Модели данных и формы представления информации в интеллектуальных системах

С. Т. Главацкий*, И. Г. Бурыкин‡

В современных интеллектуальных системах обработка данных выполняется над цифровым представлением предметной области. При этом для одной и той же предметной области могут использоваться различные формы представления данных. В данной работе рассматривается переход от исходного набора оцифрованных сведений к структуре цифрового представления в контексте классической схемы «Модель – Алгоритм – Программа» А.А. Самарского. Вводится структура цифрового представления как результат организации исходных сведений в соответствии с выбранной моделью. Если X обозначает исходный набор оцифрованных сведений, а R — совокупность отношений и иных структурных связей, задаваемых выбранным способом представления, то структура имеет вид $S = (X, R)$. Рассматривается роль структуры цифрового представления как одного из факторов выбора метода обработки, на основании которого выбирается алгоритм его реализации. Для задач науки о данных рассматривается последовательность «цифровое представление – структура – метод – алгоритм – программа».

Ключевые слова: искусственный интеллект, наука о данных, цифровое представление данных, структура данных, базы данных, метод обработки, алгоритм, математическое моделирование.

1. Постановка задачи

Классическая схема «Модель – Алгоритм – Программа», связанная с работами А.А. Самарского, описывает переход от исследуемого объекта к его компьютерной обработке: строится математическая модель, выбирается алгоритм ее обработки, затем создается программная реализация [1].

Главацкий Сергей Тимофеевич — канд. физ.-мат. наук, доцент, и.о. зав. каф. теоретической информатики мех.-мат. ф-та МГУ, e-mail: sergey.glavatsky@math.msu.ru, ORCID: 0000-0003-1857-6158.

Glavatsky Sergei Timofeevich — Candidate of Physical and Mathematical Sciences, Associate Professor, Acting Head of the Chair of Theoretical Informatics, Faculty of Mechanics and Mathematics, Lomonosov Moscow State University.

‡*Бурыкин Илья Геннадиевич* — старший преподаватель каф. теоретической информатики мех.-мат. ф-та МГУ, e-mail: ilia.burykin@math.msu.ru, ORCID: 0000-0003-1310-3060.

Burykin Ilia Gennadievich — Senior Lecturer, Chair of Theoretical Informatics, Faculty of Mechanics and Mathematics, Lomonosov Moscow State University.

Опубликовано на условиях лицензии CC-BY 4.0
Published under CC-BY 4.0

В задачах науки о данных обработка выполняется над цифровым представлением предметной области. Для одной предметной области могут использоваться различные формы представления. До выбора конкретной модели существует набор оцифрованных сведений, отобранных для решения задачи; выбор модели определяет способ их организации и отношения между ними. В работе рассматривается связь этой структуры с выбором метода обработки и алгоритма его реализации [2–4].

2. Цифровое представление и его структура

Под цифровым представлением предметной области будем понимать совокупность оцифрованных сведений, содержащих необходимые для решения задачи факты об объектах, их свойствах и отношениях. Обозначим множество таких сведений через X. Оно не является конкретной моделью данных: в него входят отобранные сведения, но способ их организации еще не задан. На практике X может формироваться из различных исходных источников: баз данных, в том числе унаследованных, журналов ERP-систем, наборов файлов XLS и других разнородных источников. Исходная организация данных в этих источниках не определяет структуру их представления для решаемой задачи – из источников выделяются необходимые сведения, которые затем включаются в X.

После выделения исходных сведений X выбирается способ их представления. В зависимости от задачи, это может быть реляционная, графовая, векторная или другая модель. Выбранная модель задает организацию элементов X и отношения между ними. Обозначим совокупность отношений и иных структурных связей через R. Тогда структура цифрового представления имеет вид

$$S = (X, R).$$

Здесь X определяет состав представленных сведений, а R — их организацию в выбранной модели.

Различие структур существенно для обработки. Поэтому одна и та же совокупность исходных сведений может создавать различные условия для последующего анализа.

Структура, однако, не определяет метод однозначно. Один и тот же набор данных и одна и та же структура могут использоваться в различных задачах. Поэтому при выборе метода учитываются постановка задачи, содержание данных и структура их представления.

