

УДК 002.001; 002:001.8

В.В. Краснопрошин, О.А. Маркова, А.Н. Вальвачев

СИСТЕМА ПОНЯТИЙ В ИНФОРМАТИКЕ

Рассматриваются основные определения информатики: задача, данные, информация, знания, алгоритм, коммуникация, система. Представляется и обосновывается единый комплекс новых определений, построенный на системном подходе и использовании фундаментальных процессов мозга – целеполагания и прогнозирования результата решения. Показывается соответствие новых определений современной парадигме обработки данных и тенденциям развития компьютерных систем.

Введение

Информатика как наука сложилась сравнительно недавно (около 40 лет назад), поэтому в настоящее время она находится в стадии становления. В широком смысле информатика трактуется как область человеческой деятельности, связанная с процессами получения и преобразования данных с помощью компьютеров и других средств вычислительной техники [1]. В зависимости от ситуации информатика рассматривается как совокупность определенных средств преобразования информации, фундаментальная наука, отрасль производства, прикладная дисциплина. Информатика как совокупность средств преобразования информации включает технические средства, программные продукты, математические методы, модели и алгоритмы.

Как и в любой другой науке, в информатике важнейшую роль играет исходная терминология, качество которой со времен Платона и Аристотеля определяется выполнением следующих основных требований:

- 1) естественности (научности) исходных оснований;
- 2) логической иерархии (наследования свойств, системности) от простого к сложному;
- 3) целостности и применимости при решении задач.

В информатике основными понятиями являются: *задача, данные, информация, знания, алгоритм, коммуникация, система* [1]. Данный порядок перечисления правомерен, так как понятие задачи является первоначальным по отношению к другим, а понятие системы определяет решение. Дискуссия о понимании и определении данных терминов началась в XX в. и продолжается сейчас. Предложено множество вариантов для каждого из указанных терминов. Причина многообразия заключается в том, что информатика является молодой прикладной наукой и использует многие определения, взятые из других научных дисциплин: общей теории систем, теории информации, философии и др. Рассмотрим основные определения [1–5] и оценим возможность их применения в качестве понятийной базы информатики:

«Задача – это последовательность шагов или гипотетических операторов, выполнение которых приводит к решению».

«Данные – это отражение окружающего нас мира в знаках и сигналах, принадлежащих конечному алфавиту».

«Информация – это обозначение содержания, черпаемого нами из внешнего мира в процессе нашего приспособления к нему и приведения в соответствии с ним нашего мышления».

«Знание – это форма существования и систематизации результатов познавательной деятельности человека».

«Коммуникация – это процесс, связывающий отдельные части систем друг с другом».

«Система – это комплекс взаимодействующих компонентов».

Приведенные и другие известные определения интересны и полезны, однако они формулировались в разное время, разными авторами и для разных научных дисциплин, поэтому их крайне сложно объединить в логически завершенную систему понятий, ориентированную на информатику и применимую на практике. Использование определений по отдельности также вызывает серьезные проблемы. В частности, определение понятия задачи универсально, но не конструктивно, так как не дает ключа к декомпозиции задачи, облегчающей ее решение. Понятие системы точно и элегантно, но не дает представление об обязательных составляющих сис-

темы в целом или ее частей. В интерпретации понятий данных, информации, знаний наблюдаются вечные разночтения и путаница, что также затрудняет их применение не только при обучении студентов, но и при постановке задач разработчикам компьютерных систем. Главным недостатком упомянутых выше определений является то, что из этих или других существующих определений крайне сложно сложить целостный понятийный каркас информатики, соответствующий всем требованиям и пригодный для решения научных, образовательных и прикладных задач.

В настоящей работе сделана попытка разработки комплекса определений, уточняющих существующие, с целью их применения в качестве понятийной базы информатики на основе свойств человеческого мозга, который рассматривается как эталон эффективности, экономичности и скорости обработки информации.

1. Задача и ее модель

Согласно определению назначение информатики – поиск эффективных механизмов решения задач на базе современных технических средств. Следовательно, в качестве базового понятия целесообразно взять понятие задачи.

