УДК 519.7

Гончаров С.В. (НТУ "ХПІ")
Каніщева О.В. (НТУ "ХПІ")

АЛГЕБРА СКІНЧЕННИХ ПРЕДИКАТІВ ЯК СКЛАДОВА ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

У статті розглянуто алгебро-логічні засоби моделювання природної мови, аналіз та дослідження її алгебро-логічної структури. Алгебра предикатів та предикатних операцій рекомендується для формального опису механізму природної мови і мислення, удосконалювання баз даних і знань, експертних систем, для розробки ЕОМ паралельної дії.

Ключові слова: алгебро-логічні структури, алгебра скінченних предикатів, алгебра предикатних операцій, предикат впізнавання.

В статье рассмотрены алгебро-логические средства моделирования естественного языка, анализ и исследования ее алгебро-логической структуры. Алгебра предикатов и предикатных операций рекомендуется для формального описания естественного языка и мышления, совершенствования баз данных и знаний, экспертных систем, для разработки ЭВМ параллельного действия.

Ключевые слова: алгебро-логические структуры, алгебра конечных предикатов, алгебра предикатних операций, предикат узнавания.

In article algebra-logic means of modelling of a natural language, the analysis and researches of its algebra-logic structure are considered. The algebra of predicates and predicate operations is recommended for the formal description of a natural language and way of thinking, perfection of databases and knowledge, expert systems, for development of the parallel computer.

Keywords: algebro-logic structures, algebra of final predicates, algebra of predicate operations, a cognizance predicate.

Вступ. Ця стаття присвячена алгебро-логічним засобам моделювання природної мови, аналізу та дослідженню її алгебро-логічної структури. Природна мова (ПМ), як явище інтелектуальної діяльності людини, є дуже складним об'єктом. Але, маючи формальний опис природної мови, його можна реалізувати на ЕОМ і таким чином надати машині здібність володіти природною мовою. Немає іншої галузі знань, яка б у вищому ступені, ніж ця, сприяла підвищенню темпів комп'ютеризації та інформатизації суспільства. Якщо можна буде знайти з аналізу мови більш досконалу алгебраїчну мову, то це збільшить можливості розроблювачів інформаційних систем, нових інформаційних технологій. Алгебро-логічний апарат, який можна буде знайти в природній мові, дозволятиме розширювати можливості розроблювача, що займається створенням нових інформаційних технологій. Отже, концептуально-методологічний підхід до ПМ (з математичної точки зору) дозволяє сприймати її як деяку алгебру, а тексти – як формули цієї алгебри.

Постановка задачі. У статті розглянуто алгебро-логічні засоби моделювання природної мови, аналіз та дослідження її алгебро-логічної структури. Запропоновано використання алгебри предикатів та предикатних операцій для формального опису механізму природної мови і мислення, удосконалювання баз даних і знань, експертних систем, для розробки ЕОМ паралельної дії.

Основний матеріал. У роботі використовується апарат алгебри предикатів [1, 2]. Останню трактуємо як алгебру, носієм якої є множина M усіх предикатів U_m, де U – непорожня множина будь-яких змінних, яку називатимемо універсумом, тобто $U = \{x_1, x_2, \ldots, x_m\}$. Множина U може бути як скінченною, так і нескінченною. У першому випадку простір U_m називатимемо скінченним, а в іншому – нескінченним. Тут $x_1, x_2, \ldots, x_m$ – будь-які місця предметів. Тому їх інакше називатимемо предметними змінними. Якщо предмет a знаходиться на місці x_i $\left(i = \overline{1, m}\right)$, то говоритимемо, що змінна

x_i приймає значення a та маємо такий запис: $x_i = a$. Якщо $a_1, a_2, \ldots, a_m \in U$ та $x_1 = a_1, x_2 = a_2, \ldots, x_m = a_m$, то пишемо $(a_1, a_2, \ldots, a_m) \in U_m$ і кажуть, що предметний вектор (набір) належить предметному простору U_m. Число m називатимемо вимірністю простору U_m.

