УДК 004.82
MSC 03B70

Вестник СПбГУ. Прикладная математика. Информатика... 2022. Т. 18. Вып. 1

Выделение общих свойств объектов для создания логических онтологий*

Т. М. Косовская, Н. Н. Косовский

Санкт-Петербургский государственный университет, Российская Федерация,
199034, Санкт-Петербург, Университетская наб., 7–9

Для цитирования: *Косовская Т. М., Косовский Н. Н.* Выделение общих свойств объектов для создания логических онтологий // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2022. Т. 18. Вып. 1. С. 37–51. https://doi.org/10.21638/11701/spbu10.2022.103

Предлагается подход к формированию онтологии, основанный на описаниях объектов в терминах языка исчисления предикатов. При таком подходе объект представлен как множество своих элементов, на котором задан набор предикатов, характеризующих свойства этих элементов и отношения между ними. Описанием объекта является конъюнкция литералов, истинных на элементах объекта. Под онтологией понимается ориентированный граф с описаниями подмножеств в качестве вершин, такой что элементы множества в конце дуги обладают свойствами элементов множества в начале этой дуги. Предлагаются следующие формулировки задачи построения онтологии: 1) все предикаты двузначные и заданы подмножества исходного множества объектов; 2) все предикаты двузначные и требуется найти подмножества исходного множества; 3) среди предикатов имеются многозначные и заданы подмножества исходного множества объектов. Основным инструментом построения такого графа является выделение элементарной конъюнкции литералов предикатных формул, изоморфной подформулам некоторых формул. Дается определение изоморфизма элементарных конъюнкций атомарных предикатных формул. Для каждой из предложенных задач формулируются алгоритмы построения онтологии. Приводится иллюстративный пример.

Ключевые слова: логическая онтология, предикатные формулы, изоморфизм предикатных формул.

1. Введение. В философии под онтологией понимают абстракции, принимающие во внимание одни признаки предметов (или отношения между частями этих предметов) и игнорирующие другие.

Для математиков онтология — это ориентированный граф, в каждой вершине которого находится множество объектов, обладающих одними и теми же свойствами (или отношениями между частями объектов этого множества). Причем все объекты, находящиеся в конце дуги, обладают свойствами, присущими всем объектам в начале дуги и некоторыми дополнительными [1].

Создание онтологии предметной области является одним из направлений представления знаний для использования их в системах искусственного интеллекта и извлечения знаний из баз данных [2–5]. При решении задач искусственного интеллекта широко применяются подходы, основанные на различных методах дискретной математики, например графах [6] и нечетких множествах [7].

Если объект представлен как совокупность своих элементов, обладающих конкретными свойствами и связанных заданными отношениями, то описание объекта мо-

* Работа выполнена при финансовой поддержке Санкт-Петербургского государственного университета (проект № 73555239).

жет быть задано формулой исчисления предикатов, в частности элементарной дизъюнкцией литералов [8].

Для выделения из множества таких объектов подмножества, обладающего общими свойствами, в [8, 9] было предложено построение многоуровневого описания исходного множества объектов, существенно уменьшающего вычислительную сложность задач распознавания и анализа сложных объектов. Фактически эти описания являются иерархическими, но для них (в отличие от онтологий) описания объектов на верхних уровнях являются частными случаями описаний на нижних уровнях. В настоящей работе предлагается использовать идеи такого построения для формирования онтологии.

Основным инструментом создания данного описания служит выделение элементарной конъюнкции литералов предикатных формул, изоморфной подформулам некоторых других элементарных конъюнкций. В п. 2 даются определения и формулируется утверждение, необходимые для понимания дальнейшего изложения.

В п. 3 приводятся математические постановки двух задач построения онтологии для объектов, описываемых двузначными предикатами, среди аргументов которых нет числовых переменных, и приводятся алгоритмы их решения.

В п. 4 рассматривается постановка задачи построения онтологии для множества объектов, описания которых содержат многозначные предикаты. Эта постановка сводится к применению в их описаниях двузначных предикатов, среди аргументов которых имеются количественные переменные. Описываются алгоритм решения такой задачи и иллюстративный пример.

2. Основные определения. В основе дальнейшего изложения лежит понятие изоморфизма элементарных конъюнкций предикатных формул. Приведем определение этого понятия.

Определение 1. *Две элементарные конъюнкции атомарных формул исчисления предикатов $P(a_1, \dots, a_m)$ и $Q(b_1, \dots, b_m)$ называются изоморфными, т. е.*

$$P(a_1, \dots, a_m) \sim Q(b_1, \dots, b_m),$$

если существуют такая элементарная конъюнкция $R(x_1, \dots, x_m)$ и подстановки аргументов $a_{i_1}, \dots, a_{i_m}$ и $b_{j_1}, \dots, b_{j_m}$ формул $P(a_1, \dots, a_m)$ и $Q(b_1, \dots, b_m)$ соответственно вместо всех вхождений переменных $x_1, \dots, x_m$ формулы $R(x_1, \dots, x_m)$, что результаты этих подстановок $R(a_{i_1}, \dots, a_{i_m})$ и $R(b_{j_1}, \dots, b_{j_m})$ совпадают с формулами $P(a_1, \dots, a_m)$ и $Q(b_1, \dots, b_m)$ соответственно с точностью до порядка литералов.

При этом полученные подстановки $(a_{i_1} \to x_1, \dots, a_{i_m} \to x_m)$ и $(b_{j_1} \to x_1, \dots, b_{j_m} \to x_m)$ называются унификаторами формул $P(a_1, \dots, a_m)$ и $Q(b_1, \dots, b_m)$ соответственно с формулой $R(x_1, \dots, x_m)$.

Отметим, что аргументами элементарных конъюнкций $P(a_1, \dots, a_m)$, $Q(b_1, \dots, b_m)$ могут быть как предметные переменные, так и предметные константы. При этом понятие изоморфизма элементарных конъюнкций является отношением эквивалентности, и формулу $R(x_1, \dots, x_m)$ можно считать обобщенным описанием этого класса: $R(x_1, \dots, x_m) \sim P(a_1, \dots, a_m)$ и $R(x_1, \dots, x_m) \sim Q(b_1, \dots, b_m)$.

Кроме того, понятие изоморфизма элементарных конъюнкций атомарных формул исчисления предикатов отличается от понятия равносильности этих формул, так как они могут иметь существенно различные аргументы. По сути дела для изоморфных формул существуют такие перестановки их аргументов, что они задают одно и то же отношение между своими аргументами.

В [10] доказано, что задача проверки формул на изоморфность полиномиально эквивалентна задаче проверки изоморфизма графов, являющейся «открытой» задачей, для которой не доказаны ни ее принадлежность классу **P**, ни ее NP-полнота. Ласло Бабай в 2017 г. [11] предложил квазиполиномиальный алгоритм ее решения. Полученная им оценка $2^{O((\log n)^3)}$ до сих пор проверяется.

