СЕМАНТИЧЕСКИЕ ОСНОВЫ 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 examines the fundamental concepts of computer science — "facts," "data," "information," and "knowledge," which form the basis for designing intelligent systems and knowledge bases. The distinctions between these entities are demonstrated, and formal definitions of the semantics of facts, data, and information are provided. Special attention is given to logical-semantic models, contextual interpretation, and relational dependencies between the attributes of a subject domain. The article also explores aspects of transforming facts into knowledge and the role of these processes in intelligent systems.

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

Annotatsiya: Maqolada intellektual tizimlar va bilimlar bazasini loyihalashning asosi bo‘lgan informatikaning fundamental tushunchalari — "ma’lumot (svedeniya)," "data," "axborot," va "bilim" tushunchalari ko‘rib chiqiladi. Ushbu tushunchalar o‘rtasidagi farqlar yoritilgan hamda ma’lumot, data va axborotning semantik jihatdan rasmiy ta’riflari keltirilgan. Mantiqiy-semantik modellarga, kontekstual talqinga va predmet sohasidagi atributlar o‘rtasidagi relyatsion bog‘lanishlarga alohida e’tibor qaratilgan. Ma’lumotlarning bilimgacha aylanish jarayonlari va ularning intellektual tizimlardagi o‘rni ham tahlil qilingan.

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

Аннотация. В статье рассматриваются фундаментальные понятия информатики — «сведение», «данные», «информация» и «знание» — которые являются основой проектирования интеллектуальных систем и баз знаний. Показаны различия между этими сущностями, приведены

ISSN 3030-3907

формальные определения семантики сведений, данных и информации. Особое внимание уделено логико-семантическим моделям, контекстной интерпретации и реляционным зависимостям между атрибутами предметной области. Рассмотрены аспекты превращения сведений в знания и роль этих процессов в интеллектуальных системах.

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

Введение

Современное развитие искусственного интеллекта (ИИ) и интеллектуальных информационных систем выводит на первый план проблему формализации знаний и построения эффективных баз знаний. Такие системы преобразуют сведения в данные, данные — в информацию, а информацию — в знания, обеспечивая принятие решений, диагностику, прогнозирование, автоматическое управление и другие интеллектуальные функции. Однако глубинные семантические различия между понятиями «сведение», «данные», «информация» и «знание» часто игнорируются. Это приводит к разрывам между теоретическими моделями и практическими процедурами построения баз знаний. Первая глава диссертации посвящена раскрытию природы этих понятий и разработке формальной семантической основы их взаимосвязи. Исторически теория информации развивалась преимущественно в рамках статистического подхода (Хартли, Шеннон), где информация рассматривалась как мера устранения неопределённости. Этот подход, однако, не затрагивает семантический содержательный аспект сведений. Кибернетический подход, предложенный Н. Винером, расширил трактовку информации, включив в неё цель, управление и взаимодействие субъекта с объектом. Согласно Винеру, информация — это содержание, полученное из внешнего мира для адаптации. Таким образом, семантика и ценность информации становятся важнейшими характеристиками. Развитие семантических теорий (Ю.А. Шрейдер, логико-семантические модели) привело к формированию понятия тезауруса, отражающего знания субъекта о мире. Тезаурус определяет способность воспринимать и интерпретировать сведения — то есть фактически определяет уровень и объём информации, получаемой из данных. Информационная сущность в таком понимании зависит от содержания сведений, контекста, в котором они интерпретируются, целей получателя, семантических связей между понятиями, структуризации предметной области.

Методы

Введено определение сведений как триады:

$$(p)\delta(x)(p)$$

ISSN 3030-3907

где X — опорное множество объектов; $\delta \in X$ — подмножество, для которого фиксируется определённое свойство; $x \in \delta$ — семантический указатель объекта; p — контекст или мера достоверности.

Сведение отражает локальное свойство объекта в заданном контексте.

Если о том же объекте фиксируется несколько свойств, получаем данные:

$$\{(p1)\delta 1(x);(p2)\delta 2(x);...\}.$$

Данные выражают глобальные свойства объекта, объединяя множество сведений.

