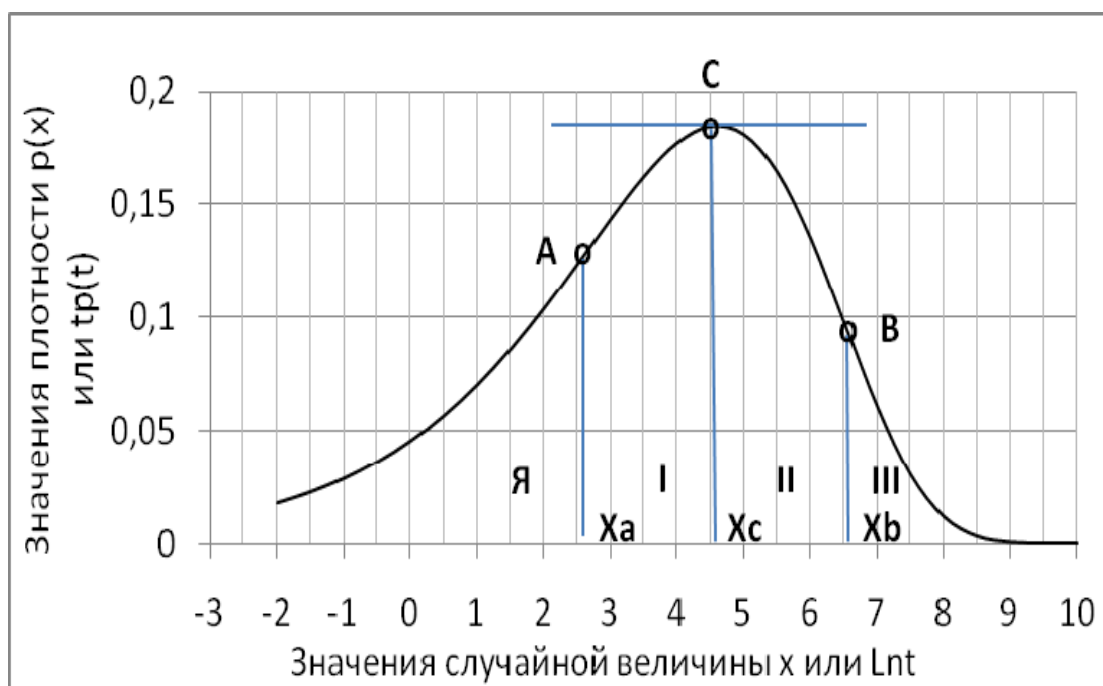


График плотностей $p(x)$ и $tp(t)=f(lnt)$ при равных значениях параметров

И вывод такой простой формулы оказался возможным лишь после построения обобщенных распределений! Полученное нами соотношение непротиворечиво и универсально, так как оно справедливо для всех распределений, образующих вторую систему непрерывных распределений при значениях параметра формы $u \leq 1/2$. Несмотря на это, **равенство (18) не является законом рассеяния публикаций** (как и «закон» Бредфорда), поскольку не содержит информации о количестве журналов и долях статей в каждой зоне, о том, сколько может быть зон рассеяния, как вычислить границы ядра и зон рассеяния, чему равна величина n . Вся эта информация содержится в обобщенной плотности $p(t)$. Поэтому **вторая система непрерывных распределений по праву является универсальным законом рассеяния публикаций**, а формула (18) – лишь следствие свойств универсального закона. Она отражает соотношение между абсциссами характерных точек на кривой распределения. Значения же функции распределения в характерных точках в этой формуле не задействованы. Поэтому, не зная теоретического рангового распределения с его значениями параметров, нельзя вычислить число журналов, входящих в ядро и зоны рассеяния, величину n , а также доли статей в ядре и зонах рассеяния, которые выражаются через функцию распределения.

Таким образом, приведение плотности (11) к форме плотности (12) позволило выявить новую информацию о существовании трех характерных точек рангового распределения при значениях параметра формы $u \leq 1/2$. При $u \geq 1$ характерных точек на кривой рангового распределения не существует. К тако-

му типу кривых относится закон Ципфа, который является частным случаем обобщенной плотности (11) с параметрами формы $u=1, \beta < 0$. Его ни в коем случае нельзя использовать для аппроксимации статистических ранговых распределений, что убедительно показывает приведенный выше график. Представленные в виде зависимости $tp(t)=f(lnt)$ ранговые распределения имеют моду и две точки перегиба. Для закона же Ципфа произведение ранга на относительную частоту равно постоянной величине. Следовательно, горизонтальная касательная к кривой распределения в точке C и есть закон Ципфа, а все ранговые распределения находятся под этой прямой.

Распределение (14) задает новую форму представления ранговых распределений (слов частотного словаря; журналов, упорядоченных по убыванию числа опубликованных в них статей по заданной тематике или числу обращений к ним, книг по числу их выдач; ученых в зависимости от количества публикаций или числа ссылок на них и множества других объектов исследования, упорядоченных по убыванию некоторого признака), а именно: по горизонтальной оси откладываются логарифмы рангов, а по вертикальной – произведения рангов на относительные частоты. В результате получается одновершинная кривая распределения, позволяющая быстро оценить, является ли выборка однородной и достаточен ли ее объем [2, 5], а также вычислить координаты характерных точек в случае однородной выборки.

Необходимо отметить, что для ранговых распределений характерна высокая энтропия. Предложенный метод приведения второй системы распределений к форме первой позволил уменьшить энтропию

распределений, в том числе ранговых, и извлечь информацию из ранговых распределений в виде трех характерных точек [7].

Можно утверждать, что универсальные законы старения и рассеяния публикаций, а также ранговые распределения лексических единиц, разных наименований книг, заданные обобщенными плотностями, являются фундаментальными закономерностями информатики, математической лингвистики и библиотекведения. Поскольку «...рассеяние научной информации является краеугольным камнем всей научно-информационной деятельности, а изучение этого свойства научной информации – важнейшей проблемой информатики» [8, с. 93], то эту проблему необходимо разрешать весьма серьезными средствами. К таким средствам относятся рассмотренные выше обобщенные распределения. Они могут быть также успешно использованы в теории вероятностей и математической статистике, библиометрии, наукометрии, экономике (эконометрии), социологии, в системах менеджмента качества и во всех других областях знания, где требуется высокая точность аппроксимации статистических распределений, в том числе ранговых.

Плотности (11), (12), (13) задают три основные системы непрерывных распределений [2]. Докажем, что каждая система соответствует требованиям своего класса случайных величин.

Рассмотрим **первую систему** непрерывных распределений, заданную обобщенной плотностью (12)

$$p(x) = Ne^{k\beta x} (1 - \alpha ue^{\beta x})^{\frac{1}{u}-1}.$$

Эта система обладает тем свойством, что при увеличении всех значений случайной величины X на постоянную величину C форма кривой распределения, т.е. графика плотности $p(x)$ не изменяется. Обозначим новое значение случайной величины X через X^* , при этом

$$X^* = X + C. \quad (19)$$

Тогда распределение новой случайной величины X^* определится по формуле

$$p(x^*) = p(x) \frac{dx}{dx^*}. \quad (20)$$

Поскольку на основании (19) $dx/dx^* = 1$, то из формулы (20) следует равенство $p(x^*) = p(x)$. Подставляя сюда вместо $p(x)$ плотность (12), с учетом равенства (19) получим

$$p(x^*) = Ne^{k\beta(x^*-C)} (1 - \alpha ue^{\beta(x^*-C)})^{\frac{1}{u}-1}. \quad (21)$$

Последнюю плотность можно привести к виду

$$p(x^*) = N^* e^{k\beta x^*} (1 - \alpha^* ue^{\beta x^*})^{\frac{1}{u}-1}, \quad (22)$$

где

$$N^* = N/e^{k\beta C}; \quad \alpha^* = \alpha / e^{\beta C}. \quad (23)$$

Таким образом, смещение случайной величины X на постоянную C приводит к изменению параметра сдвига α и вместе с ним нормирующего множителя

N . Параметры формы k , u , β не изменяются, т.е. не изменяется форма кривой распределения, что и требовалось доказать. Поскольку случайные величины X и X^* связаны функциональной зависимостью, причем с ростом X растет и X^* , то их функции распределения равны

$$F(x^*) = F(x). \quad (24)$$

Формулы (21) и (22) позволяют прогнозировать распределение случайной величины X . Чтобы рассчитать новые значения плотности распределения с учетом смещения C , в случае первой системы непрерывных распределений достаточно сдвинуть на C значения случайной величины без изменения значений плотности распределения.

Рассмотрим далее случай, когда последующие значения случайной величины X образуются из предыдущих путем их умножения на постоянную величину C : $X^* = X \cdot C$.

Тогда $X = X^*/C$, $dx/dx^* = 1/C$. Плотность $p(x^*)$ получается из плотности $p(x)$ при прежних значениях параметров α , k , u , но при этом параметр формы β и нормирующий множитель N уменьшаются в C раз: $\beta^* = \beta/C$, $N^* = N/C$. С уменьшением параметра β кривая распределения становится более пологой и длинной. Плотность распределения $p(x^*)$ задается формулой

$$p(x^*) = N^* e^{k\beta^* x^*} (1 - \alpha ue^{\beta^* x^*})^{\frac{1}{u}-1}. \quad (25)$$

Итак, умножение случайной величины X на постоянную величину C приводит к уменьшению одного из параметров формы – β . Два других параметра k , u остаются прежними. Формулу (25) также можно использовать для прогнозирования распределений первой системы при условии, когда случайная величина X увеличивается в C раз.

Перейдем ко **второй системе** непрерывных распределений.

Распределение случайной величины T задается обобщенной плотностью (11)

$$p(t) = Nt^{k\beta-1} (1 - \alpha ut^\beta)^{\frac{1}{u}-1}.$$

Пусть все значения случайной величины T увеличатся в C раз. Требуется найти распределение случайной величины $T^* = T \cdot C$. Поскольку $t = t^*/C$, $dt/dt^* = 1/C$, то

$$p(t^*) = p(t) \frac{dt}{dt^*} = \frac{p(t)}{C} \quad (26)$$

или

$$p(t^*) = \frac{N}{C^{k\beta}} t^{*k\beta-1} \left(1 - \frac{\alpha}{C^\beta} ut^{*\beta}\right)^{\frac{1}{u}-1}. \quad (27)$$

Введя обозначения

$$N^* = N / C^{k\beta}, \quad \alpha^* = \alpha / C^\beta, \quad (28)$$

последнюю плотность перепишем в виде

$$p(t^*) = N^* t^{*k\beta-1} (1 - \alpha^* ut^{*\beta})^{\frac{1}{u}-1}. \quad (29)$$

Увеличение случайной величины T в C раз приводит к уменьшению параметра α и нормирующего множителя N . При этом плотность распределения $p(t)$ уменьшается в C раз (см. формулу (26)), а произведение $tp(t)$, а также среднее значение $tp(t)$ остаются без изменения. Это значит, что форма кривой распределения $tp(t) = f(\ln t)$ не изменяется, поскольку не изменяются параметры формы k, u, β . При этом справедливы равенства: $F(t^*) = F(t)$, $t^*p(t^*) = tp(t)$.

Пусть далее случайная величина T возводится в степень C : $T^* = T^C$. Отсюда находим

$$T = T^{*1/C}, \quad dt / dt^* = (1/C)t^{*1/C-1}.$$

Плотность $p(t^*)$ равна $p(t^*) = p(t)dt/dt^*$, или

$$p(t^*) = N^* t^{*k\beta^*-1} \left(1 - \alpha u t^{*\beta^*}\right)^{\frac{1}{u}-1}, \quad (30)$$

где $N^* = N/C$, $\beta^* = \beta/C$.

Как видим, в этом случае уменьшились в C раз нормирующий множитель и параметр формы β . Неизменными остались параметр α и параметры формы k, u .

Плотности (29) и (30) позволяют прогнозировать распределения второй системы.