3. Структура, метод и алгоритм

Следующим этапом после определения структуры является выбор метода обработки. Метод задает общий способ решения задачи определенного типа, а алгоритм — конкретную последовательность вычислительных действий, реализующую выбранный метод:

$$\text{структура представления} \to \text{метод} \to \text{алгоритм} \to \text{программа}.$$

Изменение структуры данных может привести к изменению метода их обработки. После выбора метода алгоритм его реализации определяется с учетом размера и свойств данных, требований к памяти и времени выполнения, возможности распараллеливания и характеристик вычислительной среды. Поэтому один и тот же метод может иметь несколько алгоритмических реализаций.

Для современных интеллектуальных систем это особенно существенно из-за больших объемов данных и высокой стоимости вычислений. Развитие foundation models сопровождается разработкой алгоритмических решений, направленных на снижение вычислительных затрат и требований к памяти [5]. Одновременно неадекватное представление может ограничить обработку независимо от эффективности алгоритма [6].

Постановка задачи и содержание данных влияют на выбор метода, а размер и свойства данных и вычислительные ограничения — на выбор алгоритма.

4. Связь с классической схемой

В классической схеме А.А. Самарского математическая модель используется для перехода от исследуемого объекта к его компьютерной обработке [1]. Для задач науки о данных этот переход можно детализировать последовательностью

$$\begin{aligned}\text{предметная область} \to \text{цифровое представление} \to \text{структура} \\ \to \text{метод} \to \text{алгоритм} \to \text{программа}.\end{aligned}$$

Цифровое представление фиксирует необходимые для решения задачи сведения о предметной области, а выбранная модель задает структуру их организации. Метод определяет способ решения задачи, алгоритм — конкретную вычислительную процедуру, а программа реализует ее в вычислительной среде. Такая детализация не изменяет классическую схему, а выделяет этапы, связанные с организацией цифровых данных.

Для задач баз данных и интеллектуального анализа структура данных выступает одним из условий последующей вычислительной обработки.

Актуальность рассмотрения структуры цифрового представления данных соотносится с современными направлениями управления данными [7]. В частности, к ним относятся семантически обогащённый слой данных, составной семантический слой, графы знаний, онтологии и подготовка данных для задач искусственного интеллекта. Эти направления затрагивают различные аспекты организации данных, представления отношений между объектами, формализации смысла и подготовки данных к использованию в интеллектуальных системах. Рассматриваемое в настоящей работе понятие структуры цифрового представления позволяет связать эти аспекты в более общей постановке — как элементы перехода от предметной области к цифровым данным, предназначенным для последующей обработки.

Заключение

Недостаточное внимание к формированию структуры цифрового представления данных может быть одной из причин неэффективности ряда интеллектуальных систем. При разработке таких систем значительное внимание уделяется выбору методов, алгоритмов и вычислительных средств, тогда как организация данных нередко рассматривается как вспомогательный этап. Между тем, именно структура определяет форму, в которой данные поступают на последующую обработку. Она также обеспечивает доступность существенных для задачи свойств данных и отношений между ними.

Исходя из вышеизложенного, формирование структуры цифрового представления следует рассматривать как первостепенный этап построения интеллектуальной системы. Неадекватная задаче структура данных может ограничить возможности их последующей обработки независимо от эффективности выбранного алгоритма. Таким образом, выбор структуры данных должен рассматриваться как часть решения поставленной задачи, а не только как технический этап организации и хранения данных.

Исследование выполнено в рамках государственного задания МГУ имени М.В. Ломоносова.

Список литературы

[1] Б. Н. Четверушкин, А. П. Михайлов, “Триада Самарского. К 100-летию со дня рождения академика А. А. Самарского”, *Вестник Российской академии наук*, **89**:2 (2019), 187–193. DOI: 10.31857/S0869-5873892187-193.

[2] С. Т. Главацкий, И. Г. Бурыкин, “Преподавание фундаментальных основ искусственного интеллекта как реализация концепции нового научного знания”, *Интеллектуальные системы. Теория и приложения*, **26**:1 (2022), 146–152.

[3] С. Т. Главацкий, Р. Р. Айдагулов, В. В. Борисенко, И. Г. Бурыкин, И. Б. Кожухов, В. В. Тензина, А. В. Шокуров, “Математические методы и алгоритмы теоретической информатики в науке о данных и искусственном интеллекте”, Математика в созвездии наук: тезисы докладов, 2024, 604–607.

