



УДК 004.822:514

МОДЕЛИ И АЛГОРИТМЫ ИНТЕГРАЦИИ ЗНАНИЙ НА ОСНОВЕ ОДНОРОДНЫХ СЕМАНТИЧЕСКИХ СЕТЕЙ

Ивашенко В.П.

*Белорусский государственный университет информатики и радиоэлектроники,
г. Минск, Республика Беларусь*

ivashenko@bsuir.by

Рассматриваются модели и алгоритмы интеграции знаний, представленных в унифицированном виде однородными семантическими сетями, имеющими теоретико-множественную интерпретацию, учитывающие неполноту, нечёткость, неопределённость представленных знаний и их изменение со временем

Ключевые слова: представление знаний, онтология, семантическая сеть, интеграция, нечёткие множества.

ВВЕДЕНИЕ

В системах, основанных на знаниях [Гаврилова и др., 2000], одной из важных задач является задача интеграции знаний, заключающаяся в выявлении фрагментов семантической сети, представляющих одни и те же знания, и их слиянии, что позволяет исключить избыточные элементы семантической сети, способные затормозить до предела работу системы, или сократить их количество, увеличить число ассоциаций для конкретных понятий, уточнить структуру модели предметной области. Задача интеграции необходима для обеспечения производительности и развития интеллектуальных систем в условиях неполноты знаний [Нариньяни, 2000], которая в свою очередь неизбежна в открытых, обучающихся интеллектуальных системах.

Интеграция знаний заключается в погружении новых знаний в состав знаний, уже известных интеллектуальной системе, являясь необходимым этапом понимания информации, которая либо поступает в интеллектуальную систему извне, либо генерируется (порождается) в процессе обработки информации. Процесс понимания в интеллектуальной системе заключается не только и не столько в помещении добавляемых новых знаний в память, но – в выявлении синонимов, связывании и отождествлении поступающих фрагментов с уже известными. Таким образом, интеграция знаний является основой функционирования любой интеллектуальной системы и во многом определяет качество интеллектуальной системы, будучи последним и основным этапом понимания поступающих (приобретаемых) знаний.

Решение задачи интеграции связано с рядом трудностей на разных уровнях рассмотрения интеллектуальной системы:

- на уровне языков и моделей представления знаний;
- на уровне моделей, методов и алгоритмов, применяемых для решения задачи интеграции;
- на уровне архитектуры программных средств, которыми осуществляется интеграция знаний.

На уровне языков и моделей представления трудности обусловлены следующими недостатками:

- необходимость выбора среди нескольких моделей представления знаний, между которыми нет однозначного предпочтения;
- неоднородность моделей представления знаний, приводящая к тому, что в базе появляются синонимичные структуры разных типов, требующие от пользователя их согласования;
- ограниченные возможности представления знаний в некоторых языках и моделях;
- отсутствие разделения понятий и терминов (исключения – DOGMA и Gellish),
- отсутствие поддержки монотонного расширения базы знаний (исключения – OWL, OWL2 и языки, построенные на основе классических логических моделей).

На уровне архитектуры программных средств присутствуют следующие недостатки:

- ограничения на расширение базы знаний;
- в силу зависимости от сторонних средств, ограниченность иерархической таксономической структурой онтологии возможности интеграции знаний или отсутствие таких возможностей, необходимость выбора средств интеграции пользователем;

- отсутствие общих стандартов совместимости разработанных фрагментов баз знаний;
- ограниченность средств поиска и каталогизации разработанных фрагментов баз знаний;
- ограниченность или отсутствие развитых средств взаимодействия с внешней средой.

К языкам представления знаний в разных системах на сегодняшний день относятся: Conceptual Graph [Sowa et al., 2008], Frame-logic [Michael Kifer et al., 1995], Knowledge Interchange Format [Genesereth et al., 1992], Integrated DEFinition for Ontology Description Capture Method (IDEF5) [IDEF5, 1994], Common Algebraic Specification Language [CoFI:CASL-Summary, 2004], Concept maps/UML [Novak et al., 2008] [ISO24707], RDF/RDFS [W3C:RDFS, 2004], DARPA Agent Markup Language [DAML, 2006], CycL [CycL, 2002], Ontology Inference Layer [W3C:DAML+OIL, 2001], RDF/OWL-Lite, RDF/OWL-DL, RDF/OWL Full [W3C:OWL, 2004], Topic Maps [ISO13250], XTM/LTM/CML/GML [XTM, 2001], Common Logic [ISO24707], Developing Ontology-Grounded Methods and Applications [Mustafa Jarrar et al., 2008], Formal English [Martin Ph., 2002], Gellish [Van Renssen, 2005], RDF/Rule Interchange Format [W3C:RIF, 2010], Open Biomedical Ontologies [Smith et al., 2007], RDF/OWL2 [W3C:OWL2, 2009], RDF/OWL2 EL, RDF/OWL2 RL и др.

Для поиска в базах знаний и онтологиях [Хорошевский, 2008] используются такие языки, как: RDQL, squish, SPARQL [W3C:SPARQL, 2008], KQML, DMX, Datalog, TMQL, ERROL, RuleML, RQL, OQL, TQL, VERSA, DQL и др.

Среди средств, которые могут рассматриваться в качестве основы для разработки баз знаний, можно выделить: оболочки экспертных систем (CLIPS (FuzzyCLIPS, DYNACLIPS, WxCLIPS) [CLIPS, 1991], SOAR, OPS83, RT-EXPERT, MIKE, BABYLON, WindExS, ES; ACQUARE, Easy Reasoner, ECLIPSE, EXSYS Professional, SIMER+MIR, АТ ТЕХНОЛОГИЯ, CAKE v2.0) [Гаврилова и др., 2000]; инструментальные пакеты для разработки экспертных систем (G2, ART, KEE, Knowledge KRAFT); системы, ориентированные на обработку онтологий [Sowa et al., 2008] – Protégé, WebOnto, OntoEdit, WebODE, OilEd, OntoLingua. К существующим на настоящий момент средствам интеграции онтологий можно отнести: Optima, Prompt, Ontolingua, Chimaera [McGuinness et al., 2000], ONION [Gangemi et al., 1996], COMA++ [Aumueller et al., 2005] и др.

Достоинствами приведённых средств являются: поддержка представления знаний различного вида различными моделями представления знаний в рамках одной системы; наличие средств визуального проектирования баз знаний; наличие средств верификации базы знаний, включая проверку на непротиворечивость; возможность монотонного расширения базы знаний, наличие средств интеграции баз знаний; наличие средств

поддержки обмена данными с внешней средой, включая средства обмена данными в реальном времени.

Для преодоления трудностей семантической интеграции [Doan and Halevy, 2005], [Кудрявцев, 2008] (отображения онтологий (ontology mapping)) и интеграции знаний (knowledge integration) в базах знаний используются следующие подходы:

- сравнение и выравнивание онтологий (ontology matching & alignment),
- интеграция онтологий (ontology merging),
- семантическое сравнение (semantic matching),
- семантическая унификация (semantic unification).

Для интеграции онтологий можно выделить методы:

- структурно-синтаксические:
 - анализ внутренней структуры
 - экстенционально-статистические
 - анализ внешней структуры (метаструктурный анализ)
 - терминологические (лексические)
- логико-семантические

К наиболее развитым подходам и методам интеграции можно отнести ONION (ONtology composition [Mitra, 2001]), формальный концептуальный анализ (FCA [Ganter, 1999], [Stumme, 2001]) и варианты его развития для нечётких (FOGA [IEEE, 2006]) и неопределённых множеств, методы использующие элементы семантического анализа – СТХMATCH [Bouquet, 2003] и S-match. Однако все эти методы, несмотря на использование некоторыми из них нечётких и неопределённых множеств, плохо приспособлены или не приспособлены к интеграции знаний в условиях наличия НЕ-факторов [Нариньяни, 2000]. Перечисленные методы, и методы, основанные на мерах близости (CUPID [Maldavan, 2001] и т.п.), не обеспечивают достаточной формальной строгости и непротиворечивости онтологий или баз знаний, получаемых в результате.

Поэтому возникает противоречие: существует необходимость точного решения задачи интеграции, позволяющего исключить привлечение эксперта для получения качественного результата, при отсутствии моделей, методов и алгоритмов интеграции, которые бы обеспечивали точность этого решения в условиях необходимости интеграции неполных, нечётких и изменяющихся со временем знаний (при наличии НЕ-факторов).

Следует отметить, что интеграции знаний рассматривается как развитие задачи интеграции данных, учитывающих их семантические аспекты. На уровне интеграции знаний основные трудности интеграции вызваны наличием различных моделей представления знаний и их неоднородностью. Для их преодоления строятся обёртки, оболочки (медиаторы), отображения одних моделей

представления на другие. Как правило, среди моделей представления существует модель глобального (универсального) представления, на которую можно отобразить все другие. Это позволяет достаточно просто свести количество отображений к необходимому минимуму. При наличии модели глобального представления существует два подхода, ведущих своё происхождение от интеграции данных и основывающихся на выборе базовых моделей представления: выбор глобальной или локальных моделей представления. Предпочтительнее выбор модели глобального представления в качестве базовой, так как отображение в локальные модели будет зависеть только от числа медиаторов (посредников) отображения, в ином случае – не все данные могут быть отображены в глобальную модель, а только те, которые представлены и представимы в локальных моделях. Если глобальная (универсальная) модель позволяет представлять семантически различающиеся знания единственным образом, то можно говорить об универсальной модели унифицированного представления знаний. В работе предлагается подход, в котором используется отображение знаний на глобальную модель унифицированного семантического представления знаний (рисунок 1).

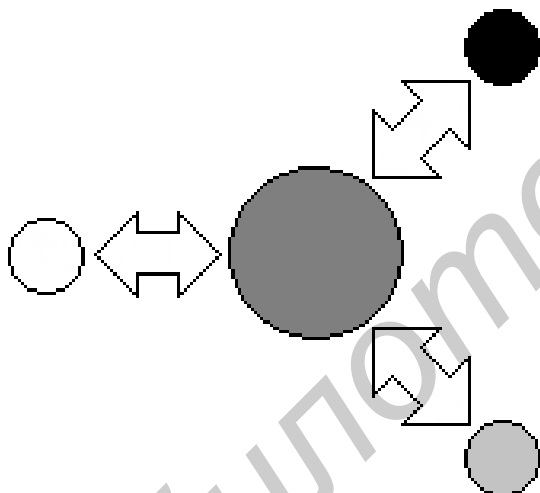


Рисунок 1 – Отображение (трансляция) знаний на модель унифицированного представления знаний

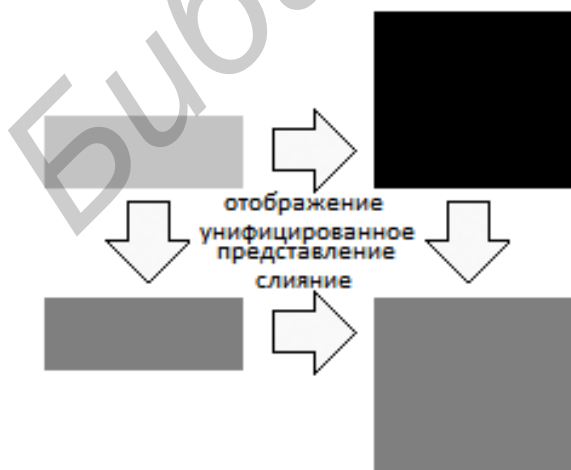


Рисунок 2 – Схема интеграции знаний

В этом случае интеграция знаний может быть сведена к следующей схеме (рисунок 2): 1) перейти к унифицированному представлению знаний путём трансляции на языки модели унифицированного представления знаний, в результате имеющиеся отображения одних знаний на другие можно заменить отношением слияния (склеивания) элементов семантической сети; 2) осуществить слияние синонимов, т.е. элементов дублирующегося представления знаний.

В результате интеграция знаний в языках модели унифицированного семантического представления знаний сводится к слиянию элементов семантической сети (рисунок 3).



Рисунок 3 – Интеграция знаний в модели унифицированного семантического представления знаний

1. Модель унифицированного семантического представления знаний

Семантическая модель интеграции использует унифицированное представление знаний и обеспечивает интеграцию sc-моделей баз знаний и их фрагментов [Ивашенко, 2012]. Унифицированное представление знаний обеспечивается моделью унифицированного семантического представления знаний, которая является частным случаем такой модели представления знаний, как семантические сети и задаётся семейством совместимых sc-языков, использующих унифицированный способ семантического кодирования Semantic Computer code (SC-код) [Голенков и др, 2001] и поддерживающих представление знаний различного вида [Ивашенко, 2003], [Ивашенко, 2004], [Ивашенко, 2011a]. Особенности SC-кода являются: простой алфавит, содержащий узлы и дуги, простой синтаксис, базовая теоретико-множественная интерпретация. Семантика sc-языка задаётся на основе модели ситуативных множеств, являющейся развитием моделей L-нечётких множеств и неоднородных нечётких множеств, предложенных Дж. Гогеном и А. Кофманом [Кофман, 1982].

