

В. А. Бубнов

ИНФОРМАТИКА И ИНФОРМАЦИЯ: знаково-символьный аспект



ИЗДАТЕЛЬСТВО

БИНОМ

В. А. Бубнов

ИНФОРМАТИКА И ИНФОРМАЦИЯ: знаково-символьный аспект



Москва
БИНОМ. Лаборатория знаний

ОГЛАВЛЕНИЕ

Предисловие	3
Введение	7
Глава 1. Информация и ее измерение	9
1.1. Понятие «информация»	9
1.2. Количественная оценка информации и процесс познания окружающей действительности	28
1.3. Количественные измерения информации на основе формулы Шеннона	60
Глава 2. Логические основы символического исчисления ..	89
2.1. Символическое исчисление Джорджа Буля	89
2.2. Связь исчисления Буля с теорией вероятностей	103
2.3. Математические и логические операции	121
2.4. Математическое мышление как одна из форм внематематического мышления	149
2.5. Механизация мышления и системы искусственного интеллекта	181
Глава 3. Алгоритмы	201
3.1. Вычислительный алгоритм как пример конструктивного процесса	201
3.2. Нормальные алгоритмы А. А. Маркова	211
3.3. Алгоритмы теории графов при анализе поэтических текстов	219

3.4. Использование математической статистики при формальном анализе авторской индивидуальности литературного произведения.....	239
3.5. Компьютерный анализ исторических явлений.....	248
3.6. Алгебраические алгоритмы решения логических уравнений.....	278
Литература.....	313

ПРЕДИСЛОВИЕ

Современная наука, продолжая традиции науки Нового времени, содержит в себе скрытый до времени парадокс. С одной стороны, научное творчество требует полной свободы, с другой — основным «продуктом» научной деятельности являются технологии, которые пределы этой свободы так или иначе ограничивают. Если же прислушаться к словам А. Уайтхеда: «Civilization advances by extending the number of important operation which we can perform without thinking about them», то получается, что неизбежным следствием научного прогресса является умаление духа свободы. Более того, можно предположить, что настанет момент возникновения парадокса «самоприменимости», когда объектом технологий станет сама наука (не это ли сейчас и происходит?).

Кроме очевидного ограничения свободы технология как таковая обладает одной принципиальной особенностью — для ее реализации нужна искусственная среда. Действительно, делая очередной шаг технологического процесса («without thinking about it»), мы должны быть полностью уверены, что не возникнет никаких «сюрпризов», способных нарушить этот процесс. Только в этом случае возможна полная автоматизация такого процесса. Понятно, что реальный мир не предоставляет нам такой возможности, и человек самостоятельно создает новую, технологическую реальность, сначала в мыслях, а потом и наяву. Стержень такой реальности становится некий универсальный объект, сам по себе лишенный каких-либо определенных качеств, но служащий «строительным материалом», из которого вылепливается весь остальной мир. При таком развороте все уникальные свойства реального объекта тают, и он становится одним из образов универсальности. Объекты такого сконструированного мира становятся совершенно «ручными» и не представляют опасности ни для какой формальной или формализованной деятельности.

Подобная схема может показаться надуманной, однако она с поразительной точностью реализуется в самых разнообразных ситуациях.

Приведем несколько примеров.

Промышленная революция Западной Европы вызвала к жизни сотни самых разнообразных технологий, для осуществления которых потребовались не только многие тысячи машин и механизмов, но и радикальное изменение природной среды. Однако эта среда издавна была носителем смысла: дождь и ветер, зной и мороз, засуха и наводне-

ния были не только событиями, но и намеками на некую высшую реальность, или даже были ее прямыми проявлениями (известный мидивист Й. Хёйзинга называл такие проявления «коротким замыканием»). Разумеется, это создавало большие неудобства для развития технологической цивилизации. И тогда на свет явилась идея «материи» в ее современной трактовке, мало напоминающей «материю как восприимчивую всему» Демокрита и Лукреция. Все окружающие предметы стали мыслиться полностью материальными, что делало возможным их безбоязненную машинную обработку. Разумеется, постоянно существует опасение, что такое допущение не вполне соответствует реальному положению вещей и что, возможно, в объектах существует «нечто иное», способное разрушить не только технологии, но и саму мысль о позитивности и надежности технологического мира.

