

## Терминологический граф контекстно-ориентированной ОНТОЛОГИИ

Г. В. Каньгин, М. С. Полтинникова

СИ РАН - филиал ФНИЦС РАН 190005, Санкт-Петербург, ул. 7-я Красноармейская, д. 25/14  
e-mail: g.kanygin@gmail.com, maria.poltinnikova@gmail.com

Поступила в редколлегию 02.06.2021

**Аннотация**—Статья продолжает цикл работ о математической модели контекстно-ориентированной онтологии: описывается построение терминологического графа КО онтологии и связанные с этим построением метод вычисления и алгоритмы. Кроме того, описывается новое правило согласования контекстов: правило контекстного домена и алгоритм построения контекстного домена в процессе построения терминологического графа. Построение терминологического графа пошагово проиллюстрировано на примере концептуального разъяснения определения понятия ветвления. В примере показано, как работает разделение генерируемого графа на разные смысловые подграфы с помощью контекстов. Приведён результат работы тестовой программы DIAGOGUE, которая использует описанные алгоритмы.

**КЛЮЧЕВЫЕ СЛОВА:** Графовая контекстно-ориентированная онтология, математическая модель, ветвление, терминологический граф, правило контекстного домена, концептуальное разъяснение.

### 1. ВВЕДЕНИЕ

В управлении социально-экономическими системами широкое применение находят системы управления знаниями, в которых строятся онтологии предметных областей, помогающие управлять большим объёмом структурированной информации. Построение предметных онтологий моделируется в стиле объектно-ориентированного программирования и требует постоянного взаимодействия экспертов предметной области и программистов, создающих языки онтологий [1–4]. Основная трудность такого взаимодействия состоит в том, чтобы построить корректный и удобный для экспертов изоморфизм между естественно-языковым изложением и формальным языком спецификаций, на основе которого строится онтология [1, 2].

Наряду с информатикой трудности формализации естественно-языковых знаний встречаются исследователи, работающие в области качественных методов социологического исследования [5–9]. Разработка компьютерных инструментов, способных усовершенствовать смысловое взаимодействие социолога и его информантов в процессе выражения и переформулирования экспертного знания привели к созданию компьютерных программ анализа качественных данных, например, MAXQDA, Atlas ti и другие.

Мы предлагаем строить математическую модель онтологии так, чтобы понятия вводились только в паре с контекстами, в которых они используются [10–13]. Предложенные в этих работах графовые контекстно-ориентированные онтологические методы управления знаниями представляют собой компьютерные техники, позволяющие экспертам излагать свой опыт в виде произвольных словесных высказываний, и согласовывать их посредством онтологических отношений.

Одним из основных алгоритмов контекстно-ориентированной онтологии является алгоритм построения терминологического графа на базе различных правил согласования контекстов. Статья описывает этот алгоритм и демонстрирует пример его работы. Кроме того, описывается правило согласования контекстов – правило контекстного домена.

## 2. ОПРЕДЕЛЕНИЕ КО ОНТОЛОГИИ

Опишем формальные термины и структуры графовой контекстно-ориентированной онтологии [10].

*Понятие* — это аналитическое обозначение объекта, которое выражено словом или словосочетанием и обозначается буквой с индексом, например,  $t_n$ .

*Словарь* — это неупорядоченное множество понятий  $V = \{t_1, t_2, \dots, t_q\}$ , где  $q$  — количество понятий в словаре.

Предполагается, что автор онтологии явно указывает ситуацию или контекст, в котором он рассматривает новое аналитическое обозначение, поэтому описание объектов реализуется упорядоченными парами понятий множества  $V$ : первое понятие для обозначения объекта, второе для обозначения контекста, в котором этот объект рассматривается. Первое понятие пары будем называть *термином*, а второе *контекстом*. Одно и то же понятие может играть роль как термина, так и контекста, в зависимости от его места в конкретной паре. Множество упорядоченных пар понятий словаря обозначим буквой  $P$ .

В процессе определения нового понятия в некотором контексте автор должен объяснять, через какие уже введённые понятия он раскрывает смысл этого нового понятия. Такие объяснения порождают двухуровневые ориентируемые деревья.

**Определение.** Граф  $E$  назовем *ветвлением*, если он представляет собой или отдельную вершину  $p_0 \in P$  (вырожденный случай), или двухуровневое ориентированное дерево с корневой вершиной  $p_0 = (t_x, t_y) \in P$  и листьями  $p_1 = (t_{x_1}, t_{y_1}), \dots, p_s = (t_{x_s}, t_{y_s}) \in P$ .

