

О НЕКОТОРЫХ ВОПРОСАХ ИДЕНТИФИКАЦИИ ЗНАНИЙ В
ИНТЕЛЛЕКТУАЛЬНЫХ СИСТЕМАХ

Булкин В.И., Шаронова Н.В.

Постановка проблемы в общем виде и ее связь с важными научными результатами. Анализ современных литературных данных свидетельствует, что понятие «знание» определяется в разных источниках по-разному, но большинство авторов соглашались с тем, что знания имеют ряд особенностей по сравнению с данными. К этим особенностям относятся:

- внутренняя интерпретируемость;
- структурированность;
- связность;
- семантическая метрика;
- активность.

В настоящее время из существующих моделей представления знаний наиболее популярны *логические, сетевые, продукционные, фреймовые модели* [2]. В рассматриваемой задаче идентификации знаний в интеллектуальных системах предметной областью обычно называется множество предметов и процессов, которые составляют основу необходимой для решения задачи обработки информации.

Общеизвестно, что языки, предназначенные для описания предметных областей, называются языками представления знаний (ЯПЗ), при этом считается, что универсальным языком представления знаний является естественный язык. Однако использовать его в системах машинного представления знаний сложно, так как он тяжело поддается формализации из-за таких присущих ему явлений, как нерегулярности, полисемия, омонимия, пресуппозиция и т.д. Главным препятствием является отсутствие формализации семантики естественного языка, которая имела бы достаточно эффективную операционную поддержку.

Анализ последних исследований и публикаций, в которых описываются попытки решения подобных задач, показывает, что для представления знаний в информационных системах используются различные формальные языки и, в частности, язык исчисления предикатов. Он имеет однозначную формальную семантику и операционную поддержку в виде совершенного механизма вывода. Кроме того, для описания различных предметных областей предлагается использовать как классические исчисления предикатов, так и различные его модификации, например, многосортное исчисление предикатов первой степени [2].

Постановка задачи. Для представления отношений желательно использовать язык уравнений, т.е. некоторую алгебраическую систему. Преимущество уравнений заключается в том, что с их помощью можно описывать любые отношения на множестве элементов предметной области. Все переменные в уравнении являются равноправными и любые из них могут выступать как в роли независимых, так и в роли зависимых переменных. Алгоритмы и созданные на их основе программы описывают функционирование систем со входами и выходами, а уравнения алгебры описывают функционирование многополюсных систем. В первом случае входы и выходы строго зафиксированы, а во втором случае входные сигналы могут подаваться на любые входы, а выходные сигналы могут сниматься также с любых оставшихся полюсов [3].

В качестве универсального математического языка для описания дискретных, детерминированных и конечных объектов, систем и процессов будем использовать алгебру ко-

нечных предикатов [3 – 6], хорошо зарекомендовавшую себя для решения задач обработки текстовой информации.

Любая алгебра представляет собой совокупность множества M с заданными в нем операциями

$$S = \{f_{11}, f_{12}, \dots, f_{1n}, f_{21}, f_{22}, \dots, f_{2n}, \dots, f_{m1}, f_{m2}, \dots, f_{mn}\},$$

где $A = \langle M, S \rangle$, где M – носитель, а S – сигнатура алгебры [1].

Алгебра предикатов в качестве носителя использует множество U с заданными на нем операциями \vee - дизъюнкции, \wedge - конъюнкции и $\bar{}$ - отрицания [5]. В алгебре предикатов используются следующие формулы:

- символы 0 и 1 называются формулами;

- выражения вида $a_i^{x_j}$ где, $i = 1, k, j = 1, n$ называются формулами;

- если выражения A и B – формулы, то выражения $(A \vee B)$, $(A \wedge B)$ тоже формулы.

Предикатом, заданным на U^n , называется любая функция $t = P(x_1, x_2, \dots, x_n)$, отображающая множество U^n в множество $\Sigma = \{0, 1\}$, где символы 0 и 1 являются логическими булевыми элементами. Элемент 0 имеет логическое значение ложь (False) и может обозначаться латинской буквой F . Элемент 1 имеет логическое значение истина (True) и может быть обозначен латинской буквой T . Переменные x_1, x_2, \dots, x_n называются предметными переменными, а их значения называются предметами. Показатель n указывает на степень предиката. Если $n = 1$, то предикат называется унарным, если $n = 2$, то бинарным, если $n = 3$, – тернарным. При произвольном n предикат называют n -арным. Если множество U конечно, то предикат называется конечным, иначе бесконечным.