Для построения онтологии потребуется выделение из двух заданных элементарных конъюнкций максимальных подформул, изоморфных друг другу.

Определение 2. *Элементарная конъюнкция называется максимальной формулой, изоморфной подформулам двух заданных элементарных конъюнкций, если она изоморфна некоторым подформулам этих элементарных конъюнкций, но после добавления в нее любого литерала она не изоморфна ни одной подформуле хоть одной из них.*

В дальнейшем изложении не будем использовать предикатные формулы, вид которых отличается от элементарной конъюнкции литералов. Поэтому вместо длинного термина «элементарная конъюнкция литералов» часто будем применять термин «формула».

Приведем **пример** максимальной формулы, изоморфной подформулам двух заданных элементарных конъюнкций.

Пусть

$$P(a,b,c) = p_1(a,b) \ \& \ \neg p_1(a,c) \ \& \ p_2(b,c,a),$$

$$Q(a,b,c) = p_1(b,a) \ \& \ \neg p_1(a,c) \ \& \ p_2(a,c,b).$$

Эти элементарные конъюнкции имеют общую подформулу

$$R_1(a,c) = \neg p_1(a,c)$$

с одним литералом и подформулу

$$R_2(x,y) = p_1(x,y) \ \& \ p_2(y,c,x),$$

изоморфную подформулам формул $P(a,b,c)$ и $Q(a,b,c)$, так как

$$P(a,b,c) = R(a,b) \ \& \ p_1(a,c),$$

$$Q(a,b,c) = R(b,a) \ \& \ p_1(a,c).$$

Добавление в формулу $R_2(x,y)$ литерала $\neg p_1(x,z)$ превращает ее в изоморфную формуле $P(a,b,c)$, но не изоморфную формуле $Q(a,b,c)$. Добавление же в нее литерала $\neg p_1(y,z)$ превращает ее в изоморфную формуле $Q(a,b,c)$, но не изоморфную формуле $P(a,b,c)$.

Таким образом, как литерал $\neg p_1(x,z)$, так и элементарная конъюнкция $R(x,y) = p_1(x,y) \ \& \ p_2(y,c,x)$ являются максимальными формулами, изоморфными подформулам формул $P(a,b,c)$ и $Q(a,b,c)$.

В [12] приведен алгоритм выделения максимальной формулы, изоморфной подформулам двух заданных элементарных конъюнкций. Отметим, что в отличие от задачи проверки формул на изоморфность задача выделения из двух заданных элементарных конъюнкций максимальных подформул, изоморфных друг другу, является NP-трудной.

Для дальнейшего изложения важным также будет понятие «наибольшее общее свойство элементов из заданного множества».

Пусть задано множество объектов Ω, в котором каждый объект ω является совокупностью своих элементов $\omega = \{\omega_1, \dots, \omega_t\}$. На этих объектах задан набор предикатов $p_1, \dots, p_n$, характеризующих свойства элементов из ω и отношения между ними.

Если все предикаты $p_1, \dots, p_n$ двузначные, то *логическим описанием* $S(\omega)$ *объекта* ω называется конъюнкция всех литералов, истинных на ω. Заметим, что аргументами логического описания $S(\omega)$ объекта ω являются константы.

Обобщенным логическим описанием $P_\omega(\overline{x}_\omega)$ *объекта* ω называется элементарная конъюнкция, задающая класс эквивалентности по отношению изоморфизма формул во множестве логических описаний объектов из Ω. При этом аргументами обобщенного логического описания $P_\omega(\overline{x}_\omega)$ объекта ω могут быть как переменные, так и константы.

Логическим описанием $S(\Omega)$ *множества* Ω называется множество обобщенных логических описаний $P_\omega(\overline{x}_\omega)$ объектов $\omega \in \Omega$:

$$S(\Omega) = \{P_\omega(\overline{x}_\omega) : \omega \in \Omega\}.$$

Заметим, что логическое описание $S(\Omega)$ множества Ω не содержит изоморфных формул.

Определение 3. *Наибольшим общим свойством элементов из множества* Ω *называется наибольшая по количеству литералов формула, для которой в каждом описании объекта из* Ω *имеется изоморфная подформула.*

Утверждение. Если 1) множества Ω_1 и Ω_2 являются подмножествами множества Ω, причем в каждом из них вместе с конкретным объектом содержатся и все объекты, описания которых изоморфны его описанию; 2) элементарные конъюнкции $P_1(\overline{x}_1)$ и $P_2(\overline{x}_2)$ являются наибольшими общими свойствами элементов из множеств Ω_1 и Ω_2 соответственно; 3) $P_1(\overline{x}_1)$ изоморфна некоторой подформуле формулы $P_2(\overline{x}_2)$, то $\Omega_1 \subset \Omega_2$.

3. Постановки задач построения логической онтологии с двузначными предикатами.

Задача 1. Пусть все предикаты $p_1, \dots, p_n$, характеризующие свойства элементов из ω и отношения между ними, по которым требуется построить онтологию, двузначны.

Во множестве Ω выделены (возможно, пересекающиеся) подмножества $\Omega_1, \dots, \Omega_m$, причем в каждом из них вместе с конкретным объектом содержатся и все объекты, описания которых изоморфны его описанию. Эти подмножества соответствуют некоторым известным свойствам объектов, которые будут присутствовать в онтологии.

Требуется построить ориентированный граф, у которого вершина с нулевым заходом помечена множеством Ω и соответствует множеству неизоморфных формул, каждая из которых изоморфна некоторым описаниям объектов из Ω.

Если из вершины Ω_k выходят дуги к вершинам, помеченным множествами $\Omega_{k_1}, \dots, \Omega_{k_r}$, то

1) $\Omega_k = \cup_{i=1}^{r} \Omega_{k_i}$;

2) для каждого $i = 1, \dots, r$ каждая формула в описании объектов из Ω_{k_i} изоморфна одной из формул в описании Ω_k, т. е. объекты из Ω_{k_i} обладают всеми свойствами, общими для всех объектов из Ω_k;

3) если $i \neq j$, то наибольшие общие свойства элементов из Ω_{k_i} и Ω_{k_j} различны.

Для решения этой задачи выделим максимальные общие свойства $Q_1(\overline{x}_1), \ldots, Q_m(\overline{x}_m)$ объектов из $\Omega_1, \ldots, \Omega_m$ и перенумеруем их в порядке убывания количества литералов в элементарных конъюнкциях $P_1(\overline{x}_1), \ldots, P_m(\overline{x}_m)$.