Для ветвления  $E$  будем использовать следующее обозначение:  $E : p_0 \rightarrow \{\dots\}$  или  $E : (t_x, t_y) \rightarrow \{\dots\}$  в случае отдельной вершины и  $E : p_0 \rightarrow \{p_1, \dots, p_s\}$  или  $E : (t_x, t_y) \rightarrow \{(t_{x_1}, t_{y_1}), \dots, (t_{x_s}, t_{y_s})\}$  в случае дерева. Здесь  $s$  — количество вершин на втором уровне,  $s < q$ .

**Определение.** Пару  $(t_x, t_y)$  назовем *головной*, пару  $(t_{x_k}, t_{y_k})$  — *разъяснением*, а множество пар  $\{(t_{x_k}, t_{y_k})\}, k = 1, \dots, s$ , — *множеством разъяснений*. Тогда  $t_x$  — *термин головной пары*,  $t_y$  — *контекст головной пары*,  $t_{x_k}$  — *термин разъяснения*,  $t_{y_k}$  — *контекст разъяснения*.

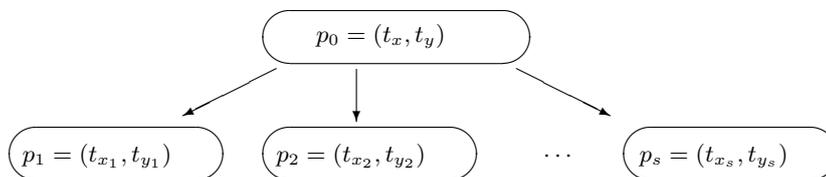


Рис. 1. Двухуровневое ориентированное дерево, задающее ветвление.

**Пример 1.** Обозначения в определении понятия *ветвление* можно описать с помощью двухуровневого ориентированного дерева:

(обозначение, ветвление)  $\rightarrow \{(E, \text{ветвление}); (p_0 \rightarrow \{\dots\}, \text{ветвление}); (p \rightarrow \{p_1, p_2, \dots, p_s\}, \text{ветвление}); ((t_x, t_y) \rightarrow \{(t_{x_1}, t_{y_1}), \dots, (t_{x_s}, t_{y_s})\}, \text{ветвление}); ((t_x, t_y) \rightarrow \{\dots\}, \text{ветвление})\}$ .  
Здесь пара (обозначение, ветвление) – головная пара, множество разъяснений состоит из пяти разъяснений. Контекст в головной паре и в разъяснениях одинаковый.

Введём следующие обозначения:  $\mathcal{E}$  — множество всех введённых автором ветвлений;  $\mathcal{P}$  — множество всех пар понятий, являющихся вершинами ветвлений множества  $\mathcal{E}$ ,  $\mathcal{C}$  — множество всех введённых автором контекстов.

**Определение.** Множество ветвлений  $\mathcal{E}$  задаёт *контекстно-ориентированный тезаурус* сокращенно *КО тезаурус*. КО тезаурус в совокупности с алгоритмами построения его разнообразных подграфов называется *контекстно ориентированной онтологией КО онтологией*.

### 3. ТЕРМИНОЛОГИЧЕСКИЙ ГРАФ КО ОНТОЛОГИИ

Терминологический граф представляет собой подграф онтологии, который начинается с усечённого по некоторому правилу ветвления. "Детали конструктора" для построения терминологического графа представляют собой ветвления, заданные пользователем в тезаурусе. На входе построения Т-графа задаётся корневой узел  $p_0 = (x_0, y_0)$  и  $f_1(y_0)$  — *правило согласования контекстов*, по которому выбираются контексты второго уровня. Программа построения Т-графа ищет ветвление  $E_1$  с головной парой  $p_0$  и, если необходимо, отсекает его в силу правила  $f_1$ . Получается граф  $E_1(f_1)$ . Этот граф будет началом в построении терминологического графа. Далее просматриваются вершины из множества разъяснений графа  $E_1(f_1)$ , с которыми выполняются те же действия, что и с вершиной  $p_0$  и так до тех пор, пока есть, что добавить.