Очень схожая революция осуществилась и в ином, идеальном мире — мире математике. Накопленное за всю историю этой науки многообразие объектов в конце XIX века было подвергнуто под единое понятие «множества». Технологический замысел этой метаморфозы стал ясным с появлением «программы Бурбаки», следствием которой стало стремление к жесткой организации интеллектуальной работы: всякая новая мысль с неизбежностью должна была облачиться в одеяния теоретико-множественных структур. Со временем появились и хорошо известные устройства, эти структуры воплощающие.

Разумеется, в рамках такого построения остается вопрос: насколько искусственный в целом мир множеств адекватно отражает математический универсум? Или по-иному: насколько теоретико-множественная технологичность оказывается самодостаточной? Ответ на этот вопрос, принадлежит прежде всего математике, но, в силу внутренней «честности» этой дисциплины, он представляет интерес и для оценки более широкого круга явлений.

Имеет смысл привести хотя бы один пример из многочисленных «ответов» на сформулированный вопрос, которыми богата математика XX века.

Возможно, одним из первых явных отрицательных «ответов» стало появление дескриптивной теории множеств. Контекст это события вкратце таков. Чтобы сделать математический анализ самодостаточным (и, в определенной смысле, «технологичным»), необходимо замкнуть на себя совокупность множеств, с которыми он работает. Описание этой совокупности включало в себя начальные элементы — отрезки для прямой или прямоугольники для плоскости и три операции: объединение счетного числа множеств, пересечение счетного числа множеств

и дополнение. Полученные множества называют борелевскими. Полное замыкание получалось тогда, когда проекция плоского борелевского множества на прямую сама являлась борелевским множеством. Желание *именно такого* результата привело к знаменитой ошибке, полностью нивелировавшей идею замкнутости — технологичности. Проекция борелевского множества не всегда оказывалась борелевской — возникли новые множества, за ними другие: «открылась бездна — звезд полна, звездам числа нет — бездне дна»...

Несмотря на такие развороты, наука продолжала генерировать универсальные понятия, целью которых была сглаживание природного ландшафта для продвижения очередной технологии.

Исключительно мощным всплеском этого процесса стало развитие компьютерной индустрии и появление самого термина «информация». И снова все развивалось по уже готовому сценарию. Сначала компьютерные технологии заявили свои права на определенную «экологическую» нишу в обществе традиционных технологий, а потом идея информации стала отвоевывать для них одну область за другой.

Сегодня, похоже, мы наблюдаем глобальное разделение: компьютерные технологии развиваются преимущественно на Западе, Восток (прежде всего, Китай) проявляет интерес к осмыслению самого понятия «информации», разумея при этом среду распространения технологий в том же самом смысле, в каком «материя» явилась средой распространения материальных технологий. Технологии не могут существовать без этой среды — автомобиль без шоссе и дорог обречен ездить только по заводскому двору. Возможно, конфуцианская мудрость уловила этот момент.

К сожалению, компьютерные технологии, особенно в сочетании с технологиями материальными, обладают колоссальными возможностями сделать наш мир подобным антиутопиям Оруэлла. В этом смысле информационные технологии могут оказаться «последними технологиями», а наш мир может вполне приблизиться к черте, за которой начинается Апокалипсис. Многие дальновидные мыслители, предчувствуя такой поворот, стараются выработать линию поведения, отдаляющую неизбежное.

И здесь следует вновь обратиться к опыту математики как наиболее чуткой и определенной дисциплине. Создатель дескриптивной теории множеств Н. Н. Лузин ясно осознавал ее бездонные глубины и ограниченность разнообразных «эффективных» конструкций — его высказывания на эту тему поразительны. Он хорошо осознавал, что объекты реальности (в самом широком ее понимании) имеют особен-

ность быть такими, какими они есть, и постижение их в этом качестве нуждается в особой методологии. Он учил, в частности, что нужно *одновременно* и доказывать утверждение, и отыскивать к нему контр-пример. Именно такое, одновременное движение в противоположных направлениях с *заранее неизвестным исходом* способно «прорваться» к сути самого объекта, по выражению М. Хайдеггера, к его «für sich Sein». Стремительный взлет Московской математической школы был во многом связан с такими прорывами.

Говоря более приближенным к информатике языком, предлагалось наряду с «системой команд», являющихся обязательным атрибутом технологий, строить еще и систему «отказов», т. е. способов действий в ситуации, когда команды перестают работать.