Наконец, рассмотрим **третью систему** непрерывных распределений, заданную плотностью (13)

$$p(y) = \frac{N}{y} (\ln y)^{k\beta-1} \left[1 - \alpha u (\ln y)^\beta\right]^{\frac{1}{u}-1}.$$

Возведем случайную величину Y в степень C и запишем равенство: $Y^* = Y^C$. Отсюда находим: $Y = Y^{*1/C}$, $dy / dy^* = (1/C)y^{*1/C-1}$. Плотность $p(y^*)$ задается формулой $p(y^*) = p(y)dy/dy^*$, или

$$p(y^*) = \frac{N^*}{y^*} (\ln y^*)^{k\beta-1} \left[1 - \alpha^* u (\ln y^*)^\beta\right]^{\frac{1}{u}-1}, \quad (31)$$

где $N^* = N/C^{k\beta}$, $\alpha^* = \alpha/C^\beta$.

Из полученных формул видно, что возведение случайной величины Y в степень C не изменяет параметров формы k, u, β . Изменяется лишь параметр α , а вместе с ним и нормирующий множитель N . Обобщенная плотность (31) позволяет прогнозировать распределения третьей системы.

Итак, все три обобщенные плотности удовлетворяют требованиям случайных величин своего класса.

Осталось найти **дополнительные системы**. Поскольку они должны служить связующими звеньями между первой и второй основными системами, второй и третьей, третьей и четвертой [2], то их необходимо получить на базе найденных выше четырехпараметрических распределений. Рассмотрим первую систему непрерывных распределений, заданную плотностью (12)

$$p(x) = N e^{k\beta x} \left(1 - \alpha u e^{\beta x}\right)^{\frac{1}{u}-1}.$$

Пусть случайная величина X связана со случайной величиной T зависимостью $X = \ln(T-L)$, где L – параметр сдвига. Тогда первая производная

$dx/dt = 1/(t-L)$ и плотность $p(t)$ задается пятипараметрической формулой [2]

$$p(t) = N(t-L)^{k\beta-1} \left(1 - \alpha u(t-L)^\beta\right)^{\frac{1}{u}-1}. \quad (32)$$

ДОПОЛНИТЕЛЬНЫЕ СИСТЕМЫ НЕПРЕРЫВНЫХ РАСПРЕДЕЛЕНИЙ

В результате исследований автором установлено, что **дополнительными для первой системы** непрерывных распределений могут быть две плотности

$$p(t) = N(t-L)^{k-1} \left[1 - \alpha u(t-L)\right]^{\frac{1}{u}-1}, \quad (33)$$

$$p(t) = N \left[1 - \alpha u(t-\bar{t})^2\right]^{\frac{1}{u}-1}, \quad (34)$$

которые являются частными случаями плотности (32). В первой из них параметр $\beta=1$, во второй $\beta=2$.

Величина \bar{t} – среднее значение. Вторая плотность задает семейство симметричных распределений. Распределения, заданные плотностью (33), также симметричны при условии $ku=1$. Две приведенные плотности задают основную часть семейства распределений Пирсона. Они являются связующим звеном первой и второй систем непрерывных распределений. Отметим, что частным случаем плотности (34) является нормальный закон при $u \rightarrow 0$.

Аналогично во **вторую систему** непрерывных распределений войдут дополнительные плотности

$$p(y) = \frac{N(\ln y - l)^{k-1}}{y} \left[1 - \alpha u(\ln y - l)\right]^{\frac{1}{u}-1}, \quad (35)$$

$$p(y) = \frac{N}{y} \left[1 - \alpha u(\ln y - \overline{\ln y})^2\right]^{\frac{1}{u}-1}, \quad (36)$$

которые следуют из плотностей (33), (34) при $T = \ln Y$. Они являются переходными между второй и третьей системами. Из плотности (36) при $u \rightarrow 0$ следует логарифмически нормальный закон.

Так же можно ввести дополнительные системы для третьей плотности.

Легко проверить, что дополнительные системы распределений соответствуют свойствам случайных величин своих классов.

В результате каждая из трех систем непрерывных распределений содержит по три плотности. Следовательно, всего имеется девять обобщенных плотностей, включающих более 50 типов распределений, которых достаточно для аппроксимации подавляющего большинства статистических распределений.

Итак, этап построения систем непрерывных распределений нами пройден. Далее необходимо всесторонне их исследовать: дать классификацию обобщенных распределений; выразить нормирующие множители всех типов распределений через их параметры; рассмотреть возможные формы кривых распределения и их зависимость от значений параметров формы; разработать методы вычисления законов распределения и оценок параметров; разработать алгоритмы и компьютерные программы; проверить по-

строенную теорию на большом статистическом материале; дать сравнительный анализ построенных моделей, методов, алгоритмов, программ и тех, что широко используются в настоящее время. Главной и наиболее сложной из этого перечня задач является разработка метода вычисления закона распределения.

ОБЩИЙ УСТОЙЧИВЫЙ МЕТОД ВЫЧИСЛЕНИЯ ЗАКОНА РАСПРЕДЕЛЕНИЯ И ОЦЕНОК ПАРАМЕТРОВ

Наконец, осталось разработать общий, единый для трех систем непрерывных распределений метод **вычисления закона распределения и оценок параметров**, причем, метод должен быть устойчивым к выбросам на концах статистического распределения.

Еще раз отметим, что существующие методы позволяют вычислять оценки параметров тех распределений, которые **заранее выбраны** в качестве нулевой гипотезы. Далее эмпирические моменты (или другие величины) приравниваются к теоретическим и решается система уравнений, число которых равно числу параметров выбранного распределения. Если после проверки по критериям согласия нулевая гипотеза не подтверждается, выбирается другое распределение и все расчеты повторяются. Такой подход для обобщенных распределений, разработанных автором данной статьи, неприемлем по двум причинам.

Во-первых, свойства случайной величины однозначно определяют нужную систему распределений – обычно первую, либо вторую. Здесь ошибиться трудно. Далее нашей задачей является **вычисление** в этой системе **закона распределения и оценок его параметров**.

Во-вторых, автором предложен метод, с помощью которого поставленная задача решается в **два этапа**. На первом этапе по двум показателям, зависящим от параметров формы k, u , устанавливается тип аппроксимирующей кривой и вычисляются оценки параметров формы **путем решения системы двух уравнений с двумя неизвестными**, что значительно упрощает расчеты. На втором этапе вычисляются оценки двух других параметров по относительно простым формулам (с учетом известных оценок параметров k, u). Заметим, что этот метод требует предварительного группирования статистических данных.

Отныне для установления теоретического закона распределения непрерывной случайной величины по ее статистическому распределению не требуется выдвижения многочисленных гипотез об аппроксимирующей кривой и проверки каждой из них по критериям согласия. Система непрерывных распределений выбирается в зависимости от свойств случайной величины, а тип распределения и оценки его параметров определяются расчетом.

При этом основная сложность заключается в разработке двух показателей для установления типа распределения и вычисления оценок параметров формы. Для разработки таких показателей система четырехпараметрических непрерывных распределений случайной величины X сводится к системе двух-

параметрических распределений случайной величины Z . Затем используются взаимосвязи между случайными величинами X и Z и их плотностями распределения $p(x), p(z)$.

Рассмотрим обобщенную плотность $p(x)$, которая задает первую основную систему непрерывных распределений (см. формулу (25)). Введем два показателя – асимметрии B и островершинности H , которые зависят от двух параметров формы k, u . По этим показателям однозначно устанавливается тип аппроксимирующего распределения и находятся оценки параметров k, u с помощью номограммы (см. ниже).

Для обобщенной плотности $p(x)$ показатели B, H задаются формулами

$$\left. \begin{aligned} B &= M[p(x)(x - M(x))] = f(k, u) \\ H &= S_3 / S_1^3 = f(k, u) \end{aligned} \right\}, \quad (37)$$

где

$$S_r = M[p(x)]^r = f(\beta, k, u). \quad (38)$$

Исследования показали, что величина H задана на интервале $\sqrt{2} < H < 2$, а величина B – на интервале $-1/4 < B < 1/4$.

Если вычислить для разных типов распределений значения показателей B, H при различных значениях параметров k, u , то по этим данным можно построить номограмму (бинарную сетку) [2, 3], которая применима к трем основным системам непрерывных распределений, заданным первыми плотностями. При этом плотности $p(t)$ и $p(y)$ должны быть приведены к форме плотности $p(x)$, т.е. представлены в виде $tp(t)=f(\ln t), y \ln p(y)=f(\ln y)$.

Для дополнительных систем автором построена другая номограмма, которая является продолжением первой [9].

Чтобы найти закон распределения, достаточно вычислить по статистическому распределению оценки показателей B, H . Эти показатели однозначно определяют тип распределения, приведенного к форме плотности $p(x)$. Более того, с их помощью легко находят оценки параметров k, u непосредственно из номограммы (*Приложение 1*). Более точно они вычисляются по программе автора. Оставшиеся два параметра (α и β) вычисляются по специальным формулам [2]. Метод обеспечивает вычисление как одновершинных частотных законов распределения, так и ранговых. Здесь уместно отметить, что показатели B, H играют роль критериев согласия теоретического и статистического распределений.

При известных оценках параметров **рангового распределения** легко вычисляются координаты характерных точек по заранее выведенным формулам, что позволяет находить ядро журналов или книжного фонда, а точнее, ядро профессиональных информационных потребностей читателей t_A (на приведенном рисунке абсцисса точки $x_A = \ln t_A$, откуда $t_A = e^{x_A}$); зоны рассеяния I, II, III; оптимальный объем фонда (активная его часть) $t_B = e^{x_B}$; а также вычислять информационную полноту комплектования

оптимального фонда или, другими словами, вероятность удовлетворения информационных потребностей пользователей этим фондом. Например, в случае справедливости закона Вейбулла ядро фонда удовлетворяет информационные потребности пользователей на 31,75%, а оптимальный объем фонда – на 92,71%. Эти величины зависят от параметров аппроксимирующего рангового распределения, а в итоге – от статистического рангового распределения и представляют собой накопленную долю книговыдач, приходящуюся на ядро фонда t_A и оптимальный объем фонда t_B . В данном случае она выражена в процентах. Здесь следует уточнить, что под ядром фонда в нашем случае понимается наиболее востребованная его часть, как и в случае с ядром журналов.

Наличие систем непрерывных распределений и методов вычисления закона распределения и оценок параметров позволяет утверждать, что решение множества практических задач отныне сводится к накоплению достаточного количества статистических данных. Именно в них содержится ценная информация, которая может быть извлечена путем вычисления закона распределения и его всестороннего исследования. Если же речь идет о виде аппроксимирующего распределения, то здесь вовсе не возникает вопросов – система непрерывных распределений (первая или вторая) выбирается в соответствии со свойствами случайной величины, а тип и оценки параметров наилучшего аппроксимирующего распределения находятся путем расчета. Поэтому некоторые задачи оказываются заранее решенными, и только для вычисления оценок параметров теоретических распределений требуется обработка статистических данных. Решение множества теоретических задач также значительно упрощается.

Дальнейшие исследования показывают, что общий устойчивый метод по точности аппроксимации не уступает методу наибольшего правдоподобия Р. Фишера, к тому же он значительно проще последнего. Существенным преимуществом методов автора является наличие двух показателей, позволяющих **вычислять** тип аппроксимирующего распределения и оценки параметров формы в заданной системе непрерывных распределений. Метод Фишера не располагает такими показателями [7]. Их еще предстоит разработать.

УНИВЕРСАЛЬНЫЙ МЕТОД МОМЕНТОВ

Метод моментов был предложен К. Пирсоном в 1895 г. (назовем его классическим методом моментов). Главное его достоинство – простота. Он позволяет вычислять закон распределения и оценки параметров по статистическому распределению, правда, только при условии, когда наверняка известно, что теоретический закон входит в семейство кривых Пирсона. Следует отметить, что данное семейство непрерывных распределений слишком узкое и за его пределами находится бесконечное множество других теоретических и статистических распределений. Кроме того, метод моментов Пирсона может быть использован лишь при условии, если существуют

моменты вплоть до четвертого порядка. Этим недостатком лишен предложенный автором настоящей статьи универсальный метод моментов.