Модель ситуативных (событийных, нестационарных [Ивашенко, 2012]) множеств может быть задана следующей шестёркой компонентов:

$$\langle Universe, [0;1], Events, r, h, sM \rangle \quad (1)$$

где *Universe* – универсальное множество объектов предметной области, *Events* – множество элементарных событий, $r \subseteq Events \times Events$ – отношение доступности (следования во времени)

событий, $h \in (2^{\text{Events}})^{\text{Universe}}$ – функция, задающая множество событий существования каждого элемента универсального множества, sM – семейство пар множеств событий существования ситуативного множества и соответствий (нечёткой) ситуативной принадлежности элементов универсального множества ситуативному множеству, отображающих элементы ситуативных множеств, множества событий и соответствующие им наборы степеней нечёткой принадлежности высших порядков на множество степеней нечёткой принадлежности $[0;1]$.

В соответствии с введённой моделью ситуативных множеств для описания динамических предметных областей, используются понятия нестационарной принадлежности и непринадлежности.

Выбор семантических сетей в качестве основы для представления знаний обусловлен их качественным отличием от строковых языков представления знаний в аспекте ассоциативности. Семантические сети позволяют качественно увеличить связность концептов и абстракций, обеспечивая экспоненциальный рост количества $\dot{V}(\langle p, q \rangle)$ различных связанных фрагментов от числа обозначений p и их связей q , являющихся готовыми ответами на вопросы пользователя, в отличие от полиномиального роста их количества в случае представления знаний в виде строк, при котором в худшем случае поиск требует просмотра всех символов в строке:

$$\frac{p*(p+1)}{2}. \quad (2)$$

$$\dot{V}(\langle p, q \rangle) \geq \frac{(\lceil (q-p)/(2*p) \rceil + p - 1)!}{(\lceil (q-p)/(2*p) \rceil - 1)! * (p!)}. \quad (3)$$

Модель унифицированного семантического представления знаний задаётся следующими компонентами.

$$\langle SClanguages, R_{SC}, F_{SC} \rangle \quad (4)$$

$$F_{SC} = A_{SC} \cup I_{SC} \cup N_{SC} \cup S_{SC} \cup E_{SC} \cup K_{SC}$$

$SClanguage$ – множество sc-языков, R_{SC} – отношения на множестве sc-языков (sc-подъязыка и трансляции), F_{SC} – функции sc-языков, A_{SC} – алфавитные функции sc-языков, I_{SC} – функции отношений инцидентности, N_{SC} – синтаксические предикаты sc-языков, S_{SC} – семантика sc-языков, E_{SC} – ключевые элементы sc-языков, K_{SC} – спецификация sc-языков и их ключевых элементов (отображение на множество онтологий sc-языков).

Semantic Code ориентирован на универсальность, т.е. на представление любых видов знаний. Semantic Code является средством

унификации представляемых знаний. На множестве sc-языков (языков, представленных в SC-коде) определены отношения sc-подъязыка и трансляции. Подъязык, являющийся пересечением выделенного семейства совместимых специализированных sc-языков, рассматривается как интегрированный sc-язык представления знаний (SCK). Основным принципом построения sc-языков является представление понятий, соответствующих основным классам объектов, описываемых sc-языком, и отношений между этими объектами ключевыми узлами такого sc-языка: каждому sc-языку однозначно сопоставляется конечное множество ключевых узлов (элементов) этого языка. Каждый ключевой узел задаёт ограничения на собственную семантическую окрестность в информационных конструкциях (текстах) этого языка. Множество ключевых элементов интегрированного sc-языка представления знаний является объединением множеств ключевых элементов остальных sc-языков выделенного семейства. Характеристиками sc-языка являются: мощность множества ключевых узлов sc-языка; семейство множеств собственных семантических окрестностей ключевых элементов языка; наличие функциональных зависимостей между собственными окрестностями множеств ключевых элементов sc-языка; соотношение алгоритмических сложностей поиска или вычисления элементов собственных семантических окрестностей на основании существующих зависимостей. Построена семантическая онтология существующих языков и моделей представления знаний, в которой указано место унифицированной модели представления знаний и её соотношение с существующими моделями представления знаний.

Введенные в sc-языки ключевые элементы поддерживают соответствующее семантически эквивалентное представление концептов OWL 2 QL и OWL 2 EL [W3C:OWL2, 2009], а также поддерживают представление множеств, мультимножеств, ситуативных множеств, отношений, включая ролевые отношения – строгие подмножества отношения принадлежности, простых, целых и рациональных чисел, логических формул. Остальные концепты могут быть определены с помощью логического sc-языка.

Модель унифицированного семантического представления знаний поддерживает представление знаний в условиях влияния различных НЕ-факторов: неполноты, неопределённости, нечёткости и пр. Далее приведены примеры представления нечётких и изменяющихся во времени знаний.

2. Интеграция знаний

Модель унифицированного семантического представления знаний, позволяет представлять знания различного вида, в т.ч. необходимые для решения задач интеграции, включая отношение совпадения. На основе модели унифицированного

семантического представления знаний разработана модель спецификации знаний, которая позволяет рассматривать различные типы фрагментов баз знаний [Ивашенко, 2009а] и их спецификации, задающие требования к фрагменту базы знаний (семантической сети) и позволяющие осуществить верификацию и оценить его качество. Модель спецификации знаний задаётся множеством формальных моделей онтологий фрагментов баз знаний Z , множеством морфизмов, соответствий между ними, и множеством отношений между парами формальных моделей и множеством соответствий и морфизмов.

$$\left\langle Z \cup 2^{\{\omega(z) \mid z \in Z\}^2}, 2^\Omega \right\rangle; \quad (5)$$

$$\Omega = \bigcup_{\{x\} \cup \{y\} \in Z} (\{x\} \times \{y\}) \times 2^{\omega(x) \times \omega(y)},$$

где 2^Ω – множество отношений модели спецификации; ω – функция элементов онтологической модели. Онтологическая модель каждого фрагмента базы знаний из множества Z задаётся тройкой $\langle G, R, O \rangle$, где G – непустое конечное множество обозначений спецификации фрагмента; R – ориентированное конечное множество отношений на обозначениях фрагмента; O – ориентированное конечное множество функций интерпретации обозначений.

$$\omega(z_i) = G_i \cup \{r \mid r = R_{ij}\} \cup \{o \mid o = O_{ij}\} \cup \{k \mid k \in R_{ij}\} \cup \{p \mid p \in O_{ij}\} \cup \{a \mid \langle a, v \rangle \in O_{ij}\}. \quad (6)$$

В рамках модели спецификации знаний построены онтологии sc-языков, используемые для описания типов фрагментов баз знаний и их интеграции. Среди требований качества по логико-семантическим признакам выделены требования полноты, непротиворечивости и неизбыточности. По методологическим признакам выделены требования к конструкциям предметно-специализированных языков, к конструкциям, описывающим определяющие признаки типа фрагмента базы знаний (семантической сети), и к конструкциям, описывающим отображения элементов и отношения интеграции фрагментов баз знаний. Для вычисления мер в соответствии с выделенным языком спецификации и записываемыми в его текстах требованиями используются функции вида

$$\left\{ L \times 2^\Phi \mid L \in SCLanguages \right\}, \quad (7)$$

где Ω – объединение отношений модели спецификации; ζ – отображение множества текстов языка спецификации на множество их онтологических моделей; Φ – упорядоченное множество оценок (например $\Phi = [0, 1]$).

Справедлива коммутативная диаграмма, описывающая взаимосвязь между фрагментами баз знаний и их моделями в моделях интеграции и

спецификации знаний, при заданных отображениях полного представления g :

$$\begin{array}{ccc} z_1 & \xrightarrow{h \in R_{NMH}((z_1, z_2))} & z_2 \\ \downarrow g & & \downarrow g \\ j_1 & \xrightarrow{f \in R_F((j_1, j_2))} & j_2 \end{array}; z_1, z_2 \in Z; j_1, j_2 \in J, \quad (8)$$

где J – множество фрагментов баз знаний, представленных в модели унифицированного семантического представления знаний, R_F – отношение слияния, R_{NMH} – отношение гомоморфного отображения.

Путём интеграции осуществляется переход от некоторого исходного фрагмента семантической сети к требуемому оптимизированному фрагменту базы знаний, имеющим более высокое качество [Ивашенко, 2011а].

В задаче интеграции двух фрагментов баз знаний в качестве исходных данных используются два фрагмента семантической сети и дополнительная метаинформация о свойствах обозначений, принадлежащих этим фрагментам. Результатом решения этой задачи является нахождение интегрированной фрагмента семантической сети, такого, что каждое обозначение из исходных интегрируемых фрагментов, имеет единственное представление в виде соответствующего обозначения интегрированном фрагменте [Ивашенко, 2009а]. Будем говорить, что осуществляется слияние двух обозначений исходных фрагментов тогда и только тогда, когда каждому из обоих соответствует в интегрированном фрагменте единственный знак. При интеграции в качестве дополнительной метаинформации может использоваться информация о внешних обозначениях (идентификаторах) понятий (обозначений) или информации, заданная базовой или ключевой рефлексивной семантикой обозначений из этих фрагментов. В условиях неполноты информации можно выделить два типа стратегий слияния обозначений: безопасные (выполняемые однозначно и непротиворечиво) и небезопасные. Безопасная стратегия включается в любую стратегию слияния.

При интеграции двух фрагментов между множеством их элементов (обозначений) выявляется некоторое множество пар потенциально синонимичных элементов (обозначений). Методика точного решения этой задачи включает как решение этой задачи вручную, путём получения соответствующего ответа от разработчика, так и автоматизированное решение на основе известной базовой теоретико-множественной семантики элементов и ключевых узлов sc-языка. Исходя только из базовой теоретико-множественной интерпретации, уже в некоторых фрагментах базы знаний можно легко установить факт наличия или отсутствия синонимии. Зная множество таких пар и множество всех возможных пар всех элементов в заданном фрагменте базы знаний, легко вычислить

множество и число пар потенциально синонимичных элементов такой базы знаний.

Имея информацию о потенциально синонимичных обозначениях, можно оценить множество и количество всевозможных структурно различимых вариантов слияния потенциально синонимичных элементов (обозначений) в элементы результирующего фрагмента. Это множество обозначим величиной $I(G)$, где G – симметричный ориентированный граф, множество рёбер $E(G)$ которого является множеством всех пар потенциально синонимичных обозначений исходных баз знаний, а множество вершин $V(G)$ – множеством всех обозначений исходных фрагментов

$$E(G) \subseteq (V(G))^2. \quad (9)$$

Для того, чтобы определить множество $I(G)$, используем специальную операцию

$$A \overset{\text{def}}{\cup} B = \bigcup_{(P,Q) \in A \times B} \{P \cup Q\}, \quad (10)$$

которая является ассоциативной и коммутативной.

Чтобы задать множество $I(G)$, введём семейство всевозможных множеств рёбер ориентированных графов $C(G)$ на множестве вершин графа G , каждая компонента связности которых является полным подграфом [Ивашенко, 2009b], тогда

$$I(G) = \left(\bigcup_{e_{ij} \in E(G)} \{\emptyset, \{e_{ij}\}\} \right) \cap C(G). \quad (11)$$

Число Q элементов множества $I(G)$ можно рассматривать в виде критерия качества: чем меньше это число, тем выше качество и наоборот. Однако, уже для небольших фрагментов баз знаний, это число может оказаться большим и трудным для расчёта, поэтому на практике более целесообразно использовать или логарифм этого числа или логарифм его оценок, или связанные с ними величины (число пар потенциальных синонимов).

$$Q = |I(G)|. \quad (12)$$

В работе [Ивашенко, 2009b] для числа Q установлены соотношения для верхней и нижней границы.

Интеграция онтологий и фрагментов баз знаний осуществляется через слияние обозначений. Рассмотрим m объектов области значений семантической интерпретации некоторого знака. Тогда, на множестве этих объектов, число неоднозначных и однозначных постоянных семантических интерпретаций (семантик) знака равно $2^{2^m} - 1$.

Чтобы осуществить слияние обозначений (воплощений знака) необходимо, чтобы пересечение областей значения их семантик было непустым. В частности – теоретико-множественных семантик.

Если пересечение областей значения семантик знаков является пустым, то слияние таких

обозначений невозможно и такие обозначения различны.

Тогда как для выявления различных обозначений достаточно одного контр-примера, для выявления подлежащих слиянию знаков необходимо не только установить, что пересечение областей значения их семантик не является пустым, но и доказать, что эти знаки совпадают. Для этой цели используются различные утверждения о совпадении в виде аксиом и теорем.

