ФОРМАЛИЗАЦИЯ СЕМАНТИКИ ИНФОРМАЦИИ

Мухамедиева Дилноз Тулкуновна
Профессор кафедры Цифровых технологий и искусственного интеллекта, Национальный исследовательский университет "Ташкентский институт инженеров ирригации и механизации сельского хозяйства"
Email: dilnoz123@rambler.ru
Orcid ID: 0000-0002-4561-4474
Рустамов Ербол Насимович
Исследователь кафедры Цифровых технологий и искусственного интеллекта, Национальный исследовательский университет "Ташкентский институт инженеров ирригации и механизации сельского хозяйства"
Email: rustamoverbol24@gmail.com
Orcid ID: 0009-0000-1976-0059

Abstract: The article is devoted to the analysis and formalization of the semantic nature of information as a fundamental concept of computer science. Based on logical-semantic, cybernetic, and pragmatic approaches, it examines the structure of information, its relationship with facts, data, and knowledge, as well as the role of context, thesaurus, and the goals of the subject in interpreting informational messages. It is shown that traditional statistical models (Hartley–Shannon) provide only a syntactic assessment of signals, whereas the semantic interpretation of information reveals its meaningful, value-based, and functional essence. The key aspects of information semantics necessary for constructing knowledge bases and intelligent information systems are formalized.

Keywords: facts, data, information, database, knowledge base, logic, supporting set.

Annotatsiya: Maqola informatikaning asosiy tushunchalaridan biri bo‘lgan axborotning semantik tabiatini tahlil qilish va formalashtirishga bag‘ishlangan. Mantiqiy-semantik, kibernetik va pragmatik yondashuvlar asosida axborotning tuzilishi, uning ma’lumot, data va bilim bilan bog‘liqligi, shuningdek, axborotni talqin qilish jarayonida kontekst, tezaurus va subyekt maqsadlarining roli ko‘rib chiqiladi. An’anaviy statistik modellarning (Xartli—Shennon) faqat signallarning sintaktik bahosini berishi, semantik talqin esa axborotning mazmuniy, qadriyatli va funksional mohiyatini ochib berishi ta’kidlanadi. Bilimlar bazasi va intellektual axborot tizimlarini qurish uchun zarur bo‘lgan axborot semantikasi asosiy jihatlari formalashtirilgan.

Kalit so‘zlar: ma’lumot, data, axborot, ma’lumotlar bazasi, bilimlar bazasi, mantiq, tayanch to‘plam.

ISSN 3030-3907

Аннотация. Статья посвящена анализу и формализации семантической природы информации как фундаментального понятия информатики. На основе логико-семантического, кибернетического и прагматического подходов рассматривается структура информации, её отношения со сведениями, данными и знанием, а также роль контекста, тезауруса и целей субъекта в процессе интерпретации информационных сообщений. Показано, что традиционные статистические модели (Хартли—Шеннон) дают только синтаксическую оценку сигналов, тогда как семантическая трактовка информации раскрывает её содержательную, ценностную и функциональную сущность. Формализованы ключевые аспекты семантики информации, необходимые для построения баз знаний и интеллектуальных информационных систем.

Ключевые слова. Сведения, данные, информация, база данных, база знаний, логика, опорное множество.

Введение

Информация является одним из центральных понятий современной информатики, кибернетики и искусственного интеллекта. Несмотря на широкое распространение термина, его содержательная трактовка имеет многозначность и зависит от научной школы и области применения. В обыденной речи понятия «информация», «данные» и «сведения» зачастую смешиваются, однако для задач построения баз знаний, машинного обучения и интеллектуальных систем необходимо строгое семантическое разделение этих сущностей [1].

Работа посвящёна осмыслению природы информации с точки зрения её семантического содержания, роли в процессах познания и управления, а также анализа существующих теоретических моделей. Эти вопросы имеют ключевое значение, поскольку без понимания того, что представляет собой информация, невозможно корректно формализовать процесс порождения знаний [2].

Методы

Если теперь рассмотреть истинные данные $\Delta(x)$ об объекте $x \in X$, то одновременно с каждыми двумя (или более) истинными сведениями $\delta_1(x), \delta_2(x)$, из $\Delta(x)$ следует, что истинными будут и сведения $\delta_3(x)$, где $\delta_3 = \delta_1 \cap \delta_2$ - пересечение подмножеств. В математической логике это означает логическую операцию «и» (конъюнкцию) [3]:

$$\delta_3(x) = \delta_1 \,\&\, \delta_1(x). \qquad (1)$$

ISSN 3030-3907

Определения истинности логических операций «и» и «или» составляют основу процедуры интерпретации $\delta(X)$ в контексте X и являются информацией об объекте, построенной из данных $\Delta(x)$.

