

УДК 004.8:004.6

ИНТЕЛЛЕКТУАЛЬНАЯ ОБРАБОТКА ИНФОРМАЦИИ: АКСИОМАТИЧЕСКИЙ И МЕТРОЛОГИЧЕСКИЙ АСПЕКТЫ

Е. Е. Бизянов, А. А. Гутник

*Донбасский государственный технический университет, г. Алчевск, ЛНР
bpeeecs@gmail.com, hutnik.aa@gmail.com*

В статье проанализированы и уточнены определения для понятий «данные», «информация» и «интеллект» из зарубежных и отечественных источников. На сегодняшний день проблема обработки информации стала особенно актуальна в связи с повышением количества информации, которой оперируют экономические объекты и государственные организации. Были рассмотрены вопросы взаимосвязи понятий «данные», «информация» и «знания», проведен семантический анализ определений и выделены наиболее часто встречающиеся термины. Выявлено, что в понимании термина «данные» многими исследователями достигнут консенсус, в отличие от терминов «информация», «знание» и «интеллект». Проанализированы сложности, связанные с оцениванием интеллекта у компьютерных программ и систем, а также известные критерии оценивания уровня интеллекта. Рассмотрен метрологический аспект оценки интеллектуальности при обработке информации и выдвинуты критерии для такого оценивания. Проанализированы проблемы, возникающие при практической реализации средств для интеллектуальной обработки информации.

Ключевые слова: данные, информация, интеллектуальная обработка информации, критерии интеллектуальности.

INTELLIGENT INFORMATION PROCESSING: THE AXIOMATIC AND METROLOGICAL ASPECTS

E. E. Bizyanov, A. A. Hutnik

*Donbas State Technical University, Alchevsk, Luhansk People's Republic
bpeeecs@gmail.com, hutnik.aa@gmail.com*

The article analyzes and specifies existing definitions for the concepts “data”, “information” and “intelligence” using foreign and domestic sources. Currently the problem of information processing is especially relevant, resulting from the increase of amount of information that economic entities and state organizations are working with. The issues of the interrelationship of the concepts “data”, “information” and “knowledge” are considered. A semantic analysis of the definitions and the most frequently used terms is carried out. Many researchers have reached a consensus with the meaning of the term “data”, in contrast to the terms “information”, “knowledge” and “intelligence”. The complexities associated with the evaluation of intelligence in computer programs and systems are examined, and the criteria of intelligence are analyzed. The metrological aspect of evaluating intelligence during the information processing is considered and criteria for such an evaluation are established. The problems arising during the practical implementation of tools for the intelligent information processing are analyzed.

Keywords: data, information, intelligent information processing, criteria of intelligence.

На сегодняшний день объемы информации, хранящейся в базах данных информационных систем экономических объектов и поисковых систем, растут с огромной скоростью. Учитывая, что информацию нужно не только хранить, но и обрабатывать, используя только ту ее часть, которая необходима в конкретной ситуации, а также явную избыточность хранимой информации, особенно в Интернет [1–3], вопрос эффективной обработки информации становится все более острым.

В настоящее время понятие «обработка информации» закреплено в государственных и международных стандартах и включает ограниченный набор операций: хранение, поиск, анализ, оценка и воспроизведение [4]. К операциям поиска можно отнести фильтрацию, расчет характеристик информации, поиск зависимостей. К операциям обработки информации можно отнести преобразование (например, из одной формы представления в другую) и расчеты с ее использованием [4].

Однако просто выполнения перечисленных выше операций для обработки информации сегодня уже недостаточно. Ученые, инженеры, экономисты и менеджеры нуждаются в информации, которая, с одной стороны, представлена в максимально сжатой форме, а с другой – максимально детализирована. Именно поэтому внимание исследователей в настоящее время привлекает интеллектуальная обработка информации, которую также часто называют «интеллектуальным анализом данных» [4–5].