Дизъюнкцией, или логическим сложением предикатов P и Q , называется предикат $P \vee Q$, значения которого при любых $x \in U$ определяется по формуле

$$(P \vee Q)(x_1, x_2, \dots, x_n) = P(x_1, x_2, \dots, x_n) \vee Q(x_1, x_2, \dots, x_n).$$

Конъюнкцией, или логическим умножением предикатов P и Q , называется предикат, значение которого определяется по формуле

$$(P \wedge Q)(x_1, x_2, \dots, x_n) = P(x_1, x_2, \dots, x_n) \wedge Q(x_1, x_2, \dots, x_n).$$

Отрицанием предиката P называется предикат \bar{P} со значениями, определяемыми в соответствии с формулой

$$\overline{(P)(x_1, x_2, \dots, x_n)} = \overline{P(x_1, x_2, \dots, x_n)}.$$

Множество всех n -арных предикатов, заданных над множеством U^n , на котором определены операции дизъюнкция (\vee), конъюнкция (\wedge) и отрицание ($\bar{}$) предикатов, называется алгеброй n -арных предикатов [5]. Операции дизъюнкции, конъюнкции и отрицания являются базисными для алгебры предикатов.

Предикаты вида

$$x_i^a = \begin{cases} 1, & \text{если } x_i = a \\ 0, & \text{если } x_i \neq a, \text{ где } i = \overline{1, n} \end{cases}$$

называются базисными для алгебры предикатов [4], где a – любой элемент универсума U . Алгебра предикатов полна в том смысле, что любой предикат можно представить в виде суперпозиции базисных операций, примененных к базисным предикатам x_i^a . Предикат x_i^a называется узнаванием (идентификатором) предмета a по переменной x_i . Для алгебры предикатов справедливы тождества:

- закон истинности

$$\bigvee_{a \in U} x_i^a = 1,$$

- закон ложности

$$x_i^a x_i^b = 1,$$

- закон отрицания

$$\overline{x_i^a} = \bigvee_{\substack{b \in U \\ b \neq a}} x_i^b.$$

Для конечного универсума $U = \{a_1, a_2, \dots, a_k\}$ тождества алгебры предикатов можно записать следующим образом:

- закон истинности

$$x_i^{a_1} \vee x_i^{a_2} \vee \dots \vee x_i^{a_k} = 1, \text{ где } i \in \{1, 2, \dots, n\},$$

- закон ложности

$$x_i^{a_j} x_i^{a_l} = 0, \text{ где } i \in \{1, 2, \dots, n\}, j, l \in \{1, 2, \dots, k\},$$

- закон отрицания

$$\overline{x_i^{a_j}} = x_i^{a_1} \vee x_i^{a_2} \vee \dots \vee x_i^{a_{j-1}} \vee x_i^{a_{j+1}} \vee \dots \vee x_i^{a_k}, \text{ где } i \in \{1, 2, \dots, n\}, j \in \{1, 2, \dots, k\}.$$

С помощью формул алгебры предикатов можно математически описывать любое отношение, заданное на множестве U^n .

Пусть R – отношение на U^n , тогда предикат P , значения которого вычисляются по правилу

$$P(x_1, x_2, \dots, x_n) = \begin{cases} 1, & \text{если } (x_1, x_2, \dots, x_n) \in R \\ 0, & \text{если } (x_1, x_2, \dots, x_n) \notin R \end{cases}$$

называется предикатом – идентификатором отношения R . Каждому отношению R соответствует свой предикат P на U^n и наоборот. Каждому предикату P можно поставить в соответствие отношения R на U^n

Любой предикат P , заданный на множестве U^n , можно записать с помощью следующей формулы алгебры предикатов:

$$P(x_1, x_2, \dots, x_n) = \bigvee_{a_1, a_2, \dots, a_n \in R} x_1^{a_1} x_2^{a_2} \dots x_n^{a_n},$$

которая называется совершенной дизъюнктивной нормальной формой предиката Р (СДНФ предиката). При этом ведется логическое суммирование по всем наборам предметов (a_1, a_2, \dots, a_n) , входящим в состав отношения R, которое задается предикатом Р.