#### 3.1. Операция усечения ветвления.

Рассмотрим ветвление

$$E : (x_0, y_0) \rightarrow \{(x_1, y_1), \dots, (x_s, y_s)\}.$$

Когда мы имеем дело с понятием  $x_0$  в контексте  $y_0$ , нас интересуют не все разъяснения, а только те, что заданы в интересующих нас контекстах. Эти контексты задаются с помощью правила согласования контекстов  $f$  следующим образом:  $f(y_0) = \{c_1, \dots, c_r\}$ . Это значит, что из множества разъяснений  $\{(x_1, y_1), \dots, (x_s, y_s)\}$  выбирается подмножество с контекстами из множества  $C = f(y_0) \cap \{y_1, y_2, \dots, y_s\}$ , получается новое ветвление  $E(f)$ .

Опишем функцию  $E((x_0, y_0), f)$ , которая находит ветвление с началом в вершине  $(x_0, y_0)$  и отсекает его по правилу согласования контекстов  $f(y_0) = \{c_1, c_2, \dots, c_r\}$ .

Вход:  $(x_0, y_0), f$ .

- 1: Находим в тезаурусе ветвление  $(x_0, y_0) \rightarrow \{(x_1, y_1), \dots, (x_s, y_s)\}$ . Если такого ветвления нет,  $E((x_0, y_0), f) = \{(x_0, y_0)\}$ .
- 2: Выбираем контексты из множества  $C = f(y_0) \cap \{y_1, y_2, \dots, y_s\}$  и оставляем в  $E$  разъяснения с этими контекстами, получаем  $E(f)$ .

Рассмотрим структуру:

(ветвление, ИКТ)  $\rightarrow \{(\text{ветвление, мат. модель КО онтологии}); (\text{ветвление, Паскаль}); (\text{ветвление, C++})\}$ .

**Пример 2.** Если  $f(\text{ИКТ}) = \{\text{Паскаль, C++}\}$ , то после усечения ветвления получим: (ветвление, ИКТ)  $\rightarrow \{(\text{ветвление, Паскаль}); (\text{ветвление, C++})\}$ .

**Пример 3.** Если  $f(\text{ИКТ}) = \{\text{мат. модель КО онтологии}\}$ . Тогда после усечения ветвления получим (ветвление, ИКТ)  $\rightarrow \{(\text{ветвление, мат. модель КО онтологии})\}$ .

## 3.2. Терминологический граф.

В построении терминологического графа ключевую роль играет правило согласования контекстов. Именно оно определяет, какие ветвления участвуют в построении. В работе [10] оно задавалось заранее и не менялось в процессе построения. В этой работе оно вычисляется на каждом новом уровне построения терминологического графа. В процессе построения получается не только  $n$ -уровневый Т-граф, но и последовательность из правил согласования контекстов  $f_1, \dots, f_{n-1}$ . Здесь  $f_k$  — это правило, которое мы применили при построении уровня  $k + 1$ .

Рассмотрим согласование контекстов при построении нового уровня терминологического графа. Для этого опишем алгоритмически создаваемое правило согласования контекстов, применяемое при построении терминологического графа. Усечение ветвления, которое будет добавлено в Т-граф происходит автоматически по правилу: разъяснение остаётся, если контекст этого разъяснения входит в список понятий, которые имеются в узлах уже построенной части Т-графа как в качестве терминов так и в качестве контекстов. Такой список понятий называется *контекстным доменом*. А само правило — *правилом контекстного домена*.

Разумность этого правила следует из того, что перед использованием понятия  $y$  в качестве контекста, т.е. обозначения условий существования другого понятия  $x$ , автор должен определить это  $y$ , если же оно стоит в качестве термина или контекста в построенной части графа, то о его определении можно не беспокоиться. Исключение составляет построение первого уровня Т-графа, когда сам пользователь должен решить в каком контексте он разъясняет понятие с помощью терминологического графа (см. примеры 2 и 3).

## 3.3. Построение нового уровня терминологического графа.

Имеется терминологический граф  $G_k$ , построенный в соответствии с последовательностью правил  $f_1, \dots, f_{k-1}$ . Сформулируем правило  $f_k$  и нарастим в силу этого правила граф  $G_k$  до графа  $G_{k+1}$ . Обозначим эту процедуру  $L(G_k, f_1, \dots, f_{k-1})$ .

Обозначим последний уровень ветвлений  $E_1(f_{k-1}), \dots, E_r(f_{k-1})$ , предпоследний уровень разъяснений через  $(T_1, C_1), (T_2, C_2), \dots, (T_r, C_r)$ , а через  $(t_1, c_1), (t_2, c_2), \dots, (t_s, c_s)$ , где  $s \geq r$  — последний уровень разъяснений.