Целью статьи является анализ аксиоматической базы и формулирование критериев интеллектуальности для методов и моделей, используемых для обработки информации.

Аксиоматическая база

Понятие «интеллектуальная обработка информации» включает такие ключевые слова, как «интеллект» и «информация» («данные»). Дискуссии о том, что является первичным – данные или информация, информация или знания, начались достаточно давно и продолжаются до сих пор [6–7]. Для соединения этих понятий Р. Акофф в 1989 г. предложил модель, которая объединяет понятия «данные», «информация», «знания» и «мудрость» в виде иерархической пирамиды [6]. Она хотя и подверглась критике за то, что якобы упрощает переходы между уровнями пирамиды, не затрагивая при этом глубинных и сложных процессов преобразования и переходов между уровнями [7], однако в наглядной форме отражает концептуальные основы области обработки информации и преобразования ее в знания.

Данные рассматривают как сведения, обстоятельства, служащие для какого-нибудь вывода, решения [8]; факты и идеи, представленные в формализованном виде, позволяющем передавать или обрабатывать эти факты и идеи при помощи некоторого процесса [9]; необработанные зафиксированные результаты эксперимента [10]; представление информации в формализованном виде, за счет чего их можно передавать, обрабатывать и интерпретировать [11–12]; содержимое документов, закодированное для использования на компьютере [13]; наименьшую хранимую единицу информации, связанную определенным образом с каким-либо событием или явлением [13]; необработанные наблюдения о мире, собранные с минимальной контекстуальной интерпретацией [13]; результаты статистических наблюдений [13] и т. п.

Проведенный авторами анализ частоты встречаемости ключевых слов в работах [9, 14–19], а также аналогичный анализ статьи [13], приведенный в [20], показывают, что чаще всего данные связывают с «фактами», «накоплением», «использованием» и «наблюдением». Таким образом, несмотря на некоторые различия, существующие в определениях понятия «данные», сформулированных различными исследователями, есть много общих элементов, это говорит о том, что в понимании термина «данные» консенсус практически достигнут.

Переходя к понятию «информация», отметим, что для него существует более ста определений, используемых в различных отраслях науки и техники [9, 13–15, 20]. Целенаправленные исследования начались еще в XX в. в математике [16], связи [17], лингвистике и философии [21]. Рассмотрим только те определения, которые можно использовать в области обработки информации. Информация – это данные, которые были обработаны и представлены в форме, имеющей смысл для получателя [13]; сообщение, используемое отправителем для представления одной или нескольких концепций, предназначенных для увеличения знаний получателя [13]; упорядоченная коллекция разрозненных и/или несопоставимых данных [13]; любые данные, представленные в электронной форме, написанные на бумаге, высказанные на совещании или находящиеся на любом другом носителе, используемые для принятия решений

[12]; знание, которое уменьшает или устраняет неопределенность в отношении возникновения конкретного события из заданного набора возможных событий [11]. Общим в приведенных определениях является то, что информация – это обработанные данные, представленные в определенной форме (формате) и пригодные для обмена между создателями (генераторами) и потребителями (приемниками) информации, причем создателями и потребителями могут быть как люди, так и машины.

Частотный анализ слов в работах [4, 9, 14–20] показывает, что информацию чаще всего связывают с «данными», «процессами», «смыслом», «сообщением», «получателем», «отправителем» и «системой».

Как отмечалось выше, знание неразрывно связано с информацией и интеллектом. Считается, что знание – это информация, обрабатываемая за счет опыта получателя; совокупность фактов, событий, убеждений, а также правил, организованных для систематического применения [12]; обработанная информация, которая проверена практикой и используется для принятия решений [2]; структурированная и организованная информация, которая развивается внутри когнитивной системы или является частью познавательного наследия человека [13, 18]; средство для осмысленного выбора, которое находится внутри самой системы [13, 19].