Использование алгебры предикатов компенсирует необходимость использования других языков представления знаний (ЯПЗ). При использовании алгебры предикатов объекты и отношения во всех моделях представления знаний записываются в виде уравнений алгебры предикатов. Системы предикатных уравнений могут решаться с помощью универсального решателя, который представляет собой программу, написанную на некотором алгоритмическом языке высокого уровня. Кроме того, любое уравнение алгебры предикатов может быть представлено в виде переключательной цепи, что предоставляет возможность сконструировать процессор представления знаний (ППЗ) из комбинаций таких цепей [4].

Покажем на конкретных примерах, как можно представлять знания, используя алгебру предикатов. В логических моделях для представления знаний широко используются различные виды исчислений предикатов – классическое, многосортное, многоуровневое исчисление предикатов, псевдофизические логики и т.п.

В классическом исчислении предикатов отношение "z является суммой x и y", где z, x, y – целые неотрицательные числа $(x, y, z \in \{0, 1, 2, \dots, n\})$, записывается в виде предиката

$$P(x, y, z) = 1.$$

Для нахождения конкретных реализаций этого предиката используется один из языков программирования, например, Пролог. Средствами алгебры предикатов это же отношение может быть задано предикатом

$$S(x, y, z) \equiv \bigvee_{\substack{i=0 \\ j=0}}^n \bigvee_{\substack{i=0 \\ j=0}}^i x^i y^{i-j} z^i.$$

Уравнение

$$S(x, y, z) = 1$$

реализует отношение "число z равно сумме чисел x и y"[4]. Если решить это уравнение относительно переменной z, то получим

$$z^i = x^0 y^i \vee x^1 y^{i-1} \vee \dots \vee x^i y^0 \quad (0 \leq i \leq n).$$

Подставляя в это уравнение конкретные значения i, получаем реализацию предиката S(x, y, z)

$$\begin{aligned} z^0 &= x^0 y^0, \\ z^1 &= x^0 y^1 \vee x^1 y^0, \\ z^2 &= x^0 y^2 \vee x^1 y^1 \vee x^2 y^0, \\ z^3 &= x^0 y^3 \vee x^1 y^2 \vee x^2 y^1 \vee x^3 y^0, \\ &\dots \end{aligned}$$

Если внимательно проанализировать общее уравнение и его конкретные реализации, то обнаруживается, что эти уравнения не только реализуют отношение "z является суммой x и y", но и фиксируют знание о том, какие комбинации целых неотрицательных чисел x и y дают в сумме заданное число z.

Любое из вышеприведенных уравнений можно представить в виде переключательной цепи [3], которая представляет собой многополюсник (рис.1), имеющий два входа и один выход. Данная схема представляет собой переключательную цепь, состоящую из 3-х ступеней. На первой ступени расположены схемы узнавания (идентификации) чисел x и y , реализующие базисные предикаты вида $t = x^i$, $t = y^i$ ($0 \leq i \leq n$), где t - логическая переменная. Будем называть их предметными идентификаторами (ПИ). На вход этих схем подаются значения чисел x и y , а сигнал 1 формируется на выходе той схемы, которая его идентифицирует (узнает). На выходе остальных идентификаторов формируется сигнал 0. На второй ступени переключательной цепи находится схема совпадения (конъюнкции) на два входа каждая. Количество этих схем совпадает со значением $i+1$. Третья ступень нашей переключательной цепи реализована схемой разделения (дизъюнкции) и имеет $i+1$ вход и один выход.