Задача является древнейшим и одним из основных общенаучных понятий, широко используемых во всех областях знаний, однако до последнего времени анализ общей природы задач не выходил за уровень интуитивных представлений. Задача априори казалась чем-то понятным и не рассматривалась как объект изучения. В настоящее время ситуация изменилась и исследованием задач занимаются в рамках ряда новых наук: проблемологии, ноологии и мнемологии [3]. Соответственно существует множество определений, приведем самые распространенные из них [2, 3]:

«Задача – это система, обязательными компонентами которой являются: во-первых, материальный или идеальный предмет (предмет задачи), находящийся в некотором исходном состоянии; во-вторых, требование задачи и модель требуемого состояния этого предмета».

«Задача – это более или менее определенные системы информационных процессов, несогласованные или даже противоречивые, отношение между которыми вызывает потребность в преобразовании».

«Задача – это знаковая модель проблемной ситуации, выраженная с помощью знаков естественного или искусственного языка».

Приведенные и другие определения универсальны и достаточно точно описывают задачу с философской точки зрения, однако они ничего не говорят о внутренней природе задачи и не соответствуют требованиям 1 и 3, так как в них отсутствует естественно-научная основа, что затрудняет их практическое применение.

В качестве естественной основы предлагается взять результаты исследований человеческого мозга, полученные физиологами Н.П. Бехтеревой, П.К. Анохиным, П.В. Симоновым и др. Они выделили два фундаментальных процесса, реализуемых мозгом человека при решении любой задачи: *целеполагание* и *прогнозирование возможностей решения* [3, 6–8]. Под целеполаганием понимается постановка задачи, определение цели решения и соответствующих ограничений. Под прогнозированием понимается анализ имеющихся для решения оснований и ресурсов. Благодаря качеству, скорости и экономичности реализации процессов целеполагания и прогнозирования их можно считать эталоном и естественной основой для определения термина задачи.

Определение 1. *Задача – это процессы целеполагания и прогнозирования возможностей решения, реализуемые субъектом для разрешения проблемной ситуации.*

В результате успешной реализации базовых процессов субъект формирует некоторую модель, на основе которой он будет практически реализовывать второй процесс. Модель является неотъемлемой частью решения задачи любым субъектом, так как в противоположном случае в принципе отсутствует возможность выбрать, какие данные и какого типа необходимы для решения. Мозг ежесекундно генерирует множество моделей, получая соответствующие данные и обрабатывая их для принятия решений с целью поддержания гомеостаза живого существа.

Определение 2. *Модель задачи – это результат отображения процессов целеполагания и прогнозирования, релевантный задаче и формализованный в рамках некоторого «решателя» (мозга, компьютера) с помощью некоторого алфавита (химических и биологических структур, групп электрических сигналов, знаков, рисунков и т. д.).*

Определение 2 позволяет построить обобщенную модель задачи:

модель задачи = (среда, субъект, целеполагание, прогноз решения).

Уточним каждый элемент с учетом специфики информатики:

среда = (локальная, глобальная);

субъект = (живое существо, компьютер);

целеполагание = (постановка задачи, цель, ограничения);

прогноз решения = (решатель, данные, алгоритм решения).

Понятие среды существования субъекта введено из чисто практических соображений: иногда только среда определяет возможность решения задачи; в ряде случаев изначально удобно классифицировать задачи в зависимости от среды на локальные и глобальные. В роли субъекта выступает человек, группа людей, животное, насекомое, компьютер. В качестве решателя могут использоваться как собственный механизм субъекта (мозг или его аналог для живых существ, процессор у компьютеров), так и внешний (мозг других людей или другие процессоры). Наличие решателя автоматически предполагает наличие данных, отражающих необходимые для решения задачи свойства предметной области, и алгоритма обработки этих данных для получения решения.

2. Понятия данных, информации, знаний

Понятие данных используется в литературе крайне широко, свободно и разными авторами интерпретируется по-разному, часто противоположно, что вызывает множество проблем у преподавателей и обучаемых, постановщиков задач и работающих с ними программистов. Приведем наиболее употребляемые определения [1–3]:

«Данные – это первоначальные, еще не обработанные результаты эксперимента, зафиксированные в протоколе (например, время выполнения задачи, субъективный отчет испытуемого и т. п.)».