Интеллект рассматривается и как средство для создания знаний на основе информации, и как потребитель знаний и информации [23]. Различные исследователи выделяют следующие аспекты интеллекта: способность применять знания и навыки [23]; способность правильно обрабатывать информацию в сложной среде [23]; генератор адаптивного поведения для достижения целей [23]; способность, позволяющая рассуждать, решать вербальные, математические и пространственные задачи, быстро обучаться и усваивать большое количество знаний [24]; способность выполнять анализ и синтез знаний, способность упорядочения данных и знаний с выделением существенных параметров, способность обучения на примерах, способность адаптироваться при изменении фактов и знаний [5].

Естественным желанием человека является желание оценить явления и объекты окружающего мира с использованием различных шкал, критериев, методов. И если данные и информацию можно оценить как количественно, так и качественно, то со знаниями и интеллектом дело обстоит гораздо сложнее. Так, в работах [25–26] предложены критерии для оценки уровня интеллекта: возможность приобретения знаний, способность получать и создавать новые знания, возможность хранения знаний; наличие интерфейса общения с человеком, наличие базы данных для хранения знаний, возможность обмена знаниями с другими интеллектуальными системами, возможность инициации новых знаний [26].

Отметим, что до сих пор в работах по искусственному интеллекту традиционно упоминается тест Тьюринга [27], который должен бы оценивать искусственный интеллект, хотя реально оценивает интеллект создателя компьютерной системы или программы. Кроме того, известны исследования по сравнению интеллекта компьютерных систем и человека, например, в работе [26] проведено оценивание интеллекта человека, суперкомпьютеров и поисковой системы Google на основе критериев приобретения, вывода, хранения и создания знаний. Однако еще И. Кант говорил о невозможности познать объективный мир полностью и непротиворечиво [28], следовательно, объективно оценить интеллект человека или компьютерной системы мы не сможем никогда. Но это не говорит о том, что такие оценки и критерии не нужны. Постоянное совершенствование теории интеллекта неизбежно вызовет появление новых критериев интеллектуальности, а развитие вычислительной техники будет способствовать внедрению в практическую деятельность методов и моделей интеллектуальной обработки данных и интеллектуального управления техническими и экономическими объектами.

Результаты. На основании определений из [9, 11–14] сформулируем определения для понятий «данные» и «информация».

Данные – факты и наблюдения, представленные в формализованном виде, пригодном для передачи, интерпретации и обработки.

Информация – структурированные и упорядоченные данные, представленные в форме, имеющей конкретный смысл для получателя в определенной среде или системе.

На основании определений из [12–13, 18–19] сформулируем определение для понятия «знание» в области обработки информации.

Знание – это совокупность информации и правил ее использования, на основе которых можно получить образ реального объекта, а также предсказать его возможные реакции на возмущения и входную информацию.

На основании определений из [5, 23, 24] сформулируем определение «интеллект».

Интеллект – способность на определенном уровне абстрагирования на основании полученной информации и имеющихся знаний проводить анализ, делать выводы, решать формализованные и неформализованные задачи, ставить задачи, обучаясь при этом на позитивных и негативных примерах из прошлого опыта, адаптируясь к изменениям во внешней среде.

Проведенные исследования терминологии позволяют сформулировать понятие «интеллектуальная обработка информации» с учетом ее реализации в компьютерных программах. Таким образом, интеллектуальная обработка информации – это применение группы методов и моделей, реализованных в виде компьютерных программ для поиска, анализа и оценки информации с использованием предметно-ориентированной базы знаний, включая операции по наполнению и упорядочиванию базы знаний, модификации в зависимости от исходной информации и изменения условий, самообучения и самооценки, сопровождающихся созданием и накоплением новых знаний, а также выделения наиболее существенных свойств объекта.

Метрологический аспект. Исходя из последнего определения, сформулируем критерии для оценки интеллектуальности методов, моделей и программ для обработки информации:

К1. Наличие предметно-ориентированной базы знаний.

