

УДК 519.711.3

АНАЛИЗ НЕПОЛНЫХ ДАННЫХ В ЗАДАЧАХ ПОСТРОЕНИЯ ФОРМАЛЬНЫХ ОНТОЛОГИЙ*

Д.Е. Самойлов^{1,2}, В.А. Семёнова^{1,2}, С.В. Смирнов²

¹ Самарский национальный исследовательский университет им. академика С.П. Королева, Самара, Россия
dmitrysam3@gmail.com, queenbfjr@gmail.com

² Институт проблем управления сложными системами РАН, Самара, Россия
smirnov@iccs.ru

Аннотация

В статье рассматривается проблема автоматизации формирования онтологических спецификаций предметных областей на основе измерений – стержневая проблема онтологического анализа данных. Представлены модели и методы, направленные на выявление понятийной структуры и, в конечном счете, формальной онтологии исследуемой предметной области. Фундаментальные реалии накопления эмпирической информации: многократные независимые измерения каждого свойства объекта обучающей выборки; конгруэнтность части процедур измерения; дифференциация доверия к различным источникам данных - отражены в модели обобщенной таблицы «объекты-свойства». Неполнота (неточность, противоречивость, неопределенность) этой информации влечет необходимость использования для её первичной обработки моделей многозначной логики. Результат такой обработки - нестрогий формальный контекст – должен быть аппроксимирован однозначным контекстом, из которого возможен вывод формальных понятий в рамках прикладной ветви теории решёток, известной как «анализ формальных понятий». Исследуется генезис «ограничений существования свойств», которые влияют на корректность аппроксимации нестроого формального контекста. Предлагаются модели и метод учета этой дополнительной информации. Формулируются прагматически обоснованные принципы преобразования решетки формальных понятий в формальную онтологию. Приводится модельный пример использования разработанных моделей и методов онтологического анализа данных.

Ключевые слова: онтологический анализ данных, формальная онтология, анализ формальных понятий, многозначная векторная логика, ограничения существования свойств.

Цитирование: Анализ неполных данных в задачах построения формальных онтологий / Д.Е. Самойлов, В.А. Семёнова, С.В. Смирнов // Онтология проектирования. – 2016. – Т. 6, №3(21). – С. 317-339. – DOI: 10.18287/2223-9537-2016-6-3-317-339.

Введение

Сложные информационные системы результативны лишь при надежном и согласованном (в разных смыслах) представлении их предмета. Систематизация, разработка и использование соответствующих информационных моделей составляют современное содержание *онтологического подхода* в вычислительных науках.

В отличие от философии онтология в информатике описывает некоторую ограниченную сферу знания, *предметную область* (PrO). Поэтому в силу множественности наук и PrO, когда каждая из них имеет свою собственную или даже несколько конкурирующих терминологий, здесь в противоположность философии приобретает смысл употребление множественного числа для термина, т.е. говорят об «онтологиях». Более того, различают лингви-

* Журнальный вариант докладов на Международной конференции и молодежной школе «Информационные технологии и нанотехнологии» ИТНТ-2015, 2016. Рекомендован к опубликованию Программным комитетом ИТНТ.

стические и *формальные онтологии*, где последние наследуют парадигму моделей и методов представления знаний, разработанных в искусственном интеллекте [1, 2].

Формальная онтология описывает ПрО с помощью стандартизирующей терминологии – словаря, а также неоднородных связей между определенными в нём понятиями. Моделирующими примитивами онтологической спецификации обычно служат классы, отношения, функции и аксиомы, что в некотором смысле сближает эти структуры представления знаний и соответствующие компьютерные ресурсы с алгебраическими системами А.И. Мальцева [3].

Сегодня можно указать три основных пути разработки онтологий [4]:

- наиболее используемый связан с прямой формализацией опыта и знаний экспертов, которые с помощью компьютерных языковых средств либо автоформализуют свои представления о ПрО, либо фиксируют их с помощью инженера по знаниям [5];
- при наличии развитой инфраструктуры работы со знаниями актуальные онтологии могут синтезироваться в результате человеко-машинных процедур конкретизации, композиции/декомпозиции апробированных формальных онтологий, их прототипов (паттернов) разного уровня и направленности [6-8];
- третий путь связан с автоматическим «выводом» формальных онтологий из доступных данных. Эти данные рассматриваются как результат измерений *свойств объектов* исследуемой ПрО и сводятся в стандартизованные *таблицы «объекты-свойства»* (ТОС) [9, 10], анализ которых приводит к выявлению понятийной структуры ПрО. Наиболее результативные методы этого направления опираются на ветвь теории решёток – анализ формальных понятий (АФП) [11-15]. В исходном варианте АФП доставляет лишь «скелет» онтологии [16], т.е. множество *формальных понятий* и существующее на этом множестве *отношение обобщения* (is-A). Методика выявления с помощью АФП произвольных семантических отношений между формальными понятиями предложена в [17] на основе интерпретации нетаксономических связей между объектами ПрО как проявления *особых свойств-валентностей*.

Рассматривая третий путь построения формальных онтологий - *онтологический анализ данных* (ОАД), - важно подчеркнуть иную, чем в других случаях роль субъекта онтологического инжиниринга («онтолога»). Фактически его основной задачей становится выдвижение гипотез о свойствах объектов исследуемой ПрО и затем априорное комплектование арсенала измерительных процедур (органов чувств, вербальных возможностей экспертов, искусственных сенсоров, приборов, систем и т.д.), с помощью которых интересующая его ПрО будет зондироваться.

В предлагаемой статье внимание сосредоточено на двух аспектах ОАД. Во-первых, это отражение реалий сбора данных о ПрО, приводящее к необходимости использования для представления эмпирической информации моделей многозначной логики, а, во-вторых (и в связи с первым), на моделировании и методе учета в работе априори известных исследователю зависимостей между измеряемыми свойствами.

1 Формирование эмпирических данных о предметной области

1.1 Отправные тезисы

В контраст с общепринятой точкой зрения в анализе данных ОАД исходит из положения, что всякое измерение свойства объекта может дать специальный результат «None», который может свидетельствовать о семантическом несоответствии исследуемого объекта и измерительной процедуры, о нахождении значения измеряемого свойства за порогами чувствительности, вне динамического диапазона средства измерений [18].

В классическом АФП подобные эффекты достигаются в результате выполнения когнитивной процедуры, именуемой *концептуальным шкалированием* [11, 19]. Здесь ТОС именуется *многозначным формальным контекстом* и описывается кортежем

$$(G^*, M, V, I),$$

где:

- $G^* = \{g_i\}_{i=1, \dots, r}$, $r = |G^*| \geq 1$ - набор объектов исследуемой ПрО, попавших в поле зрения онтолога; иначе говоря, это множество объектов *обучающей выборки*: $G^* \subseteq G$, где G – всё мыслимое множество объектов ПрО;
- $M = \{m_j\}_{j=1, \dots, s}$, $s = |M| \geq 1$ - множество измеряемых у объектов свойств;
- V - множество значений свойств;
- I - тернарное отношение между G^* , M и V ($I \subseteq G^* \times M \times V$), определённое для всех пар из $G^* \times M$.

Неформально концептуальное шкалирование измеряемого свойства представляет собой субъективное конструирование «покрытия» динамического диапазона (домена доставляемых значений) соответствующей процедуры измерения с образованием *новых* измеряемых свойств у объектов ПрО. Такое покрытие именуется *концептуальной шкалой*. Пример концептуальной шкалы с дизъюнктивным «расщеплением» динамического диапазона процедуры измерения даёт таблица 1. После шкалирования вновь введенные свойства фактически измеряются в бинарной шкале наименований $\{X, \text{None}\}$, где X - лингвистическая константа, собирательно обозначающая любой символ шкалы динамического диапазона измерительной процедуры.