Семантика данных зависит от сущностей предметной области, структуры атрибутов, типа связи между атрибутами (функциональная, транзитивная, нетривиальная), структурированности данных (линейная, табличная, вложенная). Особую роль играют миварные модели (Варламов), где данные представляются через три фундаментальные категории - объект, свойство (атрибут), отношение.

В отличие от данных, информация определяется как данные, интерпретированные в заданном контексте и полезные для решения задачи. Одни и те же данные могут привести к одной и той же информации при различных способах представления, что отражено в примере с адресом: формулировки могут отличаться, но информационное содержание остаётся одинаковым.

Если сведения истинны, то истинны и их обобщения, конъюнкции, логические следствия. Таким образом:

$$\delta_3(x) = \delta_1(x) \wedge \delta_2(x).$$

Информация представляет собой замкнутое семейство сведений, закрытое относительно логических операций. Для табличной БД носителем информации является строка, содержащая минимальный набор сведений, от которых логически выводится вся информация об объекте.

Переход от информации к знанию основывается на учёте глобальных и локальных свойств объекта, корректном выборе контекста интерпретации, структуризации предметной области, установлении причинно-следственных связей между атрибутами, замыкании информации в устойчивую структуру (семейство, класс), применении процедур логического вывода.

Знание — это информация, помещённая в информационное измерение (контекст), где субъект может использовать её для решения задач.

Определение 1. Всякое сведение предполагает свой первоначальный источник. Причем, всякое сведение описывает два множества объектов. Одно – X опорное множество тех объектов, которые являются допустимыми для

ISSN 3030-3907

регистрации приемника. Другое множество δ тех объектов из X, для которых приемником фиксируются вполне определенные свойства. Тем самым, второе множество есть часть первого, т.е. подмножество, $\delta \subset X$.

Далее, всякое сведение должно иметь *указатель* (имя) объекта, о котором сообщается в данном сведении. Обозначим через x тот объект, который указан в сведении, x будет указателем сведения.

Тогда x является элементом как опорного множества X, так и подмножества δ, т.е. $x \in X$, $x \in \delta$. Будем называть указатель объекта *семантическим* указателем сведения. По своей сути эти указатели ассоциируют рассматриваемый контекст.

Семантика всякого сведения предполагает наличие следующих четырех величин: опорного множества X объектов, семантического указателя x одного из объектов X, т.е. $x \in X$, подмножества δ объектов из X, т.е. $\delta \subset X$ и рассматриваемый контекст p, которая характеризует достоверность выполнения главного условия $x \in \delta$ [1].

В связи с этим, всякое сведение об объекте $x \in X$ будем обозначать триадой;

$$(p)\delta(x). \qquad (1)$$

Теперь перейдем к анализу более сложного понятия «данные об объекте». Для этого рассмотрим конкретный пример – прогноз погоды: завтра днем в Ташкенте ожидается солнечная погода, температура воздуха 25^0-28^0 тепла, к вечеру возможна гроза. Эти данные включают несколько (а именно три) сведения типа (1) об одном и том же объекте – дневной погоде в Ташкенте на завтра. Здесь опорное множество X объектов – это различные состояния дневной погоды в различных местах и в различные дни. Семантический указатель x – это дневная погода в Ташкенте на завтра. Здесь имеем три сведения вида (1): $p_1\delta_1(x)$, $p_2\delta_2(x)$, $p_3\delta_3(x)$, где δ_1 означает «солнечная погода», δ_2 - «погода с дневной температурой $25^o - 28^o$ тепла», δ_3 – «погода, когда вечером происходит гроза». Очевидно, что $\delta_1 \subset x$, $\delta_2 \subset x$, $\delta_3 \subset x$. Семантические достоверности здесь p_1, p_2, p_3 – это скалярные достоверности событий $x \in \delta_1, x \in \delta_2, x \in \delta_3$. Наряду с отдельными сведениями об объекте вида (1), имеем дело с наборами (семействами) сведений об одном объекте. Набор триады отражает семантику данных.

Определение 2. Наборы $p_1\delta_1(x)$, $p_2\delta_2(x)$, $p_3\delta_3(x)$ называют данными об объекте. Данные об объекте $x \in X$ обозначаются следующим образом:

$$\{(p_1)\delta_1(x);(p_2)\delta_2(x);...\}. \qquad (2)$$

ISSN 3030-3907

Понятие данные об объекте имеет существенное значение, при проектировании базы знаний (*БЗ*).