Для $i = 1, \ldots, m$ выполняем следующие действия:

1) среди формул $P_{i+1}(\overline{x}_{i+1}), \ldots, P_m(\overline{x}_m)$ находим формулы, изоморфные некоторым подформулам формулы $P_i(\overline{x}_i)$;

2) соединяем ориентированными ребрами вершину, помеченную $P_i(\overline{x}_i)$, с найденными формулами;

3) если для некоторого $j > i$, для которого добавлено ребро из $P_i(\overline{x}_i)$ в $P_j(\overline{x}_j)$, уже имеется ребро из $P_k(\overline{x}_k)$ в $P_j(\overline{x}_j)$ $(k < i < j)$, то это ребро удаляем (рис. 1);

4) к вершинам, которые в полученном графе имеют нулевой заход, добавляем ориентированные ребра от корневой вершины.

Удаление «лишних» ребер соответствует следующему. Пусть множество A со свойством $P_j(\overline{x}_j)$ изначально было определено как подмножество множества B со свойством $P_k(\overline{x}_k)$. При анализе множества A выяснилось, что оно является подмножеством множества C со свойством $P_i(\overline{x}_i)$ (левая часть рис. 1). Поэтому непосредственную связь между множествами A и B удалили (правая часть рис. 1).

Рис. 1. Удаление «лишних» ребер

Задача 2. Пусть все предикаты $p_1, \ldots, p_n$, характеризующие свойства элементов из ω и отношения между ними, по которым требуется построить онтологию, двузначные.

Заданы множество Ω и его логическое описание $S(\Omega)$, содержащее n_0 неизоморфных формул.

Требуется выделить (возможно, пересекающиеся) подмножества $\Omega_1, \ldots, \Omega_m$, обладающие общими свойствами, и построить ориентированный граф, имеющий свойства из задачи 1.

Для решения этой задачи можно воспользоваться алгоритмом построения многоуровневого описания класса [8, 9] по обучающей выборке:

1) для каждой пары элементарных конъюнкций $P_i(\overline{x}_i)$ и $P_j(\overline{x}_j)$ $(1 \leqslant i < j \leqslant n_0)$, входящих $S(\Omega)$, выделяем максимальную формулу $Q^1_{ij}(\overline{x}_{ij})$, изоморфную их подформулам.

Если среди выделенных максимальных формул имеются изоморфные, то оставляем только одну из них и перенумеровываем эти формулы без повторений $P^1_1(\overline{x}^1_1), \ldots, P^1_{n_1}(\overline{x}^1_{n_1})$. Из вершины, помеченной $S(\Omega)$, проводим ориентированные ребра к вершинам, помеченным $P^1_i(\overline{x}^1_i)$ при $1 \leqslant i \leqslant n_1$;

2) повторяем (при $l = 2, \ldots, L$) *) процесс выделения максимальных формул,

*) Процесс завершится, так как на каждой итерации длины подформул уменьшаются.

изоморфных подформулам для пар $P_i^{l-1}(\overline{x}_i^{l-1})$ и $P_j^{l-1}(\overline{x}_j^{l-1})$, $1 \leqslant i < j \leqslant n_1$, получив формулы $Q_{ij}^l(\overline{x}_{ij})$:

2а) если среди выделенных максимальных формул имеются изоморфные, то оставляем только одну из них и перенумеровываем эти формулы без повторений $P_1^l(\overline{x}_1^l), \ldots, P_{n_l}^l(\overline{x}_{n_l}^l)$,

2б) из вершин, помеченных $P_i^{l-1}(\overline{x}_i^{l-1})$ и $P_j^{l-1}(\overline{x}_j^{l-1})$, проводим ориентированные ребра к вершине, помеченной формулой $P_k^l(\overline{x}_k^l)$, изоморфной подформулам формул $P_i^{l-1}(\overline{x}_i^{l-1})$ и $P_j^{l-1}(\overline{x}_j^{l-1})$,

2в) если среди выделенных максимальных формул имеется изоморфная ранее выделенной, то вершину, соответствующую последней, «склеиваем» с вновь выделенной (вместе с ребрами, ицидентными этой вершине).

Отметим, что при выполнении алгоритма невозможно возникновение «порочного круга», при котором вершина, полученная в результате «склейки», находится в ориентированном цикле.

Действительно, пусть «склеиваются» вершины, соответствующие элементарным конъюнкциям $P_k^{l+1}(\overline{x}_k^{l+1})$ и $P_{k'}^{l'}(\overline{x}_{k'}^{l'})$. Данные формулы изоморфны и, следовательно, имеют одинаковую длину. При похождении вдоль ориентированного пути длины элементарных конъюнкций убывают, так как формула в следующей вершине изоморфна подформуле в предыдущей вершине. Следовательно, невозможен ориентированный путь между «склеиваемыми» вершинами.

На рис. 2 формула $P_k^{l+1}(\overline{x}_k^{l+1})$ изоморфна ранее полученной формуле $P_{k'}^{l'}(\overline{x}_{k'}^{l'})$ (в левой части рисунка). В результате «склейки» вершин с изоморфными формулами (правая часть рисунка) возникают дуги из $P_k^{l+1}(\overline{x}_k^{l+1})$ к ранее определенным вершинам, к которым прежде шли дуги из $P_{k'}^{l'}(\overline{x}_{k'}^{l'})$.

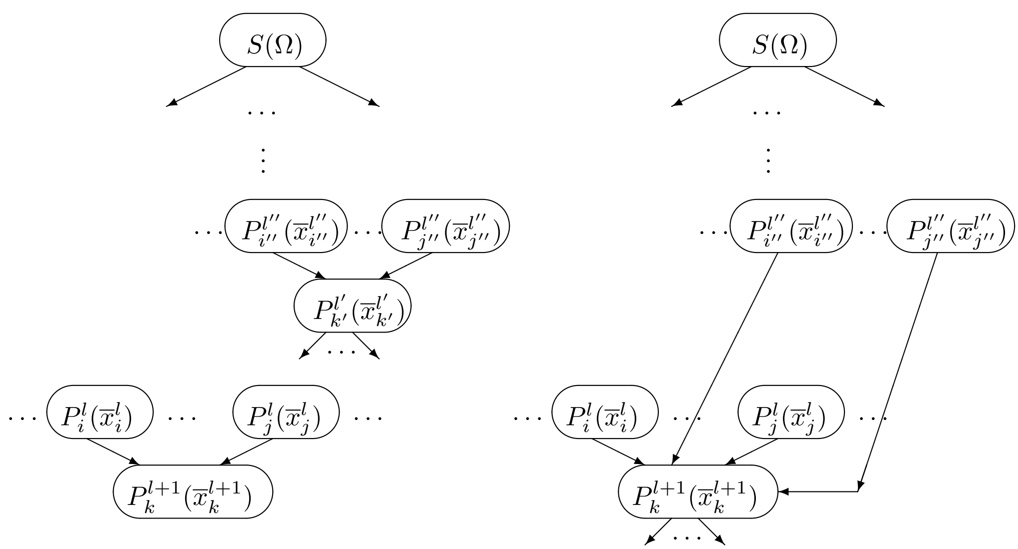

Рис. 2. «Склеивание» вершин с изоморфными формулами $P_k^{l+1}(\overline{x}_k^{l+1})$ и $P_{k'}^{l'}(\overline{x}_{k'}^{l'})$