Опишем свойства этого отношения совпадения. Если два обозначения совпадают, то не существует множества, которому одновременно принадлежит один из них и не принадлежит другой. Если два обозначения множеств совпадают, то множества равны. Выявить неравенство множеств просто: достаточно найти элемент, который принадлежит одному множеству и не принадлежит другому. Процедура выявления равенства множеств зависит от их свойств. Если множества имеют конечное, не очень большое число принадлежностей или непринадлежностей элементов, что характерно для понятий, то перечислив все принадлежности или все непринадлежности, число которых совпадает соответственно с его мощностью или количеством непринадлежностей, и убедившись, что среди них нет непринадлежностей элементов, принадлежащих другому множеству, можно заключить, что множества равны. Если же множества бесконечны или число принадлежностей, равно как и непринадлежностей у них, очень большое, то равенство множеств можно доказать только через утверждения об их свойствах.

Прежде, чем проводить слияние обозначений двух фрагментов семантической сети, следует провести отображение этих фрагментов (по аналогии с отображением онтологий).

Для отображения в некоторых работах [Maltese et al., 2010] используются следующие отношения между онтологическими понятиями $\{\equiv, \supseteq, \subseteq, \perp, \sqcap\}$ (таблица 1), соответственно – совпадения, обобщения, частности, строгого включения и строгого пересечения. Например, эти отношения в проекции на теоретико-множественные соответствуют отношениям равенства множеств ($=$), надмножества (\supseteq), подмножества (\subseteq), пустого пересечения множеств ($S1 \cap S2 = \emptyset$) и непустого пересечения множеств ($S1 \cap S2 \neq \emptyset$).

Таблица 1.

Обозначение		Следствие	Пояснение
справа	слева		
\sqcap	\sqcap	\neq	строго пересекаются
\supseteq	\subseteq	$-$	включение
\supset	\subset	\neq	строго включение
\perp	\perp	$-$	не пересекаются
\equiv	\equiv	\equiv	совпадение

Таблица 2.

Обозначение		Следствие	Пояснение
справа	слева		
—	—	—	соотношение не определено
≠	≠	≠	не совпадают (различны)
⊃	⊃	—	пересекаются (связность)
⊄	⊄	≠	исключение
⊈	⊈	≠	исключение пересекающихся
⊉	⊉	≠	симметричное исключение
⊋	⊋	≠	строго пересекаются
⊇	⊆	—	включение
⊇	⊇	—	включение пересекающихся
⊃	⊂	≠	строгое включение
⊃	⊃	≠	строгое включение пересекающихся
⊆	⊆	—	равенство
=	=	—	равенство пересекающихся
⊥	⊥	—	не пересекаются
≡	≡	≡	совпадение
≡	≡	≡	совпадение пересекающихся (непустых)

Для решения задачи интеграции неполных и нечётких знаний в предлагается расширенный набор отношений (таблица 2, таблица 3) {не уточнены*, различие*, связность*, исключение* (исключаемое, исключающее), исключение пересекающихся* (исключаемое пересекающееся, исключающее пересекающееся), симметричное исключение*, строгое пересечение*, совпадение*} и др. Отношение потенциальной синонимии можно выразить, как *возможная синонимия*=(не уточнены*∪связность*)/(исключение*∪исключение пересекающихся*∪ различие*)*.

Следующий фрагмент (рисунок 4) позволяет в рамках модели унифицированного семантического представления знаний и модели спецификации знаний задать правило отображения и слияния знаков баз знаний при использовании методов лексико-терминологического анализа (анализа идентификаторов) [Ивашенко, 2011b].

Таблица 3.

Обозначение		Следствие	Пояснение
справа	слева		
≠	≠	≠	неинъективность (строгая неравномощность)
≠	≠	≠	неинъективность пересекающихся
≠	≠	—	инъекция (неравномощность)
≠	≠	—	инъекция пересекающихся
≠	≠	—	строгая инъекция
≠	≠	—	строгая инъекция пересекающихся
≠	≠	—	биекция (равномощность)
≠	≠	—	равномощность пересекающихся

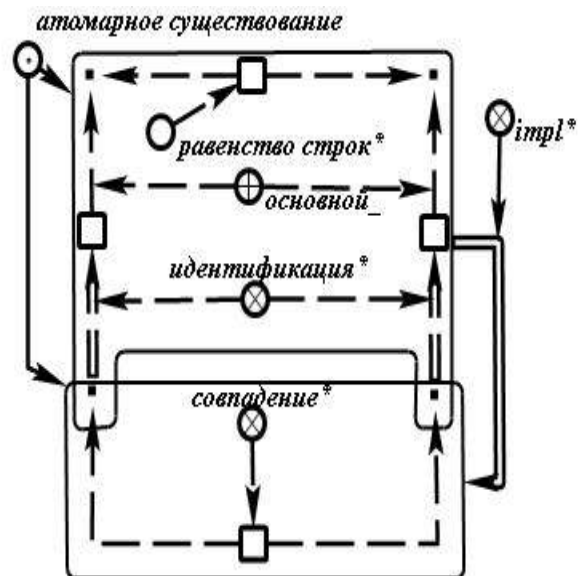


Рисунок 4 – Правило совпадения элементов с равными основными идентификаторами

Важным признаком совпадения обозначений является их принадлежность одной связке под однозначным ролевым отношением или наличие принадлежности им общего элемента под обратным однозначным ролевым отношением.

Все эти отношения можно однозначно и непротиворечиво установить в результате соответствующего базовой теоретико-множественной семантике структурного анализа, включая экстенциональный, за полиномиальное время. Более глубокий логико-семантический анализ может потребовать больших временных затрат.

Для анализа на логико-семантическом уровне важно выделять утверждения, которые приводят к совпадению знаков интегрируемых фрагментов, в результате чего можно произвести их слияние. Такими высказываниями являются высказывания о свойствах логических высказываний о единственности, высказывания о свойствах множеств без кратных вхождений элементов (канторовских множеств), высказывания об отношениях без кратных связей. Ниже приведён пример (рисунок 5), описывающий свойство логических утверждений о единственности и сводящий задачу анализа таких утверждений к задаче выявления связей отношений совпадения. Таким образом, при наличии средств логического вывода для решения задачи интеграции логико-семантическими методами достаточно описать свойства соответствующих ключевых узлов, выражающих количественные ограничения (например, единственность) или описывающих отсутствие кратных связей, через понятия совпадения sc-элементов (знаков текстов sc-языка).

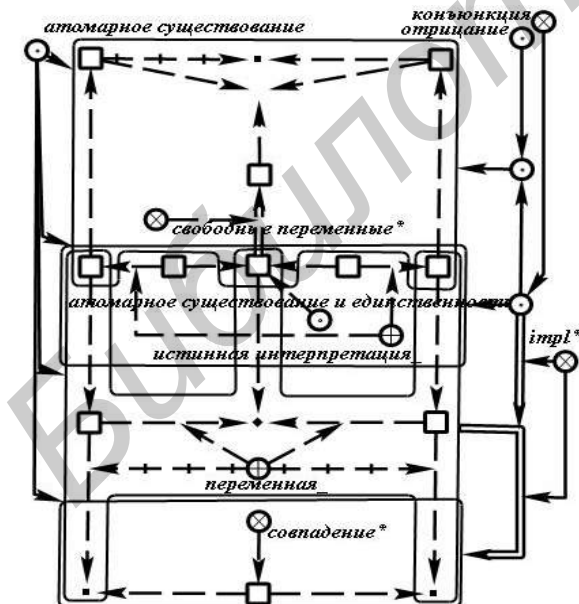


Рисунок 5 – Высказывание о слиянии значений переменных в утверждениях единственности

Алгоритм процедуры поиска множества решений задачи интеграции приведён на рисунке 7 [Ивашенко, 2011b]. Этот алгоритм использует

алгоритм сопоставления онтологий (рисунки 8, 9). Ниже (рисунок 6) приведён алгоритм выявления пар потенциальных синонимов, использующий вышеописанные правила экстенционального метода.

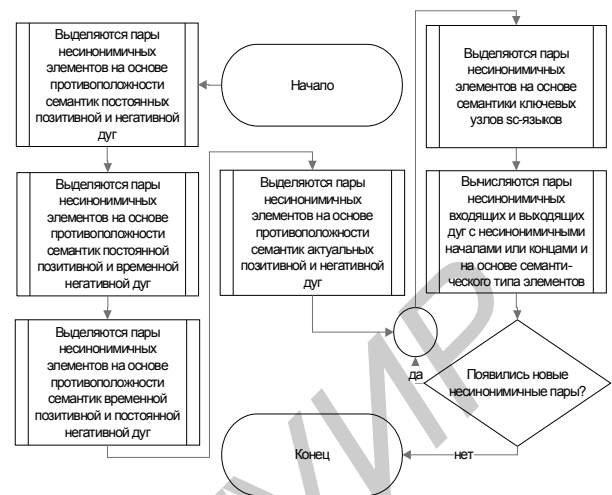


Рисунок 6 – Алгоритм вычисления множества пар различающихся (несинонимичных) элементов

При интеграции тезаурусов необходимо построить связную иерархию определяемых понятий. Для каждого понятия необходимо указать является оно определяемым или нет. Для каждого понятия необходимо указать множества понятий, на основе которых оно может быть однозначно определено, а на основе каких – нет. Для построения такой иерархии применим алгоритм, подобный обобщённому алгоритму интеграции. При поиске выполнимых интерпретаций схем обход возможных предикативных выражений осуществляется по всевозможным предварённым конъюнктивным нормальным формам, атомарные подформулы которых соответствуют алфавитным меткам, связкам инцидентности и равенству значения переменной константе и задают всевозможные размеченные привязанные с помощью констант подграфы в графе онтологии. Для сокращения перебираемых вариантов используются эвристики.

Интеграция тезаурусов (баз знаний, содержащих определения) может быть осуществлена на основе отсеивания наборов понятий, которые не могут относиться как определяющие и определяемые, в результате анализа по следующим схемам (фильтрация неоднозначно определяющих множеств понятий для каждого понятия).

Схема n-значности ($|v| = n$) свойства α задана следующим выражением.

$$\lambda(\alpha, x, v, z) = \forall w (\alpha(\langle x, w, z \rangle) \sim (w \in v)) \quad (13)$$

Схема необходимости.

$$\neg \exists U \exists V \exists Y (\neg \beta(\langle X, U, Y \rangle) \wedge \gamma(\langle Y, V, Z \rangle)) \quad (14)$$

Схема достаточности.