Есть много намеков, что стихийная деятельность по выработке системы «отказов» в самых различных областях (наряду с понятной всем разработкой «системы команд») также осуществляется. Это требует определенного и в ряде случаев значительного отступления от принятых в той или иной дисциплине канонов, действий в слабо формализованном мире, где руководящей нитью является лишь интуиция.

Как представляется, книга профессора Владимира Алексеевича Бубнова — это книга, в которой сделана попытка нащупать такую систему «отказов» в области, общей для информатики и математики. Это достигается самым естественным способом — путем расширения семантического поля вполне технологических понятий и представлений, что сразу дает объемный эффект присутствия информатики и математики в лингвистике, истории, литературе, физиологии и т. д. Границы технологий раздвигаются, образуется единство, присущее самой природе вещей.

Книга В. А. Бубнова во многом напоминает живописное полотно, написанное словами. Вглядываясь в детали, можно увидеть отдельные, вполне сознательно сделанные непривычные ходы (в логике, теории вероятности и др.), которые сглаживаются при едином взгляде на весь текст.

Вероятно, автор не ставил перед собой целей сформировать в каком-либо виде ту систему «отказов», о которой шла речь выше. Однако его книга заставляет думать именно в этом направлении, что делает ее особенно интересной и непохожей на другие книги по информатике.

*Светлой памяти моей жены —
Бубновой Лилии Марковны —
посвящается эта книга*

ВВЕДЕНИЕ

Информатика как новое научное направление определилась в конце 60-х годов прошлого столетия. Считается, что эта научная дисциплина сформировалась на базе кибернетики как науки об управлении биологическими и техническими системами.

Согласно современным представлениям, информатика — это наука о способах хранения, переработки и передачи информации. В рамках такого определения информатика должна использовать определенный математический аппарат, на базе которого создается некоторый набор программных средств, позволяющий человеку взаимодействовать с аппаратными средствами информатики в процессе переработки информации.

Каждая научная дисциплина имеет понятийный аппарат как перечень определенных понятий, с помощью которых формулируются законы и методы данной дисциплины. Для информатики это прежде всего понятие *информации*. К сожалению, для указанного понятия в литературе нет однозначного определения, а в некоторых учебниках даже утверждается, что понятию «информация» невозможно дать определение. Более того, К. Шеннон, один из основателей кибернетики, в своих работах избегает давать определение понятию «информация» и изучает только проблему распространения сигналов различной формы и в различных средах.

В предлагаемой монографии под информацией понимается то, что содержится в символе, а переработка информации связана с выполнением определенных операций над символами. Символы или, в общем случае, наборы символов, содержащие в себе информацию, могут иметь различную природу. Например, к ним можно отнести электрические сигналы, буквы и слова естественного языка, математические числа различных систем счисления и т. д.

Формальные подсчеты объемов «перерабатываемых» символов по своей сути являются количественными измерениями информации.

Переработка информации человеком в процессе познания окружающей действительности в форме слов естественного языка представляет

процесс мышления. Попытки древних исследователей формализовать законы человеческого мышления привели к созданию умозрительной логики.

Математизация умозрительной логики способствовала созданию основ символического исчисления, на базе которого возникла математическая логика. В рамках математической логики оказываются возможными механизация мышления и создание систем искусственного интеллекта.

Перечисленные проблемы раскрываются в данной монографии «Информатика и информация: знаково-символьный аспект», которая отражает содержательную часть дисциплины «Информатика», изучаемой студентами гуманитарных специальностей государственного бюджетного образовательного учреждения ВПО «Московский городской педагогический университет».

ИНФОРМАЦИЯ И ЕЕ ИЗМЕРЕНИЕ

1.1. Понятие «информация»

Слово «информация» вошло в международный обиход от латинского слова «information», что в переводе означает осведомление, просвещение. Такой перевод дает право трактовать понятие «*информация*» как *сообщение, осведомляющее о положении дел, о состоянии чего-либо*.

Указанным толкованием пользовались всегда, когда делались заявления, давались указания или предостережения, касающиеся отношений между людьми. В таких случаях не было нужды долго размышлять, что такое «*информация*», чтобы знать, когда это слово можно употреблять, а когда нет.