4. Постановки задач построения логической онтологии с многозначными предикатами. Заметим, что можно рассматривать не обязательно двузначные предикаты $p_1, \ldots, p_n$. Они могут быть конечнозначными (задающими свойство попадания числового признака объекта в конечный промежуток или, например, цвет элемента объекта) или даже бесконечнозначными (задающими, например, длину, вес и т. п.). Многозначный n-местный предикат q^n со значениями из множества D_{q^n} можно трактовать как $(n+1)$-местный предикат q^{n+1}, определяемый равносильностью $q^{n+1}(\overline{x}; y) \Leftrightarrow q^n(\overline{x}) = y$, и где переменные из списка $\overline{x}$ являются переменными для имен самого объекта, а переменная y — это переменная для значений из D_{q^n}.

Фактически это означает, что среди аргументов предиката имеются не только переменные для имен элементов объекта ω или сами имена, но и числовые аргументы. Такие предикаты и числовые аргументы у них будем называть количественными.

Для того чтобы в формуле отличать переменные для имен элементов объекта ω или сами эти имена от имен числовых аргументов, при записи обозначения формулы будем писать $P(\overline{x}; \overline{a}_1; \ldots; \overline{a}_m)$, где $\overline{x}$ — список аргументов для имен элементов объекта, $\overline{a}_1; \ldots; \overline{a}_m$ — списки количественных аргументов для различных вхождений многозначных предикатов. При проверке формул на изоморфизм приходится отождествлять (ставить им в соответствие одну и ту же переменную) разные элементы разнообразных объектов (но не различные элементы одного объекта). При наличии количественных аргументов их отождествление фактически означает отождествление разных значений многозначного предиката. Поэтому при построении онтологии для работы с количественными аргументами требуется разработать отдельный алгоритм.

Продемонстрируем невозможность использования понятия изоморфизма формул при наличии многозначных предикатов. Формулы $A = p(a, b) \ \& \ p(b, c)$ и $B = p(b, c) \ \& \ p(c, a)$ с двузначным предикатом p изоморфны, так как получаются из $R = p(x, y) \ \& \ p(y, z)$ с унификаторами $(a \to x, b \to y, c \to z)$ и $(b \to x, c \to y, a \to z)$ соответственно. Но говорить об изоморфности формул $A' = p'(a, b; l_1) \ \& \ p'(b, c; l_2)$ и $B' = p'(b, c; l_3) \ \& \ p'(c, a; l_4)$ с двузначным предикатом p', определяемым посредством многозначного предиката q как $p'(x, y; l) \Leftrightarrow (q(x, y) = l)$, нельзя. Даже если множество значений q состоит из трех элементов $\{t, ?, f\}$ (как в трехзначной логике Лукасевича), то $A' = p'(a, b; t) \ \& \ p'(b, c; f) \Leftrightarrow q(a, b) \ \& \ \neg q(b, c)$, а $B' = p'(b, c; f) \ \& \ p'(c, a; t) \Leftrightarrow \neg q(b, c) \ \& \ q(c, a)$. Для двух последних формул такой формулы R и унификатора не существует, так как у формулы A' второй аргумент у литерала без отрицания и первый аргумент у литерала с отрицанием одинаковы, а у формулы B' они различны.

Задача 3. Пусть среди предикатов $p_1, \ldots, p_n$, характеризующих свойства элементов из ω и отношения между ними, по которым требуется построить онтологию, имеются многозначные.

Во множестве Ω выделены (возможно, пересекающиеся) подмножества $\Omega_1, \ldots, \Omega_m$, причем в каждом из них вместе с конкретным объектом содержатся и все объекты, описания которых изоморфны (учитывая только двузначные предикаты) его описанию.

Требуется построить ориентированный граф, у которого вершина с нулевым заходом помечена множеством Ω и соответствует множеству неизоморфных формул, каждая из которых изоморфна некоторым описаниям объектов из Ω.

Если из вершины Ω_k выходят дуги к вершинам, помеченным множествами $\Omega_{k_1}, \ldots, \Omega_{k_r}$, то

1) $\Omega_k = \cup_{i=1}^{r} \Omega_{k_i}$;

2) для каждого $i = 1, \ldots, r$ каждая формула в описании объектов из Ω_{k_i} изоморфна одной из формул в описании Ω_k, т. е. объекты из Ω_{k_i} обладают всеми свойствами, общими для всех объектов из Ω_k;

3) если $i \neq j$, то наибольшие общие свойства элементов из Ω_{k_i} и Ω_{k_j} различны.

Как при решении задачи 1, выделяем наибольшие общие свойства объектов, для каждого множества $\Omega_1, \ldots, \Omega_m$ учитывая только двузначные предикаты.

Приступаем к анализу количественных свойств объектов. В описании каждого из выделенных подмножеств $\Omega_1, \ldots, \Omega_m$ множества Ω находим наибольшее общее свойство его элементов $P_1(\overline{x}_1), \ldots, P_m(\overline{x}_m)$ и упорядочиваем их по количеству аргументов (числовых переменных) в порядке убывания $P_{k_1}(\overline{x}_{k_1}), \ldots, P_{k_m}(\overline{x}_{k_m})$.

Для $i = 1, \ldots, m$ выполняем следующие действия:

1) среди формул $P_{k_{i+1}}(\overline{x}_{k_{i+1}}), \ldots, P_{k_m}(\overline{x}_{k_m})$ устанавливаем формулы с наименьшим количеством аргументов, полученные из $P_{k_i}(\overline{x}_{k_i})$ отождествлением числовых аргументов;

2) из вершины, помеченной Ω_{k_i}, проводим дуги к вершинам, соответствующим найденным формулам.

При добавлении к выделенным подмножествам еще одного подмножества $\tilde{\Omega}$ находим наибольшее общее свойство его элементов $\tilde{P}(\overline{x}_{\sim})$. Определяем его место среди $P_{k_1}(\overline{x}_{k_1}), \ldots, P_{k_m}(\overline{x}_{k_m})$, присваиваем ему соответствующий номер i_0 и увеличиваем на единицу индексы номеров формул с $i \geqslant i_0$. Получаем $P_{k_1}(\overline{x}_{k_1}), \ldots, P_{k_{i_0}}(\overline{x}_{k_{i_0}}), \ldots, P_{k_{m+1}}(\overline{x}_{k_{m+1}})$.

Затем выполняем действия 1, 2 для $i = i_0$.