Для упрощения записи вместо (2) будем иногда писать

$$\{(p_1)\delta_1;(p_2)\delta_2;...\}(x), \qquad (3)$$

т.е. выносить общий семантический указатель за скобки (справа). Если $p_1 = p_2 = ...$, то можно такую достоверность вынести за скобки (слева):

$$(p)\{\delta_1;\delta_2;...\}(x) = (p)\Delta(x), \qquad (4)$$

здесь

$$\Delta = \{\delta_1;\delta_2;...\} = \begin{cases}\delta_1\\ \delta_2\\ ...\end{cases}. \qquad (5)$$

Формула (4) выражает процедуру *интерпретации* сведений в контексте p. Данные (набор сведений) о свойствах объекта в базе данных (*БД*) представляются в зависимости от типа *ПО*. Это связано с объективностью отражения свойств объекта в *БД*.

Из определения 1 вытекает очень интересное свойство $(p)\delta(x)$. Если тройка $(p)\delta(x)$ выражает локальные свойства сведений, то набор $\{(p_1)\delta_1;(p_2)\delta_2;...\}(x)$ выражает соответственно глобальные свойства $\delta(x)$. Из этого следует, информация и знание будут иметь такие же свойства. Такое заключение играет существенную роль при решении практических задач.

Как известно, что данные разделяются на линейные и нелинейные структуры. К линейным структурам относятся последовательные и строчные структуры данных. Элементы *последовательной* структуры данных размешаются в том порядке, который требуется при их обработке. Они обычно занимают последовательные участки памяти, без каких–либо промежутков. Порядок обработки элементов в ряде случаев не совпадает с порядком их следования при формировании последовательной структуры данных. Чаще всего требуется, чтобы значения *важных признаков* монотонно возрастали или монотонно убывали от элемента к элементу. Такие последовательные структуры данных называются *упорядоченными.* Если это условие не соблюдается, последовательная структура данных считается *неупорядоченной.* На практике, строится упорядоченность данных, так как это дает выигрыш во времени их обработки [2]. На рис.1 показана общая структура данных.

Очень интересное представление приведено в работе [2], названное миварная концепция представления данных. Семантика этой концепции заключается в следующем. Реально существует некоторая предметная

ISSN 3030-3907

область, а определенная познающая система изучает и моделирует данную область. Полученные сведения о предметной области накапливаются в некоторой системе представления данных. Для решения этой задачи вводятся три равнозначных и равно зависимых понятия: вещь (объект, сущность); свойство (атрибут, характеристика); отношение (связь, взаимодействие). Далее говорится, что этих трех выделенных понятий – категорий достаточно для представления любой информации о любой предметной области. Из рис.1 видно, что атрибут – это сведение, а данные – обозначающая информационная сущность информационной модели ПО.

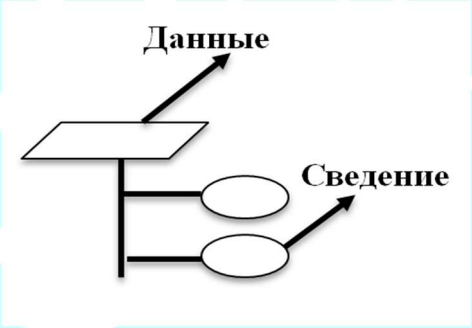


Рис.1 Семантика образования данных

Атрибуты соответствующих сущностей обозначим следующим образом. Например, первый атрибут стержневой сущности - A^1_{CC} и т.д.

Атрибуты A^1_{CC} обозначающей сущности Aoc функционально зависят от атрибута Axc, если в каждый момент времени каждое значение Aoc не имеет более одного значения в Axc, соответственно с ним посредством $O_C \to X_C$. Такие определения легко расщиряются до набора атрибутов. Будем писать $A^i_{0C} \to A^i_{0C}$, если эти атрибуты функционально зависят в отношении $O_C \to X_C$ и $A^i_{0C} \to A^j_{XC}$, если такой зависимости нет. Если одновременно имеют место зависимости $A^i_{0C} \to A^j_{XC}$ и $A^j_{XC} \to A^i_{0C}$, то $j = i$, следовательно будем писать $O_C \to X_C$, устанавливая факт взаимно однозначного соответствия. Функциональная зависимость $A^i_{0C} \to A^j_{XC}$, где $A^i_{0C} \supset A^j_{0C}$ называется простой зависимостью, в противном случае сложной зависимостью (Рис. 2).