Таблица 1 – Концептуальная шкала для свойства «Рост» объектов-мужчин («x» - **Истина**, пробел - **Ложь**)

Рост мужчины, см	Низкий	Средний	Высокий
< 168	x		
168-175		x	
> 175			x

Так или иначе, подобная «None-концепция» позволяет изменить парадигму анализа экспериментального материала о ПрО и, прежде всего, естественным образом преобразовать ТОС в совокупность оценок истинности *базовых семантических суждений* (БСС) о ПрО вида $b_{ij} = \langle \text{объект } g_i \text{ обладает свойством } m_j \rangle$:

$$(1) \quad \|b_{ij}\| = \begin{cases} \text{Истина, если результат измерения свойства } m_j \text{ у объекта } g_i \text{ есть } X; \\ \text{Ложь, в противоположном случае.} \end{cases}$$

Именно на обработку таких данных ориентирован АФП, в котором используются следующие обозначения и модели [11]:

- $K = (G^*, M, I)$ – формальный контекст, где G^* и M имеют уже указывавшийся смысл, I – бинарное соответствие «объекты-свойства», т.е. совокупность оценок вида (1);
- операторы Галуа φ , ω (общая нотация «'») для контекста K :
 $\varphi(X) = X' = \{m_j \mid m_j \in M, \forall g_i \in X: (g_i, m_j) \in I\}$ - общие свойства объектов, составляющих $X \subseteq G^*$, или Галуа-проекция X на M ;
 $\omega(Y) = Y' = \{g_i \mid g_i \in G^*, \forall m_j \in Y: (g_i, m_j) \in I\}$ - объекты, которые обладают всеми свойствами из $Y \subseteq M$, или Галуа-проекция Y на G^* ;
- (X, Y) – формальное понятие, у которого $X \subseteq G^*$ - объем, $Y \subseteq M$ - содержание, причем $X = Y'$, $Y = X'$;
- $B(K)$ - множество формальных понятий контекста K ;

- $(B(K), \leq)$ – замкнутая решетка понятий, отражающая бинарное отношение частичного порядка на множестве формальных понятий: $(X_1, Y_1) \leq (X_2, Y_2)$, если $X_1 \subseteq X_2$, или эквивалентно $Y_1 \supseteq Y_2$, - известное как отношение *обобщения понятий*, или *наследования свойств*.

ОАД в полной мере опирается на модели АФП.

1.2 Обобщенная таблица «объекты-свойства»

На практике качество исходного эмпирического материала таково, что, по крайней мере, для части БСС оценка истинности расплывчата (например, формируется экспертом на основе опыта или интуиции), и для оценивания таких суждений естественнее употреблять истинностные значения, вводимые *многозначными логиками*. Однако тогда возникает вопрос о модели, объясняющей происхождение и определяющей способ вычисления подобных оценок.

В общем плане ясно, что многозначность оценок истинности БСС вызывает неполнота данных о ПрО (неточность, противоречивость, неопределённость и т.п.), которая не находит отражения в стандартной структуре ТОС. Причины неполноты вызываются *реалиями накопления эмпирической информации*. Фундаментальными факторами здесь являются:

- выполнение, как правило, многократных независимых измерений свойства $m_j \in M$ у объекта $g_i \in G^*$;
- использование для измерения одного и того же свойства m_j нескольких различных процедур (конгруэнтных источников информации);
- дифференциация доверия к различным процедурам измерения.

Поэтому в качестве адекватной модели исходных данных предлагается *обобщенная* ТОС, описываемая кортежем

$$(2) \quad (G^*, M, Se, Pr, A),$$

где (см. рисунок 1):

- $Se = \bigcup_{i=1}^r Se_{(i)}$ - множество всех выполненных при зондировании ПрО серий измерений, $|Se| = \sum_{i=1}^r |Se_{(i)}| = m$, и $Se_{(i)} = \{se_{(i)k}\}_{k=1, \dots, q_{(i)}}$, $q_{(i)} \geq 1$, $i = 1, \dots, r$ - множество серий измерений, которым подвергнут объект $g_i \in G^*$;
- $Pr = \bigcup_{j=1}^s Pr_{(j)}$ - арсенал всех используемых при зондировании ПрО процедур измерения, $|Pr| = \sum_{j=1}^s |Pr_{(j)}| = n$, и $Pr_{(j)} = \{pr_{(j)k}\}_{k=1, \dots, p_{(j)}}$, $p_{(j)} \geq 1$, $j = 1, \dots, s$ - множество конгруэнтных процедур измерения свойства $m_j \in M$, причем всякая процедура $pr_{(j)k}$ характеризуется степенью доверия к её результатам $t_{(j)k}$;
- $A = (a_{ij})_{i=1, \dots, m; j=1, \dots, n}$ - матрица результатов серий измерений Se свойств M у объектов из выборки G^* , выполненных с помощью процедур измерения Pr . Элементами этой матрицы могут быть константы **X** и **None**, а также ещё две лингвистические константы. Константа **Failure** фиксирует отказ, сбой измерительного средства, воздержание при голосовании и т.п., т.е. тот нередко наблюдаемый на практике «результат» работы измерительной процедуры, который собирательно можно квалифицировать как «отказ от выполнения измерения». Константа **NM** (*not measured*) указывает, что в действительности в рассматриваемой серии измерений отдельная процедура измерения не использовалась (введение этого формального результата необходимо, помимо прочего, для сохранения двумерного характера обобщенной таблицы «объекты-свойства»).

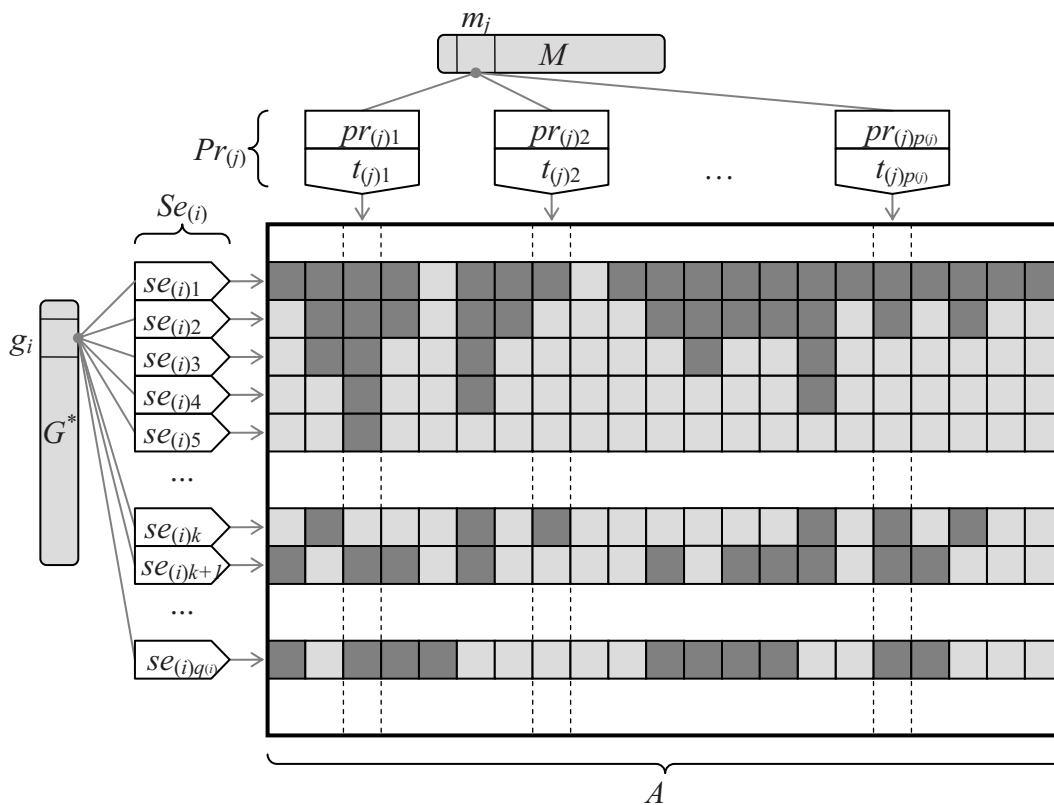


Рисунок 1 – Структура обобщенной таблицы «объекты-свойства» на основе двумерной матрицы (темные ячейки матрицы соответствуют результатам измерения **X**, **None** и **Failure**, а светлые – результату **NM**)

1.3 Построение нестроого формального контекста и задача его однозначной аппроксимации

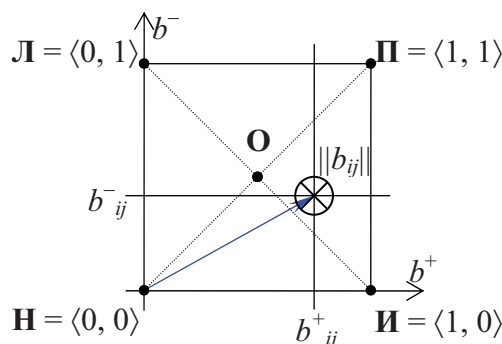
Модель (2) позволяет вычислять «мягкие», многозначные оценки истинности БСС о ПрО.

Например, это вычисление можно осуществить на основе нечёткой логики. Однако на сегодня большей мощностью моделирования «человеческого подхода» к подобным оценкам обладают *векторные логики*, для которых хорошо развит понятийный и аналитический аппарат, обобщающий положения и классической, и нечеткой, и различных многозначных логик [20]. В частности, в логике класса V^{TF} истинность БСС будет оцениваться *вектором* ⟨Истина, Ложь⟩ (рисунок 2):

$$(3) \quad \|b_{ij}\| = \langle b^+_{ij}, b^-_{ij} \rangle; b^+_{ij}, b^-_{ij} \in [0, 1],$$

где компонент, или аспект истинности, b^+_{ij} – *Истина*, формируется свидетельствами (выражающими личный опыт и знания онтолога, поступающими к нему от экспертов, находимыми им в литературе, приобретаемыми им в специально поставленных экспериментах и т.п.), *подтверждающими* БСС, а компонент (аспект) b^-_{ij} – *Ложь*, - *отрицающими* БСС. Ожидание большей адекватности *нестроого* [20] соответствия I , заданного матрицей векторов (3), реальному процессу осмысления ПрО может быть объяснено тем, что традиционные константы **Истина** и **Ложь**, которыми онтолог интуитивно стремится оценить истинность БСС, зачастую определяется для него независимым комплексом свидетельств так, что **Ложь** не выводима из отсутствия (дефицита) **Истины**, а **Истина** – из отсутствия (недостачи) **Лжи**.