5. Иллюстративный пример: онтология четырехугольников. Рассмотрим множество четырехугольников. Каждый объект (четырехугольник) задается последовательностью четырех различных имен вершин, соединенных друг с другом в цикл.

Каждый объект характеризуется длинами своих сторон и величинами его внутренних углов*). Описание каждого четырехугольника содержит бесконечнозначные предикаты: $|xy|$ — длина отрезка $[x, y]$ и $\angle x$ — внутренний угол при вершине x.

В связи с этим введем в рассмотрение двузначные предикаты, определяющие соответственно отношения «длина стороны xy равна l» и «внутренний угол при вершине x равен a»:

$$\mathrm{length}(x, y; l) \Leftrightarrow |xy| = l, \qquad \mathrm{angle}(x; a) \Leftrightarrow \angle x = a.$$

Для удобства дальнейшего изложения будем использовать привычные обозначения: $|xy| = l$ и $\angle x = a$. Здесь x и y — предметные аргументы, l и a — количественные аргументы.

Пусть в выборке даны четырехугольники, описания которых изоморфны (с точностью до значений количественных аргументов) четырехугольникам, изображенным на рис. 3, *a–h*.

Так как в выборке присутствуют не только выпуклые четырехугольники, следует ввести еще один предикат $W(x, y, z, w) \Leftrightarrow$ «один из углов четырехугольника (x, y, z, w) больше развернутого». Он позволит на первом же шаге разделить все множество объектов на выпуклые и невыпуклые, так как у первых будет присутствовать литерал вида $\neg W(x, y, z, w)$, а у вторых — литерал вида $W(x, y, z, w)$ (рис. 4).

В этом примере не будем дальше рассматривать множество невыпуклых четырехугольников. В множестве выпуклых четырехугольников все описания содержат литерал $\neg W(x, y, z, w)$ с двузначным предикатом W, поэтому для дальнейшего построения онтологии не будем принимать его во внимание.

---

*) Все углы не равны π (что исключает треугольники с поставленной точкой на одной из сторон и отрезки с поставленными на них двумя точками) и сумма всех внутренних углов равна 2π (что исключает пересечение сторон, не имеющих общих концов).

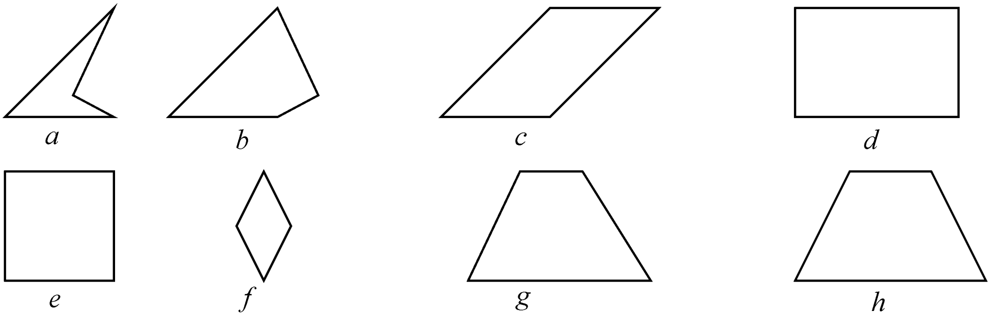

Рис. 3. Вид четырехугольников в выборке

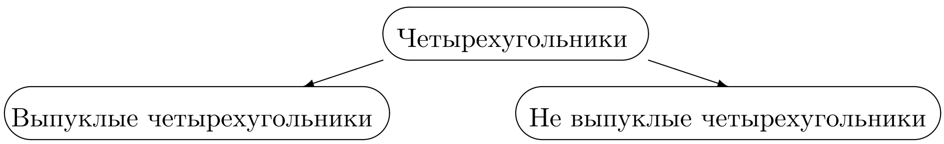

Рис. 4. Начало построения онтологии

Ниже количество входящих в формулу $A(\overline{x};\overline{a})$ числовых переменных будем обозначать посредством N_A.

В общем виде описание выпуклого четырехугольника (т. е. формула, истинная для каждого выпуклого четырехугольника) в выбранной сигнатуре имеет вид

$$\begin{gathered}Q(x,y,z,w;l_1,l_2,l_3,l_4;a,b,c,d) \Leftrightarrow\\ \Leftrightarrow |xy| = l_1 \,\&\, |yz| = l_2 \,\&\, |zw| = l_3 \,\&\, |wx| = l_4 \,\&\\ \&\, \angle x = a \,\&\, \angle y = b \,\&\, \angle z = c \,\&\, \angle w = d,\end{gathered}$$

количество входящих в это описание числовых переменных $N_Q = 8$.

Для параллелограммов:

$$\begin{gathered}P(x,y,z,w;l_1,l_2;a) \Leftrightarrow Q(x,y,z,w;l_1,l_2,l_1,l_2;a,\pi-a,a,\pi-a) \Leftrightarrow\\ \Leftrightarrow |xy| = l_1 \,\&\, |yz| = l_2 \,\&\, |zw| = l_1 \,\&\, |wx| = l_2 \,\&\\ \&\, \angle x = a \,\&\, \angle y = \pi-a \,\&\, \angle z = a \,\&\, \angle w = \pi-a,\end{gathered}$$

количество входящих в это описание числовых переменных $N_P = 3$.

Для прямоугольников:

$$\begin{gathered}R(x,y,z,w;l_1,l_2;) \Leftrightarrow\\ \Leftrightarrow Q(x,y,z,w;l_1,l_2,l_1,l_2;\tfrac{\pi}{2},\tfrac{\pi}{2},\tfrac{\pi}{2},\tfrac{\pi}{2}) \Leftrightarrow\\ \Leftrightarrow |xy| = l_1 \,\&\, |yz| = l_2 \,\&\, |zw| = l_1 \,\&\, |wx| = l_2 \,\&\\ \&\, \angle x = \tfrac{\pi}{2} \,\&\, \angle y = \tfrac{\pi}{2} \,\&\, \angle z = \tfrac{\pi}{2} \,\&\, \angle w = \tfrac{\pi}{2},\end{gathered}$$

количество входящих в это описание числовых переменных $N_R = 2$.

Для квадратов:

$$\begin{gathered}S(x,y,z,w;l;) \Leftrightarrow Q(x,y,z,w;l,l,l,l;\tfrac{\pi}{2},\tfrac{\pi}{2},\tfrac{\pi}{2},\tfrac{\pi}{2}) \Leftrightarrow\\ \Leftrightarrow |xy| = l \,\&\, |yz| = l \,\&\, |zw| = l \,\&|wx| = l \,\&\\ \&\, \angle x = \tfrac{\pi}{2} \,\&\, \angle y = \tfrac{\pi}{2} \,\&\, \angle z = \tfrac{\pi}{2} \,\&\, \angle w = \tfrac{\pi}{2},\end{gathered}$$

количество входящих в это описание числовых переменных $N_S = 1$.