Вход:  $G_k$  и  $f_1, \dots, f_{k-1}$ .

1:  $f_k(c_j) = f_{k-1}(C_i) \cup \{t_j, c_j\}, i = 1, \dots, r$ , а  $c_j$  — контекст разъяснения из  $E_i$ .

2: В цикле для каждого  $j = 1 \dots s$  строится  $g_j = E((t_j, c_j), f_k)$ , каждый  $g_j$  склеивается своей вершиной  $(t_j, c_j)$  с листом  $(t_j, c_j)$  графа  $G_k$ .

Выход: Получаем граф  $G_{k+1}$ , если хотя бы один из  $g_j$  не нулевой, и  $G_k$  в противном случае.

**Пример 4.** Если задать  $f_1$  как в примере 3 и в тезаурусе имеется структура (ветвление, мат. модель КО онтологии)  $\rightarrow$  {(определение, ветвление); (обозначение, ветвление)}, то будет построено правило  $f_2$  (мат. модель КО онтологии) = {ИКТ, ветвление, мат. модель КО онтологии} и терминологический граф, изображенный на рисунке 2.

Далее опишем алгоритм построения терминологического графа уровня  $n$ , для чего зададим вершину графа  $(t_0, c_0)$  и правило согласования контекстов  $f_1$ .

Вход:  $(t_0, c_0), f_1$ .

G1: Строим второй уровень терминологического графа:  $G_2 = E((t_0, c_0), f_1)$ .  $n := 2$  и переходим к шагу Gn.

Gn: В цикле  $n := n + 1$  и  $L(G_n, f_1, \dots, f_{n-1})$  для построения  $f_n$  и  $G_{n+1}$ .

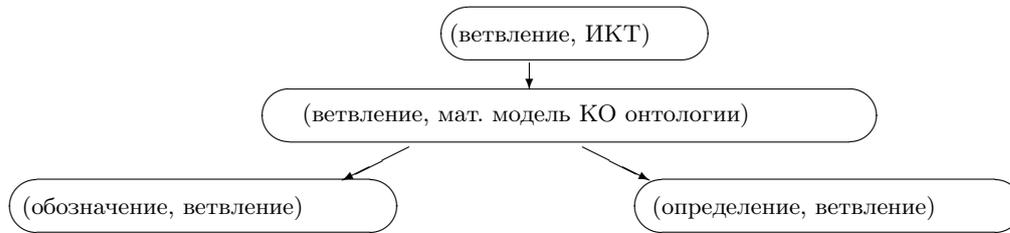


Рис. 2. Трёхуровневый граф.

Выход: Выход из цикла происходит, если нового уровня нет, результат — это  $G_n, f_1, \dots, f_{n-1}$ .

Мы не рассматриваем ситуацию, когда построение приводит к 0-графу  $(t_0, c_0)$ , поскольку считаем, что пользователь осмысленно задаёт правило  $f_1$  и получает хотя бы два уровня.

Алгоритм конечен, если в процессе построения отсутствуют циклы. В противном случае вычисления необходимо принудительно закончить и предложить пользователю скорректировать определения каких-либо понятий, поскольку закливающиеся определения некорректны. Вообще говоря получается, что мы имеем не алгоритм, а метод вычисления.

#### 4. ПРИМЕР ПОСТРОЕНИЯ Т-ГРАФА

Построим граф, разъясняющий понятие **ветвление** в рамках математической модели КО онтологии. Переходы вдоль рёбер графа поясняют понятие **ветвление** в контексте **мат. модель КО онтологии** через понятия, которые приводят к общематематическим понятиям, что решает задачу построения терминологического графа определения понятия **ветвление** и отсекает смыслы, используемые, например, в программировании.

Начальный узел этого построения (**ветвление, ИКТ**) и начальное правило согласования контекстов  $f_1(\text{ИКТ}) = \{\text{мат. модель КО онтологии}\}$ . Получаем двухуровневый терминологический граф.

Следующая структура уже определялась в примере 3: (**ветвление, мат. модель КО онтологии**)  $\rightarrow$  {(**обозначение, ветвление**); (**определение, ветвление**)}. Правило согласования контекстов  $f_2(\text{мат. модель КО онтологии}) = \{\text{ветвление, ИКТ, мат. модель КО онтологии}\}$ .

В построении четвёртого уровня графа имеем:  $f_3^1(\text{ветвление}) = \{\text{мат. модель КО онтологии, ветвление, ИКТ, определение}\}$  для одной ветки и  $f_3^2(\text{ветвление}) = \{\text{мат. модель КО онтологии, ветвление, обозначение, ИКТ}\}$  для другой.