Этапы построения нестрогого формального контекста K – формального контекста с нестрогим соответствием I – сводятся к следующему.



- ⊗ - вектор истинности $\|b_{ij}\| = \langle b^+_{ij}, b^-_{ij} \rangle$, $b^+_{ij}, b^-_{ij} \in [0, 1]$ базового семантического суждения b_{ij} ;
- - истинностные константы V^{TF} -логики:
 - \mathbf{H} - «Неопределённость», \mathbf{L} - «Ложь», \mathbf{I} - «Истина», \mathbf{P} - «Противоречие»;
 - $\mathbf{O} = \langle 0.5, 0.5 \rangle$ - «Ожидание [оценки]» (дополнительная константа)

Рисунок 2 – Оценка истинности базового семантического суждения в векторной V^{TF} -логике

Первым шагом является переход от первичных данных, структурированных в виде матрицы A , к их смысловой интерпретации в виде нестрогого соответствия «серии измерений – процедуры измерений» I :

$$(4) \quad \|b_{ij}\|_{i=1,\dots,m; j=1,\dots,n} = \left. \begin{array}{ll} \mathbf{I} = \langle 1, 0 \rangle, & \text{если } a_{ij} = \mathbf{X}; \\ \mathbf{L} = \langle 0, 1 \rangle, & \text{если } a_{ij} = \mathbf{None}; \\ \mathbf{O} = \langle 0.5, 0.5 \rangle, & \text{если } a_{ij} \in \{\mathbf{Failure}, \mathbf{NM}\}. \end{array} \right\},$$

где \mathbf{I} , \mathbf{L} и \mathbf{O} – истинностные константы V^{TF} -логики (см. рисунок 2).

Учёт значений **Failure** потребовал введения в анализ истинностной константы «Ожидание [оценки]» $\mathbf{O} = \langle 0.5, 0.5 \rangle$, которая имеет аналоги в прагматически ориентированных многозначных логиках (см., например, дополнительное значение истинности «круглый информационный ноль» **Крин** в [21]) и равноудалена от известных в V^{TF} -логике истинностных констант «Истина» \mathbf{I} , «Ложь» \mathbf{L} , «Неопределённость» $\mathbf{H} = \langle 0, 0 \rangle$ и «Противоречие» $\mathbf{P} = \langle 1, 1 \rangle$ (см. рисунок 2). Пожалуй, «Ожидание» - наиболее адекватная оценка БСС и в случае, когда необходимое измерение не проводилось.

Далее нестрогое соответствие I трансформируется в следующем порядке.

Во-первых, каждая оценка истинности БСС в I совмещается со степенью доверия к источнику данных, которые были положены в основу этой оценки. Используется простейшая модель учета степени доверия к измерительным процедурам в виде скалярных мер $t_{(j)} \in [0, 1]$ и элементарная формула пересчета вектора истинности БСС b_{ij} : $\|b_{ij}\| := t_j \cdot \|b_{ij}\|$ (фактически это соответствует применению для моделирования дифференциации доверия онтолога к источникам информации вектора истинности V^{TF} -логики (t_j^+, t_j^-) , где $t_j^+ = t_j^- = t_j$, и объединению свидетельств об истинности каждого БСС и уровня доверия к источнику этого БСС по схеме так называемого «00-композиционного совмещения» на основе композиционного умножения согласно t -норме $x \bullet y = xy$ [20]).

Во-вторых, согласно полученному соответствию I определяются оценки истинности БСС по результатам всех серий измерений, выполненных для отдельно взятого объекта каждой измерительной процедурой. Когда для объекта $g_i \in G^*$ произведена лишь одна серия измерений, т.е. $|Se_{(i)}| = 1$, то требуемая оценка де-факто уже имеется. Иначе оценки истинности во всех строках соответствия I , соответствующих элементам серии $Se_{(i)}$, должны быть совмеще-

ны. Другими словами, обработка данных многократного измерения свойства объекта приводит к необходимости интеграции независимых свидетельств истинности отдельного БСС.

Для совмещения истинностных значений в векторной логике имеется несколько допустимых схем. Наиболее адекватной смысловой деятельности онтолога следует признать схему «усиления-усреднения» - частный случай так называемого «11-композиционного совмещения» на основе композиционного сложения согласно триангулированной s -норме $x \oplus y = \min(1, x + y)$ [20].

Важно отметить, что совмещая оценки, полученные на основе работы отдельно взятой измерительной процедуры применительно к определённому объекту, следует безусловно *отбрасывать* оценки, которые были помещены в I на основе «не-измерений» NM , т.к. их появление вызвано исключительно формальными потребностями структуризации обобщенной ТОС, поэтому их учёт приведёт к некорректному результату совмещения. Единственным изъятием из этого правила будет случай, когда результат NM был «получен» во всех без исключения сериях измерений $Se_{(i)}$.

Наконец, *в-третьих*, для каждого свойства $m_j \in M$ в формируемом соответствии I «свертывается» информация столбцов, соответствующих конгруэнтным процедурам $Pr_{(j)}$. И эту интегрирующую операцию для онтолога естественно производить по схеме «усиления-усреднения» при совмещении свидетельств векторной логики.

В результате получаем нестрогий формальный контекст (G^*, M, I) , где итоговое нестрогое соответствие I имеет размерность $r \times s$, а его элемент вида (3) можно рассматривать как векторный показатель принадлежности свойства m_j априори неизвестному множеству M_i свойств объекта g_i . В терминах V^{TF} -логики M_i есть нестрогое подмножество универсума M .

На сегодня не существует эффективных методов вывода понятийной структуры Про непосредственно из «мягких» формальных контекстов. Например, сложный в теоретическом и вычислительном плане метод, использующий оператор замыкания нечёткого множества [22], представляет лишь академический интерес, поскольку генерирует гигантское количество нечетких понятий даже для малоразмерных «разреженных» нечётких контекстов. Поэтому результативные методы основаны на формально возможном разложении логически многозначного соответствия «объекты-свойства» по его *однозначным* (с элементами инцидентности, определёнными в бинарной шкале наименований {Истина, Ложь}) *альфа-сечениям* с установлением онтологом порога доверия к исходным данным (см., например [23, 24]).

Применительно к нестрогости соответствию I имеем [20]:

$$I = \bigcup_{\alpha^+, \alpha^- \in [0,1]} \langle \alpha^+, \alpha^- \rangle \cdot I^{(\alpha)},$$

$$I^{(\alpha)} = (\|b^{(\alpha)}_{ij}\|)_{i=1,\dots,r; j=1,\dots,s}, \quad \|b^{(\alpha)}_{ij}\| = \left\{ \begin{array}{ll} \text{Истина,} & \text{если } b_{ij}^+ \geq \alpha^+ \wedge b_{ij}^- \leq \alpha^-; \\ \text{Ложь} & \text{в противоположном случае} \end{array} \right\},$$

где α -сечение $I^{(\alpha)}$ - обычное бинарное отношение «объекты-свойства» векторного уровня $\alpha = \langle \alpha^+, \alpha^- \rangle$.

На практике при задании онтологом порога доверия α к исходным данным сечение $I^{(\alpha)}$ используют в качестве *приближения*, так называемой « α -аппроксимации» исходного «мягкого» соответствия I . Затем к полученной аппроксимации применяются (с различными дополнениями) апробированные методы вывода понятий.

Тем не менее, и этот подход в общем случае оказывается *некорректным*. Например, рост конкретного мужчины в однозначном формальном контексте, полученном путём α -сечения, может быть оценён и как низкий, и как высокий, поскольку приём α -аппроксимации не учитывает зависимости между измеряемыми свойствами.

2 Ограничения существования свойств

2.1 Зависимости между свойствами

Действительно, стандартная процедура альфа-сечения не учитывает какие бы то ни было связи между элементами множеств-носителей нестрогого отношения «объекты-свойства».

Поскольку об объектах ПрО согласно общей парадигме анализа данных нам ничего неизвестно (кроме предположения о возможности выделить объект в ПрО и осуществить измерения его свойств), то рассматриваемая «слепота» процедуры альфа-сечения индифферентна по отношению к обучающей выборке объектов.