К2. Идентификация входной информации (способность определять формат(ы) входной информации).

К3. Самообучаемость и самооценка (способность оценивать положительные и отрицательные результаты, полученные при обработке информации, и запоминать их в базе знаний (при необходимости)).

К4. Декомпозиция (способность выделять подзадачи для достижения цели обработки информации, выбирая при этом метод и модель, наилучшим образом подходящую для решения каждой из подзадач).

К5. Интерпретируемость (необходимость наличия человека как источника знаний в исследуемой предметной области для полной или частичной интерпретации результатов обработки информации).

К6. Модифицируемость используемых методов и моделей обработки (методы и модели обработки информации должны подвергаться постоянному улучшению по мере накопления фактов и знаний).

Проблемы практической реализации. В завершение рассмотрим ряд проблем, с которыми сталкиваются на практике разработчики программного обеспечения для интеллектуальной обработки информации.

Знания, как известно, со временем могут устаревать, аннулироваться, модифицироваться. По мере наполнения базы знаний время, затрачиваемое на поиск, изменение, удаление и упорядочение знаний будет постоянно возрастать. Для решения определенной проблемы человеческий мозг «забывает» ненужные в данный момент факты и знания, а в компьютерных системах управления базами данных используются механизмы индексации и кеширования. Однако компьютер хранит информацию до тех пор, пока мы ее не удалим принудительно. Следовательно, для систем интеллектуальной обработки информации необходимы механизмы, выявляющие устаревшие, давно не использующиеся знания и обеспечивающие их удаление (или архивацию). Конечно, наилучшим способом такой «чистки» было бы автоматическое удаление и обновление, однако на практике это реализовать достаточно сложно, так как требуется разработка новых методов и моделей для оценки знаний и управления ими. Поэтому одним из возможных путей решений могло бы стать наличие человека, по крайней мере, на первых этапах обучения системы интеллектуальной обработки информации.

Другой проблемой является разработка эффективного программного обеспечения для реализации интеллектуальной обработки информации. Увеличение тактовой частоты современных микропроцессоров и увеличение объемов оперативной и дисковой памяти часто создает у разработчиков иллюзию бесконечных и неограниченных возможностей. Однако то, что выглядит в математической записи, как простая формула, например, $\sigma_{Z>1000}(R_1 \cup R_2) \otimes R_3$, а в программе на языке высокого уровня может быть записано в виде оператора SQL, окончательно преобразуется в десятки, сотни байт машинного кода на уровне исполнимого файла. Следовательно, необходимо оптимизировать не только базы знаний, но и механизмы реализации хранящихся в них правил.

Заключение. Проведенный анализ существующей аксиоматической базы в области интеллектуальной обработки информации показал, что на сегодняшний день существует множество определений для понятий «данные», «информация», «знание» и «интеллект», что позволило уточнить указанные определения, а также сформулировать определение для понятия «интеллектуальная обработка информации». Для оценки интеллектуальности методов и моделей обработки информации предложены критерии, которые учитывают основные признаки интеллектуальности. Анализ проблем, которые возникают при практической реализации систем интеллектуальной обработки информации, показал необходимость уделять особое внимание постоянной актуализации знаний и механизмов их обработки.