Суть его заключается в том, что он разработан для первой системы непрерывных распределений, заданной четырехпараметрической плотностью (12)

$$p(x) = Ne^{k\beta x} (1 - \alpha u e^{\beta x})^{\frac{1}{u}-1}. \quad (39)$$

Для вычисления закона распределения второй и третьей систем непрерывных распределений последние должны быть приведены к форме первой системы. Для приведения второй системы к форме первой необходимо использовать замену переменных $X = \text{Ln}T$. Тогда получим плотность (14)

$$tp(t) = Ne^{k\beta \text{Ln}t} (1 - \alpha u e^{\beta \text{Ln}t})^{\frac{1}{u}-1}, \quad (40)$$

где $\text{Ln}t = x$, $tp(t) = p(x)$. Последняя плотность обладает всеми свойствами плотности (12). Поскольку случайная величина X имеет все моменты, то таким же свойством обладает и случайная величина $\text{Ln}T$. Отсюда следует, что при выравнивании статистических распределений обобщенными плотностями, которые относятся к разным системам непрерывных распределений, центральные моменты вычисляются по разным формулам.

Итак, рассмотрим первую обобщенную плотность первой системы непрерывных распределений, все параметры которой больше нуля:

$$p(x) = \frac{\beta(\alpha u)^k \Gamma(k + \frac{1}{u})}{\Gamma(k) \Gamma(\frac{1}{u})} e^{k\beta x} (1 - \alpha u e^{\beta x})^{\frac{1}{u}-1}. \quad (41)$$

Введем случайную величину Z , которая связана со случайной величиной X зависимостью

$$Z = \alpha u e^{\beta X} \quad (42)$$

и найдем ее закон распределения.

На основании (41), (42) и формулы

$$p(z) = p(x) (dx/dz)$$

имеем:

$$p(z) = \frac{\Gamma(k + \frac{1}{u})}{\Gamma(k) \Gamma(\frac{1}{u})} z^{k-1} (1-z)^{\frac{1}{u}-1} \quad (43)$$

Это известное бета-распределение. Оно зависит от двух параметров формы k, u .

Пусть нам известны оба этих параметра. Тогда оценку параметра α можно найти из формулы (42). Для этого прологарифмируем ее

$$\ln Z = \ln \alpha u + \beta X \quad (44)$$

и заменим величины $\ln Z$ и X их математическими ожиданиями. В результате получим

$$M(\ln Z) = \ln \alpha u + \beta M(X), \quad (45)$$

откуда найдем

$$\alpha u = e^{M(\ln Z) - \beta M(X)}. \quad (46)$$

Величина $M(\ln Z)$ зависит от двух параметров k, u . Она может быть вычислена теоретически. Величина $M(X)$ зависит от четырех параметров, но она заменяется оценкой, т.е. средней величиной \bar{x} , рассчитанной по статистическому распределению. Здесь уместно отметить, что произведение ou вычисляется также по формуле (46) в случае ранее рассмотренного устойчивого метода.

Теперь осталось найти формулы для вычисления оценок трех параметров: β, k, u . Для этого на базе формул (44), (45) составим равенство для центральных моментов случайных величин $\ln Z$ и X :

$$M[\ln Z - M(\ln Z)]^r = \beta^r M[X - M(X)]^r. \quad (47)$$

Вводя другие обозначения

$$\mu_r^{(z)} = M[\ln Z - M(\ln Z)]^r \quad \mu_r^x = M[X - M(X)]^r,$$

формулу (47) перепишем в виде

$$\mu_r^{(z)} = \beta^r \mu_r^x. \quad (48)$$

При $r=2$ из (48) следует формула для вычисления параметра β .

$$\beta = \sqrt{\frac{\mu_2^{(z)}}{\mu_2^x}}. \quad (49)$$

Наконец, осталось найти оценки двух главных параметров: k, u . Они могут быть найдены из формулы (48). Введем два показателя – асимметрии и островершинности, зависящие от двух параметров k, u ,

$$\beta_1 = \frac{\mu_3^z}{\mu_2^z}; \quad \beta_2 = \frac{\mu_4^z}{\mu_2^z}. \quad (50)$$

Это – критерии Пирсона, но примененные к обобщенной плотности $p(x)$ с четырьмя параметрами.

Показатели β_1 и β_2 могут быть вычислены по центральным моментам случайной величины $\ln Z$ либо случайной величины X . В результате получим формулы, зависящие от двух параметров k, u . На основе таких формул, найденных для распределений всех типов, строится номограмма (Приложение 2), позволяющая устанавливать тип аппроксимирующей кривой распределения и вычислять оценки параметров k, u по критериям β_1, β_2 , вычисленным по статистическому распределению.

Универсальный метод моментов, разработанный для первой системы непрерывных распределений (плотности $p(x)$), может быть применен ко второй и третьей системам непрерывных распределений. Но для этого их необходимо привести к форме первой системы непрерывных распределений. Например, плотность (11) должна быть приведена к форме (40). Плотность (13) – третья система распределений – должна быть приведена к форме $u p(y) \ln y = f(\ln y)$.

Отсюда следует, что центральные моменты каждой системы непрерывных распределений вычисляются по своим формулам, но при этом тип распределения и оценки параметров формы k, u находятся по одной и той же номограмме. Отметим, что часть номограммы

Приложения 2 ниже прямой $\beta_2 = 3 + 1.5\beta_1$ относится к дополнительным системам непрерывных распределений с параметром $\beta=1$.

МЕТОД НАИМЕНЬШИХ КВАДРАТОВ (МЕТОД ГАУССА)

Этот чрезвычайно простой метод целесообразно использовать в том случае, когда закон распределения может быть представлен уравнением прямой.

Рассмотрим закон Вейбулла, который является частным случаем обобщенного распределения (11) при $u \rightarrow 0$. Его плотность и функция распределения задаются формулами

$$p(t) = \alpha \beta t^{\beta-1} e^{-\alpha t^\beta}; \quad F(t) = 1 - e^{-\alpha t^\beta}.$$

Закон Вейбулла во многих случаях хорошо аппроксимирует статистические ранговые распределения.

Представим функцию распределения в виде уравнения прямой

$$\ln \ln(1/(1-F(t))) = \ln \alpha + \beta \ln t.$$

Вводя новые обозначения, можем записать

$$Y = A + \beta X. \quad (51)$$

Чтобы проверить применимость закона Вейбулла для аппроксимации статистического рангового распределения, необходимо вычислить величины

$$Y = \ln \ln(1/(1-F(t))) \quad \text{и} \quad X = \ln t, \quad (52)$$

где t – ранг события, т.е. его порядковый номер в списке по убывающим относительным частотам. Если эмпирические точки ложатся на прямую (51), необходимо вычислить оценки величин $A = \ln \alpha$ и β по методу наименьших квадратов (см., например, список литературы п. 3, 6):

$$\beta = \frac{\overline{XY} - \bar{X} \cdot \bar{Y}}{\overline{X^2} - (\bar{X})^2}; \quad A = \bar{Y} - \beta \bar{X}.$$

Оценка параметра α равна

$$\alpha = e^A.$$

Для вычисления координат трех характерных точек используются формулы, справедливые в случае закона Вейбулла

$$t_c = \left(\frac{1}{\alpha}\right)^{\frac{1}{\beta}}; \quad n = \left(\frac{3 + \sqrt{5}}{2}\right)^{\frac{1}{\beta}}; \quad t_A = \frac{t_c}{n}; \quad t_B = t_c \cdot n;$$

$$F(t_A) = 0,3175; \quad F(t_C) = 0,6321; \quad F(t_B) = 0,9271.$$

ЗАКЛЮЧЕНИЕ

Эффективность статистических методов в теоретических и прикладных исследованиях в решающей степени зависит от точности аппроксимации статистических распределений. Наибольшую точность аппроксимации можно получить при использовании теории обобщенных распределений автора настоящей статьи, в которой изложены некоторые сведения о названной теории. Кроме того, теория включает

также систему дискретных распределений [4], взаимосвязанную с системой кривых роста новых событий, номограммы для графического определения типа аппроксимирующей кривой и оценок параметров, а также серию компьютерных программ для работы с указанными системами. Применение этой теории на практике значительно облегчает задачу нахождения закона распределения по статистическим данным.

В этом случае нет необходимости выдвигать гипотезы о предполагаемом аппроксимирующем распределении. В зависимости от свойств случайной величины выбирается система непрерывных распределений (как правило, первая или вторая) и по статистическому распределению вычисляются два показателя – асимметрии и островершинности – по формулам, справедливым для данной системы. Далее они приравниваются к соответствующим теоретическим показателям, которые зависят лишь от двух параметров формы k , u , хотя обобщенная плотность содержит, как правило, четыре параметра. С помощью номограммы по двум показателям (V , H) или (β_1, β_2) устанавливается тип теоретического распределения и находятся оценки параметров k , u : в первом приближении – в ручном режиме, а более точные их значения, а также значения параметров α , β – в автоматизированном режиме по программам автора.

Следует отметить, что одной точке на номограмме с заданными значениями показателей V , H и параметров формы k , u соответствует не единственное распределение, а множество распределений с различными значениями параметров α , β . Например, во второй системе распределений одинаковые значения параметров формы $k=1$, $u \rightarrow 0$ имеют такие распределения, как показательное ($\beta=1$), Релея ($\beta=2$) и Вейбулла ($\beta > 0$). Оценки параметров α , β вычисляются по специальным формулам с учетом найденных оценок параметров формы k , u .

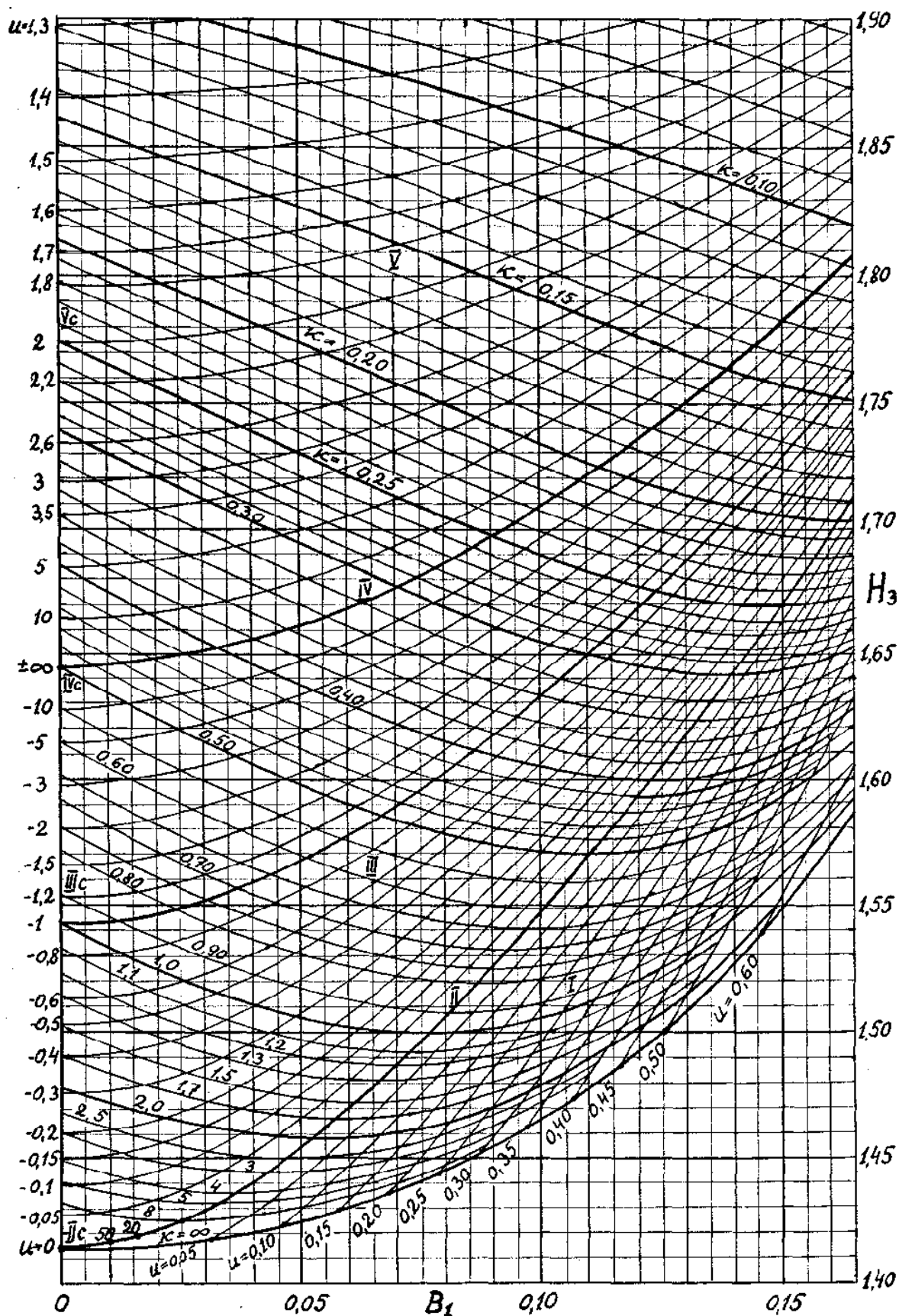
Обобщенные распределения включают как частные случаи множество известных распределений, в том числе семейство кривых Пирсона, и могут претендовать на роль универсальных законов распределения не только теории вероятностей и математической статистики, но и информатики, математической лингвистики, библиотековедения, библиометрии, наукометрии, эконометрии, экономики, социологии и многих других областей знания. Применение обобщенных распределений и общего устойчивого метода вычисления закона распределения по статистическим данным гарантирует высокую экономическую эффективность статистических методов во всех практических приложениях. Так, использование обобщенных распределений в системах менеджмента качества позволяет с высокой точностью оценивать возможности технологических процессов и поддерживать их в статистически управляемом состоянии при любом законе распределения технологических погрешностей, что обеспечивает значительное снижение уровня брака.