«Данные – это факты, из которых путем анализа и сопоставления можно получить полезную информацию».

«Данные – это набор дифференцирующих признаков (измерений) и их значений».

Приведенные определения говорят о главном: данные – это первичный материал, который отражает нужные для решения задачи свойства предметной области, однако они не дают возможности ответить на ряд существенных вопросов, в частности, почему выбраны одни свойства, а не другие? Соответственно применить эти определения на практике крайне сложно. Для снятия этой проблемы предлагается следующее определение.

Определение 3. *Данные – это конечное множество фактов, полученных в результате наблюдения или эксперимента и структурированных в рамках модели решаемой задачи.*

Определение 3 предполагает наличие даже самой простой модели задачи, в соответствии с которой собираются данные. Отсутствие модели говорит о бесполезности данных, т. е. данные являются функцией от модели задачи:

данные = f_1 (модель задачи).

Подобная однозначность позволяет избавиться от таких интерпретаций данных, как «информационное сырье», «информационный полуфабрикат», «первичная информация» и т. п. Для приведения данных в форму, пригодную к обработке, необходимо их формализовать в соответствии со свойствами решателя. После формализации данные переходят в категорию информации.

Понятие информации введено в первой половине XX в. Л. Хартли и К. Шенноном и относится к числу величайших достижений науки. Само понятие информации относится к фундаментальным понятиям предельного уровня общности и, как многие подобные понятия, не имеет строго определения. Остается раскрыть его содержание, просто объяснив, что оно означает в свете приведенных выше определений.

Определение 4. *Информация – это данные, релевантные модели задачи и формализованные в соответствии со свойствами решателя для обработки алгоритмом.*

Информация представляет собой материал, готовый для обработки «решателями» определенных типов и часто представляющий самостоятельную ценность (коммерческие базы данных). В общем случае информацию можно рассматривать как функцию от данных и решателя:

$$\text{информация} = f_2(\text{данные, решатель}).$$

Здесь есть один не учитываемый многими нюанс – при решении современных задач информация часто носит распределенный характер. Так, мозг человека, решая даже самую простую задачу, задействует глубоко структурированные биохимические механизмы и информацию из различных частей обоих полушарий. Следовательно, в общем случае можно считать, что информация обладает свойством фрагментарности и распределенности. В то же время при решении задач используется целостная информация, т. е. фрагменты информации должны быть перед обработкой объединены в нечто целое. В информатике термин, определяющий полную информацию, авторами не обнаружен, поэтому предлагается подходящий по смыслу термин «результатирующая информация».

Определение 5. *Результатирующая информация – это информация, необходимая и достаточная для решения задачи в рамках данной модели, решателя и алгоритма.*

На основании определения 5 построим следующую модель:

$$\text{результатирующая информация} = f_2([\text{данные}^1, \text{решатель}^1], \dots, [\text{данные}^n, \text{решатель}^n]).$$

Очевидно, что термин «результатирующая информация» носит вспомогательный характер, но он очень полезен на практике, так как устраняет путаницу при разработке программных приложений распределенными по разным странам группами: группы работают с информацией, центр – с результирующей информацией.

Для формирования результирующей информации необходим некоторый коммуникационный механизм, обеспечивающий субъекту доступ к фрагментам информации и их доставку. Термин «коммуникация» (от лат. *communicatio*) появился в научной литературе в начале XX в. Несмотря на длительность исследований, общая теория коммуникации находится в стадии становления, о чем говорит, в частности, наличие более 40 вариантов определений термина «коммуникация» [5]. Приведем три определения, составляющих основу понятия коммуникации в инженерии знаний [1, 5]:

«Под коммуникацией понимается механизм, посредством которого становится возможным существование и развитие человеческих отношений – все символы разума вместе со способами их передачи в пространстве и сохранения во времени».