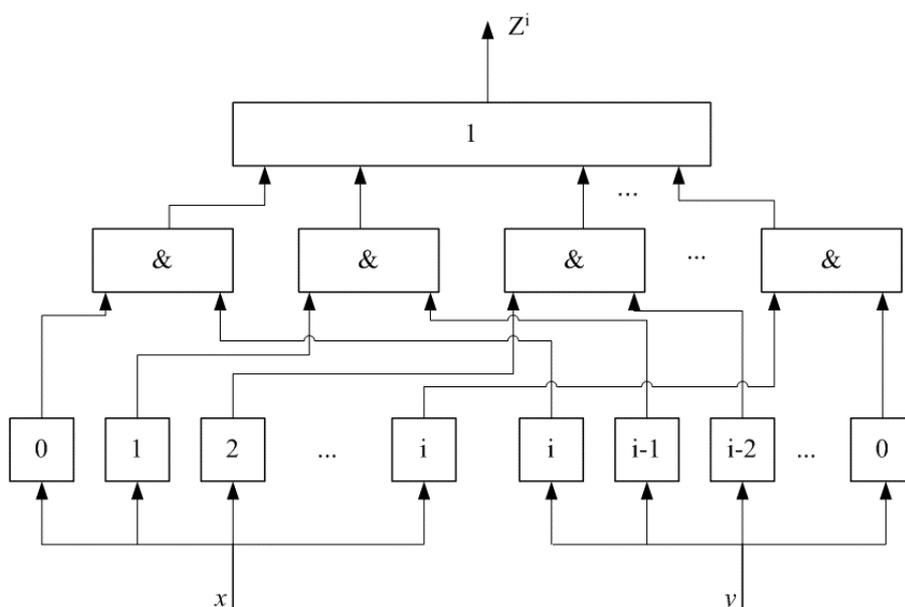


Рис. 1

Схема работает следующим образом. Если на вход многополюсника подаются два числа x и y , то на выходе сигнал 1 появляется только в том случае, если сумма этих чисел равна i . Например, для $i = 3$ сигнал 1 на выходе схемы будет появляться лишь в том случае, если на вход будут подаваться следующие пары чисел

- $x = 3, y = 0;$
- $x = 2, y = 1;$
- $x = 1, y = 2;$
- $x = 0, y = 3.$

При подаче на входы переключательной цепи любых других комбинаций ($x = 3, y = 1$, например) на выходе цепи будет сформирован сигнал 0.

Рассмотрим теперь пример использования алгебры предикатов для описания предметной области в терминах многосортной логики [2]. В многосортной логике универсум U разбивается на подмножества или, как их называют, сорта, которые отвечают классам сущностей. В отличие от классической логики, кванторы всеобщности и существования ограничиваются областью определения сорта (класса сущностей). Следовательно, прежде чем описывать предметную область, необходимо разбить универсум U на классы эквивалентностей. В многосортной логике это осуществляется простым описанием имен сортов,

например, Деталь, Станок, Операция и т. д. На языке алгебры предикатов любое множество $U = \{a_1, a_2, \dots, a_n\}$ может быть записано в виде уравнения $x^{a_1} \vee x^{a_2} \vee \dots \vee x^{a_n} = 1$, где x - предметная переменная, характеризующая данное множество. Совокупность всех корней этого уравнения совпадает с множеством U . Уравнение является формальной записью утверждения $x \in U$.

На языке алгебры предикатов можно описать любую систему подмножеств M универсума U [4]. Если обозначить имена всех подмножеств некоторого универсума $U = \{a_1, a_2, \dots, a_n\}$ следующими буквами $m_0, m_1, \dots, m_{2^n-1}$, то связь подмножеств с их именами может быть задана уравнением вида

$$\theta(x, u) = 1, \tag{1}$$

где $x \in U$, а переменная u ($u \in \{m_1, m_2, \dots, m_{n-1}\}$) играет роль переменного подмножества универсума U . Любую систему M подмножеств универсума U можно записать, добавив к зависимости (1) уравнение

$$\varphi(u) = 1, \tag{2}$$

где φ - предикат, задаваемый условиями

$$\varphi(u) = \begin{cases} 1, & \text{если } u \in M \\ 0, & \text{если } u \notin M \end{cases}$$