Иначе дело обстоит с множеством измеряемых свойств. Однозначная аппроксимация нестрогого соответствия «объекты-свойства» способна «обнажить» противоречия между БСС, которые проявляются для онтолога как *априори недопустимые сочетания свойств объектов*. В общем случае эти противоречия не могут быть обнаружены в формальном контексте с мягкими оценками истинности БСС.

В концептуальном плане источником проблемы является присущая ОАД когнитивная асимметрия «объектов» и «свойств». Формально «объекты» независимы от онтолога, описывающего ПрО. Напротив «свойства» - результат продуцирования онтологом гипотез о ПрО на основе достигнутого им ранее миропонимания или, точнее говоря, имеющейся у него совокупности *априорных знаний*. Онтолог является «владельцем» арсенала процедур измерения свойств, обладая достаточно полной информацией об этих инструментах. В частности, он может (и должен) знать об определённых зависимостях между результатами выполнения различных процедур измерения применительно к одному и тому же объекту, т.е. о *зависимостях между свойствами объектов*.

Общие модели характерных типов связей между свойствами предложены в [25] в форме бинарных отношений «ограничений существования» свойств (ОСС). Пара свойств $m_j, m_k \in M, j \neq k$ для любого объекта ПрО (и, следовательно, для $\forall g_i \in G^*$) может обладать:

- *обусловленностью*, если, обладая свойством m_j , объект g_i непременно обладает свойством m_k (хотя обратное может быть неверно), т.е. $C(m_j, m_k) \leftrightarrow \forall g_i \in G^*: m_j \in \{g_i\}' \rightarrow m_k \in \{g_i\}'$, где $\{g_i\}'$ – множество свойств объекта g_i (см. подраздел 1.1);
- *несовместимостью*, если, обладая свойством m_j , объект g_i заведомо не обладает свойством m_k , и наоборот, т.е. $E(m_j, m_k) \leftrightarrow \forall g_i \in G^*: m_j \in \{g_i\}' \rightarrow m_k \notin \{g_i\}'$.

При таких зависимостях объект обучающей выборки может обладать лишь *нормальным подмножеством* множества измеряемых свойств [16, 25]. Подмножество измеряемых свойств $Z \subseteq M$ нормально тогда и только тогда, когда оно замкнуто и совместимо: Z замкнуто, если оно содержит все свойства, обусловленные любым элементом Z , т.е. $\forall m_j \in Z (\exists m_k \in M: C(m_j, m_k) \rightarrow m_k \in Z)$; Z совместимо, если любые два элемента Z не связаны отношением несовместимости, т.е. $\forall m_j \in Z (\exists m_k \in M: E(m_j, m_k) \rightarrow m_k \notin Z)$.

Свойства симметрии бинарных отношений E и C вполне очевидны как и то, что пара несовместимых свойств не может принадлежать множеству пар свойств с обусловленностью. Более важным в модельном и методическом обеспечении ОАД оказывается условная *транзитивность несовместимости относительно обусловленности* свойств (рисунок 3а):

$$\forall m_j, m_k, m_l \in M, j \neq k, j \neq l, k \neq l: C(m_j, m_k) \& E(m_k, m_l) \rightarrow E(m_j, m_l).$$

2.2 Генезис ограничений существования свойств

Как уже отмечалось, в общем плане источником ОСС являются априорные знания субъекта онтологического анализа, релевантные исследуемой ПрО. Очевидно, что исследование такого «источника» ограничений весьма проблематично.

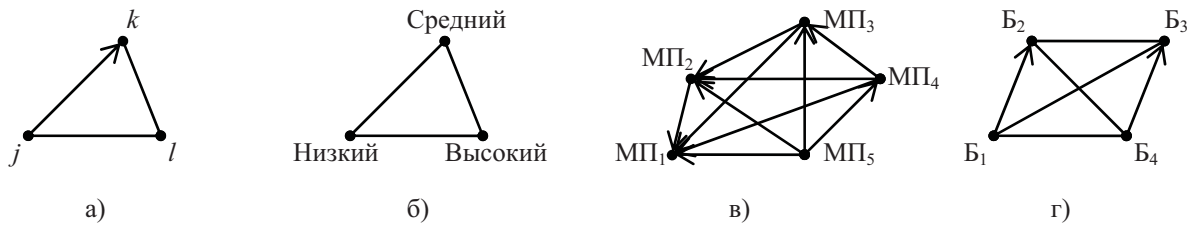


Рисунок 3 – Примеры отношений несовместимости и обусловленности измеряемых свойств:
 • - свойство; / - несовместимость; → - обусловленность

Однако в рассматриваемом методическом комплексе анализа данных упоминалась фундаментальная когнитивная процедура - концептуальное шкалирование, - где онтолог фактически вводит качественно новую информацию о ПрО, прямо формирующую ОСС.

Так, таблица 1 является примером *номинальной шкалы*, или шкалы наименований [9], устанавливающей строго дизъюнктивно покрытие исходного домена значений шкалируемого свойства «Рост мужчины, см»; образцом более тонкого подхода здесь могло бы быть использование нечёткой шкалы наименований. В любом варианте ясно, что вводимым номинальной шкалой свойствам объектов ПрО присуща *парная несовместимость E*. В этом смысле результат номинального шкалирования согласно таблице 1 иллюстрирует рисунок 3б.

Эффекты применения других типов шкал можно показать на примерах из [26], разумеется, не исчерпывающих способы выражения исследователем своего субъективного восприятия ПрО.

Порядковую шкалу целесообразно использовать для сохранения упорядоченности значений в домене значений измеряемого свойства.

Домен свойства «Материальное положение (МП)» может быть описан следующими выражениями (от тяжелого до благополучного) [26]:

- 1) денег не хватает даже на питание;
- 2) на питание денег хватает, но не хватает на покупку одежды и обуви;
- 3) на одежду и обувь денег хватает, но приобретение бытовой техники позволить не можем;
- 4) денег вполне хватает на приобретение бытовой техники, но не можем купить новый автомобиль;
- 5) денег хватает на всё, кроме таких дорогих приобретений как квартира, дом;
- 6) материальных затруднений не испытываем, при необходимости могли бы приобрести квартиру, дом.

Для онтолога естественной шкалой для такого свойства будет таблица 2, которая устанавливает между вновь введёнными свойствами бинарное отношение обусловленности C : $i < k \leftrightarrow C(МП_k, МП_i)$ – см. рисунок 3в.

Таблица 2 - Шкала материального положения

≤	МП ₁	МП ₂	МП ₃	МП ₄	МП ₅	МП ₆
1	×					
2	×	×				
3	×	×	×			
4	×	×	×	×		
5	×	×	×	×	×	
6	×	×	×	×	×	×

В настоящее время весьма популярны шкалы с разделением и порядком, описанные в [26] на примере закрытого вопроса «Чувствуете ли Вы себя в безопасности? (Б)» с вариантами ответа:

- 1) безусловно да;
- 2) скорее да;
- 3) скорее нет;
- 4) безусловно нет.

Субъективное понимание значений этого домена значений может быть выражено *двурядковой шкалой*, доставляемой таблицей 3.

Таблица 3 - Шкала безопасности

	Б ₁	Б ₂	Б ₃	Б ₄
1	×	×		
2		×		
3			×	
4			×	×

В этом примере онтолог субъективно расширяет имеющиеся данные о ПрО, вводя следующие бинарные отношения между вновь вводимыми свойствами (рисунок 3г):

- $E = \{(B_1, B_3), (B_1, B_4), (B_2, B_3), (B_2, B_4)\}$;
- $C = \{(B_1, B_2), (B_4, B_3)\}$.

2.3 Двухуровневая модель ограничений существования свойств

Различные задачи онтологического инжиниринга побуждают рассматривать характерные группы зависимых свойств.

Например, в [25] при построении онтологий лишь на основе знания свойств объектов ПрО и зависимостей между свойствами отправным пунктом анализа являются взаимообусловленные пары свойств. В [16] при расширении существующих онтологий в центре внимания оказываются нормальные подмножества учитываемых свойств объектов.

В [27] исследовалась проблема вывода формальных понятий из противоречивых данных с учётом зависимости между свойствами, возникающей в результате их номинального шкалирования. Как уже отмечалось, этот наиболее распространенный когнитивный приём порождает во множестве измеряемых свойств объектов группы свойств с парной несовместимостью. Назовем такие группы свойств *сопряжёнными* (ГСС). Фактически, каждая такая группа в целом представляет некоторое *протосвойство*, домен значений которого дизъюнктивно расщеплён в результате шкалирования.

Каждая ГСС предопределяет одну из двух возможностей: либо все измеряемые свойства, входящие в группу, у объекта обучающей выборки должны отсутствовать, либо объект обучающей выборки должен обладать каким-либо одним и только одним свойством из данной группы.