Будь-яку підмножину T простору U_m називатимемо m-місним відношенням, яке задано на U_m. Для формульного запису таких відношень використовуватимемо явний спосіб завдання скінченного алфавітного оператора, що є базою для апаратурного методу розв'язання рівнянь алгебри скінченних предикатів [1, 3].

Нехай T – множина усіх відношень на U_m, Q – множина усіх предикатів на U_m. Відношення T та предикат Q називатимемо відповідними одне одному, якщо при будь-яких $x_1, x_2, \ldots, x_m$ маємо:

$$Q(x_1, x_2, \ldots, x_m) = \begin{cases} 0, \text{якщо} & x_1, x_2, \ldots, x_m \notin T, \\ 1, \text{якщо} & x_1, x_2, \ldots, x_m \in T. \end{cases} \tag{1}$$

Згідно з (1) можливий перехід від будь-якого відношення T до відповідного йому предиката Q. Предикат Q, що знаходимо за (1), називатимемо характеристичною функцією відношення T.

Предикатом розпізнавання предмета $a \in U$ за змінною x_i $(i = \overline{1, m})$ називатимемо умову

$$a(x_i) = \begin{cases} 0, \text{якщо} & a \neq x_i, \\ 1, \text{якщо} & a = x_i. \end{cases} \tag{2}$$

Предикат $a(x_i)$ розглядатимемо як предикат $a(x_1, x_2, \ldots, x_i, \ldots, x_m)$ із P, усі аргументи якого, крім x_i, неістотні. Вираз у вигляді $a(x_i)$, де $(i = \overline{1, m})$, $a \in U$, замінимо на x_i^a: тут a називатимемо показником для змінної x_i.

Алгеброю предикатів над M називатимемо множину T з базисними елементами x_i^a $(i = \overline{1, m}, a \in U)$ та базисними операціями: диз'юнкція, кон'юнкція, заперечення. Виключення з базису даної алгебри операції заперечення дозволяє отримати диз'юнктивно-кон'юнктивну алгебру.

Природну мову уявлятимемо як математичний об'єкт (якась алгебра). При цьому смисл (зміст) думок можна висловити реченнями і текстами, що з точки зору їх математичної природи будемо розглядати як предикати. Таким чином, кожне речення виражає деяку функцію з двійковим значенням, тобто задає деякий предикат $P(x) = \lambda$. Незалежною змінною x даної функції буде змінна ситуація, залежною – істинна змінна λ. Після підставлення замість змінної x конкретної постійної ситуації $x = a$ задане речення стає істинним $(\lambda = 1)$ або хибним $(\lambda = 0)$. Це залежить від того, чи відповідає зміст цього речення, чи ні ситуації a, до якої воно віднесено. Таким чином, у статті алгебра скінченних предикатів розглядається як інструмент дослідження, а не як його предмет.

Далі розглянемо поняття алгебри предикатів. Візьмемо деяку непусту множину U, елементи якої будемо називати предметами. Сама ж множина U називається універсумом предметів. Візьмемо, далі, m деяких непустих, необов'язково різних підмножин $A_1, A_2, \ldots, A_m$ універсума U. Декартовий добуток $S = A_1 \times A_2 \times \ldots \times A_m$ множин $A_1, A_2, \ldots, A_m$ називається предметним простором S з координатними предметними осями $A_1, A_2, ..., A_m$ над універсумом U. Число осей m називається розмірністю простору S. Вводимо множину $V = \{x_1, x_2, ..., x_m\}$ різних змінних $x_1, x_2, ..., x_m$, які називаються предметними змінними простору S. Множина V називається універсумом змінних простору S. Значеннями змінної x_i $(i = \overline{1, m})$ є елементи множини A_i, так що $x_1 \in A_1, x_2 \in A_2, ..., x_m \in A_m$. Множини $A_1, A_2, ..., A_m$ називаються областями зміни змінних $x_1, x_2, ..., x_m$.