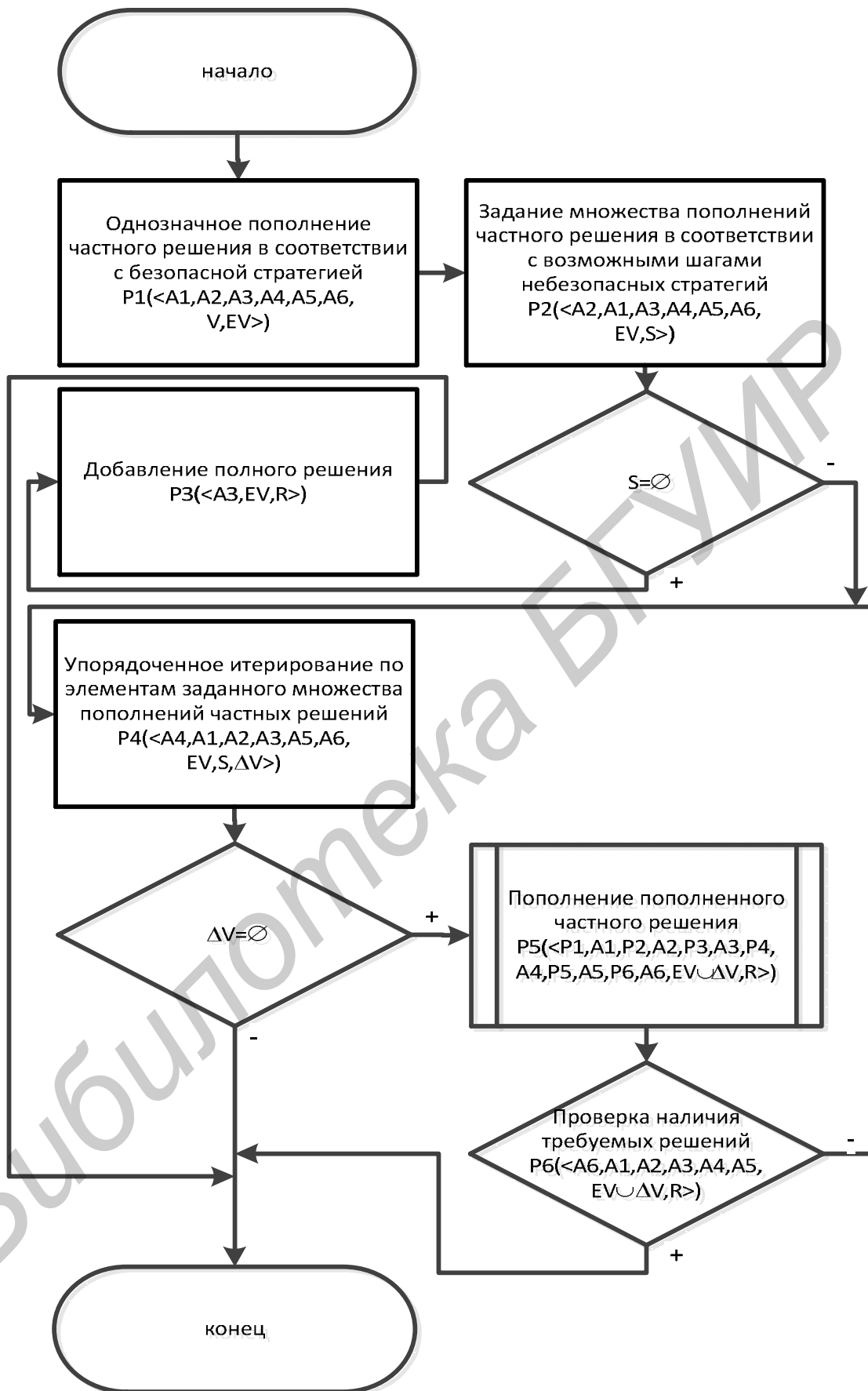


Рисунок 7 – Алгоритм процедуры поиска множества решений задачи интеграции

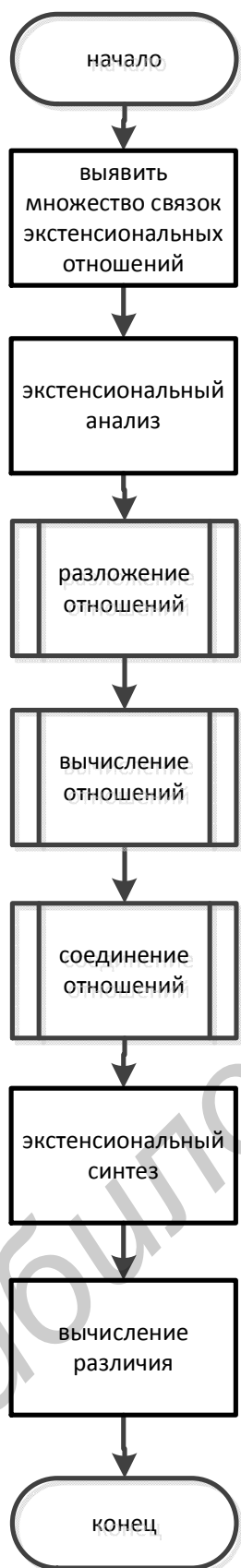


Рисунок 8 – Алгоритм процедуры сопоставления онтологий на основе экстенционального анализа троек sc-узлов

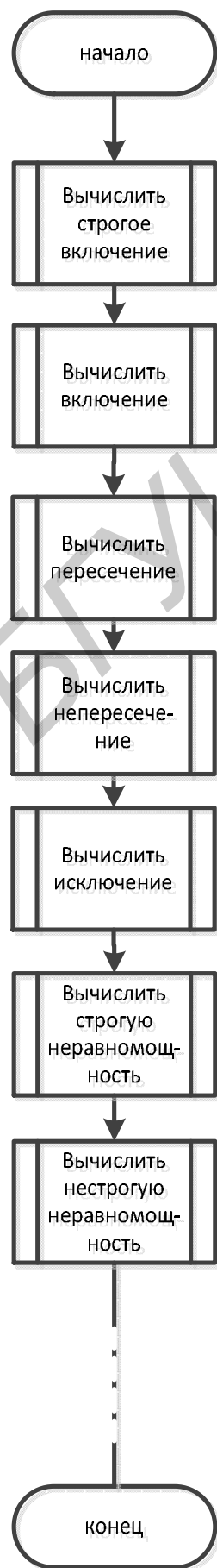


Рисунок 9 – Алгоритм процедуры вычисления отношений сопоставления онтологий на основе экстенционального анализа

Задача

В декартовой плоскости p даны точки A и C и окружность O с центром A и радиусом δ ($= 9600$).

$A \in p$

$C \in p$

$O \subset p$

Ученик провёл прямую a из точки C , касающуюся окружности O в точке H , отложил на касательной a от точки C отрезок BC длиной α ($=5600$) и построил треугольник ABC , периметр которого γ , вычисленный учеником, оказался целым числом.

$|AH| = \delta$

$\delta = 9600$

$H \in O$

$H \in a$

$AH \perp a$

$C \in a$

$BC \subset a$

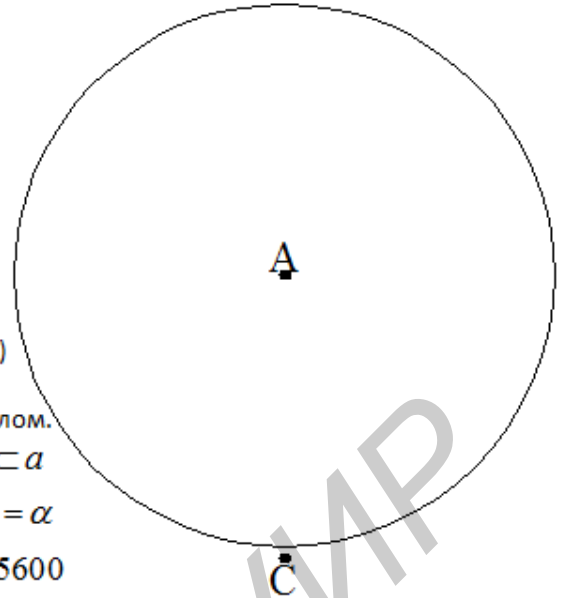
$|BC| = \alpha$

$\alpha = 5600$

ABC

$|\partial ABC| = \gamma$

$\gamma \in \mathbb{Z}$



Постройте этот треугольник в декартовой плоскости p и вычислите его периметр, если известно, что в этой декартовой плоскости точка A имеет координаты $\langle 0, 9126 \rangle$, а точка $C - \langle 0, -784 \rangle$.

$A \stackrel{p}{\cong} \langle 0, 9126 \rangle$

$C \stackrel{p}{\cong} \langle 0, -784 \rangle$

Рисунок 10 – Условие задачи на построение треугольника

$$\neg \exists U \exists V \exists Y (\beta(\langle X, U, Y \rangle) \wedge \neg \gamma(\langle Y, V, Z \rangle)) \quad (15)$$

Схема прямой однозначности.

$$\exists U \exists V \exists Y (\lambda(\langle \beta, X, \{U\}, Y \rangle) \wedge \neg \lambda(\langle \gamma, Y, \{V\}, Z \rangle)) \quad (16)$$

Схема обратной однозначности.

$$\exists U \exists V \exists Y (\lambda(\langle \gamma, Y, \{V\}, Z \rangle) \wedge \neg \lambda(\langle \beta, X, \{U\}, Y \rangle)) \quad (17)$$

Здесь U, V, X, Y, Z являются множествами элементов: X – определяющий набор понятий; Z – определяемый набор понятий; U, V, Y – элементы окрестности X и Z ; α, β и γ – предикативные выражения определяющих и определяемых свойств.

Для работы с исходными текстами онтологий фрагментов баз знаний была разработана модель гипертекстового представления знаний, которая ориентирована на массовую разработку фрагментов баз знаний. Были реализованы соответствующие программные средства. В соответствии с технологией решение задачи интеграции сводится к следующей схеме: 1) данные и знания представляются на некотором языке, удобном для

диалога с пользователем или взаимодействия с внешней средой, например на языке SCn в гипертекстовой модели представления знаний (рисунок 11); 2) затем они транслируются на язык модели унифицированного представления знаний (рисунок 12), где осуществляется их интеграции в соответствии с алгоритмом поиска решений задачи интеграции знаний; 3) при поиске решения на основе анализа сходства и различия формируется сопоставление элементов интегрируемых фрагментов семантической сети, выявляются отображения и соответствия слияния на основе чего формируется результат решения задачи интеграции знаний.

Рассмотрим решение задачи интеграции на примере анализа решения планиметрической задачи на построение треугольника. В этой задаче неустранимая неполнота знаний связана с естественным взаимодействием субъектов в процессе обучения. Условие задачи приведено на рисунке 10. Для ввода в базу знаний это условие транслируется на язык SCn (рисунок 13). Затем осуществляется трансляция на язык модели унифицированного семантического представления знаний (рисунок 16).



Рисунок 11 – Гипертекстовое представление знаний

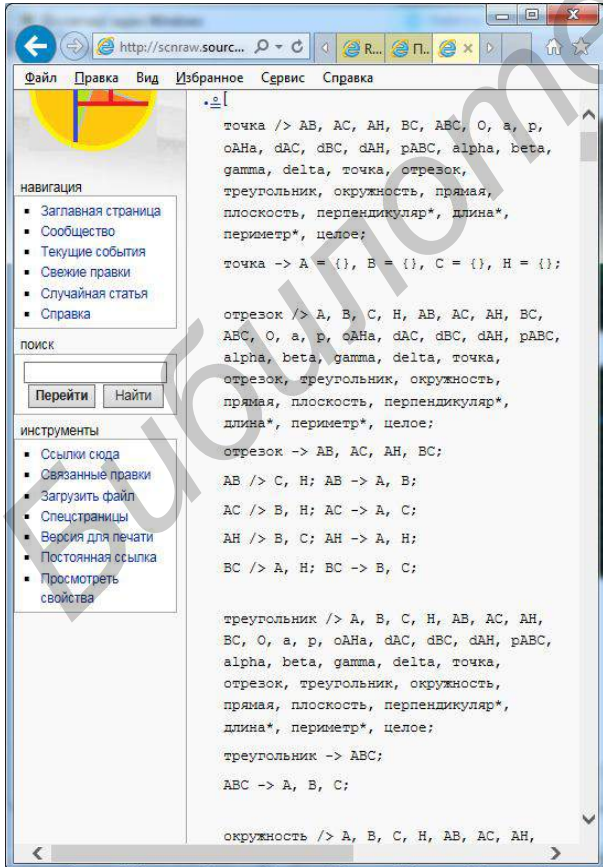


Рисунок 12 – Представление на языке SCs

Пример с треугольником

•.⊃[
геометрическое понятие	⊃ перпендикуляр*
⊃ точка	⊃ оАна
⊃ А	⊃ АН
⊃ В	⊃ а
⊃ С	⊃ длина*
⊃ D	⊃ dAC
⊃ отрезок	⊃ AC
⊃ АВ	⊃ alpha
⊃ С	⊃ dBC
⊃ Н	⊃ BC
⊃ А	⊃ beta
⊃ В	⊃ dАН
⊃ AC	⊃ АН
⊃ В	⊃ delta
⊃ Н	⊃ периметр*
⊃ А	⊃ pABC
⊃ С	⊃ ABC
⊃ АН	⊃ gamma
⊃ В	⊃ целое
⊃ С	⊃ alpha
⊃ А	⊃ beta
⊃ Н	⊃ gamma
⊃ BC	⊃ delta
⊃ А	⊃ А
⊃ Н	⊃ В
⊃ В	⊃ С
⊃ С	⊃ D
⊃ треугольник	⊃ АВ
⊃ ABC	⊃ AC
⊃ А	⊃ АН
⊃ В	⊃ BC
⊃ С	⊃ ABC
⊃ окружность	⊃ O
⊃ O	⊃ а
⊃ А	⊃ p
⊃ В	⊃ оАна
⊃ С	⊃ dAC
⊃ Н	⊃ dBC
⊃ прямая	⊃ dАН
⊃ а	⊃ pABC
⊃ А	⊃ точка
⊃ В	⊃ отрезок
⊃ С	⊃ треугольник
⊃ Н	⊃ окружность
⊃ плоскость	⊃ прямая
⊃ p	⊃ плоскость
⊃ А	⊃ перпендикуляр*
⊃ В	⊃ длина*
⊃ С	⊃ периметр*
⊃ Н	⊃ целое

Рисунок 13 – Условие задачи в гипертекстовом представлении

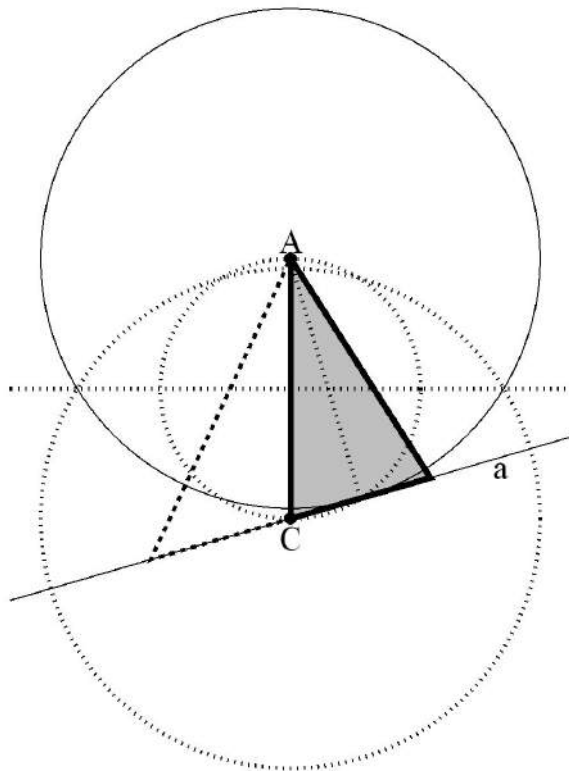


Рисунок 14 – Два треугольника с вершиной А и с равными основаниями на одной касательной а