Условимся, что всякое измеряемое свойство, не входящее ни в одну из ГСС, понимается как «нерасщепленное» протосвойство, или *одиночная* ГСС (ОдГСС в противоположность *множественным* ГСС - МнГСС). Отсюда естественно принять, что ОСС задаются на множестве протосвойств. Тем самым формируется *двухуровневая модель* ОСС, которую, как будет пояснено в следующем разделе, целесообразно использовать в ОАД.

Укрупнённо эта модель определяется кортежем

$$(M_P, E_P, C_P),$$

где:

- M_P - множество актуальных для онтолога протосвойств объектов ПрО, $1 \leq |M_P| \leq |M|$, $M_P = M_{P1} \cup M_{P2}$, $M_{P1} \cap M_{P2} = \emptyset$; M_{P1} – подмножество нерасщеплённых протосвойств, т.е. множество ОдГСС; M_{P2} – подмножество расщепленных протосвойств, т.е. множество МнГСС: $M_{P2} = \{Gr_1, \dots, Gr_{|M_{P2}|}\}$, - причем в каждой МнГСС Gr_i составляющие ее измеряемые свойства несовместимы, т.е. $(\forall m_j, m_k \in Gr_i, j \neq k) \rightarrow E(m_j, m_k) = \text{Истина}$;
- E_P – пары несовместимых протосвойств, $E_P \subseteq M_P \times M_P$, $|E_P| \leq C_{|M_P|}^2$ (число сочетаний);
- C_P – пары обусловленных протосвойств, $C_P \subseteq M_P \times M_P$, $|C_P| \leq A_{|M_P|}^2$ (число размещений).

3 Рациональная аппроксимация нестроого формального контекста

Итак, с целью применения эффективных методов вывода понятийной структуры ПрО из нестроого формального контекста необходимо обеспечить его корректную α -аппроксимацию при наличии ОСС.

Формально эту задачу можно свести к построению одноместного предиката «*Альфа-сечение корректно*» с векторным аргументом $\alpha = \langle \alpha^+, \alpha^- \rangle$, $\alpha^+, \alpha^- \in [0, 1]$, где условие подтверждения истинности каждого эмпирического БСС b_{ij}

$$\|b_{ij}\| \geq \alpha^+ \wedge \|b_{ij}\| \leq \alpha^- ,$$

(или, напротив, необходимой ложности этого БСС) должно быть совмещено с выполнением всех актуальных ОСС. И далее найти область (возможно, она окажется пустой) существования порогов доверия α , доставляющих такому предикату значение **Истина**.

Разумеется, в общем случае построить такой предикат и выявить указанную область весьма непросто; пример построения такого предиката для случая, когда ОСС вызваны номинальным шкалированием некоторой части измеряемых свойств, можно найти в [27]. Однако, даже допуская возможность такого решения, связать онтолога необходимостью выбирать порог только *из ограниченной области* весьма *непрактично*. Например, при таком подходе для онтолога в общем случае не будут выполняться интуитивные ожидания на смягчение и ужесточение порога доверия к исходным данным.

Взамен предлагается следующая *эвристика*: онтолог свободен в выборе порога, а соответствующее порогу и в общем случае недопустимое множество свойств у каждого объекта обучающей выборки последовательно *сокращается* за счёт локального ужесточения порога доверия для отсекаемого на каждом шаге свойства, нарушающего ограничения существования. Основным критерием выбора свойства для отсекаемого служит *минимальное ужесточение порога доверия к данным, установленного онтологом*.

Предлагаемый эвристический метод получения корректной бинарной аппроксимации нестроого формального контекста *результативен*.

Действительно, метод реализуется для каждого объекта обучающей выборки сепаративно, и на некотором шаге возникнет одно из следующих условий:

- либо множество свойств объекта станет удовлетворять ограничениям существования (заметим, что пустое множество свойств им удовлетворяет, правда тогда объект с таким «набором» свойств приходится квалифицировать как неопознанный);
- либо будет констатировано, что в исходных «мягких» оценках истинности БСС имеется *неустранимое противоречие*: относительно объекта имеются изначально истинные в классической логике БСС, нарушающие априорные ОСС.

Выполнять предлагаемый метод целесообразно как *двухэтапную* процедуру, укрупнённо представленную на рисунке 4.

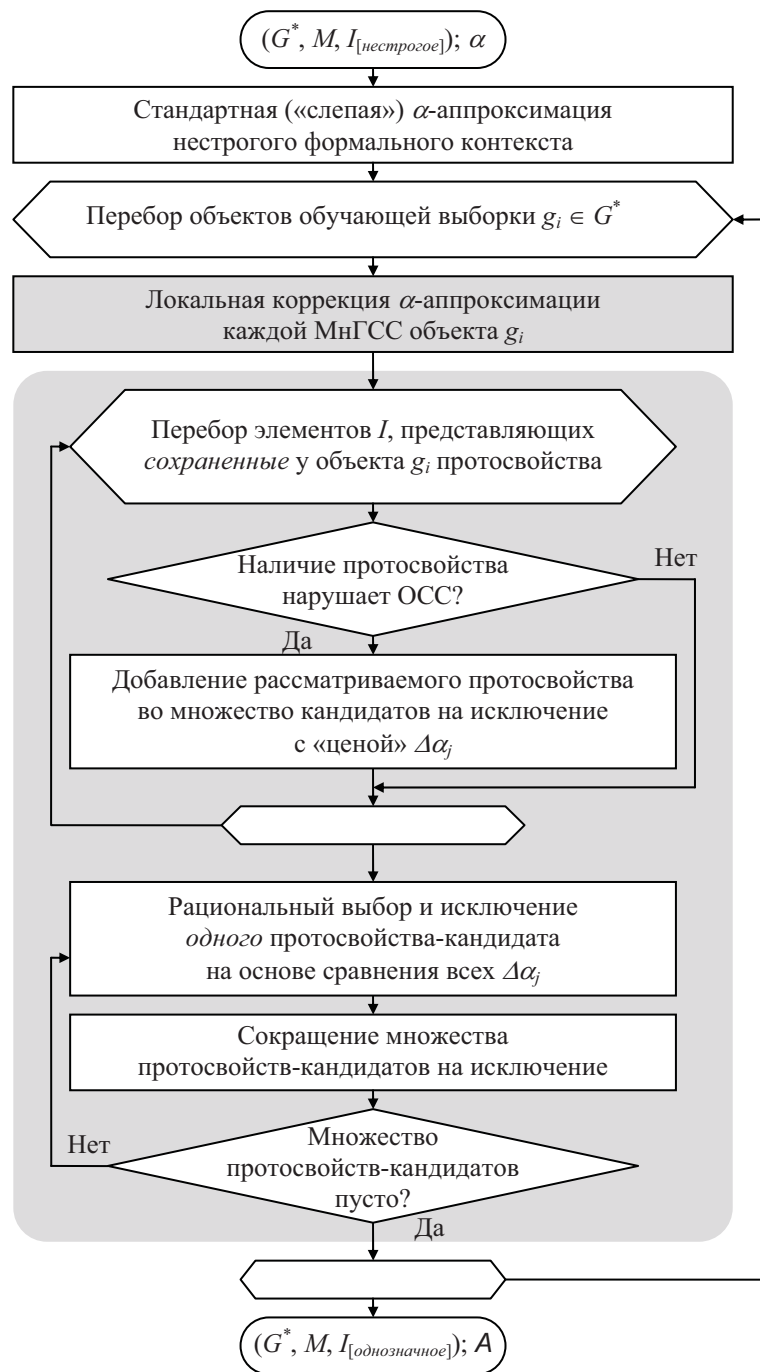


Рисунок 4 – Схема построения рациональной α -аппроксимации нестроого формального контекста (A - множество реально использованных порогов доверия для аппроксимации различных фрагментов контекста)

Сначала предлагаемым способом обрабатываются все МНГСС с принципиальным условием сохранения у каждого объекта обучающей выборки не более одного свойства из каждой группы (соответствующий блок выделен на рисунке 4 тёмной заливкой). Затем результаты первой стадии обрабатываются тем же способом совместно с исходными данными о «нерасщеплённых» протосвойствах (на рисунке 4 этот фрагмент схемы размещён на тёмном фоне).

Основанием для такой организации процедуры коррекции являются два момента:

- во-первых, упомянутая в подразделе 2.1 «транзитивность несовместимости относительно обусловленности». Конкретно реализации предлагаемого метода в один этап препятствует уникальная ситуация, когда согласно ограничениям существования некоторые из протосвойств являются *обусловленными* и в то же время *«расщеплёнными»*. Каждое такое сочетание порождает n оригинальных вариантов бинарного отношения обусловленности на множестве *измеряемых свойств*, где n – количество измеряемых свойств, на которые «расщеплено» обусловленное протосвойство. Поэтому одноэтапная стратегия реализации предложенного метода по необходимости должна предусматривать конструирование и обработку прямого произведения всех вариантов ограничений существования измеряемых свойств с последующим сопоставлением получаемых результатов, что нельзя признать рациональным;
- во-вторых, обособленная локальная коррекция α -аппроксимации каждой МНГСС имеет резервы *эффективной реализации* предлагаемого метода за счёт «однородности» ОСС – отсутствие каких-либо сложных схем ограничений кроме парной несовместимости измеряемых свойств (подробности можно найти в [28]).