Якщо $a_1 \in A_1, a_2 \in A_2, \ldots, a_m \in A_m$ й $x_1 = a_1, x_2 = a_2, ..., x_m = a_m$, то пишуть $(a_1, a_2, ..., a_m) \in S$ і говорять, що предметний вектор $(a_1, a_2, ..., a_m)$ належить до простору $S = A_1 \times A_2 \times ... \times A_m$. Елементи $a_1, a_2, ..., a_m$ вектора $(a_1, a_2, ..., a_m)$ називаються його компонентами (першим, другим, ..., m-тим). Предметний простір S можна розглядати як сукупність всіх векторів виду $(x_1, x_2, ..., x_m)$, компоненти яких задовольняють умові $x_1 \in A_1, x_2 \in A_2, ..., x_m \in A_m$. Будь-яка підмножина P простору S називається відношенням, утвореним в (або інакше: заданим на) просторі S. Відношення має розмірність m. Говорять, що воно m-місне. Відношення, задані на одному й тому ж просторі S, називаються однотипними. Тип відношення визначається набором змінних $x_1, x_2, ..., x_m$ та набором множин $A_1, A_2, ..., A_m$. Відношення $\varnothing$, що не містить жодного вектора, називається порожнім, відношення S, у якому є різноманітні вектори, – повним.

Предикатом, заданим на декартовому добутку $A_1, A_2, ..., A_m$, називається будь-яка функція $P(x_1, x_2, ..., x_m) = \xi$, яка відображає декартовий добуток $A_1 \times A_2 \times ... \times A_m$ множин $A_1, A_2, ..., A_m$ у множину $\Sigma = \{0, 1\}$. Символи 0 та 1 називаються булевими елементами, Σ – множина всіх булевих елементів. Змінна $\xi = \{0, 1\}$, яка є значенням предиката P, називається булевою. Предикат $P(x_1, x_2, ..., x_m)$ на $A_1 \times A_2 \times ... \times A_m$ називається скінченним, якщо всі множини $A_1, A_2, ..., A_m$ скінченні, та нескінченним – у противному випадку. Ця ж термінологія переноситься й на відповідні предикатам відношення. Змінні $x_1, x_2, ..., x_m$ називаються аргументами предиката P.

Припустимо, що L – множина всіх відношень на S, M – множина всіх предикатів на S. Між всіма відношеннями множини L й всіма предикатами множини M, заданими на S, існує взаємно однозначна відповідність. Відношення P з L і предикат P з M називаються відповідними один одному, якщо при будь-яких $x_1 \in A_1, x_2 \in A_2, ..., x_m \in A_m$

$$P(x_1, x_2, ..., x_m) = \begin{cases} 1, \text{ если } (x_1, x_2, ..., x_m) \in P, \\ 0, \text{ если } (x_1, x_2, ..., x_m) \notin P. \end{cases} \tag{3}$$

Зворотний перехід від предиката P до відношення P здійснюється за правилом:

$$\begin{aligned} &\text{якщо } P(x_1, x_2, ..., x_m) = 1, \text{ то } (x_1, x_2, ..., x_m) \in P; \\ &\text{якщо } P(x_1, x_2, ..., x_m) = 0, \text{ то } (x_1, x_2, ..., x_m) \notin P. \end{aligned} \tag{4}$$

Множина всіх векторів $(x_1, x_2, ..., x_m)$, що задовольняють рівнянню $P(x_1, x_2, ..., x_m) = 1$, утворює відношення P, яке називається областю істинності предиката P. Предикат $P \in M$ називається характеристичною функцією відношення $P \in L$. Алгеброю предикатів називається будь-яка алгебра, задана над носієм M. Операції диз'юнкції, кон'юнкції й заперечення над предикатами визначаються наступними рівностями: для будь-яких $x_1 \in A_1, x_2 \in A_2, ..., x_m \in A_m$

$$(P \vee Q)(x_1, x_2, ..., x_m) = P(x_1, x_2, ..., x_m) \vee Q(x_1, x_2, ..., x_m);$$
$$(P \wedge Q)(x_1, x_2, ..., x_m) = P(x_1, x_2, ..., x_m) \wedge Q(x_1, x_2, ..., x_m); \quad (5)$$
$$(\neg P)(x_1, x_2, ..., x_m) = \neg(P(x_1, x_2, ..., x_m)).$$

Символи $\vee, \wedge, \neg$, які знаходяться ліворуч від знака рівності, визначають операції над предикатами, праворуч – операції над значеннями предикатів, тобто над булевими елементами.