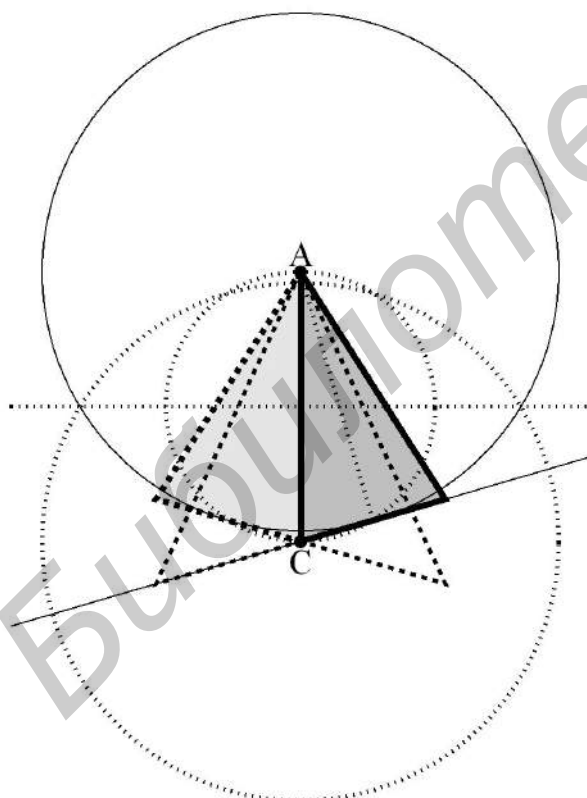


Рисунок 15 – Четыре треугольника с основаниями равной длины на двух касательных к окружности и вершиной в точке А

точка /> AB, AC, AH, BC, ABC, O, a, p, oAHa, dAC, dBC, dAH, pABC, alpha, beta, gamma, delta, точка, отрезок, треугольник, окружность, прямая, плоскость, перпендикуляр*, длина*, периметр*, целое;

точка -> A = {}, B = {}, C = {}, H = {};

отрезок /> A, B, C, H, AB, AC, AH, BC, ABC, O, a, p, oAHa, dAC, dBC, dAH, pABC, alpha, beta, gamma, delta, точка, отрезок, треугольник, окружность, прямая, плоскость, перпендикуляр*, длина*, периметр*, целое;

отрезок -> AB, AC, AH, BC;

AB /> C, H; AB -> A, B;

AC /> B, H; AC -> A, C;

AH /> B, C; AH -> A, H;

BC /> A, H; BC -> B, C;

треугольник /> A, B, C, H, AB, AC, AH, BC, O, a, p, oAHa, dAC, dBC, dAH, pABC, alpha, beta, gamma, delta, точка, отрезок, треугольник, окружность, прямая, плоскость, перпендикуляр*, длина*, периметр*, целое;

треугольник -> ABC;

ABC -> A, B, C;

окружность /> A, B, C, H, AB, AC, AH,

Рисунок 16 – Результат трансляции условия задачи на язык SCs

В соответствии с алгоритмом поиска множества решений задачи интеграции вызывается процедура однозначного пополнения решения, которая применяет правила слияния по равенству основных идентификаторов и которая осуществляет вызов алгоритма сопоставления элементов интегрируемых фрагментов. В результате из исходных данных (рисунок 17) будет получено сопоставление (рисунок 18), которое способно обеспечить выявление различия всех сопоставляемых узлов семантической сети, тогда как существующие модели не могут обеспечить этого (рисунок 19). Далее, при решении задачи, на построение будет сформирована следующая семантическая сеть, описывающая четыре варианта построения треугольника (рисунки 15, 20). В силу неполноты информации, неизвестно какой именно вариант треугольника дан в условии. Процедура однозначного пополнения решения или процедура добавления полного решения (рисунок 7) в

результате анализа противоречий может отбросить два варианта, например, если была предположена возможность слияния с треугольниками этих двух вариантов на ранних этапах поиска (рисунки 14, 21, 22, 23). В результате останутся два варианта построения треугольника (рисунок 24). Два варианта гипотетического слияния треугольников и будут результатом решения соответствующей задачи

интеграции (рисунки 25, 26). При неоднократном повторении решения этой задачи на построения разные результаты построения могут быть интегрированы, т.е. сколько бы раз конкретно первый или второй треугольник не был бы построен их количество, и количество их обозначений не изменится. В памяти системы будет только два обозначения треугольников (рисунок 27).

	A	B	C	H	AB	AC	AH	BC	ABC	a	O	P	alpha=5600 beta=10000	gamma	delta=9600	dAC	dAH	dBC	pABC	oAHa	точка	отрезок	треугольник	прямая	окружность	плоскость	целое	длина*	периметр*	перпендикуляр*	геометрическое понятие				
A	III				m	m	m		m	m	m										m														
B		III			m	m	m		m	m	m											m													
C			III		m	m	m		m	m	m											m													
H				III			m			m	m											m													
AB					III																	m													
AC						III										m						m													
AH							III										m					m													
BC								III										m				m													
ABC					U	U		U	III												m		m												
a								U		III											m		m												
O											III															m									
p							U	U	U	U	III															m									
alpha												III							m								m								
beta													#	III		m											m								
gamma															III					m							m								
delta																#	#										m								
dAC																	III																		
dAH																		III																	
dBC																			III																
pABC																					III														
oAHa																						III													
точка																							+	+	+	+	+	+	+	+	+	+		m	
отрезок																							+	+	+	+	+	+	+	+	+	+		m	
треугольник																							+	+	+	+	+	+	+	+	+	+		m	
прямая																							+	+	+	+	+	+	+	+	+	+		m	
окружность																							+	+	+	+	+	+	+	+	+	+		m	
плоскость																							+	+	+	+	+	+	+	+	+	+		m	
целое																							+	+	+	+	+	+	+	+	+	+		m	
длина*																							+	+	+	+	+	+	+	+	+	+		m	
периметр*																							+	+	+	+	+	+	+	+	+	+	+		m
перпендик.*																							+	+	+	+	+	+	+	+	+	+	+		m
геометрич.																							+	+	+	+	+	+	+	+	+	+	+	+	III

Рисунок 17 – Исходное сопоставление обозначений в условии задачи

	A	B	C	H	AB	AC	AH	BC	ABC	a	O	p	alpha=5600	beta=10000	gamma	delta=9600	dAC	dAH	dBC	pABC	oAHa	точка	отрезок	треугольник	прямая	окружность	плоскость	целое	длина*	периметр*	перпендикуляр*	геометрическое понятие		
A	III				m	m	m	m	m	m	m	m					m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	
B	#	III			m	m	m	m	m	m	m	m					m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	
C	#	#	III		m	m	m	m	m	m	m	m					m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	
H	#	#	#	III			m			m	m	m					m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	
AB	#	#	#	#	III												m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	
AC	#	#	#	#	□	III											m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	
AH	#	#	#	#	III	III											m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	
BC	#	#	#	#	□	□	III										m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	
ABC	#	#	#	#	□	□	III										m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	
a	#	#	#	#	□	□	□	□	III								m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	
O	#	#	#	#	III	III	III	III	III								m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	
p	#	#	#	#	□	□	□	□	□	□	□	III					m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	
alpha	#	#	#	#	#	#	#	#	#	#	#	#	III				m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	
beta	#	#	#	#	#	#	#	#	#	#	#	#	#	III			m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
gamma	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III		m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
delta	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
dAC	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
dAH	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
dBC	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
pABC	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
oAHa	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
точка	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
отрезок	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
треугольник	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
прямая	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
окружность	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
плоскость	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
целое	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
длина*	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
периметр*	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
перпендик.*	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
геометрич. понятие	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	#	III	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m

Рисунок 18 – Результат сопоставления, полученный по алгоритму процедуры сопоставления онтологий, каждая пара обозначений различна

	A	B	C	H	AB	AC	AH	BC	ABC	a	O	p	alpha=5600	beta=10000	gamma	delta=9600	dAC	dAH	dBC	pABC	oAHa	точка	отрезок	треугольник	прямая	окружность	плоскость	целое	Длина*	периметр*	перпендикуляр*	геометрическое понятие					
A	III				m	m	m			m	m											m															
B		III			m	m	m			m	m												m														
C			III			m	m			m	m												m														
H				III			m			m	m												m														
AB					III																		m														
AC						III											m						m														
AH							III											m					m														
BC								III												m			m														
ABC					U	U	U	U	III												m			m													
a								U		III												m															
O											III															m											
p					U	U	U	U	U	U	U	III															m										
alpha													III						m									m									
beta													#	III			m											m									
gamma														III						m								m									
delta													#	#	III		m											m									
dAC																	III													m							
dAH																		III													m						
dBC																				III											m						
pABC																					III										m						
oAHa																						III										m					
точка																							III										m				
отрезок																								U	III								m				
треугольник																								U	U	III							m				
прямая																								U	U	U	III							m			
окружность																								U	U	U	U	III						m			
плоскость																								U	U	U	U	U	III					m			
целое																								U	U	U	U	U	U	III				m			
Длина*																								U	U	U	U	U	U	U	III			m			
периметр*																								U	U	U	U	U	U	U	U	III			m		
перпендик.*																								U	U	U	U	U	U	U	U	U	III		m		
геометрич.																								U	U	U	U	U	U	U	U	U	U	III		m	