Для ромбов:

$$Rh(x, y, z, w; l; a) \Leftrightarrow Q(x, y, z, w; l, l, l, l; a, \pi - a, a, \pi - a) \Leftrightarrow$$
$$\Leftrightarrow |xy| = l \ \& \ |yz| = l \ \& \ |zw| = l \ \& \ |wx| = l \ \&$$
$$\& \ \angle x = a \ \& \ \angle y = \pi - a \ \& \ \angle z = a \ \& \ \angle w = \pi - a,$$

количество входящих в это описание числовых переменных $N_{Rh} = 2$.

Для трапеций в общем виде*):

$$T(x, y, z, w; l_1, l_2, l_3, l_4; a, b) \Leftrightarrow Q(x, y, z, w; l_1, l_2, l_3, l_4; a, \pi - a, b, \pi - b) \Leftrightarrow$$
$$\Leftrightarrow |xy| = l_1 \ \& \ |yz| = l_2 \ \& \ |zw| = l_3 \ \& \ |wx| = l_4 \ \&$$
$$\& \ \angle x = a \ \& \ \angle y = \pi - a \ \& \ \angle z = b \ \& \ \angle w = \pi - b \ \&$$
$$\& \ l_2 \neq l_4,$$

количество входящих в это описание числовых переменных $N_T = 6$.

Для равнобедренных трапеций:

$$TI(x, y, z, w; l_1, l_2, l_4; a) \Leftrightarrow Q(x, y, z, w; l_1, l_2, l_1, l_4; a, \pi - a, \pi - a, a) \Leftrightarrow$$
$$\Leftrightarrow |xy| = l_1 \ \& \ |yz| = l_2 \ \& \ |zw| = l_1 \ \& \ |wx| = l_4 \ \&$$
$$\& \ \angle x = a \ \& \ \angle y = \pi - a \ \& \ \angle z = \pi - a \ \& \ \angle w = a \ \&$$
$$\& \ l_2 \neq l_4,$$

количество входящих в это описание числовых переменных $N_{TI} = 4$.

Упорядочим формулы, определяющие выделенные выше объекты, в порядке убывания числа количественных переменных:

Объект	Трапеция	Равнобедренная трапеция	Параллелограмм	Ромб	Прямоугольник	Квадрат
Число количественных переменных в описании	6	4	3	2	2	1

Наибольшее число количественных переменных (шесть) имеется в описании трапеции. Среди определяющих ее формул есть те, которые соответствуют равнобедренной трапеции. Среди них нет формул, определяющих какие-либо другие объекты. Поэтому имеем часть онтологии, изображенную на рис. 5, *а*.

Наибольшее число количественных переменных (три) среди оставшихся описаний — в описании параллелограмма. Среди определяющих его формул имеются те, которые задают ромб и прямоугольник (с числом количественных переменных 2). Поэтому имеем часть онтологии, изображенную на рис. 5, *б*.

Во множествах формул, задающих ромб и прямоугольник, есть общее подмножество, задающее квадрат (с числом количественных переменных 1). Поэтому имеем онтологию, изображенную на рис. 5, *в*.

*) На вопрос о том, является ли параллелограм частным случаем трапеции, в разной литературе имеются различные ответы. Будем считать, что это не так, и поэтому добавим конъюнктивный член $l_2 \neq l_4$. Если его не писать, то $P(x, y, z, w; l_1, l_2; a) = T(x, y, z, w; l_1, l_2, l_1, l_2; a)$, что соответствует другой точке зрения и будет построена другая онтология. В частности, на рис. 5, *в* будет добавлена стрелка от «Трапеция» к «Параллелограмм».

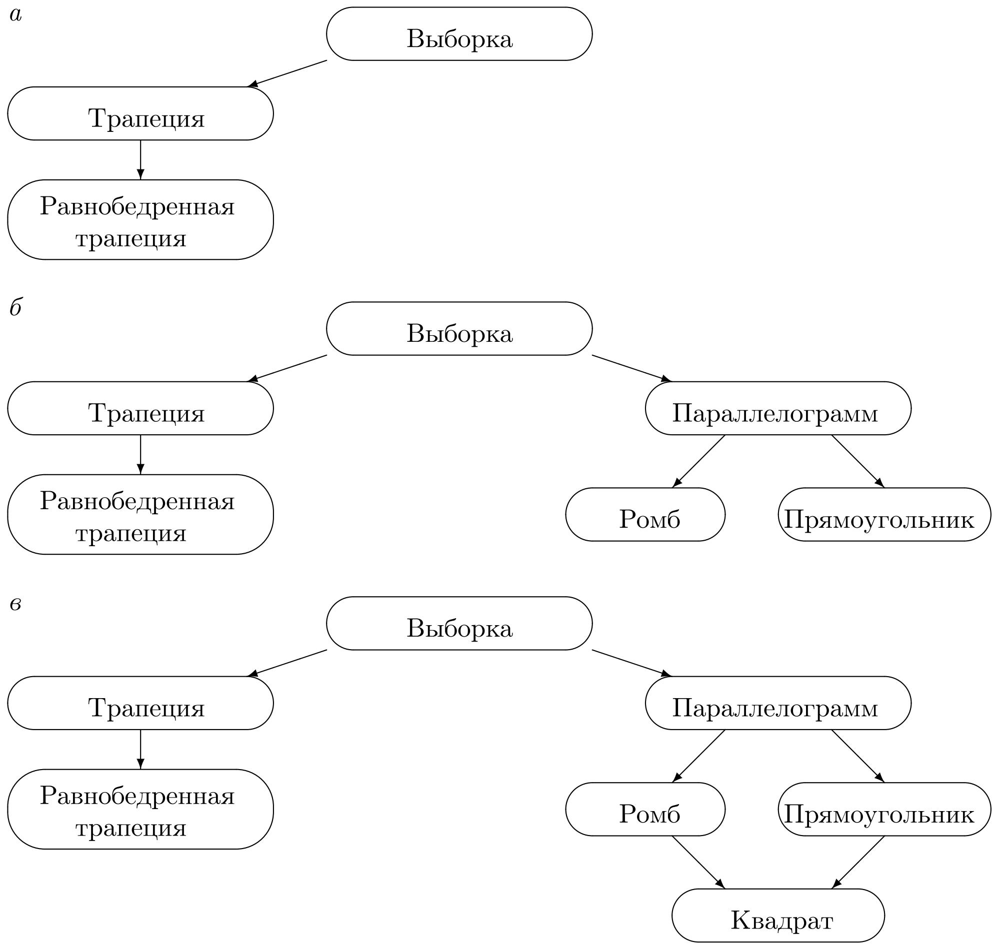

Рис. 5. Построение онтологии (объяснение в тексте)

Продемонстрируем возможность расширения онтологии при выделении в выборке еще одного множества.