Но широкое использование теории обобщенных распределений можно реализовать лишь путем включения ее в учебные программы высших учебных заведений, что обещает дать большой экономический эффект во всех областях ее применения.

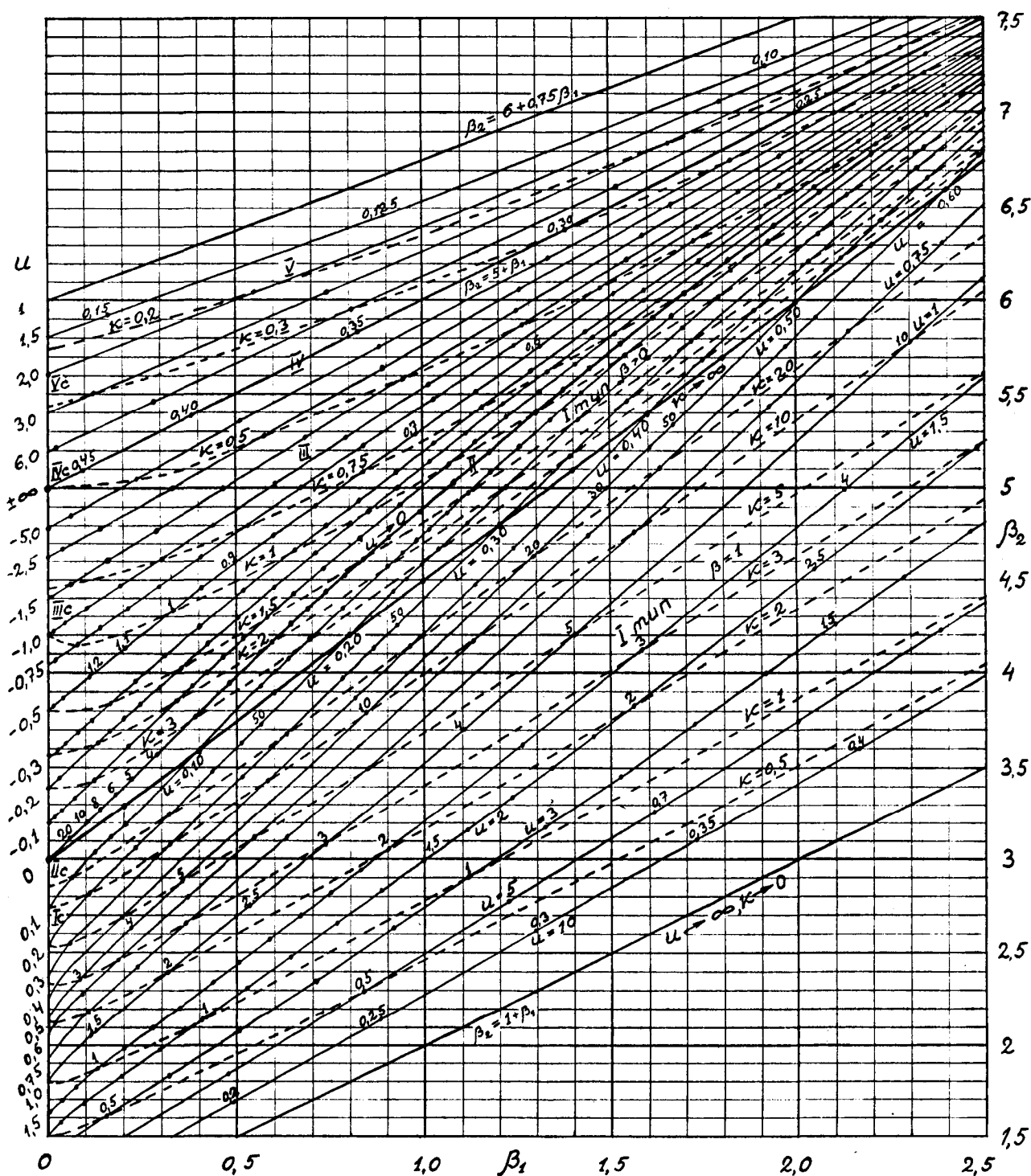
СПИСОК ЛИТЕРАТУРЫ

1. Нешиной В.В. Как вычислить закон распределения случайной величины? // Материалы XI Междунар. конф. «Медико-социальная экология личности: состояние и перспективы», 17–18 мая 2013г., г. Минск / отв. ред. В.А. Прокашева. – Минск: Изд. центр БГУ, 2013. – С. 484–486.
2. Нешиной В.В. Элементы теории обобщенных распределений: монография. – Минск: РИВШ, 2009. – 204 с.
3. Нешиной В.В. Законы Ципфа, Бредфорда и универсальные модели // Научно-техническая информация. Сер. 2. – 2010. – № 1. – С. 26–33; Neshitoy V.V. Zipf's and Bradford's laws and universal models // Automatic Documentation and Mathematical Linguistics. – 2010. – Vol. 44, № 1. – P. 30–37.
4. Нешиной В.В. Методы статанализа в библиотечно-информационной деятельности: вычисление дискретных распределений и кривых роста: учеб.-метод. пособие. – Минск: РИВШ, 2012. – 134 с.
5. Нешиной В.В. Форма представления ранговых распределений // Ученые записки Тартуского гос. ун-та. – 1987. – Вып. 774. – С. 123–134.
6. Нешиной В.В. Методы вычисления границ ядра и зон рассеяния публикаций // Научно-техническая информация. Сер. 2. – 2013. – №11. – С. 1–11; Neshitoy V.V. Methods for Calculating the Boundaries of the Core and Zones of Scattering of Publications // Automatic Documentation and Mathematical Linguistics. – 2013. – Vol. 47, № 5–6. – P. 169–179.
7. Нешиной В.В. Метод наибольшего правдоподобия, устойчивый метод и энтропия // Научно-техническая информация. Сер. 2. – 2012. – № 5. – С. 27–33.
8. Михайлов А.И., Черный А.И., Гиляревский Р.С. Основы информатики. – М.: Наука, 1968. – 756 с.
9. Нешиной В.В. Методы статанализа в библиотечно-информационной деятельности: вычисление непрерывных распределений: учеб.-метод. пособие. – Минск: Бел. гос. ун-т культуры и искусств, 2010. – 61 с.

Номограмма для вычисления закона распределения и оценок параметров по общему устойчивому методу



Номограмма для вычисления закона распределения и оценок параметров по универсальному методу моментов



Материал поступил в редакцию 25.12.15.

Сведения об авторе

НЕШИТОЙ Василий Васильевич – доктор технических наук, профессор, профессор кафедры информационных ресурсов УО «Белорусский государственный университет культуры и искусств», г. Минск
 e-mail: neshitoi_vv@tut.by

АВТОМАТИЗАЦИЯ ОБРАБОТКИ ТЕКСТА

УДК 004 : [025.4.05 : 026.072]

А. А. Крулев

Системный рубрикатор и тезаурус ведомственной технической библиотеки: особенности, возможности и перспективы

Рассматриваются вопросы формирования и редактирования словаря предметных рубрик и тезауруса научно-технической библиотеки Крыловского государственного научного центра. Отмечается, что независимо от места и способа публикации, хранения и использования документа, ключевые слова являются важным элементом упорядочивания массива информации. Формальный подход к составлению списка ключевых слов и предметных рубрик документа «опасен» тем, что документ не будет обнаружен при информационном поиске даже при наличии информативного заглавия.

Ключевые слова: рубрикатор, тезаурус, информационное обслуживание, информационный поиск, электронная библиотека

При тематическом поиске в современных автоматизированных информационно-библиотечных системах (АИБС) используются предметные рубрикаторы, а также нормативные словари дескрипторов и ключевых слов – тезаурусы, являющиеся неотъемлемой частью лингвистического обеспечения функционирования АИБС.

Словари предметных рубрик используются практически во всех библиотеках. При незначительных различиях в методике составления все словари предназначены для алгоритмизации действий предметизатора при отборе, нормализации и систематизации лексики.

Существует несколько способов формирования предметного рубрикатора: на основании предметных рубрик УДК и ББК, государственного рубрикатора НТИ и т.п. Нередко при формировании собственных рубрикаторов, используют списки и словари предметных рубрик по отраслям или темам.

При создании тезауруса может использоваться свободное индексирование: когда из текста документа выбираются ключевые слова, также возможно включение слов, отсутствующих в тексте, но при этом отражающих смысл документа.

Определенная сложность в формировании предметного рубрикатора и тезауруса научно-технической библиотеки (НТБ) Крыловского государственного научного центра (г. Санкт-Петербург) связана с тем, что последний «Тезаурус информационно-поисковый по судостроению» – собрание научно-технических терминов, составляемое на основе ана-

лиза словарного состава текущих текстов научных публикаций, был выпущен в 1978 г. Организацией выпуска занимался ЦНИИ «Румб», который прекратил свое существование в 1994 г., так и не успев издать актуализированный справочник. Работа по систематизации записей, естественно, не прекратилась, но носила скорее инерционный характер. Кадровые и корпоративные проблемы, в результате которых статус и штатный состав научно-технической библиотеки были неопределенными, также негативно сказались на этом участке работы.

В 2012 г. в Крыловском центре был создан Информационно-издательский центр, который объединил НТБ и издательский отдел. После создания этого подразделения работа по информационному обеспечению приобрела качественно иной вид. Одной из основных причин этого стало внедрение в НТБ Крыловского государственного научного центра автоматизированной информационно-библиотечной системы «Руслан», предоставленной по контракту ООО «Открытые библиотечные системы» в 2013 г. Система «Руслан» обеспечивает автоматизацию всех основных процессов обработки литературы и обслуживания читателей в библиотеках различного профиля. На базе этой АИБС была создана Электронная библиотека предприятия (ЭБ).