[4] С. Т. Главацкий, И. Г. Бурыкин, “Совершенствование методики подготовки студентов математических специальностей для решения задач искусственного интеллекта и науки о данных”, Программирование и вычислительная математика: сборник материалов конференции, посвященной 100-летию со дня рождения Н. П. Трифонова, 2025, 50–54. DOI: 10.55959/MSU012316-6-2025-50-54.

[5] M. Xu, D. Cai, W. Yin, S. Wang, X. Jin, X. Liu, “Resource-efficient Algorithms and Systems of Foundation Models: A Survey”, *ACM Computing Surveys*, **57**:5 (2025), 110. DOI: 10.1145/3706418.

[6] E. S. Raymond, *The Cathedral and the Bazaar*, Version 3.0, accessed 14.10.2026, http://www.catb.org/~esr/writings/cathedral-bazaar/cathedral-bazaar/ar01s06.html.

[7] Gartner, Inc., *Hype Cycle for Data Management*, Gartner ID: G00850550. Published 21 July 2026, accessed 14.10.2026, https://www.gartner.com/en/documents/8152229.

Статья поступила 9 октября 2026 г.

Data Models and Forms of Information Representation in Intelligent Systems

S. T. Glavatsky, I. G. Burykin

Modern intelligent systems perform computational processing over a digital representation of a subject domain. The same subject domain can be represented in different forms. This paper considers the transition from an initial set of digitized information to a data structure in the context of the classical “Model – Algorithm – Program” scheme associated with A.A. Samarskii. A digital representation structure is introduced as the result of organizing the initial information according to a selected model. If X denotes the initial set of digitized information and R denotes relations and other structural connections specified by the representation, then the structure is written as $S = (X, R)$. The role of the structure of digital representation as one of the factors in choosing a processing method, on the basis of which an algorithm for its implementation is selected, is considered. For data science problems,

the sequence “digital representation – structure – method – algorithm – program” is considered.

Keywords: artificial intelligence, data science, digital data representation, data structure, databases, processing method, algorithm, mathematical modeling.

References

[1] B. N. Chetverushkin, A. P. Mikhailov, “Triad of Samarskii. The 100th Anniversary of Academician A. A. Samarskii”, *Herald of the Russian Academy of Sciences*, **89**:2 (2019), 187–193. DOI: 10.31857/S0869-5873892187-193 (In Russian).

[2] S. T. Glavatsky, I. G. Burykin, “Teaching the Fundamental Foundations of Artificial Intelligence as an Implementation of the Concept of New Scientific Knowledge”, *Intelligent Systems. Theory and Applications*, **26**:1 (2022), 146–152 (In Russian).

[3] S. T. Glavatsky, R. R. Aidagulov, V. V. Borisenko, I. G. Burykin, I. B. Kozhukhov, V. V. Tenzina, A. V. Shokurov, “Mathematical Methods and Algorithms of Theoretical Informatics in Data Science and Artificial Intelligence”, Mathematics in the Constellation of Sciences: Abstracts of Reports, 2024, 604–607 (In Russian).

[4] S. T. Glavatsky, I. G. Burykin, “Improving the Training Methodology for Students of Mathematical Specialties to Solve Problems of Artificial Intelligence and Data Science”, Programming and Computational Mathematics: Proceedings of the Conference Dedicated to the 100th Anniversary of the Birth of N. P. Trifonov, 2025, 50–54. DOI: 10.55959/MSU012316-6-2025-50-54 (In Russian).

[5] M. Xu, D. Cai, W. Yin, S. Wang, X. Jin, X. Liu, “Resource-efficient Algorithms and Systems of Foundation Models: A Survey”, *ACM Computing Surveys*, **57**:5 (2025), 110. DOI: 10.1145/3706418.

[6] E. S. Raymond, *The Cathedral and the Bazaar*, Version 3.0, accessed 14.10.2026, http://www.catb.org/~esr/writings/cathedral-bazaar/cathedral-bazaar/ar01s06.html.

[7] Gartner, Inc., *Hype Cycle for Data Management*, Gartner ID: G00850550. Published 21 July 2026, accessed 14.10.2026, https://www.gartner.com/en/documents/8152229.

Received on October 9, 2026