Имеется еще одно интересное множество четырехугольников — дельтоиды. Их описания образуют множество формул, которое может содержаться в описании четырехугольников в другой выборке и не совпадать ни с одним из перечисленных выше. Типичный представитель дельтоидов изображен на рис. 6.

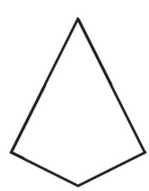

Рис. 6. Дельтоид

Описания этих объектов удовлетворяют формулам

$$D(x,y,z,w;l_1,l_2;a,b,c) \Leftrightarrow Q(x,y,z,w;l_1,l_2,l_2,l_1;a,b,c,a) \Leftrightarrow$$
$$\Leftrightarrow |xy| = l_1 \ \& \ |yz| = l_2 \ \& \ |zw| = l_2 \ \& \ |wx| = l_1 \ \& \ \angle x =$$
$$= a \ \& \ \angle y = b \ \& \ \angle z = c \ \& \ \angle w = a$$

при $l_1 = l_2$ и $l_1 \neq l_2$ (и как следствие при $a = c$ и $a \neq c$ соответственно).

У дельтоида число количественных переменных $N_D = 5$. Поэтому в онтологии это множество четырехугольников сформируется после множества трапеций. Кроме того, если $l_1 = l_2$, то не только $a = c$, но и $a = c = \pi - b$. Следовательно, ромб является частным случаем дельтоида. Поэтому имеем онтологию, изображенную на рис. 7.

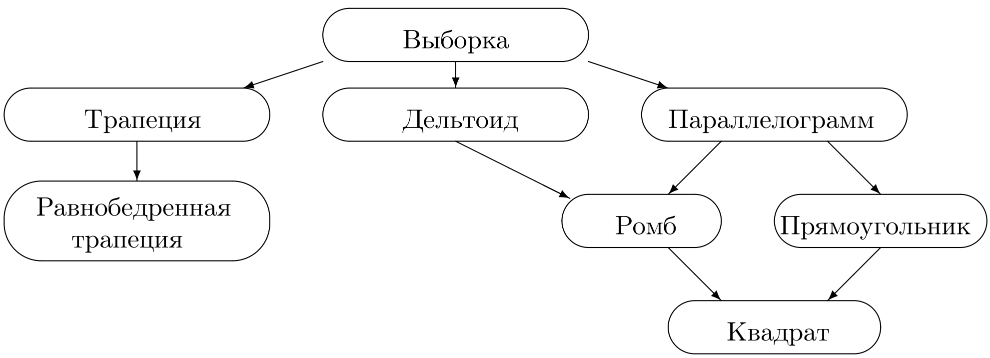

Рис. 7. Расширение онтологии

6. Заключение. В статье описываются постановки трех задач построения логической онтологии в зависимости от того, заданы или нет множества объектов, для которых требуется установить наследственные связи. Для постановки задачи, в которой такие множества заданы, рассмотрены случаи двузначных и многозначных предикатов.

Задача построения логической онтологии, в постановке которой заранее не известны множества объектов, входящих в онтологию и описанных с помощью многозначных предикатов, является наиболее сложной. По-видимому, ее можно свести к построению онтологии, учитывающей только двузначные признаки, а затем внутри каждого из полученных множеств строить онтологию на основе многозначных признаков. При этом возникает задача объединения онтологий.

В дальнейшем внимание будет уделено детальной проработке решения последней задачи.

Литература

1. *Noy N. F., McGuinness D. L.* Ontology development 101: A guide to creating your first ontology. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880. Stanford. March 2001.

2. *Бениаминов Е. М.* Некоторые проблемы широкого внедрения онтологий в IT и направления их решений // Симпозиум «Онтологическое моделирование»: сб. науч. трудов. Москва: ИПИ РАН, 2008. С. 71–82.

3. *Когаловский М. Р., Паринов С. И.* Семантическое структурирование контента научных электронных библиотек на основе онтологий // Современные технологии интеграции информационных ресурсов: сб. науч. трудов. СПб.: Президентская библиотека, 2011. Вып. 2. С. 1–13.

4. *Дяченко О. О., Загорулько Ю. А.* Подход к коллективной разработке онтологий и баз знаний // Знания — Онтологии — Теории. Всерос. конференция с международным участием / гл. ред. Д. Е. Пальчунов. Новосибирск: Ин-т математики им. С. Л. Соболева СО РАН, 2013. С. 141–149.

5. *Mykhailiuk A., Petrenko M.* Machine learning and ontologies as two approaches for building intellectual information systems // Intern. J. "Information Technologies & Knowledge". 2019. Vol. 13. N 1. P. 55–75.

6. *Карпов А. Г., Клемешев В. А., Куранов Д. Ю.* Определение работоспособности системы, структура которой задана графом // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2020. Т. 16. Вып. 1. С. 41–49. https:// doi.org//10.11702/spbu10.2020.104

7. *Гончарова А. Б.* Постановка предварительного медицинского диагноза на основе теории нечетких множеств с использованием меры Сугено // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2019. Т. 15. Вып. 4. С. 529–543. https://doi.org/10.21638/11702/spbu10.2019.409

8. *Kosovskaya T.* Predicate calculus as a tool for AI problems solution: Algorithms and their complexity // Intelligent System. Pt 3. Open access peer-reviewed. Ed. vol. Chatchawal: Chatchawal Wongchoosuk Kasetsart University, 2018. P. 1–20.

9. *Косовская Т. М.* Подход к решению задачи построения многоуровневого описания классов на языке исчисления предикатов // Труды СПИИРАН. 2014. № 3 (34). С. 204–217.

10. *Косовская Т. М., Косовский Н. Н.* Полиномиальная эквивалентность задач, изоморфизм предикатных формул и изоморфизм графов // Вестник Санкт-Петербургского университета. Математика. Механика. Астрономия. 2019. Т. 6 (64). Вып. 3. С. 430–439. https://doi.org/10.21638/11701/spbu10.2019.308

11. *Babai L.* Graph isomorphism in quasipolynomial time (Version 2.1. Unfinished Revision of Version 2 Posted on arXiv May 23, 2017). URL: http://people.cs.uchicago.edu/ laci/17groups/version2.1.pdf (дата обращения: 21.03.2019).

12. *Косовская Т. М., Петров Д. А.* Выделение наибольшей общей подформулы предикатных формул для решения ряда задач искусственного интеллекта // Вестник Санкт-Петербургского университета. Прикладная математика. Информатика. Процессы управления. 2017. Т. 13. Вып. 3. С. 250–263. https://doi.org/10.21638/11702/spbu10.2017.303

Статья поступила в редакцию 21 августа 2021 г.
Статья принята к печати 1 февраля 2022 г.