Электронная библиотека Крыловского центра – это информационная система, предназначенная для хранения электронных документов и обеспечивающая доступ к ним через локальную сеть предприятия.

Для удобства читателей был разработан сайт научно-технической библиотеки, работа с электронной библиотекой ведется в браузере. С помощью данного сервиса можно получить доступ к следующим электронным ресурсам:

- Электронный каталог научно-технической библиотеки: массив вторичной информации (записей) для поиска содержательных единиц фонда ЭБ и стандартизованного доступа к ним вне зависимости от места хранения (рис. 1);

- Электронная коллекция: множество полнотекстовых электронных документов, обладающих однотипными формальными признаками и содержащих фактографическую (объектографическую) информацию (рис. 2);

- Актуализируемый перечень периодических изданий, поступающих в НТБ по подписке;

- Раздел «Новости», в котором регулярно обновляются сведения о поступающих в фонд изданиях и размещаются интерактивные ссылки на электронные ресурсы.

Два последних раздела не нуждаются в комментариях, поскольку их назначение очевидно и свидетельствует о приоритетности сервисного подхода к информационному обеспечению.

Возможности, доступные для пользователей в первых двух разделах сайта, требуют более детального рассмотрения.

Помимо основных параметров (или их сочетания), по которым читатель может идентифицировать документ, а именно (см. рис. 1): заглавие, автор и дата публикации, читателю предлагается обозначить «тематику» интересующего документа. Для ускорения поиска вводимые слова можно сокращать (при этом необходимо поставить символ *).

Нередко читатель сам не вполне четко представляет, какой именно документ ему необходим, поэтому указание определенной рубрики и списка ключевых слов является для посетителя электронной библиотеки значительным подспорьем в информационном поиске, а при заказе информационной или библиографической справки эти параметры неocenимы также для работников НТБ.

Прежде чем перейти к краткому обзору практической работы по пополнению и корректировке системного рубрикатора, следует кратко описать еще один раздел сайта НТБ Крыловского центра – Электронная библиотека (рис. 2).

При простоте интерфейса раздел является одним из важнейших инструментов проведения информационного поиска, особенно когда возникают сложности при формулировке конкретного запроса. Поиск в электронной библиотеке, в отличие от каталога (рис. 1), проводится по всему тексту документа. В результате запроса в окне браузера появляется список документов, где слово или словосочетание, введенное в строку поиска, выделено жирным шрифтом (с учетом морфологии: число, падеж и т.д.). Этот результат можно было бы назвать формальным (простой релевантный поиск). Более того, этот алгоритм идентичен поиску в общедоступных поисковых системах, которые безоговорочно располагают тысячекратно превосходящим полнотекстовым контентом. Основным преимуществом электронной библиотеки является предлагаемая автоматическая классификация записей.

В правой части окна браузера читателю демонстрируются следующие критерии: авторы-составители, дата публикации, тематика, тип документа, язык документа (рис. 3).

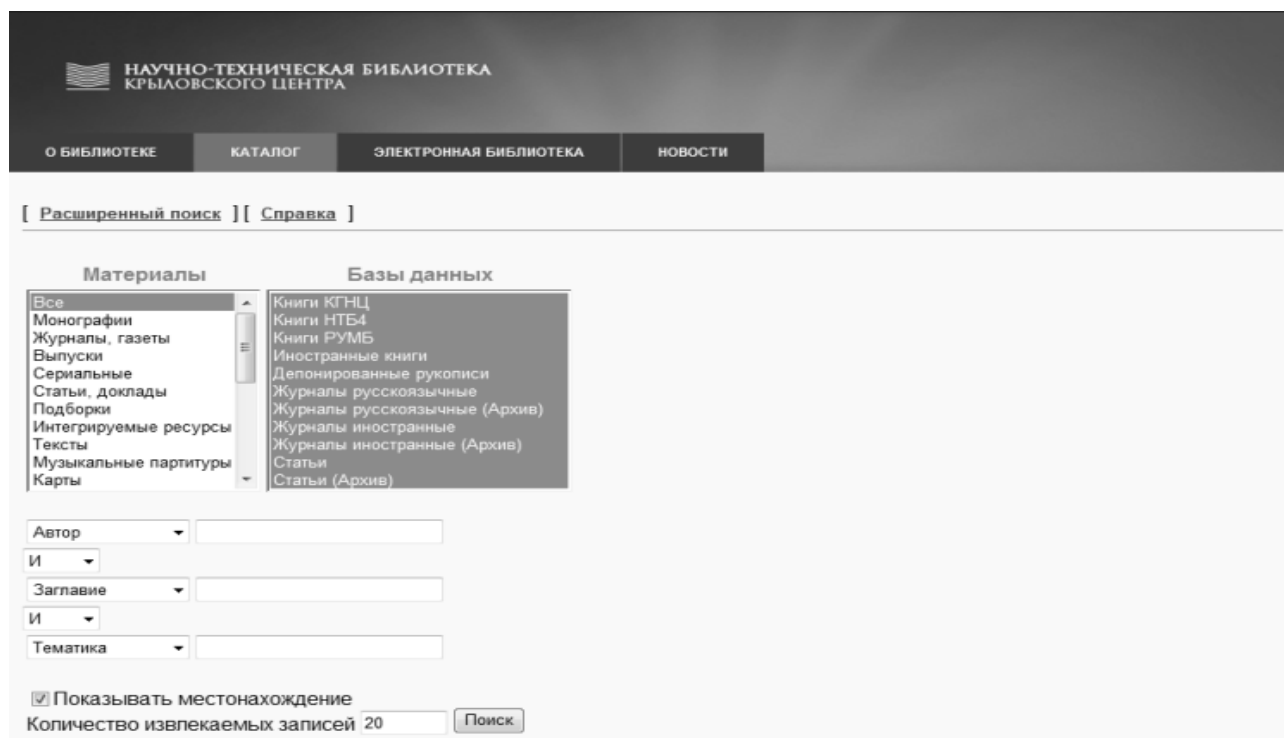


Рис. 1. Электронный каталог НТБ Крыловского центра. Интерфейс страницы сайта для читателей

НАУЧНО-ТЕХНИЧЕСКАЯ БИБЛИОТЕКА КРЫЛОВСКОГО ЦЕНТРА

О БИБЛИОТЕКЕ КАТАЛОГ ЭЛЕКТРОННАЯ БИБЛИОТЕКА НОВОСТИ

Личный кабинет Выход из системы Расширенный поиск Атрибутный поиск Русский English



ЭБ Электронная библиотека Поиск [Расширенный поиск](#) ?

Поиск по электронным документам

Статистика использования

Найдено документов **295** 1 2 3 4 5 6 7 8 9 10 11 ... 30 [Следующая](#)

Статистика слов: звук: 2092

1  **T.61 № 2 // Акустический журнал** 

Дополнительные сведения: , 2015
 Тип документа: Выпуск
 Тип файла: PDF

БИОМЕДИЦИНСКАЯ АКУСТИКА ВОССТАНОВЛЕНИЕ ПРОСТРАНСТВЕННЫХ РАСПРЕДЕЛЕНИЙ СКОРОСТИ ЗВУКА И ПОГЛОЩЕНИЯ В ФАНТОМАХ МЯГКИХ БИОТКАНЕЙ ПО ЭКСПЕРИМЕНТАЛЬНЫМ ДАННЫМ УЛЬТРАЗВУКОВОГО ТОМОГРАФИРОВАНИЯ Буров В.А., Зотов Д.И., Румянцева О.Д. УЛЬТРАЗВУКОВОЙ...

Автор-соавтор
[Гладилин А.В. \(8\)](#)
[Кирпичников В.Ю. \(7\)](#)
[Савенко В.В. \(7\)](#)
[Иванов В.С. \(5\)](#)
[Смоляков А.В. \(5\)](#)
[Чичин В.А. \(5\)](#)
[Ионов А.В. \(4\)](#)
[Романов В.Н. \(4\)](#)
[Баронкин В.М. \(3\)](#)
[Бовис А. \(3\)](#)
[Полный список >>>](#)

Рис. 2. Выдача результатов поиска по запросу «звук» в Электронной библиотеке Крыловского центра

Автор-соавтор	Год публикации	Тематика	Тип документа
Гладилин А.В. (8)	2015 (103)	АКУСТИКА (41)	Выпуск (162)
Кирпичников В.Ю. (7)	2014 (95)	КОЛЕБАНИЯ АКУСТИЧЕСКИЕ (17)	Статья, доклад (117)
Савенко В.В. (7)	2008 (14)	ШУМ (11)	Другой (16)
Иванов В.С. (5)	2010 (12)	ВИБРАЦИИ (10)	Полный список >>>
Смоляков А.В. (5)	2012 (9)	ГИДРОАКУСТИКА (9)	
Чичин В.А. (5)	1974 (8)	КОНСТРУКЦИИ КОРПУСА (6)	
Ионов А.В. (4)	2003 (8)	КОЛЕБАНИЯ ТЕЛ (5)	
Романов В.Н. (4)	2004 (8)	КОРПУС СУДНА (5)	
Баронкин В.М. (3)	1975 (7)	ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ (5)	Язык
Бовис А. (3)	2011 (7)	МЕХАНИКА ДЕФОРМИРУЕМЫХ ТЕЛ — Д... (4)	Русский (219)
Полный список >>>	Полный список >>>	Полный список >>>	Полный список >>>

Рис. 3. Классификация документов, содержащих в тексте или в названии слово «звук», по различным критериям

Каждый из критериев (язык, пожалуй, в меньшей степени) имеет большое значение, независимо от требований различных групп читателей к запрашиваемой информации.

Если такие критерии как автор, год публикации, тип документа и язык публикации практически не требуют предварительного анализа при создании записи, то раздел «ТЕМАТИКА» является наиболее сложным, требующим при обработке не только времени, но и определенной квалификации работника библиотеки.

При разработке или внедрении методики формирования словаря предметных рубрик и тезаурусов в различных библиотеках основное внимание уделяется автоматизации рабочих процессов для уменьшения субъективности предметизатора при отборе материала,

а также для снижения временных и интеллектуальных затрат. Противопоставить что-либо такому обоснованию внедрения той или иной методики довольно сложно и основными аргументами не в пользу автоматизации процесса являются:

- относительно небольшой поток документов, поступающих в фонд ведомственной научно-технической библиотеки;
- сложность определения рубрики для некоторых записей, даже при наличии ключевых терминов в тексте публикации;
- вероятность изменения перечня существующих тематических рубрик и, следовательно, необходимость выборочной корректировки созданных ранее записей.

На результате информационного поиска может негативно сказаться не только автоматическое, но даже просто формальное определение тематики: рубрики, подрубрики и списка ключевых слов документов.

Простейший пример. Публикация посвящена гребному винту. Даже при наличии составленного автором произвольного списка ключевых слов и аннотации к статье, определение рубрики для данной публикации может оказаться затруднительным. В таблице представлены примеры выбора предметной рубрики в зависимости от тематики публикации.

Даже тщательный лексический анализ текста публикации с использованием компьютерных методов не может гарантировать правильное и безоговорочное определение рубрики, тем более, что в практике НТБ встречаются примеры присвоения двух (реже трех) рубрик одной записи. Результаты данной работы приобретают ценность, если организован диалог с представителями подразделений, которые являются или в будущем станут клиентами НТБ.

В. В. Брежнева и Р. С. Гиляревский утверждают, что «уточнение и детализация представлений о коллективных информационных потребностях, а также выявление индивидуальных потребностей происходит в ходе интервьюирования сотрудников подразделений. Респондентов выбирают из числа наиболее авторитетных, т.е. квалифицированных, творческих и информационно активных специалистов, так как от правильности и полноты их ответов, дальновидности и ориентации в достижениях своей и смежных отраслей будет затем зависеть рациональность построения системы информационного обеспечения предприятия» [1, с. 62–63]. Добавить к этому практически нечего, кроме того, что, конечно, не каждая запись требует согласования со специалистами. Существенным подспорьем для предметизатора является анализ существующих записей, помощь в классификации которых была оказана учеными ранее. К тому же ученый сам нередко является автором той или иной публикации, в таком случае одним из определяющих моментов для классификации будет фамилия автора. В любом случае, «доверить машине» классификацию записей при, как было сказано ранее, довольно небольших объемах поступлений – нецелесообразно.