4 Построение формальной онтологии предметной области

Заключительным этапом ОАД является трансформация решётки формальных понятий, полученной с помощью базовых алгоритмов АФП, в формальную онтологию, понимаемую здесь как таксономия классов объектов ПрО с заданными на классах нетаксономическими отношениями. При этом таксономию определяет отношение обобщения на множестве формальных понятий, зафиксированной решёткой, а все нетаксономические структурные отношения между понятиями-классами будут описаны, как уже упоминалось во введении, благодаря свойствам-валентностям [17].

Вместе с тем это преобразование становится нетривиальной задачей, если учитывать перспективу генерации на основе формальной онтологии баз данных для хранения денотативных объектных моделей ПрО.

Дело в том, что найденные формальные понятия в зависимости от структуры своих объёмов имеют три разновидности:

- понятия первого вида описывают объекты, действительно существующие в анализируемой ПрО (разумеется, насколько её представляет обучающая выборка). Эти понятия определяют классы объектов, которые заслуживают именование *сущностные*, или *фундаментальные*;
- понятия второго вида лишь обобщают другие понятия. В программировании такие классы известны как *виртуальные*;
- третий вид понятий характеризуется совмещением указанных черт понятий первого и второго видов.

Прагматические соображения обычно требуют ограничиться при построении формальной онтологии лишь фундаментальными и виртуальными классами объектов, и в целом опираться на следующие принципы преобразования решётки формальных понятий в таксономию классов:

- все понятия решётки являются кандидатами в фундаментальные классы таксономии;
- в фундаментальный класс превращается минимальное (в терминологии решёток) понятие, содержащее объект в своём объеме;
- свойство сохраняется в максимальном понятии, содержащем это свойство в своём содержании;

- наибольшее понятие решётки (его признак – мощность объёма равна мощности множества объектов обучающей выборки) заведомо исключается из таксономии, если его содержание пусто;
 - наименьшее понятие решётки (его признак – мощность содержания равна мощности множества свойств) заведомо исключается из таксономии, если его объём пуст;
 - анализ понятий-кандидатов в фундаментальные классы ведётся по ярусам ближайших надпонятий («родителей»), начиная с наименьшего понятия.
- Метод, обладающий описанным качеством и результативностью, подробно описан в [29].

5 Пример

Элементы представленного подхода использовались в различных случаях, связанных с необходимостью структурированного описания ПрО в задачах поддержки принятия решений, в частности:

- при проектировании цветового оформления элементов пользовательского интерфейса инструментальных программных комплексов с целью улучшения удобства использования;
- при проектировании базирующихся на онтологиях субъективно-ориентированных интерфейсов к реляционным базам данных;
- при выделении целевых групп населения при формировании региональных государственных программ социальной поддержки;
- для описания типов внешности групп людей на основе экспертизы согласно габитоскопии – теории идентификации внешности людей, разработанной в криминалистике;
- при изучении товарных рынков.

Эти применения объединял вполне определённо выраженный характер исходных данных. Мнение о свойствах объектов столь разных ПрО выражали группы экспертов, пользователей, граждан, и соответствующие данные как дополняли друг друга, так и нередко противоречили один другому. Данных из одного источника могло оказаться значительно больше, чем из конкурирующего, а доверие заказчика исследования к различным источникам часто было не одинаковым. Для консолидации такой информации и были использованы модели и методы, представленные в статье.

Как правило, упомянутые примеры громоздки, и за редким исключением (см., например, [27, 30]) остаются в малодоступных отчетах о НИР. Здесь для иллюстрации ограничимся искусственным модельным примером.

На рисунке 5 представлена обобщённая ТОС, где имитированы данные измерений свойств $M = \{a, б, \dots, ф\}$ у объектов ПрО, из которой в обучающую выборку попали объекты, условно поименованные прописными буквами, а именно: $G^* = \{A, Б, \dots, К\}$. Повторяющиеся имена объекта в первом столбце ТОС идентифицируют различные серии измерений, выполненные для данного объекта, а повторяющиеся имена измеряемого свойства в первой строке ТОС – конгруэнтные процедуры, измеряющие это свойство. Во второй строке ТОС в процентах зафиксирована степень доверия к процедурам измерения. Незаполненные ячейки результатов работы процедур измерения в различных сериях означают, что в соответствующей серии измерение не производилось (ср. с рисунком 1).

Рисунок 6 демонстрирует эту таблицу не только с упорядоченными сериями измерений для объектов, но, прежде всего, с уточнением незаполненных элементов соответствия: одни из них получили значение **НМ** и будут учитываться при построении нестроого формального контекста, согласно формуле (4), другие при построении нестроого формального контекста использованы не будут (см. подраздел 1.3).

	а	а	б	в	в	в	г	д	е	ё	ж	ж	з	и	й	к
	70	50	75	25	80	60	75	90	80	85	85	95	60	95	85	70
A	X	X	None	Failure		None	Failure	None	None			Failure	X	None	X	None
B		X	X	None	X		Failure		None	None	X	X		Failure	X	None
C	X	None	X	X		X	None	None	None	None	None	None			X	X
D	X	None	X				X	X	None		X	X	X	X	None	X
D	X		X	None		X	X	X	None	None	None		None		None	X
D	X	X	X	None	X		X	X	None	None			X		None	X
E	X		None				None	X		None	None	Failure		None	X	X
E		None		Failure		X	Failure		Failure	X		X		Failure	None	None
F			X		None	X	Failure		X		X		Failure	X		
E		X		X	None		None	X	X	None	None			X	X	None
C				X	Failure			X	None	Failure		Failure	None	None		
G		X	Failure		X	X		X		None	X			X	X	
H	X			None		None		X				Failure	X			
I	X		None		X	X	X		X	None		X	None	X	None	
J			None		None			None	X	X	X	None	X	X	X	X
K	None	None		None	None	X	None	X	X	None	X		X	X		X
L	X	X	None	X	X	X	None	None	None	None	None	None	X	X	X	
L	X	None	X	None	X			None			X	None	X	X		None
M	None	None	None		None	X		X		Failure		Failure		X		
N	X			X	X	X	None		None	X	None	None	Failure	X	None	
O	None		None	None	None	X			None	X			Failure	X	X	None
P	X	None				None	X		None	X	X			X	None	None
Q	X	X	None	X	X	X	Failure			X	None	X	None	X		X
R				X	X	X	X		None	X		None				
R	None	None	X		X	None	None		Failure			None	X			
R	None	X		Failure	X			None	X		X	X		None	None	
B			X	X	None	X		None		None	X	X			X	None
N		X		X	X	X				X	None	X		X	None	X
E	X		X	Failure	X	X	X		X	Failure	X		Failure	X	None	None
H	None	None	None		Failure				X		X	None	X		None	
S		None	X	X	X	None	X		X		X	None		None	None	X
D	None	None	X	X	X	X	None	None	Failure		None	Failure	None	X		X
B	X	X	None	None	X	X		None		X	X	None	X		X	X
D	None	X	X	X	X	X			None	X	None				None	X
S	None	None	Failure	None	X	None		X			None	X	None			
B	X	None	None	None	None	X	X	None		X	X	None			None	X
D		X	Failure	Failure			Failure	None			None	X	None		None	X
D			Failure				X	None	Failure		None	X		None		X
M		X	X	X		None	None	X	X	X	None	None		X		
H	X	None	X		None	None	X		None	X	X				None	
S	None	X	None	None		None	X		None				Failure			None
C	None		X	X	Failure		X			X			Failure	None	None	
S	X	X			X	X		None				None	None	None	X	

Рисунок 5 – Обобщенная таблица «объекты-свойства»

Рисунок 7 показывает табличную схему ОСС, априори фиксируемую онтологом. Первая строка схемы содержит идентификаторы измеряемых свойств и «расщеплённых» протосвойств. В первом столбце схемы перечисляются идентификаторы «связок» свойств, префиксы которых указывают вид отношения между свойствами: ОдГСС; МнГСС; ПССВ - пара сопряжённых свойств-валентностей; ПСО – пара свойств с обусловленностью; ПСН – пара свойств с несовместимостью. Соответствие «связки-свойства» представлено в схеме горизонтальными «лентами» и знаками инцидентности «X» и «O». Последний знак указывает обусловленное свойство в ПСО и свойство-ковалентность в ПССВ (т.е. схема ограничивает описание нетаксономических семантических связей между понятиями в выводимых онтологиях бинарными асимметричными отношениями [15]).