Контактная информация:

Косовская Татьяна Матвеевна — д-р физ.-мат. наук, проф.; kosovtm@gmail.com

Косовский Николай Николаевич — канд. физ.-мат. наук, доц.; kosovnn@gmail.com

Extraction of common properties of objects for creation of a logic ontology*

T. M. Kosovskaya, N. N. Kosovskii

St Petersburg State University, 7–9, Universitetskaya nab., St Petersburg,
199034, Russian Federation

For citation: Kosovskaya T. M., Kosovskii N. N. Extraction of common properties of objects for creation of a logic ontology. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2022, vol. 18, iss. 1, pp. 37–51. https://doi.org/10.21638/11701/spbu10.2022.103 (In Russian)

The paper describes an approach to the formation of ontology based on descriptions of objects in terms of the predicate calculus language. With this approach, an object is represented as a set of its elements, on which a set of predicates is defined that defines the properties of these elements and the relationship between them. A description of an object is a conjunction of literals that are true on elements of an object. In the present work, ontology is understood as

* This work was carried out with the financial support of St Petersburg State University (project N 73555239).

an oriented graph with descriptions of subsets as nodes and such that the elements of a set at the end of an oriented edge have the properties of the elements of the set at the beginning of this edge. Three settings of an ontology construction problem are considered: 1) all predicates are binary and subsets of the original set of objects are given; 2) all predicates are binary and it is required to find subsets of the original set; 3) among the predicates there are many-valued ones and subsets of the original set of objects are given. The main tool for construction such a description is to extract an elementary conjunction of literals of predicate formulas that is isomorphic to subformulas of some formulas. The definition of an isomorphism of elementary conjunctions of atomic predicate formulas is given. The method of ontology construction is formulated. An illustrative example is provided.

Keywords: logic ontology, predicate formula, isomorphism of predicate formulas.

References

1. Noy N. F., McGuinness D. L. *Ontology development 101: A guide to creating your first ontology.* Stanford Knowledge Systems Laboratory Technical Report KSL-01-05 and Stanford Medical Informatics Technical Report SMI-2001-0880. Stanford, March 2001.

2. Beniaminov E. M. Nekotorye problemy shirokogo vnedreniia ontologii v IT i napravleniia ikh reshenii [Some problems of the widespread introduction of ontologies in IT and the directions of their solutions]. *Simpozium "Ontologicheskoe modelirovanie"*. Sbornik nauchnykh trudov [*Symposium "Ontological modeling"*. A collection of scientific papers]. Moscow, Institute IPI RAN Publ., 2008, pp. 71–82. (In Russian)

3. Kogalovskii M. R., Parinov S. I. Semanticheskoe strukturirovanie kontenta nauchnykh elektronnykh bibliotek na osnove ontologii [Semantic structuring of the content of scientific electronic libraries based on ontologies]. *Sovremennye tekhnologii integratsii informatsionnykh resursov.* Sbornik nauchnykh trudov [*Modern technologies for the integration of information resources.* A collection of scientific papers]. St Petersburg, President Library Publ., 2011, iss. 2, pp. 1–13. (In Russian)

4. Diachenko O. O., Zagorulko Yu. A. Podkhod k kollektivnoi razrabotke ontologii i baz znanii [An approach to the collective development of ontologies and knowledge bases]. *Znaniia — Ontologii — Teorii. Vserossiiskaia konferentsiia s mezhdunarodnym uchastiem* [*Knowledge — Ontologies — Theories. All-Russian Conference with international participation*]. Eds. by D. E. Palchunov. Novosibirsk, S. L. Sobolev Institute of Mathematics SO RAN Publ., 2013, pp. 141–149. (In Russian)

5. Mykhailiuk A., Petrenko M. Machine learning and ontologies as two approaches for building intellectual information systems. *Intern. J. "Information Technologies & Knowledge"*, 2019, vol. 13, no. 1, pp. 55–75.

6. Karpov A. G., Klemeshev V. A, Kuranov D. Yu. Opredelenie pabotosposobnosti sistemy, struktura kotoroy zadana grafom [Determining the ability to work of the system, the structure of which is given using graph]. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2020, vol. 16, iss. 1, pp. 41–49. https://doi.org//10.11702/spbu10.2020.104 (In Russian)

7. Goncharova A. B. Postanovka predvaritelnogo meditsinskogo diagnoza na osnove teorii nechetkih mnozhestv s ispolzovaniem mery Sugeno [Preliminary medical diagnostics based on the fuzzy sets theory using the Sugeno measure]. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2019, vol. 15, iss. 4, pp. 529–543.
https://doi.org//10.21638/11702/spbu10.2019.409 (In Russian)

8. Kosovskaya T. Predicate calculus as a tool for AI problems solution: Algorithms and their complexity. *Intelligent System. Pt 3. Open access peer-reviewed.* Ed. vol. Chatchawal, Chatchawal Wongchoosuk Kasetsart University Press, 2018, pp. 1–20.

9. Kosovskaya T. M. Podhod k resheniyu zadachi postroeniya mnogourovnevogo opisaniya klassov na yazyke ischisleniya predikatov [An approach to the construction of a level description of classes by means of a predicate calculus language]. *SPIIRAS Proceedings*, 2014, no. 3(34), pp. 204–217. (In Russian)

10. Kosovskaya T. M., Kosovskii N. N. Polinomialnaya ekvivalentnost zadach izomorfizm predikatnyh formul i izomorfizm grafov [Polynomial equivalence of the problems predicate formulas isomorphism and graph isomorphism]. *Vestnik of Saint Petersburg University. Mathematics. Mechanics. Astronomy*, 2019, vol. 6(64), iss. 3, pp. 430–439. https://doi.org//10.21638/11701/spbu10.2019.308 (In Russian)

11. Babai L. *Graph isomorphism in quasipolynomial time (Version 2.1. Unfinished Revision of Version 2 Posted on arXiv May 23, 2017).* Available at:
http://people.cs.uchicago.edu/ laci/17groups/version2.1.pdf (accessed: March 21, 2019).

12. Kosovskaya T. M., Petrov D. A. Vydelenie naibolshey obschty podformuly predikatnyh formul

dlya resheniya ryada zadach iskusstvennogo intellekta [Extraction of a maximal common sub-formula of predicate formulas for the solving of some artificial intelligence problems]. *Vestnik of Saint Petersburg University. Applied Mathematics. Computer Science. Control Processes*, 2017, vol. 13, iss. 3, pp. 250–263. https://doi.org//10.21638/11702/spbu10.2017.303 (In Russian)

Received: August 21, 2021.
Accepted: February 01, 2022.

Authors' information:

Tatiana M. Kosovskaya — Dr. Sci. in Physics and Mathematics, Professor; kosovtm@gmail.com

Nikolai N. Kosovskii — PhD in Physics and Mathematics, Associate Professor; kosovnn@gmail.com