Предикати будь-якого типу можна записувати у вигляді формул. Тип скінченних предикатів задаємо, вказуючи множини $V=\{x_1, x_2, ..., x_m\}$ та $A_i = \{a_{1i}, a_{2i}, ..., a_{k_i i}\}$, $i = \overline{1, m}$, k_i – число елементів у множині A_i. Над носієм M вводимо диз'юнктивно-кон'юнктивну алгебру предикатів. У ролі базисних елементів цієї алгебри використовуємо предикати 0 та 1, а також предикати x_i^a впізнавання предмета a за змінною x_i, $i = \overline{1, m}$, $a \in A_i$

$$x_i^a = \begin{cases} 1, \text{если } x_i = a, \\ 0, \text{если } x_i \neq a. \end{cases} \quad (6)$$

Символ a у записі предиката x_i^a називається його показником. У ролі базисних операцій у диз'юнктивно-кон'юнктивній алгебрі предикатів використовуються диз'юнкція й кон'юнкція предикатів. Будь-який предикат $P(x_1, x_2, ..., x_m)$ у цій алгебрі можна записати формулою у вигляді його досконалої диз'юнктивної нормальної форми (ДДНФ)

$$P(x_1, x_2, ..., x_m) = \bigvee_{(a_1, a_2, ..., a_m) \in P} x_1^{a_1} x_2^{a_2} ... x_m^{a_m}. \quad (7)$$

Вирази виду $x_1^{a_1} x_2^{a_2} ... x_m^{a_m}$ називаються конституентами одиниці предиката P. Запис $(a_1, a_2, ..., a_m) \in P$ під знаком $\vee$ означає, що береться диз'юнкція всіх конституент одиниці $x_1^{a_1} x_2^{a_2} ... x_m^{a_m}$, показники співмножників якої задовольняють умові $(a_1, a_2, ..., a_m) \in P$, де P – відношення, що відповідає предикату $P(x_1, x_2, ..., x_m)$. Це означає, що диз'юнктивно-кон'юнктивна алгебра предикатів повна, тобто, що формулами цієї алгебри можна записати будь-який предикат, а отже можна виразити аналітично будь-яке відношення довільного типу.

Кожному предикату $F(x, y)$ визначеному на $M \times N$, відповідає відображення f, що діє з M у N. Якщо для $x \in M$ існує $y \in N$, такий, що xFy, то говорять, що відображення f ставить у відповідність предмету x предмет y, і пишуть $y = f(x)$ або $y = fx$. Іншими словами: $F(x, y) = 1 \Leftrightarrow f(x) = y$. Коректне визначення відображення: $f(x) = S_x$, де $S_x = \{y \mid F(x, y) = 1\}$.

Сукупність S_a всіх образів $y \in N$ предмета $a \in M$ щодо відображення $y = f(x)$ називається повним образом предмета a щодо відображення f. Іншими словами, $S_a = \{y \mid F(a, y) = 1\}$. Останній запис мовою логіки формально записується в такий спосіб: $S_a(y) = F(a, y)$. Знайдемо повний образ деякого предмета щодо деякого відображення. Сукупність R_b всіх прообразів $x \in M$ предмета $b \in N$ відносно відображення $y = f(x)$ називається повним прообразом предмета b щодо відображення f. Іншими словами, $R_b = \{x \mid F(x, b) = 1\}$; $R_b(x) = F(x, b)$.