Нестрогий формальный контекст, полученный из рассматриваемой обобщённой ТОС приведён на рисунке 8. Его стандартную, «слепую» к ОСС α -аппроксимацию при пороге доверия $\langle 0,55; 0,5 \rangle$ представляет рисунок 9, где сохранённые элементы соответствия «объекты-свойства» следует понимать как свидетельства об истинности соответствующих БСС; иные БСС полагаются ложными.

	а	а	б	в	в	в	г	д	е	ё	ж	ж	з	и	й	к
	70	50	75	25	80	60	75	90	80	85	85	95	60	95	85	70
A	X	X	None	Failure	NM	None	Failure	None	None	NM	NM	Failure	X	None	X	None
B		X	X	None	X		Failure	None	None	None	X	X		Failure	X	None
B			X	X	None	X		None		None	X	X			X	None
B	X	X	None	None	X	X		None		X	X	None	X		X	X
B	X	None	None	None	None	X	X	None		X	X	None			None	X
C	X	None	X	X		X	None	None	None	None	None	None			X	X
C				X	Failure			X	None	Failure		Failure	None	None		
C	None		X	X	Failure		X			X			Failure	None	None	
D	X	None	X				X	X	None		X	X	X	X	None	X
D	X		X	None		X	X	X	None	None	None		None		None	X
D	X	X	X	None	X		X	X	None	None			X		None	X
D	None	None	X	X	X	X	None	None	Failure		None	Failure	None	X		X
D	None	X	X	X	X	X			None	X	None				None	X
D		X	Failure	Failure			Failure	None			None	X	None		None	X
D			Failure				X	None	Failure		None	X		None		X
E	X		None				None	X		None	None	Failure		None	X	X
E		None		Failure		X	Failure		Failure	X		X		Failure	None	None
E		X		X	None		None	X	X	None	None			X	X	None
E	X		X	Failure	X	X	X		X	Failure	X		Failure	X	None	None
F	NM	NM	X	NM	None	X	Failure	NM	X	NM	X	NM	Failure	X	NM	NM
G	NM	X	Failure	NM	X	X	NM	X	NM	None	X	NM	NM	X	X	NM
H	X			None		None		X				Failure	X	NM		NM
H	None	None	None		Failure				X		X	None	X	NM	None	NM
H	X	None	X		None	None	X		None	X	X	X		NM	None	NM
I	X	NM	None	NM	X	X	X	NM	X	None	NM	X	None	X	None	NM
J	NM	NM	None	NM	None	NM	NM	None	X	X	X	None	X	X	X	X
K	None	None	NM	None	None	X	None	X	X	None	X	NM	X	X	NM	X
L	X	X	None	X	X	X	None	None	None	None	None	None	X	X	X	
L	X	None	X	None	X			None			X	None	X	X		None
M	None	None	None	None	None	X		X		Failure		Failure	NM	X	NM	NM
M		X	X	X		None	None	X	X	X	None	None	NM	X	NM	NM
N	X		NM	X	X	X	None	NM	None	X	None	None	Failure	X	None	
N		X	NM	X	X	X		NM		X	None	X		X	None	X
O	None	NM	None	None	None	X	NM	NM	None	X	NM	NM	Failure	X	X	None
P	X	None	NM	NM	NM	None	X	NM	None	X	X	NM	NM	X	None	None
Q	X	X	None	X	X	X	Failure	NM	NM	X	None	X	None	X	NM	X
R				X	X	X	X		None	X		None				NM
R	None	None	X		X	None	None		Failure			None	X			NM
R	None	X		Failure	X			None	X		X	X		None	None	NM
S	None	None	X	X	X	None	X		X		NM	X	None	None	None	X
S	None		Failure	None	X	None		X		NM	None	X	None			
S	None	X	None	None		None	X		None	NM			Failure			None
S	X	X			X	X		None		NM		None	None	None	X	

Рисунок 6 – Уточнённая обобщенная таблица «объекты-свойства»

	а	б	в	г	д	е	ё	ж	з	и	й	к	МНГСС еёж	МНГСС зийк
ОдГСС а	X													
ОдГСС б		X												
ОдГСС в			X											
ОдГСС г				X										
ОдГСС д					X									
МНГСС еёж						X	X	X						
МНГСС зийк									X	X	X	X		
ПССВ д-г				X	O									
ПССВ в-а	O	X												
ПСО а-еёж	X												O	
ПСО еёж-зийк													X	O
ПСО а-зийк	X													O
ПСН г-зийк				X										X
ПСН г-еёж				X								X		
ПСН а-г	X			X										

Рисунок 7 – Табличная схема ограничений существования свойств

	а	б	в	г	д	е	ё	ж	з	и	й	к
A	<0,6; 0>	<0; 0,75>	<0,17; 0,37>	<0,38; 0,38>	<0; 0,9>	<0; 0,8>	<0,43; 0,43>	<0,45; 0,45>	<0,6; 0>	<0; 0,95>	<0,85; 0>	<0; 0,7>
B	<0,51; 0,09>	<0,38; 0,38>	<0,35; 0,2>	<0,56; 0,19>	<0; 0,9>	<0; 0,8>	<0,43; 0,43>	<0,66; 0,23>	<0,6; 0>	<0,47; 0,47>	<0,64; 0,21>	<0,35; 0,35>
C	<0,17; 0,43>	<0,75; 0>	<0,42; 0,13>	<0,38; 0,38>	<0,45; 0,45>	<0; 0,8>	<0,43; 0,43>	<0,12; 0,78>	<0,15; 0,45>	<0; 0,95>	<0,43; 0,43>	<0,7; 0>
D	<0,36; 0,24>	<0,64; 0,11>	<0,51; 0,04>	<0,56; 0,19>	<0,45; 0,45>	<0,13; 0,67>	<0,28; 0,57>	<0,48; 0,41>	<0,24; 0,36>	<0,63; 0,32>	<0; 0,85>	<0,7; 0>
E	<0,47; 0,12>	<0,38; 0,38>	<0,39; 0,16>	<0,28; 0,47>	<0,9; 0>	<0,67; 0,13>	<0,32; 0,53>	<0,5; 0,4>	<0,3; 0,3>	<0,59; 0,35>	<0,43; 0,43>	<0,17; 0,52>
F	<0,3; 0,3>	<0,75; 0>	<0,24; 0,31>	<0,38; 0,38>	<0,45; 0,45>	<0,8; 0>	<0,43; 0,43>	<0,66; 0,23>	<0,3; 0,3>	<0,95; 0>	<0,43; 0,43>	<0,35; 0,35>
G	<0,43; 0,17>	<0,38; 0,38>	<0,51; 0,04>	<0,38; 0,38>	<0,9; 0>	<0,4; 0,4>	<0; 0,85>	<0,66; 0,23>	<0,3; 0,3>	<0,95; 0>	<0,85; 0>	<0,35; 0,35>
H	<0,23; 0,37>	<0,38; 0,38>	<0,07; 0,48>	<0,75; 0>	<0,9; 0>	<0,4; 0,4>	<0,85; 0>	<0,66; 0,23>	<0,6; 0>	<0,47; 0,47>	<0; 0,85>	<0,35; 0,35>
I	<0,47; 0,12>	<0; 0,75>	<0,51; 0,04>	<0,75; 0>	<0,45; 0,45>	<0,8; 0>	<0; 0,85>	<0,69; 0,22>	<0; 0,6>	<0,95; 0>	<0; 0,85>	<0,35; 0,35>
J	<0,3; 0,3>	<0; 0,75>	<0,14; 0,41>	<0,38; 0,38>	<0; 0,9>	<0,8; 0>	<0,85; 0>	<0,43; 0,47>	<0,6; 0>	<0,95; 0>	<0,85; 0>	<0,7; 0>
K	<0; 0,6>	<0,38; 0,38>	<0,2; 0,35>	<0; 0,75>	<0,9; 0>	<0,8; 0>	<0; 0,85>	<0,66; 0,23>	<0,6; 0>	<0,95; 0>	<0,43; 0,43>	<0,7; 0>
L	<0,47; 0,12>	<0,38; 0,38>	<0,51; 0,04>	<0; 0,75>	<0; 0,9>	<0; 0,8>	<0; 0,85>	<0,22; 0,69>	<0,6; 0>	<0,95; 0>	<0,85; 0>	<0; 0,7>
M	<0,12; 0,47>	<0,38; 0,38>	<0,18; 0,37>	<0; 0,75>	<0,9; 0>	<0,8; 0>	<0,64; 0,22>	<0,12; 0,78>	<0,3; 0,3>	<0,95; 0>	<0,43; 0,43>	<0,35; 0,35>
N	<0,6; 0>	<0,38; 0,38>	<0,55; 0>	<0; 0,75>	<0,45; 0,45>	<0; 0,8>	<0,85; 0>	<0,23; 0,66>	<0,3; 0,3>	<0,95; 0>	<0; 0,85>	<0,7; 0>
O	<0,12; 0,47>	<0; 0,75>	<0,2; 0,35>	<0,38; 0,38>	<0,45; 0,45>	<0; 0,8>	<0,85; 0>	<0,45; 0,45>	<0,3; 0,3>	<0,95; 0>	<0,85; 0>	<0; 0,7>
P	<0,35; 0,25>	<0,38; 0,38>	<0,17; 0,37>	<0,75; 0>	<0,45; 0,45>	<0; 0,8>	<0,85; 0>	<0,66; 0,23>	<0,3; 0,3>	<0,95; 0>	<0,85; 0>	<0; 0,7>
Q	<0,6; 0>	<0; 0,75>	<0,55; 0>	<0,38; 0,38>	<0,45; 0,45>	<0,4; 0,4>	<0,85; 0>	<0,47; 0,43>	<0; 0,6>	<0,95; 0>	<0,43; 0,43>	<0,7; 0>
R	<0,12; 0,47>	<0,75; 0>	<0,43; 0,12>	<0,38; 0,38>	<0; 0,9>	<0,4; 0,4>	<0,85; 0>	<0,59; 0,31>	<0,6; 0>	<0; 0,95>	<0; 0,85>	<0,35; 0,35>
S	<0,28; 0,32>	<0,38; 0,38>	<0,34; 0,21>	<0,75; 0>	<0,45; 0,45>	<0,4; 0,4>	<0,43; 0,43>	<0,38; 0,53>	<0,1; 0,5>	<0; 0,95>	<0,43; 0,43>	<0,35; 0,35>