Когда же идеи, распространившиеся с развитием кибернетики, стали способствовать применению слова «*информация*» и к ситуациям, до того не рассматривавшимся, возникла потребность в уточнении этого понятия. К числу таких ситуаций следует отнести, например, установление связи между человеком и животным, между человеком и машиной, между машинами или в самом общем случае — между человеком и окружающей его действительностью.

Процесс уточнения понятия «*информация*» привел к различным его толкованиям, среди которых встречаются и такие, в которых данное понятие объясняется с помощью других, имеющих столь же неопределенное значение, например таких, как «*сведения*», «*содержание*», «*данные*» и т. д.

Среди различных толкований понятия «*информация*» встречается такое:

Если наименование некоторого понятия x обозначить T_x , а определяющее его выражение (содержание) через D_x , то информация — это предложение типа T_x есть D_x [45].

Другими словами, информация — это содержание, заключенное в символе, которым обозначается то или иное понятие как назва-

ние определенного объекта. Наиболее легко такое толкование понятия «*информация*» иллюстрируется анализом содержания математических символов.

Общеизвестно, что математические знаки (символы) служат для записи математических понятий, предложений и выкладок. Одними из первых математических знаков были символы для записи количества и порядка, т. е. числа — цифры, возникновение которых предшествовало появлению письменности.

Например, число π — отношение длины окружности к диаметру. В первой строке табл. 1.1 последнее предложение переписано в форме « T_x есть D_x ».

Таблица 1.1

x (понятие)	T_x (наименование понятия)	D_x (выражение понятия)
Число π	π	Отношение длины окружности к диаметру
Функция	$y = f(x)$	Закон, по которому произвольному числу x ставится в соответствие строго определенное число y
Производная функции	$f'(x)$	Предел отношения приращения функции к приращению аргумента, если последний стремится к нулю

Во второй и третьей строках табл. 1.1 приведены предложения « T_x есть D_x » для раскрытия информации, которая содержится в математических символах $f(x)$ и $f'(x)$.

Древнегреческий философ и математик Пифагор утверждал, что весь мир основан на числах и все вещи можно представить в виде чисел. Он также считал, что математические предметы и их начала — причины всего сущего.

Началами он считал числа и числовые пропорции (которые он назвал «*гармониями*»), а сочетания этих начал — элементами, понимая под этим так называемые геометрические элементы.

Идеи Пифагора и его последователей-пифагорейцев о том, что числа представляют собой рациональную сущность вещей или их подлинную природу, оказались плодотворными применительно к простым геометрическим фигурам — квадратам, прямоугольникам, а также к некоторым простым телам.

При изучении сущности геометрических фигур с помощью чисел натурального ряда пифагорейцы открыли так называемые *фигурные числа*. Суть дела в этом случае можно пояснить следующим образом.

Если указать квадрат посредством четырех точек, то его можно интерпретировать как результат добавления трех точек к одной, находящейся в правом нижнем углу. Эти три точки, соединенные отрезками прямой, образуют фигуру — прямой угол (рис. 1.1).

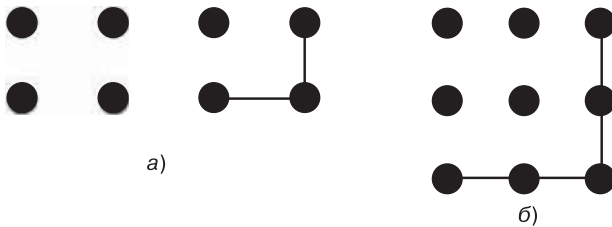


Рис. 1.1

После добавления к этому квадрату другого прямого угла, построенного на пяти точках, получим квадрат большего размера.

Легко заметить, что множество добавленных точек 1, 3, 5, 7, ... образуют прямой угол квадрата, а суммы 1, 1+3, 1+3+5, 1+3+5+7, ... являются квадратами чисел, и что если n (число точек) есть сторона квадрата, то его площадь как общее число точек будет равна n^2 .

Указанные суммы пифагорейцы называли *квадратными числами*.

Таким образом, если, по рассуждениям пифагорейцев, такие геометрические фигуры, как квадраты, состоят из квадратных чисел, то, в соответствии с данными представлениями, квадратные числа как символы содержат информацию о площадях рассматриваемых фигур.