Ветвление может обозначаться буквой  $E$  или непосредственно через узлы  $p \rightarrow \{p_1, p_2, \dots, p_s\}$  и определяется через двухуровневое ориентированное дерево специального вида в содержательном случае или нуль-дерево в вырожденном случае, поэтому можно задать следующие структуры:

(**обозначение, ветвление**)  $\rightarrow$  {( $E$ , **ветвление**); ( $p \rightarrow \{p_1, p_2, \dots, p_s\}$ , **ветвление**)}.

(**определение, ветвление**)  $\rightarrow$  {(2-х уровневое ориент. дерево, **ветвление**); (0-дерево, **ветвление**)}.

Добавим к терминологическому графу на рисунке 2 эти два ветвления. Получим терминологический граф четвёртого уровня, изображенный на рисунке 3.

Далее строится  $f_4^1(\text{ветвление}) = \{\text{мат. модель КО онтологии, ветвление, ИКТ, определение, 2-х уровневое ориент. дерево}\}$  для одной ветки и  $f_4^2(\text{ветвление}) = \{\text{мат. модель КО онтологии, ветвление, ИКТ, определение, 0-дерево}\}$  для другой и для построения  $G_5$  добавляются два ветвления (множество вершин, 2-х уровневое ориент. дерево)  $\rightarrow$  {(головная пара, **ветвле-**

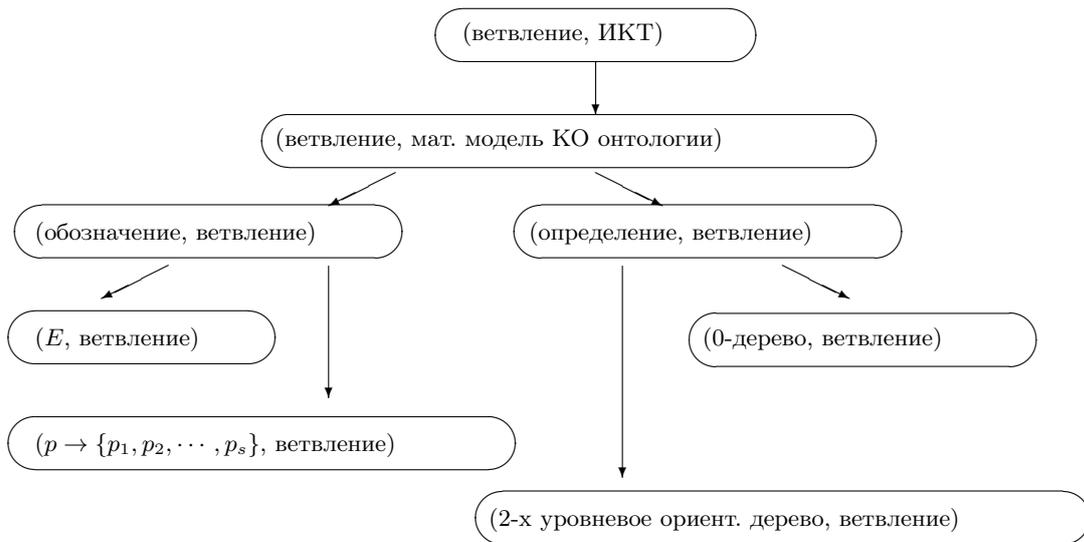


Рис. 3. Четырёхуровневый терминологический граф, выходящий из вершины (ветвление, ИКТ).