Таким образом, любую систему множеств M можно записать уравнением

$$\theta(x, u) \wedge \varphi(u) = 1.$$

Например, пусть $U = \{c, d\}$. Необходимо описать систему множеств $\{\{c\}, \{c, d\}\}$. Систему всех подмножеств U запишем с помощью уравнения (1)

$$u^{m_1} x^c \vee u^{m_2} x^d \vee u^{m_3} (x^c \vee x^d) = 1$$

Уравнение (2) примет вид

$$u^{m_1} \vee u^{m_3} = 1.$$

Тогда система множеств $\{\{c\}, \{c, d\}\}$ будет записана с помощью уравнения

$$u^{m_1} x^c \vee u^{m_3} (x^c \vee x^d) = 1.$$

Используя указанные выше соотношения, можно формально описывать отношения второго порядка, связывающие между собой множества предметов и системы множеств второго порядка, т. е. множества систем множеств предметов. Следовательно, на языке алгебры предикатов можно описывать конструкции многосортной логики. Однако средствами многосортной логики недостаточно адекватно можно описать многоуровневую сложно структурированную предметную область. Для описания структурных зависимостей «класс – подкласс», «часть – целое» между объектами предметной области используется многоуровневая логика. Многоуровневая логика может рассматриваться как расширение многосортной логики за счет введения метода структурирования типов. Отношение «элемент – множество» между объектами A и B записывается $(X/B^1)EQ(A, X)$; Отноше-

ние «класс – подкласс» между понятиями В и С – $(X/B^2) EQ(C,X)$, где EQ – предикат равенства [2].

Данные отношения можно выразить с помощью математических структур высших порядков. Рассмотрим пример математической структуры четвертого порядка. Пусть требуется описать структуру

$$\{\{\{a\}, \{a, b\}\}, \{\{b\}, \{a, b\}\}\}.$$

На языке алгебры предикатов это можно сделать следующим образом. Пусть x – объект некоторой ПрО, u – множество объектов, v – система множеств объектов. Систему множеств v можно интерпретировать как класс множества объектов, u – как подкласс. Тогда рассматриваемая структура запишется в виде системы уравнений

$$\begin{aligned} u^{m_1} x^a \vee u^{m_2} x^b \vee u^{m_3} (x^a \vee x^b) &= 1, \\ \forall (u^{m_1} \vee u^{m_3}) \vee \forall (u^{m_2} \vee u^{m_3}) &= 1. \end{aligned}$$

Здесь символом m_1 обозначен подкласс $\{a\}$, символом m_2 – подкласс $\{b\}$, символом m_3 – подкласс $\{a,b\}$. Символом 1 обозначен класс $\{\{a\}, \{a,b\}\}$, символом 2 – класс $\{\{b\}, \{a,b\}\}$ [4].

Заключение. Рассмотренные методы идентификации знаний с использованием алгебры предикатов предоставляют возможность единообразного представления знаний в виде уравнений алгебры предикатов. Любое уравнение алгебры предикатов можно реализовать аппаратно в виде переключательной цепи. Используя переключательные цепи, можно конструировать технические средства обработки и хранения знаний в виде некоторого интеллектуального процессора обработки знаний (ИПОЗ).

In article questions of identification and the formal description of knowledge in information intellectual systems with use of the algebra-logic approach that allows by transition to switching circuits are considered to design means of processing and storage of knowledge as some intellectual processor of processing of knowledge.

1. Горбатов В.А. Основы дискретной математики: Учебн. Пособие для студентов вузов.- М.: Высш. шк., 1986.-311с.
2. Кокорева Л.В., Перевозчикова О.Л., Ющенко Е.Л. Диалоговые системы и представление знаний.
3. Шабанов-Кушнаренко Ю.П. Теория интеллекта. Математические средства – Х.: Вища шк. Изд-во при Харьк. ун-те, 1984.-144с.
4. Шабанов-Кушнаренко Ю.П. Теория интеллекта. Технические средства –Х.: Вища шк. Изд-во при Харьк. ун-те, 1986.-136с.
5. Шабанов-Кушнаренко Ю.П., Шаронова Н.В. Компараторная идентификация лингвистических объектов: Монография.-К.:ІСДО, 1993. –116с.
6. Хайрова Н.Ф., Шаронова Н.В. Автоматизированные информационные библиотечные системы: задачи обработки информации: Монография – Харьков, 2002. – 120 с.