Определение 1. Понятие «информация об объекте» $x \in X$ определяется как такое семейство сведений $\delta(x)$ об этом объекте, которому принадлежат, наряду с каждым сведением из этого семейства, все и простейшие (очевидные) логические следствия. Другими словами, если $\delta(x)$ принадлежит семейству, то и всякое $\delta'(x)$ где $\delta < \delta' < X$, также принадлежит семейству, то и $\delta_3(x)$, где $\delta_3 = \delta_1 \cap \delta_2$ также принадлежит семейству. Только, и только в этом случае, если, конечно, $x \in X$, данные $\Delta(x)$ можно воспринимать как «информацию об объекте». Одним из эффективных (в практическом смысле) алгоритмов интерпретации $\Delta(x)$ является табличный метод.

Определение 2. Непустое семейство элементарных сведений о точке x_0 из X назовем элементарной информацией о точке x_0 из X и обозначим $I^x(x_0)$, если выполняются три условия:

1) из $\delta(x_0) \in I^x(x_0)$ следует, что - не пустое подмножество, т.е. $\delta \neq 0$,

2) из $\delta(x_0) \in I^x(x_0)$ следует, что любое более общее истинное сведение $\delta(x_0)$ принадлежит $I^x(x_0)$, т.е. для любого надмножества $\delta \supset \delta$ в X будет $\delta(x_0) \in I^x(x_0)$,

3) из $\delta_1(x_0), \delta_2(x_0) \in I^x(x_0)$ следует, что $(\delta_1(x_0) \& \delta_2(x_0)) \in I^x(x_0)$.

Здесь x выражает контекст интерпретации $\Delta(x)$ в $\Delta^x(x_0)$. При этом, Δ^x - будет носителем $I^x(x_0)$. Под элементарной информацией о точке понимается не просто семейство каких-либо сведений, а такое семейство, к которому относятся сведения вместе с их обобщениями, конъюнкциями и импликациями интерпретируемыми в заданном контексте. Контекст задается в зависимости от решаемой задачи.

Данные преобразуются в информацию различными способами. Рассмотрим некоторые из них;

-*контекстуализация*: известно, для какой цели данные были собраны;

-*категоризация*: известны единицы анализа или ключевые компоненты данных;

-*вычисляемость*: данные могут быть проанализированы математически или статистически;

- *корректировка*: ошибки убраны из данных;

ISSN 3030-3907

- *сжатие*: данные могут быть обобщены в более сжатую форму.

Отметим, что компьютеры могут помочь добавить полезность и трансформировать данные в информацию, но пока они редко могут помочь с контекстом, и человек должен заранее, исходя из сущности решаемой задачи задавать контекст, где интерпретируется информация о *ПО [3]*.

Определение 3. Носителем элементарной информации $I^x(x_0)$ о точке x_0 из X назовем всякое подсемейство $\Delta^x(x_0)$ сведений, входящих в состав $I^x(x_0)$ таких, что для любого сведения $\bar{\delta}(x_0)$ из $I^x(x_0)$ существует менее общее сведение $\delta(x_0)$ из $\Delta^x(x_0)$, т.е. $\delta \supset \bar{\delta}$.

Носитель, состоящий из одного сведения, называется одиночным. Если $\Delta^x(x_0)$, - носитель $I^x(x_0)$, то говорят, что информация $I^x(x_0)$, принесена $\Delta^x(x_0)$. В табличной структуре данных таким носителем является строка, т.е. запись.

Утверждение 1. Всякий носитель $\Delta^x(x_0)$ однозначно определяет принесенную им информацию $I^x(x_0)$, в x.

Доказательство проведем методом от противного. Пусть, кроме $I^x(x_0)$, имеется другая информация $\bar{I}^x(x_0)$ с тем же носителем $\Delta^x(x_0)$. Рассмотрим любое сведение $\bar{\delta}(x_0)$ из $\bar{I}^x(x_0)$, тогда, по определению носителя, существует менее общее сведение $\delta(x_0)$ из $\Delta^x(x_0)$, $\delta \supset \bar{\delta}$. Но $\Delta^x(x_0)$ является носителем $\bar{I}^x(x_0)$ в x и, следовательно, сведение $\delta(x_0)$ принадлежит $I^x(x_0)$. Тогда по определению элементарной информации любое более общее сведение, в частности $\bar{\delta}(x_0)$, т.е. принадлежит $I^x(x_0)$. Тем самым доказано, что $\bar{I}^x(x_0)$ является подсемейством $I^x(x_0)$. Аналогично доказывается, что $I^x(x_0)$, является подсемейством $\bar{I}^x(x_0)$. Следовательно, $I^x(x_0) = \bar{I}^x(x_0)$.