«Коммуникация представляет собой общение, передачу информации от человека к человеку, от одной системы к другой. Коммуникация также обозначает связь, сообщение, известие, взаимодействие, обмен информацией в обществе, создание и распространение информации, а также средство связи. Кроме того, этот термин используется для обозначения связи любых объектов материального и духовного мира».

«Коммуникация – это процессы перекодировки вербальной информации в невербальную и невербальной в вербальную сферы».

Наличие большого количества различных определений само по себе является серьезной научно-технической проблемой, решение которой вряд ли возможно в ближайшем будущем. В рамках информатики предлагается определение, построенное на основе синтеза сущностей приведенных выше определений.

Определение 6. *Коммуникация – это механизм доступа субъекта к распределенным фрагментам информации с целью формирования результирующей информации.*

Модель коммуникации соответствует традиционной модели Шеннона – Уивера:

коммуникация = (отправитель, получатель, канал связи, сообщение).

Эффективность коммуникации зависит от свойств субъекта и инфраструктуры среды. Наличие результирующей информации позволяет перейти к стадии ее обработки алгоритмом с целью получения решения задачи.

Слово «алгоритм» происходит от «algorithmi» – латинской формы написания фамилии великого математика IX в. Эль-Хорезми, который сформулировал правила выполнения арифметических действий. Определение понятия алгоритма и изучение этого понятия – предмет специальной математической дисциплины – теории алгоритмов, которая, в свою очередь, опирается на аппарат математической логики. Приведем наиболее распространенные определения [1, 3, 9]:

«Алгоритм – это последовательность команд, ведущих к какой-либо цели».

«Алгоритм – это строго определенная процедура, гарантирующая получение результата за конечное число шагов».

«Алгоритм – это правило, вызывающее действия, в результате цепочки которых происходит переход от исходных данных к искомому результату».

Проведя синтез приведенных и сформулированных выше определений, получим следующее определение.

Определение 7. Алгоритм это конечный набор команд для решателя по преобразованию результирующей информации с целью получения решения задачи.

Соответственно определение цели решения задачи является функцией от алгоритма и результирующей информации:

цель = информация = f3 (алгоритм, результирующая информация).

Каждая команда алгоритма содержит точное описание некоторого элементарного действия по преобразованию информации, а также указание на инструкцию, которую необходимо выполнить следующей. В информатике алгоритм формализуется командами программы, а решателем служит процессор компьютера.

Итогом работы алгоритма (решение, цель) является информация, которая отличается от результирующей информации. Если эта новая информация полезна для решения не только текущей, но и других задач, то ее можно интерпретировать как знание.

Определение 8. Знания – это новая информация, полученная решателем в результате применения алгоритма к результирующей информации для решения задачи и, возможно, полезная для решения других задач.

Формально знание не отличается от информации:

знание = f3 (алгоритм, результирующая информация).

Перевести информацию в категорию знания может только человек или специализированная программа, анализирующая результаты решений с помощью соответствующих критериев оценки.

Определение 8 соответствует парадигме добычи знаний Data mining и Knowledge Discovery, реализованных в системах Data Analyzer, ProClarity Analytic Platform и др.

Знания являются важнейшим результатом применения информатики как науки, однако не менее важен и другой аспект – разработка компьютерных систем для их эффективного применения при решении задач и получения новых знаний.

3. Понятия системы

Компьютерные системы являются результатом развития общей теории систем, основы которой заложены в работах Евклида, Платона, Аристотеля. В XX в. они получили развитие в работах А.А. Богданова, Н. Винера, У. Эшби, М. Месаровича, П.К. Анохина и др. Центральной концепцией общей теории систем, кибернетики, системного анализа, всей системологии является понятие системы. Приведем наиболее известные формулировки [4, 8, 9]:

«Система – это комплекс взаимодействующих компонентов или совокупность элементов, находящихся в определенных отношениях друг с другом».

«Система есть любая сущность, концептуальная или физическая, которая состоит из взаимодействующих частей».

«Система – это комплекс избирательно вовлеченных компонентов, у которых взаимодействие и взаимоотношения принимают характер взаимодействия компонентов на получение фокусированного полезного результата».