В примечании к п. 6.1.4 ГОСТ 7.59–2003 «Индексирование документов. Общие требования к систематизации и предметизации» отмечается, что термины централизованного индексирования (классификационные индексы или предметные рубрики) допускаются дополнять, уточнять и, при необходимости, изме-

нять [2]. Это лишний раз указывает на наличие определенной свободы в деятельности предметизатора, которая не только желательна, но и необходима.

Особое значение работа по присвоению записям тематических рубрик и их регулярная актуализация приобретет при внедрении разрабатываемой в НТБ методики «встроенного» обслуживания специалистов, предполагающей определение круга информационных источников, необходимых для выявления информации. Наличие актуализированного рубрикатора и тезауруса и, следовательно, классифицированных должным образом записей для этой работы сложно переоценить.

Еще одним перспективным направлением работы, также связанным с классификацией документов, является присвоение демонстрируемым на сайте электронной библиотеки документам условных характеристик: анализ цитирования публикаций (включая показатели ведущих международных баз), статистику использования и т.д.

Также требуют рассмотрения определенные моменты в работе предметизатора, возникающие при необходимости изменить присвоенной записи тематическую рубрику, или список ключевых слов. Прежде всего, следует обозначить причины возникновения такой ситуации.

Развитие направлений разработок Крыловского государственного научного центра демонстрируется на рис. 4.

Направление «Гидродинамика» возникло вместе с созданием первой лаборатории центра в 1894 г. и по прошествии 120 лет является одним из ключевых. Безусловно, фонд НТБ Крыловского центра, являющийся его «ровесницей», комплектовался и комплектуется различными документами по вопросам прикладной и теоретической гидродинамики. Оставить перечень тематических рубрик неизменным теоретически возможно, тем более, что есть таблицы УДК, рубрики тематических РЖ, упомянутый «Тезаурус информационно-поисковый по судостроению» и т.д. Так каковы же причины изменения или дополнения присвоенных записям рубрик, подрубрик и ключевых слов?

1. Специфика деятельности Крыловского центра, особенно в работе над так называемыми «сложными проектами», подразумевает совместную деятельность различных подразделений: гидромехаников, проектантов, прочнистов, энергетиков и т.д., которые используют, а по результатам выполнения проекта и «производят» специфическую информационную продукцию.

Сопоставление тематики публикации и рубрики для записи

Тематика публикации	Присваиваемая рубрика
Проектирования движителя	ТЕОРИЯ КОРАБЛЯ
Материалы для производства движителя	СУДОСТРОИТЕЛЬНЫЕ МАТЕРИАЛЫ
Кавитации движителя	ГИДРО- И АЭРОМЕХАНИКА
Шум движителя	АКУСТИКА
Технологическая прочность движителя	ТЕХНОЛОГИЯ СУДОСТРОЕНИЯ
Особенности эксплуатации движителя на судах определенных типов	ЭКСПЛУАТАЦИЯ СУДОВ

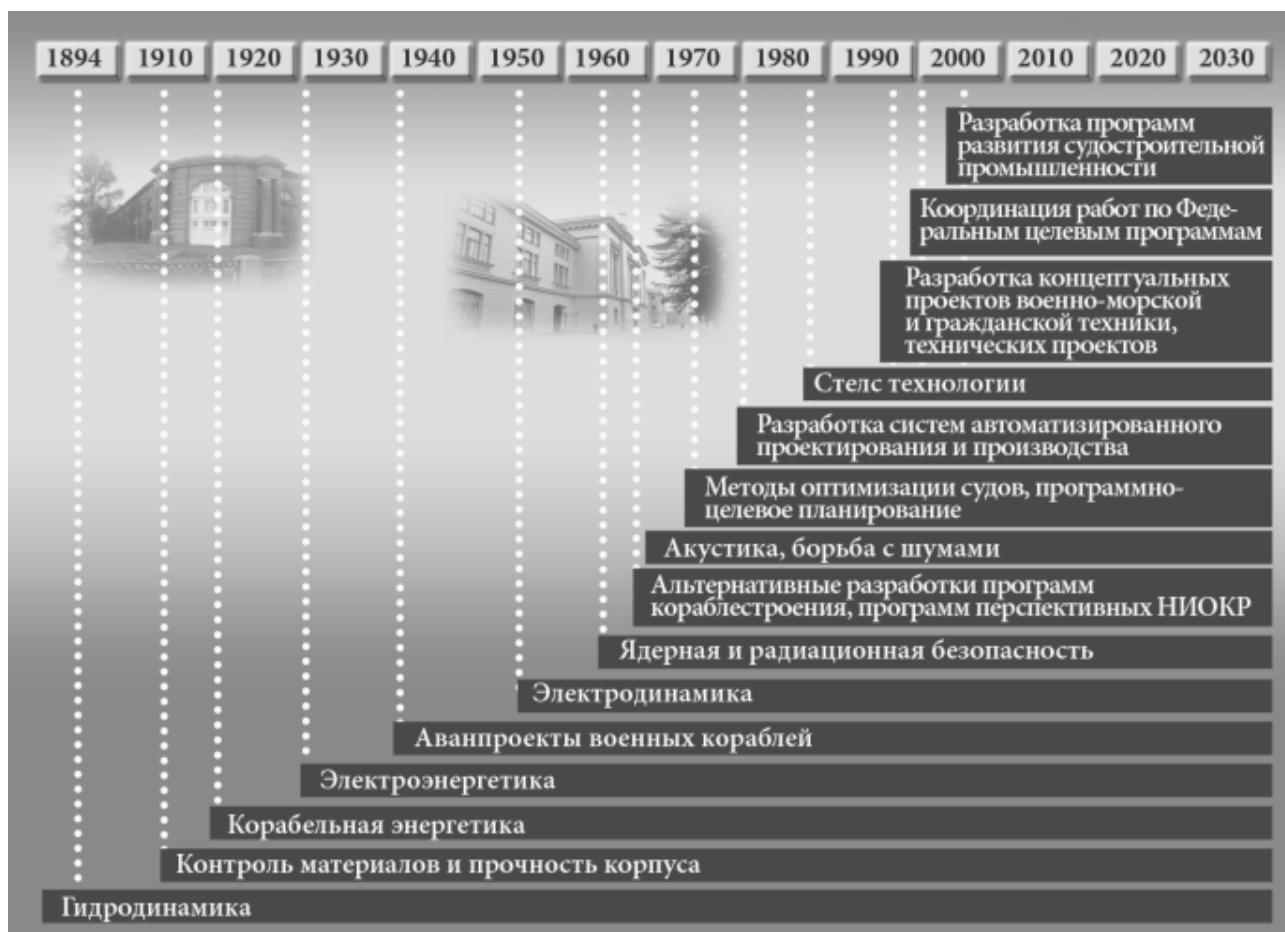


Рис. 4. Развитие направлений научной деятельности Крыловского центра

Предположим, что на стадии перспективной разработки какого-нибудь проекта отделением гидродинамики выпущен ряд публикаций по созданию того же гребного винта. Далее на стадии технического проекта отделением прочности была подготовлена статья, посвященная рекомендуемым для изготовления гребного винта материалам. После строительства и ввода в эксплуатацию объекта появились работы по специфике его эксплуатации.

Комплексный анализ публикаций, посвященных данному проекту, представляет большую ценность при работе, например, над следующим «сложным проектом».

Весьма вероятно, что единое ключевое слово (или группа ключевых слов), позволяющее объединить все публикации, посвященные конкретному объекту на различных стадиях его создания и эксплуатации, не употреблялось при создании записей. Хорошо, если название проекта фигурирует в названии публикации или в полном тексте статей, что также возможно не всегда. Дополнение списка ключевых слов в этих публикациях «объединяющим» термином – задача важная и непростая. Однако автоматизация этого действия на данный момент не представляется возможной.

2. Как видно из рис.4, новые тематические направления появлялись практически каждое десятилетие, и как правило, это связано с созданием новых

подразделений, а, следовательно, с комплектованием фонда НТБ документами новой тематики. Очевидно, что та же «Судовая энергетика» всегда представляла практический интерес для исследователей. Однако, поскольку объем публикаций поступающих в фонд был невелик, а сотрудники практически не публиковали материалы по этому научному направлению, то редкие публикации относили к разделам «Теория корабля» или «Строительная механика». С появлением профильного научного подразделения изменился не только тезаурус, но и рубрикатор. Появилась соответствующая рубрика – «Судовая энергетика». Такие изменения актуальны и сегодня.

Так, например, в 2012 г. в состав Крыловского центра вошел филиал ЦНИИ СЭТ (Центральный научно-исследовательский институт судовой электротехники и технологий) — многопрофильный специализированный научно-производственный центр, выполняющий наукоемкие научно-исследовательские, опытно-конструкторские и технологические работы в области создания судового электротехнического оборудования. Ученые подразделения публикуют научные статьи, заказывают научно-техническую литературу и т.д.

Для создания актуализированных информационных и библиографических справок, а также для оптимизации поиска, проводимого читателями, в системный рубрикатор внесена подрубрика «Электроэнергетические системы». Записи, созданные ранее, откоррек-

тированы. Даже задать условия для автоматизации данного процесса практически невозможно. Работа ведется сотрудниками следующим образом: анализ систематического каталога, существующих рубрикаторов, консультации со специалистами, и т.д.

3. Работа по классификации некоторых записей была проведена «излишне осторожно». Например, к публикациям, посвященным подводным лодкам, несмотря на отсутствие конкретной инструкции, присваивали подрубрику «*Суда подводные*». Как говорится, на всякий случай. Заменить автоматически «*Суда подводные*» на «*Подводные лодки*» нельзя, так как «пострадают» материалы про глубоководные аппараты и т.п.

4. Изменяются характер и методы научных исследований. А.В. Дутов, занимавший пост Генерального директора предприятия в 2012–2014 гг., отмечает, что «перечень фундаментально-поисковых научных исследований должен содержать формулировки научных задач для расширения знаний о фундаментальных основах и новых явлениях в области научного направления данного подразделения, прямо не связанных с конкретной прикладной задачей» [3, с. 192]. В качестве примера фундаментально-поисковых исследований приводится создание суперкомпьютерных моделей кавитационных явлений в трубопроводах, арматуре, в пристеночных областях, разрезных конструкциях и пр.

В тезаурусе НТБ есть термины: «кавитация»; «трубопроводы»; «резонанс» и т.д. Однако очевидно, что данные элементы классификации едва ли смогут удовлетворить информационную потребность специалиста, ведущего поиск информации по указанному направлению. Статьи и монографии, соответствующие критериям поиска, будут освещать научные разработки с использованием модельных экспериментов или математических расчетов без применения суперкомпьютерных технологий!

Конечно, в тезаурусе присутствуют такие термины как «методы компьютерные» или «программирование», но, принимая во внимание тот факт, что суперкомпьютерный центр был сформирован на базе отделения гидродинамики в 2011 г. все предшествующие публикации, даже посвященные суперкомпьютерным технологиям, либо созданы в организациях с другим профилем деятельности, либо носят сугубо теоретический характер. Курьеза ради можно привести такой пример. С ключевым термином «*Методы компьютерные*» в НТБ присутствуют записи, одна из которых относится к 1997 г., и сегодня может быть интересна только исследователям истории развития компьютерных технологий.

Конечно, когда речь идет об одном из самых динамично развивающихся научных направлений, год издания документа является определяющей характеристикой, и читатель, скорее всего, ограничит хронологические рамки поиска, но наличие в описании современной публикации подходящего ключевого слова значительно упростило бы его работу при анализе научных публикаций по данной тематике.