Результаты

Когда решаемая практическая задача рассматривается в информационном пространстве, тогда носитель информации играет существенную роль. Так как при проектировании базы знаний (*БЗ*), именно носитель будет определять, в каком информационном измерении (контексте) интерпретируется $I^x(x_0)$.

Надо отметить, что понятия семантической информации, семантики знаний отсутствуют и в традиционной (виннеровской) кибернетике. Это связано с тем, что в кибернетике бывают ситуации, когда каждая из исследуемых точек множества X характеризуется (если рассмотреть эту точку в информационном пространстве) большим числом взаимосвязанных

ISSN 3030-3907

разнотипных свойств, тогда определить знание об этой точке требует более глубокого (в диалектическом смысле) изучения свойств и идентификации его семантического указателя. Решение указанной проблемы требует привлечения методологии диалектического материализма, взаимодействия теории познания, коммуникации, психологии, лингвистики, знаковых систем, прикладных дисциплин математики и других наук. Актуальность этой проблемы обусловлена как необходимостью решения задачи, определением семантики свойств точки, так и связанных с созданием *БЗ*. Для информационных систем, работающих с семантическими информациями, само создание базы знаний тесно связано с *представлением знаний* для обработки, хранения и передачи [4,5].

Всякий объект обладает определенными свойствами (признаками), проявляющимися при отражении с другими объектами. Всякое свойство объекта проявляется в рамках того или иного контекста. Как уже отметили, всякое сведение должно иметь указатель (имя) объекта, о котором сообщается в данном сведении. Обозначим через x тот объект, который указан в сведении. Тогда x является элементом как опорного множества X, так и подмножества δ, т.е. $x \in X$, $x \in \delta$. В случае табличного представления $\{x_1, x_2, \ldots, x_n\} \in X$ будет выражать строку таблицы (рис.1).

	Информация в контексте x_0					
Информация в контексте δ, I^{δ}	x_0	a_{01}	a_{02}	a_{03}	…	a_{0n}
	x_1	a_{11}	a_{12}	a_{13}	…	a_{1n}
	x_2	a_{21}	a_{22}	a_{23}	…	a_{2n}
	⋮	⋮	⋮	⋮	⋮ ⋮ ⋮	⋮
	x_n	a_{n1}	a_{n2}	a_{n3}	…	a_{mn}

Рис.1. Табличное представление свойства точки $x \in X$, если множество X представляет некую область в пространстве R

Если множество, X представляет некую *ПО*, то тогда каждая клетка области X представляет определенный контекст и описывается набором свойств $\Delta = \{\delta_1, \delta_2, \ldots, \delta_n\}$ или данными об $x \in X$, т.е. $(p)\Delta(x)$ [6-7].

ISSN 3030-3907

Следовательно, каждая клетка области X представляет знание, интерпретируемое в определенном контексте, и описывается набором свойств $\Delta = \{\delta_1, \delta_2, \ldots, \delta_n\}$ или данными об $x \in X$, т.е. $(p)\Delta(x)$. Из рис.3 видно, что информация является характеристической информационной сущностью информационной модели ПО [8-10].