О роли математических знаков и важности точного определения их смысла русский математический гений Николай Иванович Лобачевский написал:

«Подобно тому как дар слова обогащает нас мнениями других, так язык математических знаков служит средством еще более совершенным, более точным и ясным, чтобы один передавал другому понятия, которые он приобрел, истину, которую он постигнул, и зависимость между частями, которую он открыл. Но так как мнения могут казаться ложными от того, что разумеют иначе слова, то всякое суждение в математике останавливается, как скоро перестаем понимать под знаком то, что оно собой представляет» [цит. по:35].






Роль употребления математических знаков отнюдь не сводится к большей краткости символической записи математических предложений по сравнению с их словесным выражением. Только на основе разработанной системы математических знаков стало возможным создание математических «исчислений», в которых математические умозаключения заменяются выкладками, производимыми по определенным формальным правилам.

Более того, использование математической символики способствовало созданию математического языка, который преобладает над естественным языком при описании математических знаний.

Выдающийся русский физиолог и психолог Иван Михайлович Сеченов (1829–1905) при изучении различных форм человеческого мышления определил форму, которую назвал *мышление символами* [61].

Действительно, окружающий мир наполнен различными предметами, которые человек может запоминать с их индивидуальными различиями. Но, в силу подмеченного опытным путем закона регистрации впечатлений по сходству, в человеческой памяти все сходные предметы, по мнению Сеченова, смешиваются в средние итоги. Эти средние продукты мышления не есть точное воспроизведение действительности, но по смыслу они представляют знаки, заменяющие собой множество однородных предметов.

Таблица 1.2

Знак	Изображаемый объект
	Египетский коршун
	Рука
	Глаз
	Дом
	Рот

Такие знаки Сеченов называл *символами первой инстанции*.

С течением веков в так называемых средних итогах того или иного предмета человек научился различать отдельные части данного предмета. Расчленение целого предмета на части и оценку математических

соотношений между частями Сеченов назвал *символами второй инстанции*.

Очевидно, что знаковое письмо древнего человека служит доказательством того, что на данном этапе своего умственного развития человечество мыслило символами второй ступени.

Действительно, в табл. 1.2, заимствованной из [11], приведено несколько примеров египетских иероглифов, которые в некоем обобщенном виде напоминают предметы, сущность которых они обозначают.

Если применительно к табл. 1.2 знак-иероглиф (наименование понятия) обозначить как T_x , а сущность описываемого им объекта — через D_x , то и в данном случае информация это « T_x есть D_x ».

Исследователи египетских иероглифов обнаружили, что кроме знаков, символически изображающих существа или предметы, имеют место знаки, изображающие абстрактные понятия. Например, желая изобразить небо, египтяне рисуют змею, пожирающую свой хвост. При этом тело змеи покрыто пестрой чешуей, потому что чешуйки изображают звезды на небе. Это животное тяжело, как земля, и тихо, как вода; как оно каждый год сбрасывает кожу, так и годичный цикл, совершив полный круг, начинается заново. То, что змея питается своим телом, означает, что все, порождаемое в мире Божественным промыслом, вновь претерпевает умирание.

Приведенный пример описания абстрактного понятия с помощью знаков-иероглифов свидетельствует, во-первых, о неоднозначном толковании и, во-вторых, о субъективном представлении этого и других абстрактных понятий.

Возможно, поэтому совершенствование процесса мышления пошло по пути классификации звуков, с помощью которых произносились иероглифы, и изображению характерных звуков новыми символами — буквами, определенный набор которых составил алфавит естественного языка.

Изучение алфавитов древних языков и анализ трансформирования букв до современного написания позволил многим лингвистам заключить, что любая буква — символ объемных знаний, отражающих принципы построения всего живого и разумного на земле.

Некоторые специалисты с альтернативным взглядом на историю возникновения письменности на Руси классифицируют письмо по следующим формам [66]:

- 1) **х' Арийская Коруна** (союз рун¹) — жреческое письмо, сборник тайных рунических образов;

¹ Руны — это скандинавские знаки на камнях.



Бубнов Владимир Алексеевич — доктор технических наук, профессор, заведующий кафедрой естественнонаучных дисциплин института математики и информатики ГБОУ ВПО «Московский городской педагогический университет».

Области научных исследований: молекулярная физика, гидродинамика, теплофизика, математическая логика, информатика, компьютерная лингвистика. В этих областях им опубликовано около двухсот научных работ и учебных пособий, а также три монографии.