Рисунок 19 – Результат сопоставления традиционными подходами

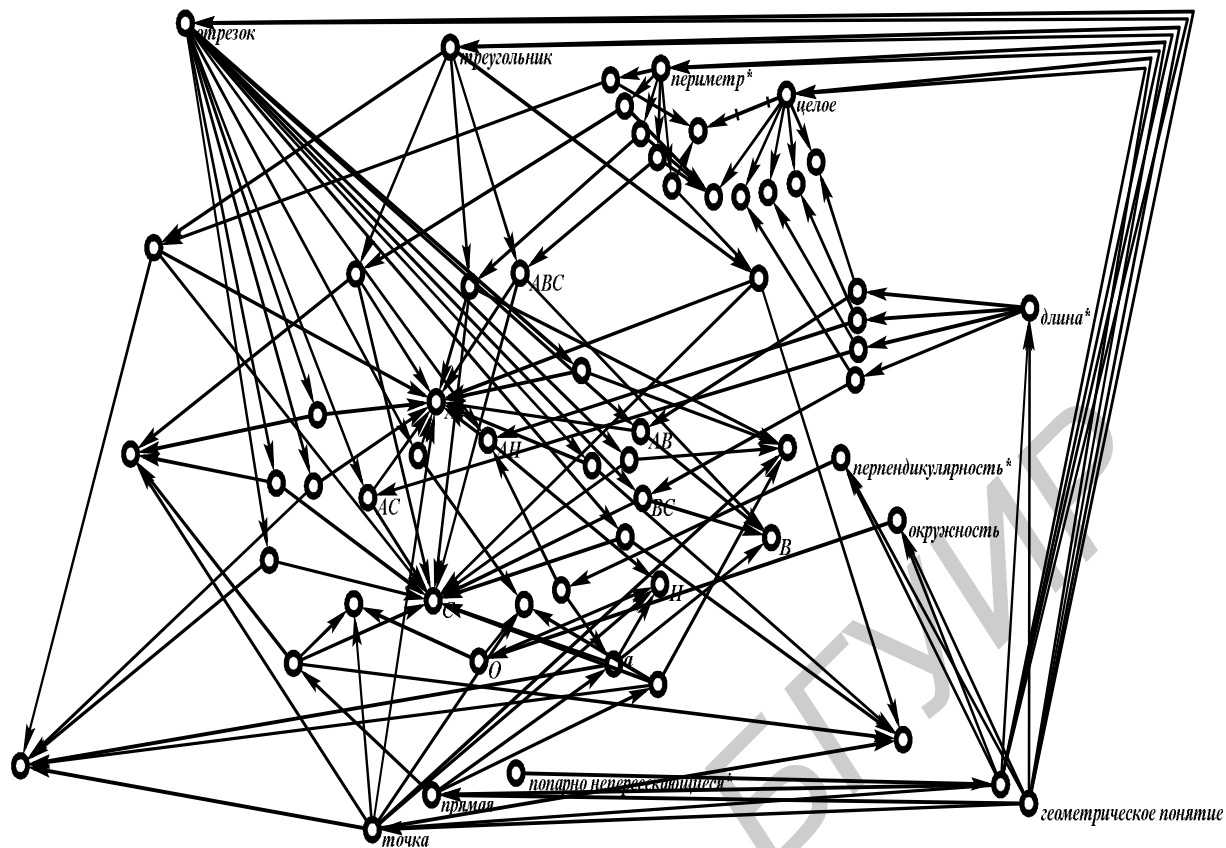


Рисунок 20 – Четыре варианта построения треугольника

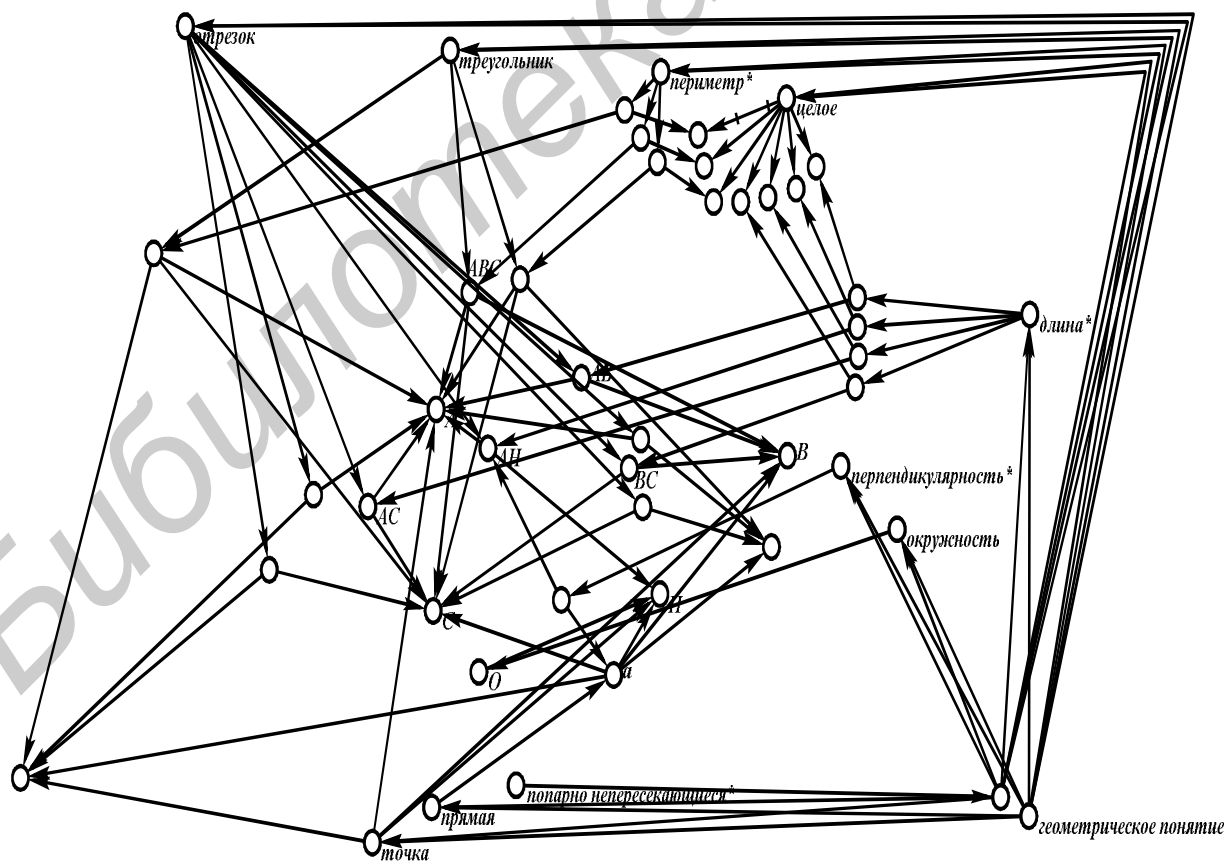


Рисунок 21 – Два варианта построения треугольника на одной касательной

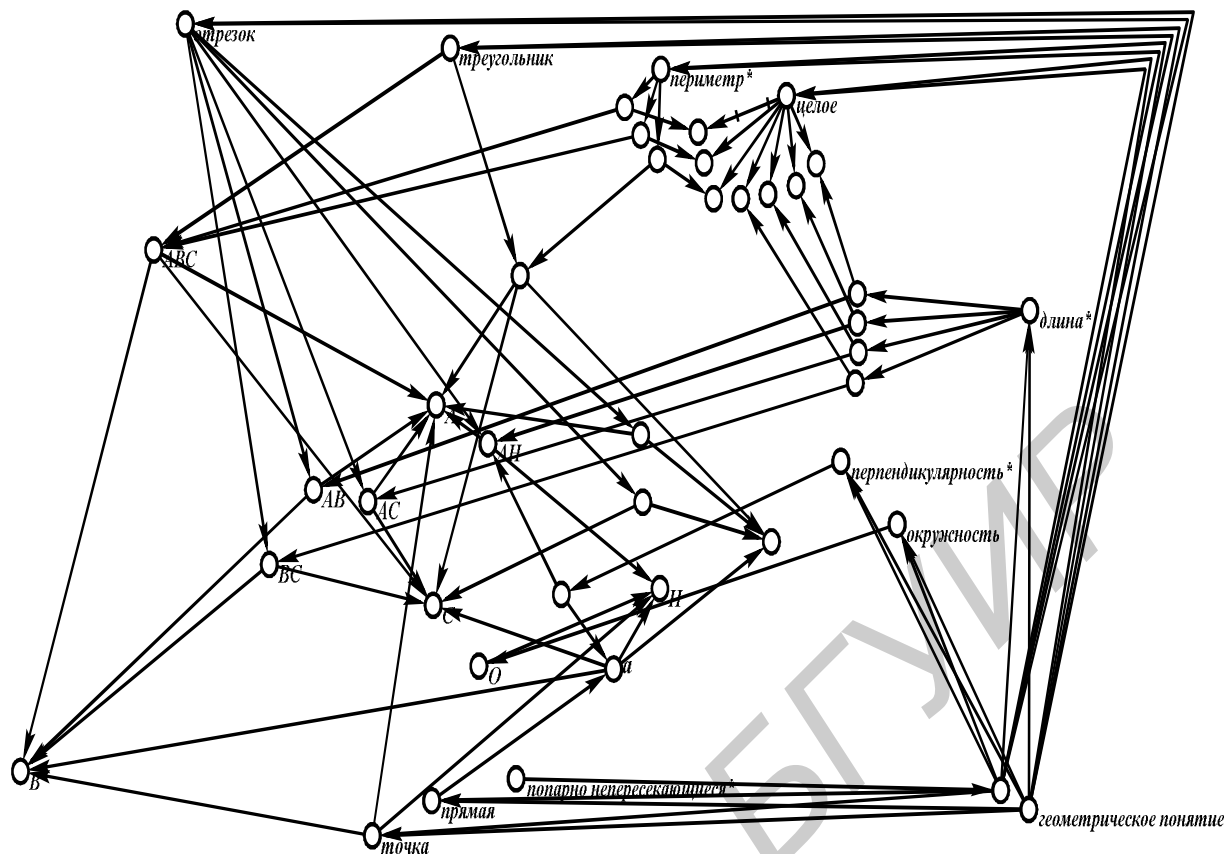


Рисунок 22 – Предполагаемое слияние с неравносторонним треугольником

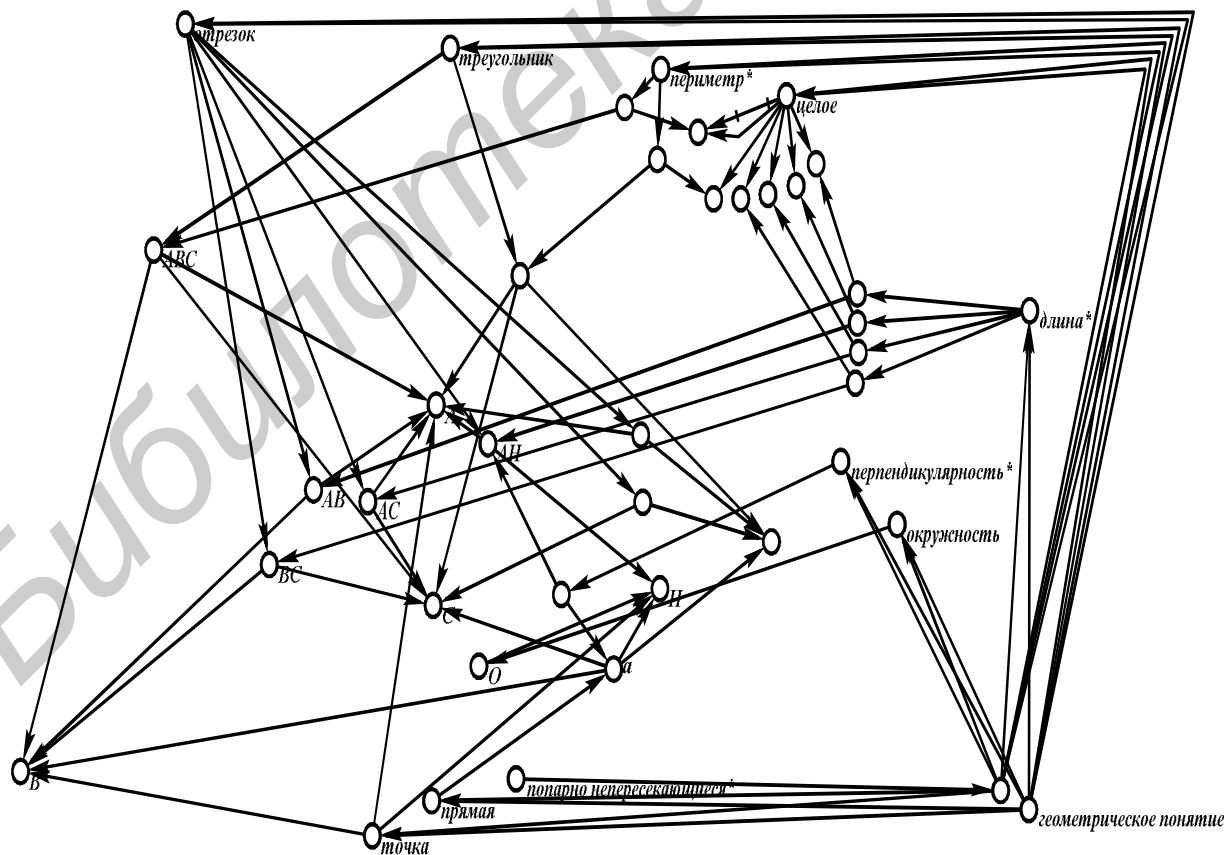


Рисунок 23 – Противоречие свойств периметра, полученное в результате слияния с неравносторонним треугольником