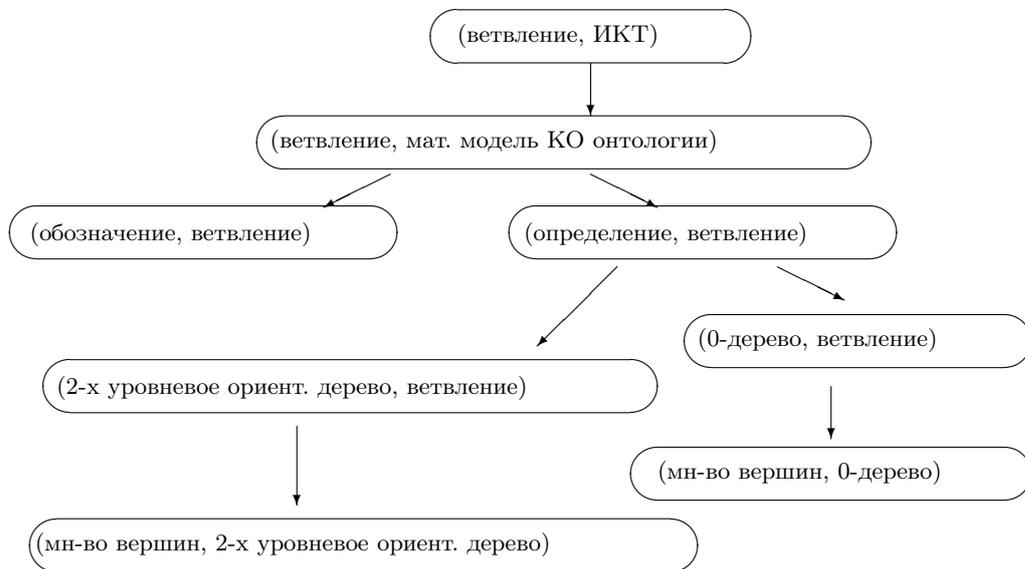
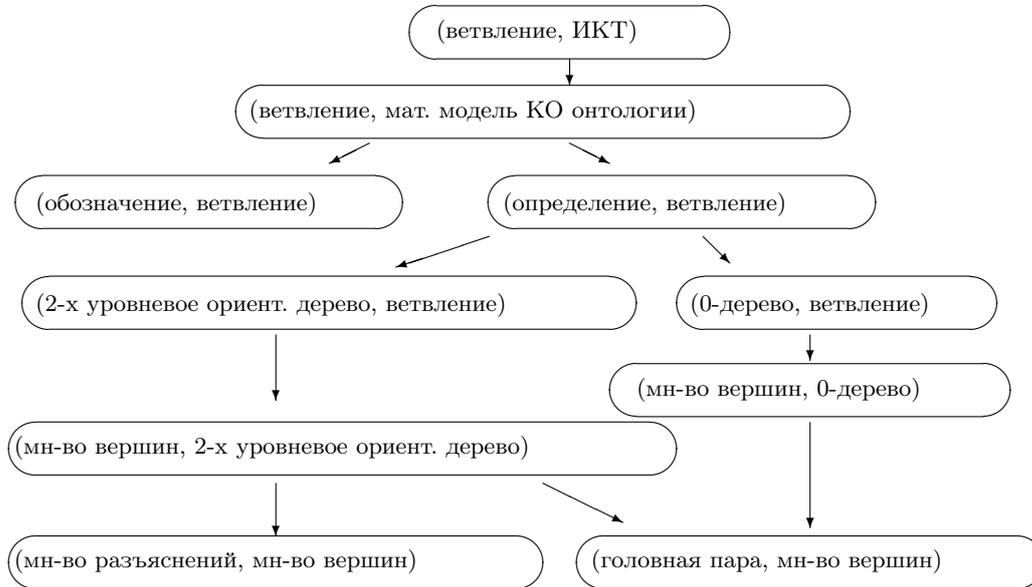


Рис. 4. Пятиуровневый терминологический граф, выходящий из вершины (ветвление, ИКТ), в котором скрыта ветка с обозначениями.

ние); (множество разъяснений, ветвление)}; (множество вершин, 0-дерево)  $\rightarrow$  {(головная пара, ветвление)}; (см. рис. 4).

Затем строится  $f_5(\text{ветвление}) = \{\text{мат. модель КО онтологии, ветвление, ИКТ, определение, 2-х уровневое ориент. дерево}\}$  для одной ветки и  $f_5(\text{ветвление}) = \{\text{мат. модель КО онтологии, ветвление, ИКТ, определение, 0-дерево}\}$  для другой и  $G_6$  с ещё двумя ветвлениями, показанными на рис. 5. Этот граф имеет допустимый замкнутый контур.



**Рис. 5.** Шестиуровневый терминологический граф, выходящий из вершины (ветвление, ИКТ), в котором скрыта ветка с обозначениями.