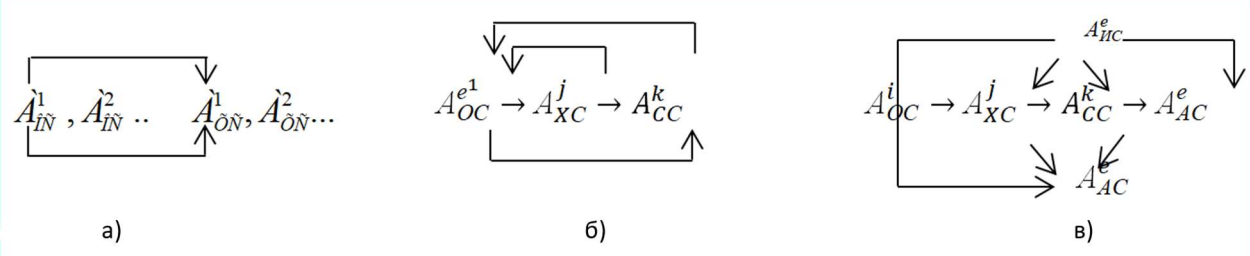


Рис.2. Основные меж-атрибутивные зависимости информационных сущностей, а) функциональная б) транзитивная в) нетривиальная.

ISSN 3030-3907

Предположим, что для трех различных наборов атрибутов $A^i_{XC}, A^j_{0C}, A^e_{CC}$ информационных сущностей выполняются следующие условия:

$$A^i_{XC} \to A^j_{0C};\ A^j_{0C} \to A^e_{CC};\ A^j_{0C} \to A^i_{XC}\ .$$

Отсюда следует, что $A^i_{XC} \to A^e_{CC}$ и $A^e_{CC} \to A^i_{XC}$. При этом функциональная зависимость $A^e_{CC} \to A^j_{OC}$ не запрещается и не требуется. Тогда A^e_{CC} транзитивно зависит от A^j_{XC} в данной информационной совокупности. Если $A^e_{CC} \to A^j_{CC}$ то $A^j_{OC} \to A^e_{CC}$ транзитивно зависит от A^i_{XC} в информационной совокупности (Рис.2б). Те атрибутивные зависимости, не подчиняющиеся этим требованиям, относятся к нетривиальным зависимостям. (Рис.2в).

Эти описанные зависимости, в принципе определяют превращения понятий «сведение» → «данные» → «информация» → «знание» друг в друга.

Сегодняшний этап человеческой цивилизации определяется всеобщей цифровизацией и интеллектуализацией социально–экономической жизни общества. В свою очередь, все это зависит от степени автоматизации обработки, хранения, управления и представления данных, информации и знаний.

Определение 3. Семантика информации – это, данные, интерпретированные в определенном контексте p_1 (необходимые пользователю, полезные для решения рассматриваемой задачи)

В дискуссиях по проблемам интеллектуальных ресурсов обсуждаются различные точки зрения на эти широко используемые понятия. Теория информации определяет информацию как меру неопределенности, устраняемую сведением. В информационных технологиях под информацией понимаются данные, которые могут быть зафиксированы приборами или людьми, обработанные и представленные в той или иной форме. В коммуникационном маркетинге информацию определяют как содержание сообщения. То есть сфера применения понятия налагает некоторый существенный отпечаток на определение понятия *информация*.

3. Результаты. Совпадают ли между собой понятия «данные об объекте» и «информация об объекте?» Оказывается, эти понятия разные. В самом деле, можно привести примеры разных данных [4].