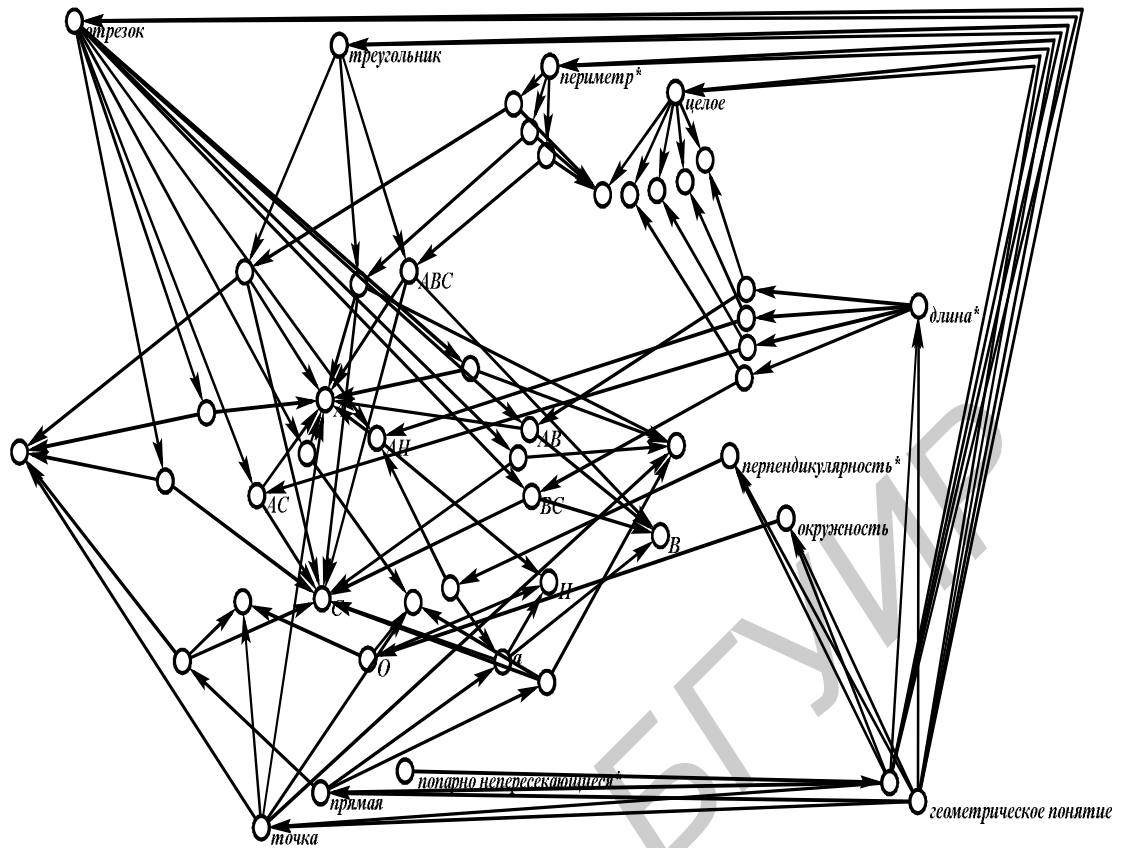


Рисунок 24 – Два допустимых варианта слияния

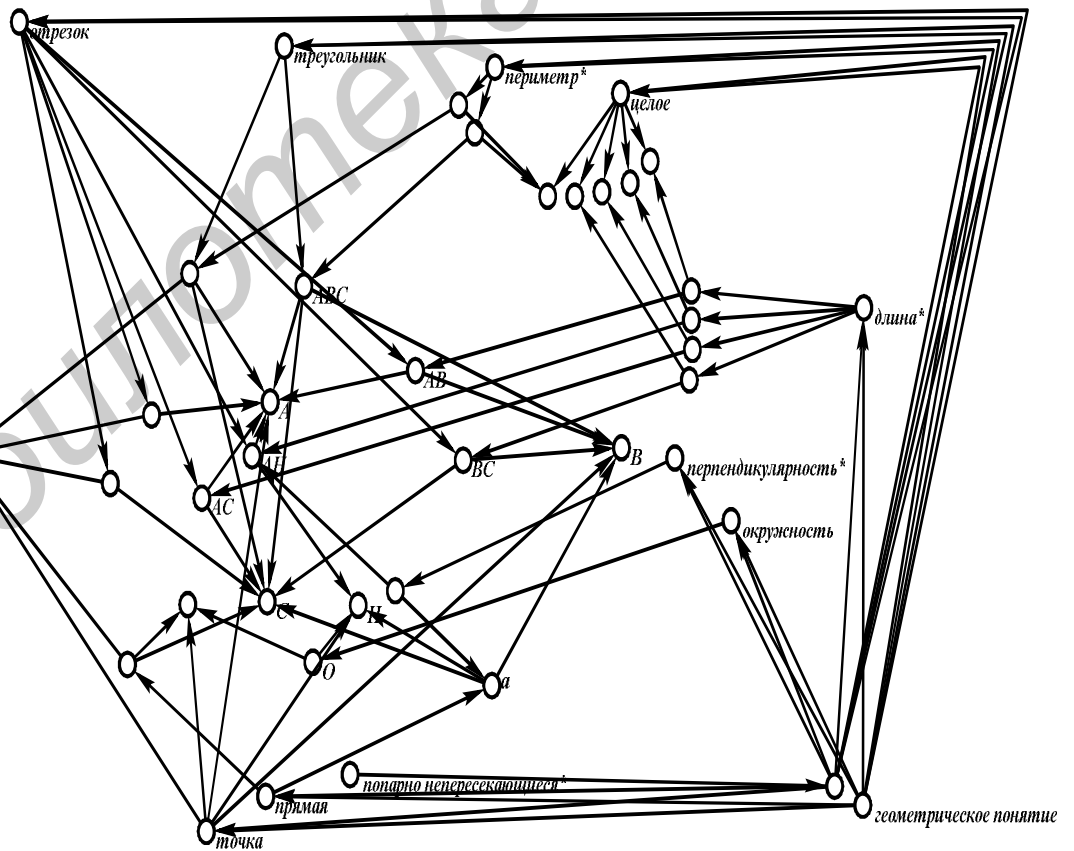


Рисунок 25 – Слияние с одним из вариантов

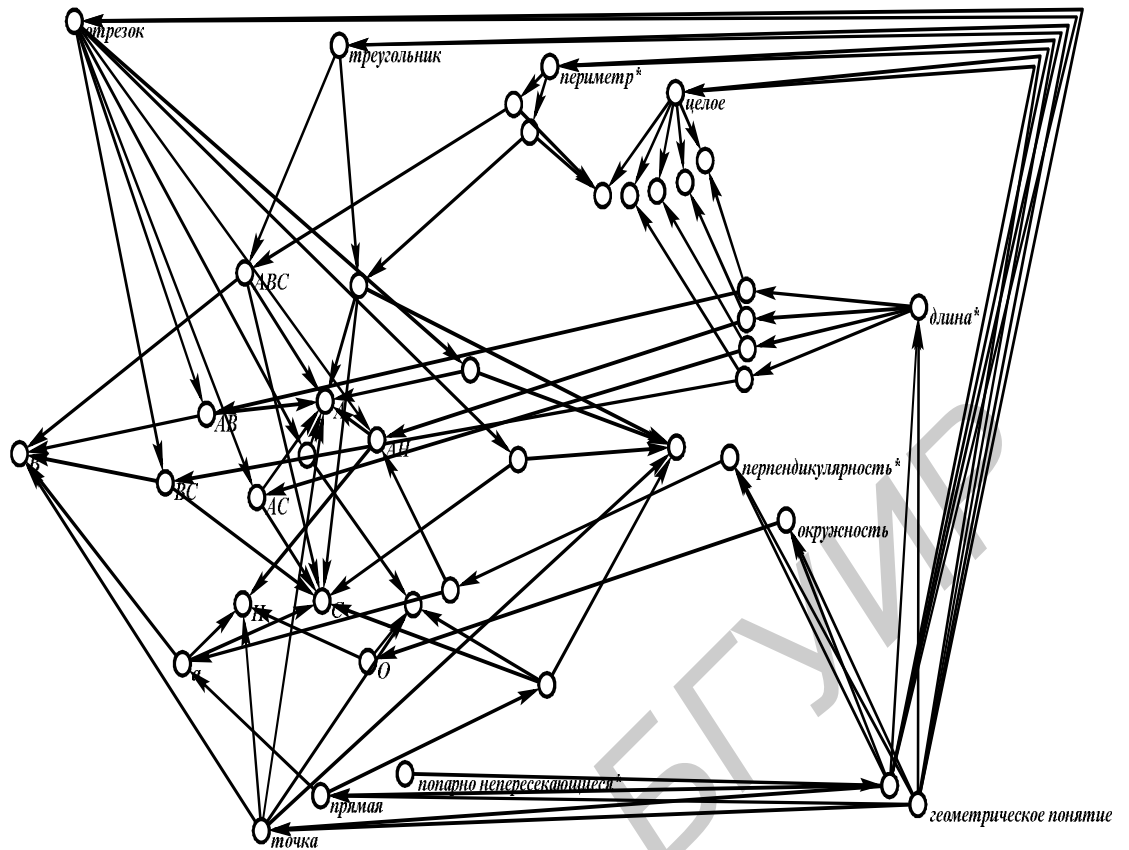


Рисунок 26 – Слияние со вторым из вариантов

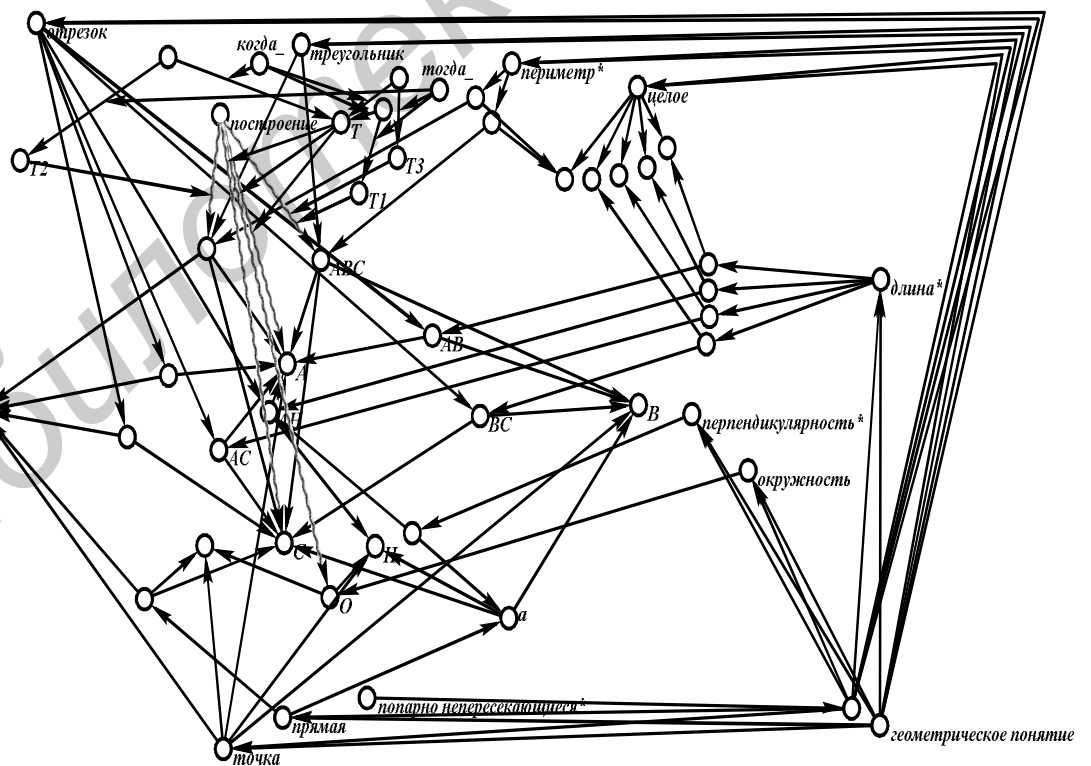


Рисунок 27 – Слияние трёх построений двух равных треугольников

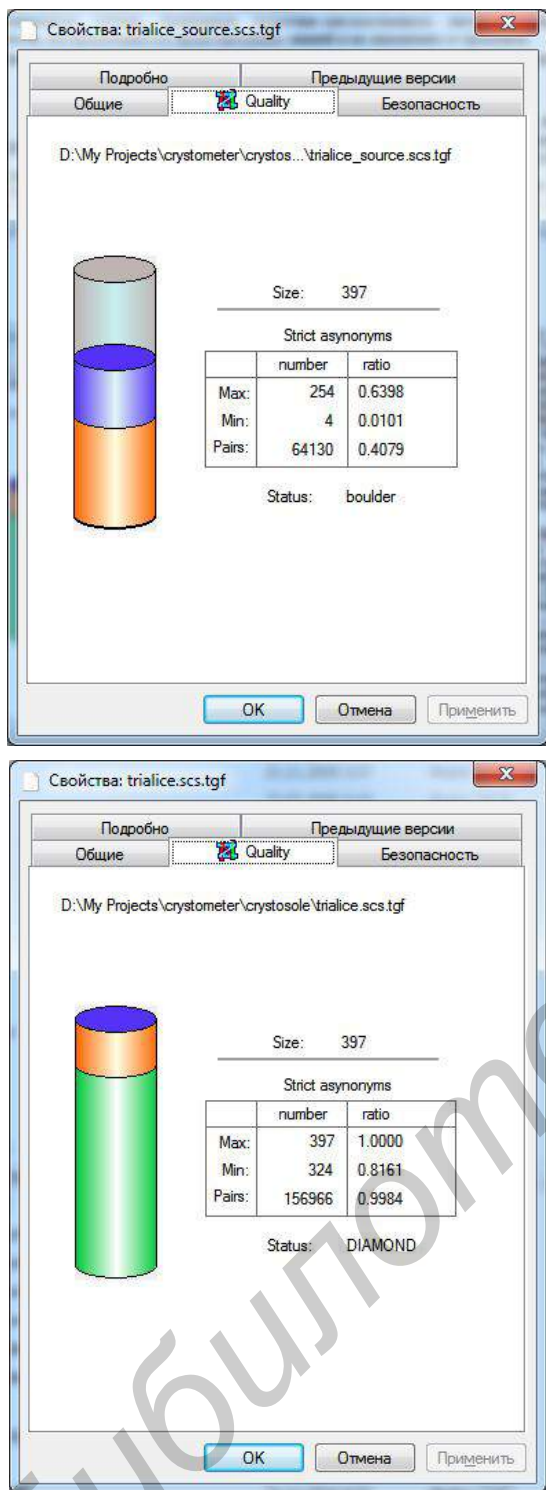


Рисунок 28 – Число пар различных обозначений

Алгоритм вычисления множества пар различающихся обозначений позволяет оценить качество фрагментов баз знаний. Так для фрагмента, который соответствует рисунку 17, количество пар различных обозначений составляет 64130, а для фрагмента на рисунке 18 – 156966 (рисунок 28).

ЗАКЛЮЧЕНИЕ

Рассмотренные модели и алгоритмы применяются для решения задач интеграции знаний в рамках технологии OSTIS [OSTIS, 2014]. Сравнение

результатов работы алгоритмов с системой ASMOV [Jean-Mary et al., 2007] показало преимущество предлагаемых решений от 1,45 раз по показателю точности и от 1,54 раз по показателю полноты. Реализация разработанных алгоритмов проведена средствами C/C++ и средствами процедурного языка обработки знаний SCP (Semantic Code Programming).