Дальнейшие разъяснения приводят, например, к ветвлениям вида:

(головная пара, множество вершин)  $\rightarrow$  {(определение, головная пара); (обозначение, головная пара)};

(определение, головная пара)  $\rightarrow$  {(термин, контекст), ветвление)};

(обозначение, головная пара)  $\rightarrow$  {( $t_x, t_y$ ), ветвление)};

(множество разъяснений, множество вершин)  $\rightarrow$  {(определение, множество разъяснений);

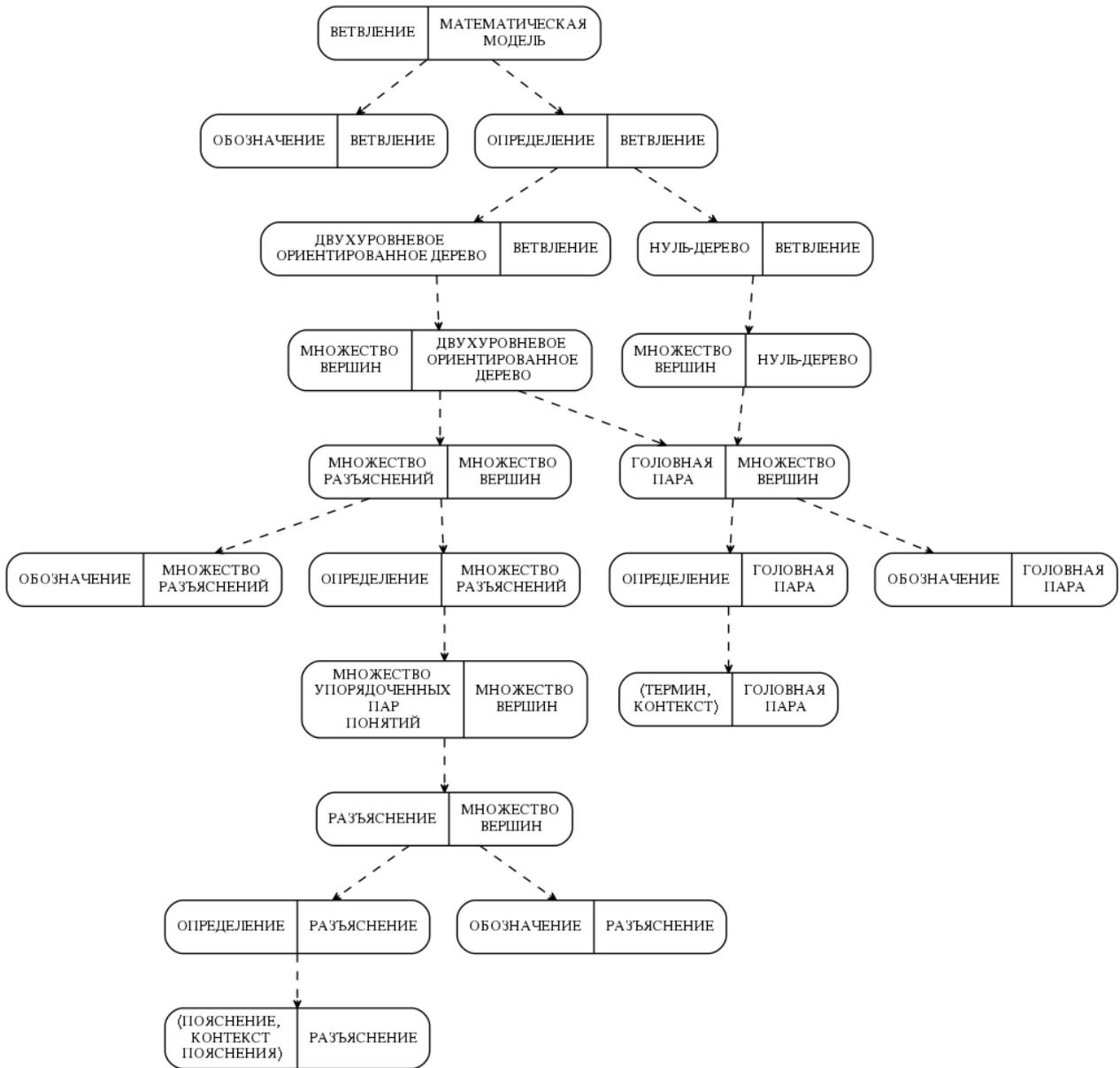
(обозначение, множество разъяснений)};

(обозначение, множество разъяснений)  $\rightarrow$  {( $\{(t_{x_k}, t_{y_k})\}, k = 1 \dots n$ , ветвление)}.

На рисунке 6 показан T-граф разъясняющий понятие ветвление в контексте математическая модель, в котором скрыта ветка с обозначениями ветвления. Этот граф построен с помощью приложения DIAGOGUE, которое реализует описываемый в данной работе алгоритм. Кроме того, в DIAGOGUE автоматически создаётся dot-файл T-графа, который отрисовывается потом с помощью программы GraphViz. DIAGOGUE позволяет скрывать ветки T-графа, которые не интересны пользователю и рисовать терминологический граф с необходимой для автора степенью подробности.