Образом множини $A \subseteq M$ відносно відображення f називається множина $B \subseteq N$, утворена з усіх образів предметів, що належать множині A.

$$B=\{y \mid \exists x \in M(F(x,y)\wedge A(x))=1\} \text{ або}$$
$$B(y)=\exists x \in M(F(x,y)\wedge A(x)) \qquad (8)$$

Перетворення виду (8) називається лінійним логічним оператором $B=F(A)$ (інакше – відповідністю Галуа). Воно характеризується адитивністю $F(A_1 \vee A_2)=F(A_1)\vee F(A_2)$ й однорідністю $F(\alpha A)=\alpha F(A)$, $\alpha \in \{0,1\}$. Вираз (8) – це загальний вид лінійного логічного оператора. Прообразом множини $B \subseteq N$ відносно відображення f називається множина $A' \subseteq M$, утворена з усіх прообразів предметів, що належать множині B:

$$A'(x)=\exists y \in N(F(x,y)\wedge B(y)). \qquad (9)$$

Припустимо, що $M=\{a_1,a_2,\ldots,a_k\}$ – фіксована множина, що складається з k елементів; A – деяка із її підмножин, $A \subseteq M$. Для множини A створюємо набір $(\alpha_1,\alpha_2,\ldots,\alpha_k)$ логічних елементів $\alpha_1,\alpha_2,\ldots,\alpha_k$ за наступним правилом: якщо $a_i \in A$, то $\alpha_i=1$; якщо ж $a_i \notin A$, то $\alpha_i=0$ $(i=\overline{1,k})$. Набір $(\alpha_1,\alpha_2,\ldots,\alpha_k)$ називається характеристикою множини A. Знайдемо характеристику деякої підмножини заданої множини. Предикат $A(x)$ на M, що відповідає множині A з характеристикою $(\alpha_1,\alpha_2,\ldots,\alpha_k)$, запишеться формулою:

$$A(x)=\alpha_1 x^{\alpha_1} \vee \alpha_2 x^{\alpha_2} \vee \ldots \vee \alpha_k x^{\alpha_k} = \bigvee_{i=1}^{k} \alpha_i x^{\alpha_i}. \qquad (10)$$

Набір $(\alpha_1,\alpha_2,\ldots,\alpha_k)$ називається також координатним представленням множини A. Пишуть: $A=(\alpha_1,\alpha_2,\ldots,\alpha_k)$.

Висновки. Таким чином, ми ввели необхідні поняття алгебри предикатів, яка є перевіреним математичним інструментарієм для запису відношень, виявлених у текстах природної мови.

Розглянутий підхід можливо використовувати для опису знань та він може надати в перспективі можливість однакового представлення знань в електронних бібліотеках у вигляді рівнянь алгебри предикатів. Будь-яке рівняння алгебри предикатів можна реалізувати апаратно у виді перемикального ланцюга. Використовуючи перемикальні ланцюги, можна конструювати технічні засоби обробки і збереження знань у виді деякого інтелектуального процесора обробки знань [4].

Також, апарат алгебри скінченних предикатів та розроблену на її основі теорію компараторної ідентифікації лінгвістичних об'єктів можна використовувати для автоматизації процесів створення та поповнення галузевих термінологічних словників [5].

ЛІТЕРАТУРА:

1. Шабанов-Кушнаренко Ю. П. Теория интеллекта: Проблемы и перспективы / Ю. П. Шабанов-Кушнаренко. – Х. : Вища шк., 1987. – 158 с.

2. Шабанов-Кушнаренко Ю. П. Компараторная идентификация лингвистических объектов / Ю. П. Шабанов-Кушнаренко, Н. В. Шаронова. – К. : ИСИО, 1993. – 116 с.

3. Хайрова Н. Ф. Автоматизированные информационные библиотечные системы: задачи обработки информации: Монография / Н. Ф. Хайрова, Н. В. Шаронова. – Харьков : ХГУ "НУА", 2002. – 120 с.

4. Гончаров С. В. Створення та поповнення електронних галузевих термінологічних словників / С. В. Гончаров, Н. В. Борисова. – Херсон : ХНТУ, 2009. – № 1 (34). – С. 204-206.

5. Канищева О. В. Идентификация знаний в электронных библиотеках / О. В. Канищева // Системный анализ и информационные технологии: тез. IX Междунар. науч.-практ. конф. – К. : ЕКМО, 2007. – С. 114.

Рецензент: д.т.н., проф. Шаронова Н.В.