Литература

1. Никошин Л. А. Избыточность информации в системах электронного документооборота // Электрон. средства и системы управления. 2005. № 2. С. 199–200.
2. Полудина В. П. Информационный шум в интернете как проблема потребления коммуникации // Журн. социологии и социал. антропологии. 2011. Т. 14. № 5. С. 386–394.
3. Смирнов Ю. В. Поисковые системы интернета и методы повышения качества обработки запросов при поиске научной информации в сети // Науч. и технич. б-ки. 2016. № 9. С. 79–87.
4. Кислова О. Интеллектуальный анализ данных: история становления термина // Украин. социолог. журн. 2011. № 1–2. С. 83–94.
5. Финн В. К. Об интеллектуальном анализе данных // Новости искусствен. интеллекта. 2004. № 3. С. 3–17.
6. Ackoff R. L. From data to wisdom // Journal of Applied Systems Analysis. 1989. V.16. P. 3–9.
7. Bernstein J. H. The Data-Information-Knowledge-Wisdom Hierarchy and its Antithesis // Proceedings North American Symposium on Knowledge Organization. 2009. V. 2. P. 68–75.
8. Ушаков Д. Н. Толковый словарь современного русского языка. М. : Аделант, 2014. 800 с.
9. Энциклопедия кибернетики : в 2 т. / отв. ред. В. М. Глушков [и др.]. К. : Гл. ред. УСЭ, 1974. Т. 1. 606 с.
10. Готтсданкер Р. Основы психологического эксперимента. М. : Изд-во Моск. ун-та, 1982. 464 с.
11. ISO/IEC 2382:2015 Information technology. Vocabulary. Geneva : ISO Central Secretariat, 2015.
12. ГОСТ 33707-2016. Информационные технологии. Словарь. М. : Стандартинформ, 2016. 206 с.
13. Zins Ch. Conceptual approaches for defining data, information, and knowledge: Research Articles // Journal of the American Society for Information Science and Technology. 2007. V. 58. I. 4. P. 479–493.
14. Чернавский Д. С. Синергетика и информация. Динамическая теория информации. 3-е изд., доп. М. : Либроком, 2009. 304 с.
15. Аверкин А. Н., Гаазе-Рапопорт М. Г., Поспелов Д. А. Толковый словарь по искусственному интеллекту. М. : Радио и связь, 1992. 256 с.
16. Фишер Р. А. Статистические методы для исследователей. М. : Госстатиздат, 1958. 267 с.

17. Шеннон К. Работы по теории информации в кибернетике. М. : Изд-во иностр. лит., 1963. 830 с.
18. Peirce C. S. Collected papers of Charles Sanders Peirce [Электронный ресурс] C. Harts-horne & P. Weiss (Eds.), V. I–VI. 1931. URL: <https://colorysemiotica.files.wordpress.com/2014/08/peirce-collectedpapers.pdf> (дата обращения: 25.05.2018).
19. Capurro R., Hjørland B. The concept of information // *Annual Review of Information Science and Technology*. 2003. V. 37 (8). P. 343–411.
20. Badia A. Data, information, knowledge: An information science analysis // *Journal of the Association for Information Science and Technology*. 2014. V. 65. I. 6. P. 1279–1287.
21. Шалютин С. М. Искусственный интеллект: гносеологический аспект. М. : Мысль, 1985. 198 с.
22. Калягина Л. В., Разумов П. Е. Категория «данные»: понятие, сущность, подходы к анализу // *Вестн. КрасГАУ*. 2014. № 4. С. 3–8.
23. Legg Sh., Hutter M. A Collection of Definitions of Intelligence // *Frontiers in Artificial Intelligence and Applications*. 2007. V. 157. P. 17–24.
24. Carroll, J. B. Human cognitive abilities: a survey of factor-analytic studies. NY : Academic Press, 1993. 819 p.
25. Лябах Н. Н., Умрихин Н. Г. Методы оценки интеллектуальности компьютерных систем // *Вестн. ун-та. Гос. ун-т управления*. 2013. № 6. С. 54–58.
26. Liu F., Shi Y., Liu Y. Intelligence Quotient and Intelligence Grade of AI // *Annals of Data Science*. 2017. V. 4. I. 2. P. 179–191.
27. Тьюринг А. Может ли машина мыслить? М. : Гос. изд-во физ.-мат. лит., 1960. 67 с.
28. Кант И. Критика чистого разума / пер. с нем. Н. О. Лосского с вариантами пер. на рус. и европ. языки. М. : Наука, 1999. 655 с.