## 5. ЗАКЛЮЧЕНИЕ

Терминологический граф позволяет собирать в единую структуру большое количество разрозненных аспектов знания в автоматизированном режиме. Эти аспекты знания представлены пользователем в виде ветвлений, которые представляют собой элементарные отношения между понятиями. Само построение может выполняться с необходимой степенью подробности.



**Рис. 6.** Результат работы программы DIAGOGUE построения Т-графа из вершины (ветвление, математическая модель), в котором скрыты разъяснения обозначений ветвления.

Пользователь получает возможность осуществлять визуальную проверку связности знания о чем-либо в процессе его формирования.

#### СПИСОК ЛИТЕРАТУРЫ

1. Рубашкин В.Ш. Онтологическая семантика. Знания. Онтологии. Онтологически ориентированные методы информационного анализа текстов. /М., Физматлит 2013. 348 с.
2. Гаврилова Т. А., Хорошевский В. Ф. Базы знаний интеллектуальных систем. / Питер СПб., 2000. 384 с.
3. Бениаминов Е.М., Лапшин В.А. Уровни представлений онтологий, языки, математические модели и проект Веб-сервера онтологий в стиле Веб 2.0. *Научно-техническая информация, Серия 2: Информационные процессы и системы*, 2012, № 3, стр. 1-10.
4. Бениаминов Е.М. Библиотеки онтологий в Веб. Состояние и перспективы. *Научно-техническая информация, Серия 2: Информационные процессы и системы*, 2018, № 5, стр. 5-8.
5. Вебер М. О некоторых категориях понимающей социологии. В кн. *Избранные произведения*. Под ред. Ю.Н. Давыдова. М.: Прогресс, 1990, стр. 494-546.
6. Алле М. Современная экономическая наука и факты. *THESIS*, 1994, том 2, вып. 4, стр. 11-19.
7. Андреенков В.Г., Орлов А.И., Толстова Ю.Н. *Анализ нечисловой информации в социологических исследованиях*. М.: Наука, 1985.
8. Charmaz K. Grounded theory: Objectivist and constructive methods. In: *Handbook of Qualitative Research*. Ed. Norman K. Denzin, Yvonna S. Lincoln Thousand Oaks, Ca.: Sage, M., 2000. pages 509-535.
9. Ядов В. А. Стратегии и методы качественного анализа данных. *Социология: методология, методы, математическое моделирование (Социология:4М)*. 1991. Том 0. № 1. стр. 14-31.
10. Каныгин Г.В., Полтинникова М.С. Контекстно-ориентированные онтологические методы в социологии. *Труды СПИИРАН*. 2016. № 5(48) стр. 107-124.
11. Каныгин Г.В., Полтинникова М.С. Контекстно-ориентированная онтология как основа документооборота в социальном управлении. *XVIII объединенная конференция «Интернет и современное общество» (IMS-2015) Компьютерная лингвистика и вычислительные онтологии: сборник научных статей*. СПб: Университет ИТМО. 2015. стр. 94-103.
12. Poltinnikova M. & Kanygin G. Graphic Ontological Methods to Compile Collective Knowledge of Social Processes. *Electronic Governance and Open Society: Challenges in Eurasia* SPb: ITMO University, 2016, pp 223-228.
13. Kanygin G., Koretckaia V. Analytical Coding: Performing Qualitative Data Analysis Based on Programming Principles *The Qualitative Report*. 2021, 26(2), 316-333.

### Terminological graph of context-oriented ontology

**Kanygin G. V., Poltinnikova M. S.**

The article continues a series of works on the mathematical model of a context-oriented ontology: the construction of a terminological graph of a CO ontology and the associated calculation method and algorithms are described. In addition, a new rule for matching contexts is described: the rule of the context domain and the algorithm for constructing the context domain in the process of constructing a terminological graph. The construction of a terminological graph is illustrated step by step using the example of a conceptual explanation of the definition of the concept of branching. The example shows how the division of the generated graph into different semantic subgraphs using contexts works. The result of the work of the DIAGOGUE test program, which uses the described algorithms, is given.

**KEYWORDS:** Graph context-oriented ontology, mathematical model, branching, terminological graph, context domain rule, conceptual explanation.