В целом можно сказать, что термин «система» употребляется, когда речь идет о чем-то собранном вместе, упорядоченном, организованном, однако ни в одном приведенном определении, исключая формулировку П.К. Анохина [10], не упоминается критерий, по которому компоненты собраны, упорядочены, организованы. Кроме того, как отмечает Дж. Клир, термин «система», к сожалению, оказался чрезмерно перегружен и имеет различный смысл при разных обстоятельствах и для разных людей [9].

В информатике систему можно описать следующим образом.

Определение 9. Система – это естественный или искусственный механизм для совместного использования решателя, алгоритма и результирующей информации с целью решения задачи.

Согласно определению 9 можно построить универсальную модель системы:

система = f4 (решатель, коммуникация, алгоритм, результирующая информация).

Если решатель имеет искусственное происхождение, то система называется компьютерной. В случае естественного происхождения решателя система является живым существом.

Если результирующая информация отображает все необходимые для решения задачи свойства предметной области, то такая система называется *детерминированной*. Если хотя бы одно существенное свойство не отображено, то такая система называется *системой с неопределенностью*.

По утверждению В. Гейнса, понятие системы стоит на самом вершине иерархии понятий, поэтому можно считать версию базовых понятий информатики, соответствующих требованиям естественности, системности, полноты и применимости, завершенной.

Заключение

В статье рассмотрены вопросы формирования простого, удобного, логически завершенного комплекса основных определений информатики, построенного на естественно-научной основе. Приведены традиционные определения, показаны их недостатки и возможности модификации в соответствии со спецификой информатики.

В качестве естественно-научной основы новых определений предложено рассматривать фундаментальные процессы целеполагания и прогнозирования, выполняемые человеческим мозгом во время решения задач. Предложены новые определения для понятий задачи, данных, информации, знаний, алгоритма, системы, ориентированные на применение в информатике и представляющие логически законченный комплекс, удовлетворяющий требованиям естественности, системности, полноты и применимости в решении как практических, так и теоретических задач.

Список литературы

1. Гиляревский, Р.С. Основы информатики: курс лекций / Р.С. Гиляревский. – М.: Экзамен, 2003. – 320 с.
2. Пойя, Д. Как решать задачу / Д. Пойя. – М.: Радио и связь, 1961. – 78 с.
3. Фридман, Л.М. Основы проблемологии / Л.М. Фридман. – М.: Синтег, 2001. – 226 с.
4. Эшби, У.Р. Общая теория систем как новая научная дисциплина / У.Р. Эшби // Исследования по общей теории систем: сб. науч. тр. – М.: Наука, 1969. – С. 125–142.
5. Шарков, Ф.И. Теория коммуникаций / Ф.И. Шарков. – СПб.: РИП–Холдинг, 2004. – 245 с.

6. Хазен, А.М. Интеллект как иерархия синтеза информации / А.М. Хазен // Новости искусственного интеллекта. – 1994. – № 1. – С. 71–98.
7. Системные исследования. Ежегодник АН СССР. – М.: Наука, 1970. – 208 с.
8. Месарович, М.Д. Теория систем и биология: точка зрения теоретика / М.Д. Месарович // Системные исследования: сб. науч. тр. – М.: Наука, 1970. – С. 137–163.
9. Klir, G.J. The general system research movement / G.J. Klir // System models for decision making. – Bangkok: Asian institute of technology, 1978. – P. 25–70.
10. Анохин, П.К. Кибернетика функциональных систем / П.К. Анохин. – М.: Медицина, 1998. – 297 с.

Поступила 16.07.07

*Белорусский государственный университет,
Минск, пр. Независимости, 4
e-mail: krasnoproshin@bsu.by*

V.V. Krasnoproshin, O.A. Markova, A.N. Valvachev

CONCEPTUAL FRAMEWORK IN INFORMATICS

The paper presents basic definitions related to informatics: problem, data, information, knowledge, algorithm, communication, system. The paper demonstrates and substantiates a unified set of new definitions based on a system approach and the use of fundamental mechanisms of brain processes, i.e. definition of objectives and prognosis of solution results. The compliance of new definitions to present-day paradigm of data processing and tendencies towards the development of computer systems is presented.