Таким образом можно сделать следующие выводы:

1. Актуализация системного рубрикатора НТБ Крыловского центра – процесс очень важный и постоянный. Для «встроенного» обслуживания, методика которого находится в стадии подготовки, значение этой работы крайней велико.

2. Автоматизация этого процесса при создании и редактировании записей, по крайней мере на современном этапе, нецелесообразна, не исключено, что и попросту невозможна.

3. Стремительное развитие современных баз научного цитирования, таких как Scopus, которые позиционируют свою деятельность как услуги исследователям, позволяющие сэкономить время при поиске информации, диктуют новые требования, которые можно реализовать на базе автоматизированных информационно-библиотечных систем ведомственных библиотек, с учетом специфики деятельности организации.

СПИСОК ЛИТЕРАТУРЫ

1. Брежнева В.В., Гиляревский Р.С. Информационное обслуживание: учеб. пособие. – СПб. : Профессия, 2012. – 368 с.
2. ГОСТ 7.59–2003. Индексирование документов. Общие требования к систематизации и предметизации // Сборник стандартов СИБИБД. – СПб. : Профессия, 2010. – С. 212-219.
3. Дутов А.В. Структурирование перспективных научных задач ФГУП «Крыловский государственный научный центр» // Труды Крыловского государственного научного центра. – 2013. – Вып.78(362). – С.191-198.

Материал поступил в редакцию 13.01.16.

Сведения об авторе

КРУЛЕВ Андрей Александрович – руководитель группы обработки научно-технической информации ОНТИ информационно-издательского центра ФГУП «Крыловский государственный научный центр», магистрант 1 курса СПбГИК.
E-mail: galaction@yandex.ru

СПРАВОЧНО-ИНФОРМАЦИОННЫЙ РАЗДЕЛ

УДК [002 : 004.89] : 351.07

С.С. Терещенко

Современные ситуационные центры: информационно-аналитическое обеспечение

Рассматриваются аналитические издания ВИНТИ РАН, отражающие вопросы ситуационного, проблемного анализа, и роль сборника «Научно-техническая информация» в развитии ситуационных центров России как части региональных аналитических служб. На примерах раскрывается значение информационного обеспечения в формировании сети аналитиков и экспертов при административном и корпоративном управлении.

Ключевые слова: *ситуационные центры, интеллектуальный анализ данных, информационно-аналитические службы*

В 2015 году исполнилось 20 лет активного создания сети региональных аналитических служб в субъектах Российской Федерации. Этот проект государственного уровня реализован в соответствии с распоряжением Президента Российской Федерации от 20 июля 1992 г. №385-рп по обеспечению подготовки аналитических материалов. Последующим распоряжением Правительства Российской Федерации от 14 января 1993 г. № 51-р, утвержден «Порядок-Приложение» взаимодействия информационно-аналитического центра Администрации Президента Российской Федерации с государственными организациями, администрациями субъектов Федерации. Это было фактически техническим заданием на разработку сети аналитических служб федерального и регионального уровня, которое обсуждалось на рабочих совещаниях ведущих аналитиков и представителей государственной власти 10 июня 1992 г. и в ноябре 1993 г. Научными руководителями проекта, включая представителя ВИНТИ РАН, были А.Н. Ракитов, И.С. Милыхин, А.Н. Райков (ИНИОН РАН, ИПУ РАН), впоследствии – Е.Г. Ясин (ВШЭГУ).

Сеть аналитических служб должна была формироваться в субъектах федерации по типовым проектным решениям с использованием потенциала центров анализа информации, экспертизы и прогнозирования, активно создаваемых в исследовательских институтах РАН. При этом возникла необходимость объединения информационно-аналитических ресурсов Государственной системы научно-технической информации и Общегосударственной автоматизированной системы (ОГАС).

В ВИНТИ РАН был накоплен значительный опыт разработки не только автоматизированных информационно-поисковых систем (АИПС) и автоматизированной системы научно-технической информации (АСНТИ), различных информационных сетей международного и национального уровней, но и активного интеллектуального обеспечения проблемного, ситуационного анализа, а также сопровождения центров анализа информации в обеспечении национальной безопасности. В качестве примера ответственного использования информации в ситуационном управлении можно привести создание базы данных в Гидрометеоцентре – Международном центре данных, задачи которого – краткосрочное и долгосрочное прогнозирование изменения климата и среды обитания с использованием информационных ресурсов отраслевой автоматизированной системы научно-технической информации (ОАСНТИ) в гидрометеорологии [1]. Ситуационный компонент при этом определялся временным интервалом, в котором подготавливаются и принимаются управленческие решения. Второй пример – это информационное сопровождение продвижений по воздушным трассам, когда быстро изменяются часовые пояса, климатические зоны, погодные условия.

Концепцию ситуационного анализа, его информационного и аналитического обеспечения, разработанную в ВИНТИ за 40 лет, можно представить следующей смысловой цепочкой (моделью): подмножество ситуаций возникает из транзитивного взаимодействия рисков, вызовов, угроз в так называемых критических системах жизнеобеспечения. Пример: энергетика, жилищно-коммунальное хозяй-

ство, агрономия в зоне рискованного земледелия, транспорт, торговля, разведка и разработка месторождений полезных ископаемых и т.д. В субъектах Федерации необходимо накапливать и использовать разные типы информации в управлении такими критическими системами [1-4].

Современные аналитические службы в администрациях субъектов Федерации ведут несколько десятков баз и банков данных политического, социального, экономического, финансового, промышленного, отраслевого, ресурсного назначения. В управлении ситуациями участвуют специалисты социального анализа, эксперты центров компетентности, аналитики центров анализа информации, представители исследовательских и проектно-конструкторских организаций, а также различных академий и университетов. Таким образом, региональная сеть аналитических служб представляет собой синергетическое образование с размытыми границами организационного подчинения и ответственности. Определенный вклад в дезорганизацию и беспорядок вносят современные слабо контролируемые сетевые коммуникации. Общее, что позволяет управлять критическими системами в регионах и отраслях народного хозяйства России, – это информационная культура, информационные и аналитические технологии, регулирование семантической сложности в разрабатываемых прогнозах, планах, стратегиях, дорожных картах, а также умение использовать гетерогенные источники информации из разных сфер деятельности, отраслей, научных направлений и дисциплин.

Разработка концепции ситуационного и проблемного анализа в ВИНТИ РАН как национальном центре научно-технической информации отражалась в публикациях сборника «Научно-техническая информация», которые были посвящены следующим вопросам:

- проектирование отраслевых автоматизированных систем НТИ и автоматизированных систем управления (АСУ), проблемы их взаимосвязи; объекты ситуационного управления: отрасли энергетики, электроники, строительства; системный анализ, целевое планирование и управление, сетевое моделирование, программный подход;
- проектирование технологий и зданий центров НТИ, центров науки по программам СЭВ; проблемный анализ образования в высшей школе России – научная организация труда как идеология упорядочения синергии в промышленности и экономике (попытка управляемого хаоса);
- разработка Положения о проблемно-ориентированных базах и банках данных (ПОБД) для управления научно-исследовательскими программами и Методического пособия по созданию ПОБД (совместно с ИПКИРом и ГКНТ СССР по науке и технике);
- стратегия развития Государственной системы НТИ и Государственной автоматизированной системы НТИ;
- создание информационного анализа, экспертизы, прогнозирования и коммерческой деятельности: функции, технология, организационная структура, кадры;

- создание сети центров анализа информации правоохранительных органов: проблемный, ситуационный, информационный анализ;
- аналитика фирменного анализа, экономический анализ, проблемный анализ;
- структурный, технологический, экспертный анализ; аналитическая и экспертная деятельность на федеральном, региональном и международном уровнях;
- антикризисное управление, стратегический анализ;
- риски в контексте аналитики и информатики; ситуационные центры и системы; технологии выявления неявных проблем;
- проблемно-ориентированные базы и банки данных и проблемы образования;
- аналитика корпоративных и административных систем управления.

Комплексный материал по ситуационным центрам опубликован в томе III Аналитики-Капитал. ВИНТИ РАН участвовал в государственном проекте «Аналитика Агропрома Российской Федерации», который предусматривает создание Аналитического Федерального центра – регионального центра субъекта Федерации – аналитического центра межотраслевого кластера (компаний) – аналитического центра районного (муниципального) уровня. В этом издании предложена одна из классификаций ситуационных центров на примере Курской области, показана работа ситуационного центра в структуре региональной аналитической службы.

В томе IV издания «Аналитика-Капитал» анализируется стратегический менеджмент и маркетинг в АПК России, представлен разработанный Рубрикатор новых направлений по материалам полевых исследований в Курской, Калужской, Тамбовской, Ярославской областях [2].

Активное участие в разработке ситуационных центров федерального уровня приняли специалисты Федерального агентства правительственной связи и информации (ФАПСИ) – в настоящее время Федеральная служба охраны (ФСО) [3]. Так, в 1993 г. Центром новых компьютерных технологий научно-технического центра ФАПСИ разработана типовая оболочка проблематизации и оценки ситуации – ТОПОС: создана система «Инвариантрон-1993», разработаны АИПС по топливно-энергетическому комплексу России, АИПС по ситуациям исполнения поручений, АИПС и интегрированная БД по чрезвычайным ситуациям (ИНФОРМ-ЧС), информационно-моделирующая система по зонам затопления, авариям на гидротехнических сооружениях, аналитическая система по материалам изучения общественного мнения; фактографическая АИПС по экономико-географической информации России, база метаданных по органам административного управления, их аналитическим массивам. Для проблемного анализа процессов мониторинга предприятий госсобственности создана система ТОПАЗ.

Эти информационные ресурсы позволили ФАПСИ создать Ситуационный центр Президента России, а также Ситуационный Центр Совета безопасности России. К 2013 году в Российской Федерации рабо-

тали 50 ситуационных центров в федеральных округах и субъектах Федерации.

Новые технологии и информационное обеспечение получил Ситуационный центр Министерства по чрезвычайным ситуациям.

С 2012 г. активно работает Ситуационный центр Министерства обороны России, а в 2013 г. этот центр участвовал в урегулировании катастрофической обстановки в Приморье, Хабаровском крае, Амурской области, Якутии, где было затоплено около одного миллиона квадратных километров территории. При этом использовались архивные и современные геофизические и гидрометеоданные и материалы институтов Российской академии наук.

В 2013 г. Президент Российской Федерации поручил Федеральной службе охраны создать ситуационные технологии и службы по мониторингу в регионах (тематические направления – экономика, политика, социальные отношения, миграционные процессы). В Российской академии наук в рамках Отделения глобальных проблем и международных отношений был создан Центр ситуационного анализа для экспертной оценки актуальных проблем российской политики с использованием информационных ресурсов более 30 институтов РАН.

Многие правительственные и президентские структуры активно используют информационные ресурсы ВИНТИ РАН и в настоящее время. Эти ресурсы и идеи начинают осваивать и исследовательские университеты.

Опыт ВИНТИ по ситуационному и проблемному анализу до 2011 г. был обобщен в исследовательской работе «Аналитика ситуационного управления критическими системами жизнеобеспечения в субъектах Российской Федерации» [4], в которой использованы результаты прикладной науки и намечены пути их внедрения в практику взаимодействия МЧС и ВИНТИ РАН. Тематически эта работа продолжает и развивает исследования Главного управления МЧС России по Удмуртской Республике и ВПО «Ижевский Государственный Технический Университет» [5], где изложена концепция сети ситуационных центров по управлению особо опасными объектами.

Специалисты Федеральной службы охраны разработали проект создания более 100 Ситуационных центров в субъектах Российской Федерации, объединяемых концепцией Стратегии национальной безопасности [3]. В этих центрах предполагается использовать информационные ресурсы ВИНТИ РАН по вопросам наук о Земле – геология, геохимия, гидрология, геофизика, астрономия, климатология, природопользование, радиохимия, химия отравляющих веществ, радиобиология, биохимия, коррозия, микробиология, биоценозы. Особый интерес представляет тематика ситуационного управления и экономии ресурсов, предложенная Президентом России. В ВИНТИ РАН это информационно-аналитическое направление разрабатывается в рамках темы «Ресурсосбережение, нанотехнологии и инновации», предполагающей создание технологий ситуационного взаимодействия научно-образовательного центра ВИНТИ РАН и Ситуационного учебного центра

Российской академии государственной службы при Президенте РФ.