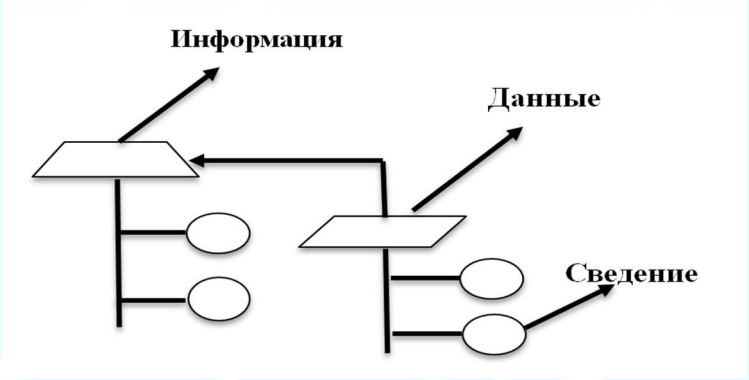


Рис.2. Порождение информации в информационной модели ПО

Исходя из того, что главным в порождении знания из информации являются сведения, имеющие глобальные и локальные свойства. Первая задача, решаемая в работе, — это формализация семантики этих понятий. Вторая задача, решаемая в работе, - разработка алгоритма оценки этих свойств сведений. Оценка этих свойств помогает алгоритмизации проектирования классов для знаний. Третья задача, решаемая в диссертационной работе, — это разработка алгоритмической технологии применения этих свойств при формировании классов для знаний. Четвертая задача, решаемая в работе, связана с построением продукционного знания о *ПО*. Пятая задача, которую необходимо решить, – это создание продукционной *БЗ* для *ПО*, основываясь на фрактальных свойствах знаний. Последняя решаемая задача в работе связана с разработкой алгоритмической технологии решения практических задач, разработанными алгоритмами и методами, дающими возможность коммерциализации разработанных *БЗ* для выбранной предметной области.

Таблица-1

Взаимосвязь сведений, данных и информации

Уровень	Сущность	Характеристика
1	**Сведения**	Локальные свойства объекта
2	**Данные**	Совокупность сведений (глобальные свойства)
3	**Информация**	Интерпретированные данные + логическое замыкание

Заключение

ISSN 3030-3907

В статье рассмотрены фундаментальные понятия сведений, данных и информации, их семантические различия и логические взаимосвязи. На основе формальных определений, предложенных работе, показана роль сведений как элементарных семантических блоков, данных — как совокупности свойств объекта, и информации — как интерпретированных данных, замкнутых относительно логических следствий и обобщений. Детализированы категории локальных и глобальных свойств понятий, введены критерии истинности сведений и способы порождения информации. Показано, что различение этих сущностей является ключевым для построения баз знаний и формализации предметной области. Работа позволяет выстроить строгую семантическую модель сведений, данных и информации.

ЛИТЕРАТУРА

1. Чичкин А.В. Математическая информатика - М.: Наука. Гл.ред. физ. – мат. лит., 1990, - 416 с.
2. Королев М.А. и др. Информационные системы и структура данных. – М.: «Статистика», 1977, 183с., Редько В.Н., Басароб И.А. Базы данных и информационные системы, - М.: Знание, 1987, 32с.
3. Поспелов Д.А., Пушкин В.Н. Мышление и автоматы. М.: Сов. Радио, 1972. -222с
4. Варламов О.О. Эволюционные базы данных и знаний для адаптивного синтеза интеллектуальных систем. Миварное информационное пространство. – М.: Радио и связь, 2002, -288с.
5. Davenport T., Prusak L. Working Knowledge: how organization manage what they know. – Boston: Harvard Business Schol Press, 1998.
6. Попов, Г.А. Метод структуризации отношений и правил логического вывода в базе знаний интеллектуальной системы образовательного назначения / В.С. Лысогорский, Г.А. Попов // VIII Всероссийская научно-техническая конференция «Повышение эффективности методов и средств обработки информации на базе математического моделирования», 26–28 апр. 2006 г.: материалы докладов. – Тамбов: ТВВАИУРЭ, 2006. – Ч. 2. – С. 167-172.,
7. Попов, Г.А. Описание состояния сложного технического объекта с помощью гибридной нейронной сети / Канушкин С.В., Коробков С.П., Попов Г.А., Чернобрывка В.В. // Материалы IXL научной конференции МФТИ «Современные проблемы фундаментальных и прикладных наук», 27–29 нояб. 2006 г.: труды конференции – М: МФТИ, 2006. – Ч. X. – С. 41.
8. Н.Е. Досанов, Е.Н. Рустамов. Фрактальность продукционных знаний. Материалы III Международной научно-практической конференции «Роль

ISSN 3030-3907

инновационных технологий в новом мире» в цели поддержки статьи «Взгляд в будущее: духовное возрождение». I том. / Алматы, 2019. С.41-49.
9. Н.Е. Досанов, Е.Н. Рустамов. ФРАКТАЛЬНОСТЬ ПРОДУКЦИОННЫХ ЗНАНИЙ. Материалы III Международной научно-практической конференции «Роль инновационных технологий в новом мире» в цели поддержки статьи «Взгляд в будущее: духовное возрождение». I том. / Алматы, 2019. С.41-49.
10. Мухамедиева Д.Т., Рустамов Е.Н. АЛГОРИТМ ОБРАБОТКИ ЗНАНИЙ. Проблемы вычислительной и прикладной математики №1(46) 2023, с.88-100.

ISSN 3030-3907