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

- [Гаврилова и др., 2000] Гаврилова, Т.А. Базы знаний интеллектуальных систем / Т. А. Гаврилова, В.Ф. Хорошевский. – СПб.: Питер, 2000.
- [Голенков и др., 2001] Голенков, В.В. Представление и обработка знаний в графодинамических ассоциативных машинах / В. В. Голенков[и др.]. – Мн.: БГУИР, 2001.
- [Гулякина и др., 2004] Гулякина Н.А. Ивашенко В.П. Интеграция знаний в информационных системах. / Н.А. Гулякина, В.П. Ивашенко // Доклады БГУИР. – 2004. – №6. – С. 113-119.
- [Ивашенко, 2003] Представление нейронных сетей и систем продукции в однородных семантических сетях. / В.П. Ивашенко // Известия Белорусской инженерной академии. – 2003. – №1(15)/1. – с.184-188.
- [Ивашенко, 2004] Применение однородных семантических сетей для представления знаний о нестационарных предметных областях. / В.П. Ивашенко // Известия Белорусской инженерной академии. – 2004. – №1(17)/3. – с.77-80.
- [Ивашенко, 2009а] Ивашенко В.П. Семантические модели баз знаний / В.П. Ивашенко Информационные системы и технологии (IST'2009): материалы V Междунар. конф.-форума в 2-х ч. Ч.2 – Минск: А.Н.Вараксин, 2009.- с.125-128.
- [Ивашенко, 2009б] Ивашенко В.П. Алгоритмы верификации и интеграции баз знаний. Вестник Брестского государственного технического университета, БрГТУ, 2009, №5.
- [Ивашенко, 2011а] Ивашенко В.П. Семантическая технология компонентного проектирования баз знаний. Материалы Международной науч.-техн. Конференции OSTIS,2011:Минск, Республика Беларусь, БГУИР 10-12 февраля 2011.
- [Ивашенко, 2011б] Ивашенко В.П. Алгоритмы операций отладки и интеграции баз знаний. Дистанционное обучение – образовательная среда XXI века: материалы VII Международной науч.-метод. конференции (Минск, 1-2 декабря 2011г.). – Минск: БГУИР, 2011, сс.227-229.
- [Ивашенко, 2012] Ивашенко В.П. Семантическая модели интеграции и отладки баз знаний. Материалы Международной науч.-техн. Конференции OSTIS,2012:Минск, Республика Беларусь, БГУИР 16-18 февраля 2012.
- [Колб, 2012] Колб, Д. Г. Web-ориентированная реализация семантических моделей интеллектуальных систем / Д. Г. Колб // Материалы международной научно-технической конференции «Открытие семантические технологии проектирования интеллектуальных систем» – Минск, 2012
- [Кофман, 1982] Кофман А. Введение в теорию нечетких множеств. М.: Радио и связь, 1982. — 432 с.
- [Кудрявцев, 2008] Кудрявцев Д. В. Практические методы отображения и интеграции онтологий. Семинар Знания и онтологии *Elsewhere*, КИИ-2008, Дубна, 2008.
- [Нариньяни, 2000] Нариньяни А. С. НЕ-факторы: неточность и недоопределенность — различие и взаимосвязь // Изв. РАН. Сер. Теория и системы управления. 2000. № 5. С. 44-56.
- [Тэрано, 1993] Прикладные нечеткие системы: Пер. с япон. / К. Асаи, Д. Ватада, С. Ивай и др.; Под ред. Т. Тэрано, К. Асаи, М. Сугэно. – М.: Мир, 1993.
- [Хорошевский, 2008] Хорошевский, В.Ф. Пространства знаний в сети Интернет и Semantic Web (Часть 1) / В. Ф. Хорошевский // Искусственный интеллект и принятие решений. – 2008. - № 1. - С.80-97.
- [Allen, 1991] Allen, J.F.: Time and time again: the many ways to represent time. International Journal of Intelligent Systems 6 (1991) 341-355.
- [Allen, 1994] Allen, J. F. Ferguson, G. Actions and Events in Interval Temporal Logic. J. Log. Comput., 1994, 4, 531-579.
- [Aumuller et al., 2005] D. Aumuller, H. Do, S. Massmann, E. Rahm. 2005. Schema and ontology matching with COMA++. Proc. of

the 2005 International Conference on Management of Data, pp. 906-908.

[Bouquet, 2003] P. Bouquet, L. Serafini and S. Zanobini, M. Benerecetti. An algorithm for semantic coordination. Proc. of Semantic Integration Workshop, colocated with the 2nd Int'l Semantic Web Conference (ISWC2003), 20-23 October 2003, Sanibel Island, Florida, US.

[DAML, 2006] DAML.org: The DARPA Agent Markup Language Homepage, Электронный ресурс. Режим доступа: <http://www.daml.org/>.

[Doan and Halevy, 2005] A. H. Doan, A. Y. Halevy. Semantic integration research in the database community: A brief survey. AI magazine, 26(1), 2005

[CLIPS, 1991] CLIPS Version 5.1 User's Guide, NASA Lyndon B. Johnson Space Center, Software Technology Branch, Houston, TX, 1991.

[CoFI:CASL-Summary, 2004] CoFI Language Design Group. Casl summary. In Casl Reference Manual, LNCS 2960 (IFIP Series), part I. Springer, 2004.

[CycL, 2002] The Syntax of CycL. Электронный ресурс. Режим доступа: <http://www.cyc.com/cycdoc/ref/cycl-syntax.html>.

[Ganter, 1999] B. Ganter, R. Wille: Formal Concept Analysis – Mathematical Foundations. Springer, Heidelberg 1999.

[Gangemi et al., 1996] A. Gangemi, G. Steve and F. Giacomelli, 1996. "ONIONS: An Ontological Methodology for Taxonomic Knowledge Integration", ECAI-96 Workshop on Ontological Engineering, Budapest, August 13th.

[Genesereth and Fikes, 1992] Michael R. Genesereth and Richard E. Fikes. Knowledge interchange format version 3.0 reference manual. Logic Group First Draft January 1992 Report Logic-92-1 Current Version June 1992.

[Giunchiglia et al., 2006] F. Giunchiglia, M. Marchese and I. Zaihrayeu (2006). Encoding classifications into lightweight ontologies. University of Trento Technical Report # DIT-06-016, March 2006.

[Jean-Mary et al., 2007] Jean-Mary Y., Kabuka, M. ASMOV: Ontology Alignment with Semantic Validation. Joint SWDB-ODDIS Workshop, September 2007, Vienna, Austria, 15-20.

[IEEE, 2006] IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, Automatic Fuzzy Ontology Generation for Semantic Web, VOL. 18, NO. 6, JUNE 2006.

[IDEF5, 1994] Information Integration for Concurrent Engineering (IICE). IDEF5 Method Report. - Knowledge Based Systems, Inc., 1408 University Drive East College Station, Texas, USA. September 21, 1994.

[ISO13250] ISO/IEC 13250:2003 Topic maps – representation and interchange of knowledge, with an emphasis on the findability of information.

[ISO24707] ISO/IEC 24707:2007 – Information technology – Common Logic (CL): a framework for a family of logic-based languages.

[Maldavan, 2001] J. Madhavan, P. A. Bernstein, E. Rahm. Generic Schema Matching with Cupid. In. Proc. of the 27th Conference on Very Large Databases, 2001.

[Maltese et al., 2010] V. Maltese, F. Giunchiglia, A. Autayeu: Save up to 99% of your time in mapping validation In Proceedings of ODBASE, 2010.

[Martin Ph., 2002] Martin Ph. (2002). Knowledge representation in CGLF, CGIF, KIF, Frame-CG and Formalized-English. Proceedings of ICCS 2002, 10th International Conference on Conceptual Structures (Springer, LNAI 2393, pp. 77-91), Borovets, Bulgaria, July 15-19, 2002.

[McGuinness et al., 2000] D. L. McGuinness, R. Fikes, J. Rice, and S. Wilder. The chimaera ontology environment. In Proceedings of AAAI, pages 1123-1124, 2000.

[Mitra, 2001] Prasenjit Mitra, Gio Wiederhold and Stefan Decker: A Scalable Framework for Interoperation of Information Sources. 1st International Semantic Web Working Symposium (SWWS '01), Stanford University, Stanford, CA, July 29-Aug 1, 2001, Jul.

[Michael Kifer et al., 1995] Michael Kifer, Georg Lausen, James Wu: Logical Foundations of Object-Oriented and Frame-Based Languages. Journal of ACM 42(4): 741-843 May (1995).

[Mustafa Jarrar et al., 2008] Mustafa Jarrar and Robert Meersman: "Ontology Engineering -The DOGMA Approach". Book Chapter (Chapter 3). In Advances in Web Semantics I. Volume LNCS 4891, Springer, 2008.

[Nagy et al., 2010] Miklos Nagy and Maria Vargas-Vera. Towards an Automatic Semantic Data Integration: Multi-Agent Framework Approach. Sematic Web. Gang Wu (ed), Chapter 7, pp.

107-134; In-Tech Education and Publishing KG; 2010, ISBN 978-953-7619-54-1.

[Novak et al., 2008] Joseph D. Novak & Alberto J. Canas, The Theory Underlying Concept Maps and How to Construct and Use Them. Florida Institute for Human and Machine Cognition. Pensacola FL, 32502. Technical Report IHMC CmapTools 2006-01 Rev 2008-01.

[OSTIS, 2014] Открытая семантическая технология проектирования интеллектуальных систем [Электронный ресурс]. – 2014. - Режим доступа: <http://ostis.net>. – Дата доступа: 2.11.2014

[Smith et al., 2007] Smith, B.; Ashburner, M.; Rosse, C.; Bard, J.; Bug, W.; Ceusters, W.; Goldberg, L. J.; Eilbeck, K. et al. (2007). "The OBO Foundry: Coordinated evolution of ontologies to support biomedical data integration". Nature Biotechnology 25 (11): 1251-1255. doi:10.1038/nbt1346.

[Sowa et al., 2008] Sowa, J. Conceptual Graphs/ John F. Sowa, F. van Harmelen, V. Lifschitz, B. Porter// eds., Handbook of Knowledge Representation, Elsevier, 2008, pp. 213-237.

[Stumme, 2001] Gerd Stumme and Alexander Maedche. FCA-merge: bottom-up merging of ontologies. In In Proceedings of 17th IJCAI, pages 225{230, Seattle (WA), USA, 2001.

[W3C:DAML+OIL, 2001] DAML+OIL (March 2001) Reference Description: W3C Note 18 December 2001. Dan Connolly, Frank van Harmelen, Ian Horrocks, Deborah L. McGuinness, Peter F. Patel-Schneider, Lynn Andrea Stein and Lucent Technologies, Inc. Электронный ресурс. Режим доступа: <http://www.w3.org/TR/daml+oil-reference>.

[W3C:RDFS, 2004] RDF Vocabulary Description Language 1.0: RDF Schema: W3C Recommendation 10 February 2004 Электронный ресурс. / eds. Dan Brickley, R.V. Guha. Режим доступа: <http://www.w3.org/TR/rdf-schema/>, свободный.

[W3C:RIF, 2010] RIF Overview: W3C Working Group Note 22 June 2010. Электронный ресурс. / eds. Michael Kifer, Harold Boley. Режим доступа: <http://www.w3.org/TR/2010/NOTE-rif-overview-20100622/>.

[W3C:OWL, 2004] OWL Web Ontology Language. Overview: W3C Recommendation 10 February 2004 Электронный ресурс. /eds. Deborah L. McGuinness, Frank van Harmelen. Режим доступа: <http://www.w3.org/TR/owl-features/>.

[W3C:OWL2, 2009] OWL 2 Web Ontology Language Document Overview: W3C Recommendation 27 October 2009. Электронный ресурс. / eds. W3C OWL Working Group. Режим доступа: <http://www.w3.org/TR/owl2-overview/>.

[W3C:SPARQL, 2008] SPARQL Query Language for RDF: W3C Recommendation 15 January 2008. Электронный ресурс. / eds. Eric Prud'hommeaux, Andy Seaborne. Режим доступа: <http://www.w3.org/TR/rdf-sparql-query/>.

[XTM, 2001] XML Topic Maps (XTM) 1.0: TopicMaps.Org Specification. Members of the TopicMaps.Org Authoring Group. Электронный ресурс. / eds. Steve Pepper, Graham Moore. Режим доступа: <http://www.topicmaps.org/xtm/>.

[Van Renssen, 2005] Van Renssen, Andries (2005). Gellish: A Generic Extensible Ontological Language. Delft University Press. ISBN90-407-2597-4. <http://repository.tudelft.nl/view/ir/uuid%3Ade26132b-6f03-41b9-b882-c74b7e34a07d/>.

MODELS AND ALGORITHMS OF INTEGRATION OF KNOWLEDGE BASED ON HOMOGENEOUS SEMANTIC NETWORKS

Ivashenko V.P.

Belarusian State University of Informatics and Radioelectronics, Minsk, Republic of Belarus

ivashenko@bsuir.by

The work concerns models and algorithms of integration of knowledge with unified representation in the form of homogeneous semantic networks. These networks have model (set-theoretic) semantic that describes incompleteness, fuzziness, uncertainty of represented knowledge in dynamic conditions and changes over time.