Таким образом можно сделать следующие выводы:

1. За последние годы значительно активизировалось ситуационное управление субъектами Российской Федерации на основе технологий Ситуационного центра Президента России и Ситуационного центра Правительства России, СИЦ МЧС России, СИЦ МО России, а также оперативно-диспетчерское управление инфраструктурными системами России (транспорт, связь, логистика, финансы, торговля, ЖКХ, образование, медицинское обслуживание и т.д.).

2. Качество ситуационного и проблемного анализа, эффективность управления часто не учитывают уровень рисков, угроз, чрезвычайных ситуаций, а темпы принимаемых решений не соответствуют скоротечности опасных и особо опасных событий. Исключительно некоммуникабельны финансовые, таможенные и силовые структуры.

3. Анализ работающих или проектируемых ситуационных информационных центров, центров экспертизы и диагностики, центров компетентности свидетельствует о недостаточном использовании этими структурами значительных государственных и корпоративных информационных и интеллектуальных ресурсов. В России созданы уникальные когнитивные базы и банки данных фундаментальной, прикладной и отраслевой науки. Поддерживаются многие ведомственные информационно-аналитические системы, позволяющие прогнозировать события, подобные аварии на Саяно-Шушенской ГЭС, затоплению города Крымска, катастрофам в Приморье и Хабаровском крае в августе 2013 г.

4. Существующие группы экспертов и аналитиков отраслевого, ведомственного, образовательного, научно-дисциплинарного уровней не мотивированы к активной совместной работе. В то же время потенциал российской научной и инженерно-конструкторской аналитики и экспертизы может эффективно и солидарно использоваться в особых случаях. Пример плодотворного сотрудничества специалистов РосАтома, ученых нескольких институтов РАН, университетов химико-физического, биологического, геофизического, океанологического профилей – совместное исследование и прогнозирование чрезвычайных событий на атомной электростанции Фукусима, Чернобыльской атомной электростанции.

5. Инфраструктура интеллектуального управления отраслями и государством в целом демонстрирует синергетический характер, который частично нивелируется традиционной культурой информационного обеспечения и сопровождения управленческих решений, характерных для академических институтов и ведущих университетов страны. К этому можно добавить аналитические возможности центральных изданий средств массовой информации.

6. В стратегическом прогнозировании, от которого зависит и качество проблемного ситуационного анализа, может использоваться разработанная в ВИНТИ РАН концепция многоуровневой, междоотраслевой, междисциплинарной аналитики. Эта технология аналитического прогнозирования и управления заложена в методике ситуационного управления

региональных аналитических служб (Президентский проект 1993-2013 гг.).

7. В июне 2013 г. на Ученом Совете ВИНТИ РАН рассматривались новые направления подготовки актуальных аналитических материалов, которые могут использоваться в отраслевых стратегиях развития. Отдельные аспекты отраслевых и научных точек роста, проблемных направлений развития страны были предварительно обсуждены на 8-й Международной конференции «Актуальные проблемы информационного обеспечения науки, аналитической и инновационной деятельности» [6]. Обсуждались отдельные концептуальные направления информационно-аналитического сопровождения Федеральных целевых программ, критических и высоких технологий, технологических платформ. Особое место занимали проблемы перехода на новый технологический уклад, где приоритет отдан нанотехнологиям, нановеществам и наноматериалам.

8. В разработанной Саратовским государственным социально-экономическим университетом «Энергетической стратегии-2030» [7] для управления и стратегического развития газовой корпорации использован зарубежный опыт ситуационного прогнозирования, обобщен опыт Гарвардской школы бизнеса, описана модель стратегического планирования, вводится понятие стратегической ситуации.

9. Создание новой современной элементной базы для приборов ситуационного анализа и сбора ситуационной информации, а также сенсоров и датчиков высокой чувствительности, планшетных персональных компьютеров, мобильных средств связи высокого быстродействия и пропускной способности позволяет разрабатывать распределенные технологии ситуационного управления. Так, для стратегического и оперативного управления «Газпромом» создан планшетный ПК за 120 млн. руб. (~iPad2), он работает в сетях 3G, имеет функции GPRS и Wi-Fi. Это мобильное рабочее место по ситуационному управлению корпорацией, зеркально копирующее содержание информации и работу базового сервера высокой надежности Газпрома. Ситуационно обрабатываются около 500 показателей, из них 50 – регулярно аналитически оцениваются в 50 группах экспертов «Бизнес-анализа». С этого мобильного устройства Председатель правления «Газпрома» с оператором и системным программистом осуществляет мониторинг проблемных и ситуационных аспектов в круглосуточном режиме. Предусматривается использование справочной научно-технической информации из баз данных по энергетике. Не исключается возможность обращения и в Банк данных ВИНТИ РАН – рубрика «Электронная Земля» (трубопроводы, ЧС, коррозия, геодинамика и сейсмические явления, балансовые показатели месторождений, рынки углеводородов, подземные газосы хранящиеся и др.).

10. В 2014 году ситуационное управление стали преподавать в Финансовом университете при Правительстве Российской Федерации. Концепция развития кластерных образовательных структур была разработана в ВИНТИ РАН [7].

СПИСОК ЛИТЕРАТУРЫ

1. Шечков Б.Н., Терещенко С.С. Основные принципы построения ОАСНТИ в гидрометеорологии // Научная организация труда и проектирование информационных центров / под ред. проф. А.И. Михайлова. – М.: ВИНТИ, 1976. – С. 191-204.
2. Стратегический менеджмент и маркетинг. – Аналитика-Капитал – Том IV. – М.: ВИНТИ РАН – ИПКИР, 2000.
3. Фаличев О. Стране нужны фабрики «мысли» // Военно-промышленный курьер. – 2013. – № 16 (484). – С. 9.
4. Терещенко И.С. Аналитика ситуационного управления критическими системами жизнеобеспечения в субъектах РФ. – М.: ВИНТИ РАН – ИПКИР, 2011. – 92 с.
5. Габричидзе Т.Г. Комплексная многоступенчатая система безопасности критически важных, потенциально опасных объектов – Ижевск: ИжГТУ, 2008. – 59 с.
6. Материалы 8-й Международной конференции «Актуальные проблемы информационного обеспечения науки, аналитической и инновационной деятельности», 28-30 ноября 2012 г. – ВИНТИ 60 лет. – 259 с.
7. Терещенко А.С. Генезис аналитического обеспечения интегрированных научно-образовательных структур (на примере научно-образовательного центра ВИНТИ РАН). – М.: ВИНТИ РАН – ИПКИР, 2012. – 96 с.
8. Смирнов В.В. Управление развитием укрупненных образовательных космплексов с использованием ситуационных центров // Труды VII-й Международ. Конференции «Управление развитием крупномасштабных систем – MISD'2013», 30 сентября-2 октября 2013 г., г. Москва. – М.: ИПУ РАН, 2013. – 447 с.

Материал поступил в редакцию 30.12.15.

Сведения об авторе

ТЕРЕЩЕНКО Сергей Сергеевич – доктор технических наук, ведущий научный сотрудник ВИНТИ РАН, Москва

ВНИМАНИЮ ЧИТАТЕЛЕЙ!

Полное четвертое издание таблиц Универсальной десятичной классификации (УДК) на русском языке

УДК. ТОМ VIII. 66 Химическая технология. Химическая промышленность. Пищевая промышленность. Металлургия. Родственные отрасли

Содержание тома:

- 661 Продукты химической промышленности
- 662 Взрывчатые вещества. Топлива
- 663 Промышленная микробиология. Промышленная микология. Технология бродильных производств. Промышленная ферментация. Пивоваренная промышленность. Производство алкогольных напитков
- 664 Производство и консервирование пищевых продуктов. Пищевая промышленность в целом
- 665 Масла. Жиры. Воски. Камеди. Клеящие вещества. Смолы
- 666 Стекольная промышленность. Керамическая промышленность. Промышленность вяжущих
- 667 Промышленность красящих веществ. Производство и применение красок, чернил, пигментов и прочих материалов
- 669 Metallургия
- Алфавитно-предметный указатель к разделу 66

К сведению читателей: Изданные ранее Таблицы УДК - Том I, Том II, Том III, Том V, Том VI (ч. I, II), Алфавитно-предметный указатель к т. VI, а также «Изменения и дополнения к таблицам УДК» (Вып. 2, 3) предоставляются только в электронном виде на CD-ROM

Для подписки необходимо направить заявку для оформления счета по адресу:

125190, Россия, Москва, ул. Усиевича, 20, НМО ВИНТИ

Телефон: 8-499-155-42-52

Факс: 8-499-943-00-60 (для НМО)

E-mail: typo@viniti.ru

ВНИМАНИЮ ЧИТАТЕЛЕЙ!

Полное четвертое издание таблиц Универсальной десятичной классификации (УДК) на русском языке

УДК. ТОМ IX. 67/69 Различные отрасли промышленности и ремесел. Строительство

Содержание тома:

класс 67

В этот класс входят такие подклассы, как:

671 Производство изделий из благородных металлов и драгоценных камней

674 Деревообрабатывающая промышленность

675 Кожевенная промышленность. меховое производство

676 Целлюлозно-бумажная промышленность

677 Текстильная промышленность

678 Промышленность высокомолекулярных веществ и др.

класс 68

681 Точная механика

682 Кузнечное ремесло

684 Мебельная промышленность

685 Обувное производство

687 Швейная промышленность

класс 69 Строительство. Строительные материалы. Строительно-монтажные работы

Алфавитно-предметный указатель ко всем представленным в томе классам

К сведению читателей: Изданные ранее Таблицы УДК - Том I, Том II, Том III, Том V, Том VI (ч. I, II), Алфавитно-предметный указатель к т. VI, а также «Изменения и дополнения к таблицам УДК» (Вып. 2, 3) предоставляются только в электронном виде на CD-ROM

**Для подписки необходимо направить заявку для оформления счета по адресу:
125190, Россия, Москва, ул. Усиевича, 20, НМО ВИНТИ**

Телефон: 8-499-155-42-52

Факс: 8-499-943-00-60 (для НМО)

E-mail: typo@viniti.ru

ВНИМАНИЮ ЧИТАТЕЛЕЙ!

Полное четвертое издание таблиц Универсальной десятичной классификации (УДК) на русском языке

УДК. ТОМ X. 7/9 Искусство. Спорт. Филология. География. История

Содержание тома:

- класс 7 Искусство. Развлечения. Зрелища. Спорт.
В этот класс входят такие подклассы, как:
 - 71 Планировка в масштабе страны. Районная планировка. Градостроительство
 - 72 Архитектура
 - 74 Рисование и черчение. Декоративно-прикладное искусство.
Художественные промыслы. Дизайн
 - 75 Живопись
 - 76 Графические искусства
 - 77 Фотография, кинематография
 - 78 Музыка
 - 79 Зрелищные искусства. Массовые развлечения. Игры. Спорт
- класс 8 Языкознание и языки. Лингвистика. Литература
- класс 9 География. Биографии. История

Алфавитно-предметный указатель ко всем представленным в томе классам

К сведению читателей: Изданные ранее Таблицы УДК - Том I, Том II, Том III, Том V, Том VI (ч. I, II), Алфавитно-предметный указатель к т. VI, а также «Изменения и дополнения к таблицам УДК» (Вып. 2, 3) предоставляются только в электронном виде на CD-ROM

**Для подписки необходимо направить заявку для оформления счета по адресу:
125190, Россия, Москва, ул. Усиевича, 20, НМО ВИНТИ**

Телефон: 8-499-155-42-52

Факс: 8-499-943-00-60 (для НМО)

E-mail: typo@viniti.ru