Рассмотрим адрес некоторого место жительства x:

$$\Delta(x) = \left\{ \begin{array}{l} \delta_1 - г..Ташкент \\ \delta_2 - ул..Ч.Айтматова \\ \delta_3 - дом№26 \\ \delta_4 - квартира\ \ 50 \end{array} \right\}(X). \qquad (6)$$

ISSN 3030-3907

Здесь опорное множество X – все местожительства в Узбекистане, δ_1 – подмножество местожительств города Ташкента и т.п. Теперь рассмотрим другое описание того же местожительства $x \in X$ **:**

$$\hat{\Delta}(x) = \hat{\Delta}(x) = \left\{ \begin{array}{l} \delta_1 - Ташкент \\ \delta_2 - к \text{ западу } от \text{ Гидромедцентра} \\ \delta_3 - ближайшее \text{ кирпичное 4-этажное здание} \\ \delta_4 - четвертый\ этаж \\ \delta_5 - 3\ подъезд \end{array} \right\}(X), \quad (7)$$

Здесь данные $\Delta(x)$ и $\hat{\Delta}(x)$ – разные, но они дают одну и ту же информацию о местожительстве. Аналогичная ситуация появляется, когда, например, одно и тоже действительное число X мы представляем в двоичной, десятичной или какой-либо другой системе исчисления. И так, «данные об объекте» и «информация об объекте» — это разные понятия. «Данные об объекте» являются как бы носителем «информации об объекте». Одна и та же информация может быть принесена разными данными, но каждые данные определяют свою единственную информацию [5-6].

Из определения 3 выходит, что в отличие от данных, информация имеет *смысл*, *соответствие* и *цель*. Информация передается по объектам посредством жестких (формальных) и мягких (неформальных) сетей. К первым относятся электронные письма и документы, традиционные письма и документы, посылки и пакеты Интернет-сети. Мягкая сеть менее видима и устойчива. Она возникает от случая к случаю во время встреч, дискуссий и т.п. К количественным мерам информации можно, кроме объема, отнести связность и число обменов. К качественным мерам – информативность и полезность.

Если рассмотреть одно истинное сведение $\delta(x)$ об объекте $x \in X$, то вместе с ним одновременно можно всегда сформулировать другие истинные сведения, являющиеся простейшими (очевидными) логическими следствиями из $\delta(x)$. Например, пусть x - конкретный человек, а $\delta(x)$ - сведение о нем, означающее «человек x короче 1.8 метров». Тогда, очевидно, будут справедливы такие сведения как $\delta'(x)$, означающее «человек x короче пяти метров» и т.д. Другими словами, если имеется одно истинное сведение $\delta(x)$, то вместе с ним истинными будут всякие другие сведения $\delta'(x)$, где δ' - над множество δ, т.е. $\delta \subset \delta' \subset X$. Что интересно, в этом примере надо заметить, когда мы определяем начальное истинное сведение $\delta(x)$ об объекте, молча допускаем множество X. На самом деле, в общем случае, X определяет контекст, в котором мы *интерпретируем* (т.е. находим)

ISSN 3030-3907

начальное истинное сведение $\delta(x)$ $x \in X$, и зависящее или вытекающее из этого сведения $\delta'(x)$, истинное сведение $\delta(x)$. Если ввести определенный порядок в процедуру *интерпретации* (или определим контекст), то мы поймем алгоритм *превращения данных в информацию [7-8].*

Заключение

В работе раскрыты ключевые понятия информатики и знания: «сведение», «данные», «информация», «знание». Представлена их формальная семантическая основа, необходимая для проектирования баз знаний и построения интеллектуальных систем. Показано фундаментальное различие между сведениями, данными и информацией. Введены строгие определения и семантические модели сведений и информации. Рассмотрены функциональные, транзитивные и нетривиальные зависимости атрибутов. Показан процесс перехода от данных к информации и далее к знаниям. Определена роль контекста и логических операций в интерпретации информации.

Эти результаты являются методологической основой для дальнейших глав диссертации, посвящённых построению продукционных знаний, оценке их свойств и разработке алгоритмов формирования базы знаний.

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. Варламов О.О. Эволюционные базы данных и знаний для адаптивного синтеза интеллектуальных систем. Миварное информационное пространство. – М.: Радио и связь, 2002, -288с.
4. Davenport T., Prusak L. Working Knowledge: how organization manage what they know. – Boston: Harvard Business Schol Press, 1998.
5. 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.
6. 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.

ISSN 3030-3907

7. 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.
8. Мухамедиева Д.Т., Рустамов Е.Н. АЛГОРИТМ ОБРАБОТКИ ЗНАНИЙ. Проблемы вычислительной и прикладной математики №1(46) 2023, с.88-100.

ISSN 3030-3907