Рисунок 8 – Нестрогий формальный контекст

	а	б	в	г	д	е	ё	ж	з	и	й	к
A	<0,6; 0>								<0,6; 0>		<0,85; 0>	
B				<0,56; 0,19>				<0,66; 0,23>	<0,6; 0>		<0,64; 0,21>	
C		<0,75; 0>										<0,7; 0>
D		<0,64; 0,11>		<0,56; 0,19>						<0,63; 0,32>		<0,7; 0>
E					<0,9; 0>	<0,67; 0,13>				<0,59; 0,35>		
F		<0,75; 0>				<0,8; 0>		<0,66; 0,23>		<0,95; 0>		
G					<0,9; 0>			<0,66; 0,23>		<0,95; 0>	<0,85; 0>	
H				<0,75; 0>	<0,9; 0>		<0,85; 0>	<0,66; 0,23>	<0,6; 0>			
I				<0,75; 0>		<0,8; 0>		<0,69; 0,22>		<0,95; 0>		
J						<0,8; 0>	<0,85; 0>		<0,6; 0>	<0,95; 0>	<0,85; 0>	<0,7; 0>
K					<0,9; 0>	<0,8; 0>		<0,66; 0,23>	<0,6; 0>	<0,95; 0>		<0,7; 0>
L									<0,6; 0>	<0,95; 0>	<0,85; 0>	
M					<0,9; 0>	<0,8; 0>	<0,64; 0,22>			<0,95; 0>		
N	<0,6; 0>		<0,55; 0>				<0,85; 0>			<0,95; 0>		<0,7; 0>
O							<0,85; 0>			<0,95; 0>	<0,85; 0>	
P				<0,75; 0>			<0,85; 0>	<0,66; 0,23>		<0,95; 0>		
Q	<0,6; 0>		<0,55; 0>				<0,85; 0>	<0,66; 0,23>		<0,95; 0>		<0,7; 0>
R		<0,75; 0>					<0,85; 0>	<0,59; 0,31>	<0,6; 0>	<0,95; 0>		
S				<0,75; 0>								

Рисунок 9 – Стандартная α -аппроксимация нестроого формального контекста при пороге доверия $\langle 0,55; 0,5 \rangle$

Также следует интерпретировать и рисунок 10, показывающий результат коррекции стандартной α -аппроксимации с учётом принятых ОСС (т.е. корректный однозначный формальный контекст), а рисунок 11 приводит использованные для этого пороги доверия, которые обусловили сохранение указанных на рисунке 10 элементов соответствия «объекты-свойства» (ранее на рисунке 4 множество этих порогов обозначено как А).

	а	б	в	г	д	е	ё	ж	з	и	й	к
A											<0,85; 0>	
B								<0,66; 0,23>			<0,64; 0,21>	
C		<0,75; 0>										<0,7; 0>
D		<0,64; 0,11>		<0,56; 0,19>								<0,7; 0>
E					<0,9; 0>	<0,67; 0,13>				<0,59; 0,35>		
F		<0,75; 0>				<0,8; 0>		<0,66; 0,23>		<0,95; 0>		
G					<0,9; 0>			<0,66; 0,23>		<0,95; 0>	<0,85; 0>	
H				<0,75; 0>	<0,9; 0>		<0,85; 0>	<0,66; 0,23>	<0,6; 0>			
I				<0,75; 0>		<0,8; 0>		<0,69; 0,22>		<0,95; 0>		
J						<0,8; 0>	<0,85; 0>		<0,6; 0>	<0,95; 0>	<0,85; 0>	<0,7; 0>
K					<0,9; 0>	<0,8; 0>		<0,66; 0,23>	<0,6; 0>	<0,95; 0>		<0,7; 0>
L									<0,6; 0>	<0,95; 0>	<0,85; 0>	
M					<0,9; 0>	<0,8; 0>	<0,64; 0,22>			<0,95; 0>		
N	<0,6; 0>		<0,55; 0>				<0,85; 0>			<0,95; 0>		<0,7; 0>
O							<0,85; 0>			<0,95; 0>	<0,85; 0>	
P				<0,75; 0>			<0,85; 0>	<0,66; 0,23>		<0,95; 0>		
Q	<0,6; 0>		<0,55; 0>				<0,85; 0>	<0,66; 0,23>		<0,95; 0>		<0,7; 0>
R		<0,75; 0>					<0,85; 0>	<0,59; 0,31>	<0,6; 0>	<0,95; 0>		
S				<0,75; 0>								

Рисунок 10 – Однозначная аппроксимация нестроого формального контекста при пороге доверия $\langle 0,55; 0,5 \rangle$ с учётом ограничений существования свойств

Рисунок 12 представляет все формальные понятия, выведенные из полученного корректного формального контекста, а рисунок 13 – специально размеченную транзитивную редукцию решётки этих формальных понятий (т.н. диаграмму Хассе [11-15], где свойство помечает *сверху* максимальное понятие, заключающее это свойство в своём содержании, а объект помечает *снизу* минимальное понятие, заключающее этот объект в своём объёме).

	а	б	в	г	д	е	ё	ж	з	и	й	к
A											<0.6, 0.5>	
B								<0.56, 0.5>			<0.6, 0.5>	
C		<0.55, 0.5>										<0.55, 0.5>
D		<0.56, 0.5>										<0.63, 0.5>
E					<0.55, 0.5>	<0.55, 0.5>				<0.55, 0.5>		
F		<0.55, 0.5>				<0.66, 0.5>				<0.55, 0.5>		
G					<0.55, 0.5>			<0.55, 0.5>		<0.85, 0.5>		
H					<0.85, 0.5>							
I						<0.75, 0.5>				<0.75, 0.5>		
J							<0.8, 0.5>			<0.85, 0.5>		
K					<0.55, 0.5>	<0.66, 0.5>				<0.7, 0.5>		
L										<0.85, 0.5>		
M					<0.55, 0.5>	<0.64, 0.5>				<0.55, 0.5>		
N	<0.55, 0.5>		<0.55, 0.5>				<0.55, 0.5>			<0.7, 0.5>		
O							<0.55, 0.5>			<0.85, 0.5>		
P							<0.75, 0.5>			<0.75, 0.5>		
Q	<0.55, 0.5>		<0.55, 0.5>				<0.55, 0.5>			<0.7, 0.5>		
R		<0.55, 0.5>					<0.59, 0.5>		<0.55, 0.5>			
S				<0.55, 0.5>								

Рисунок 11 – Локальные пороги доверия для элементов соответствия нестрого формального контекста, оценкой истинности которых в однозначной аппроксимации является **Истина**

Имя понятия	Объём	Содержание
1		а, б, в, г, д, е, ё, ж, з, и, й, к
2	A, B	й
3	B	ж, й
4	C, D	б, к
5	A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S	
6	E, K, M	д, е, и
7	F	б, е, и
8	C, D, F, R	б
9	E, F, I, K, M	е, и
10	G	д, ж, и
11	B, G	ж
12	E, G, K, M	д, и
13	E, F, G, I, J, K, L, M, N, O, P, Q	и
14	E, G, H, K, M	д
15	J, N, O, P, Q	ё, и
16	N, Q	а, в, ё, и
17	R	б, ё, з
18	J, N, O, P, Q, R	ё
19	S	г

Рисунок 12 – Множество формальных понятий

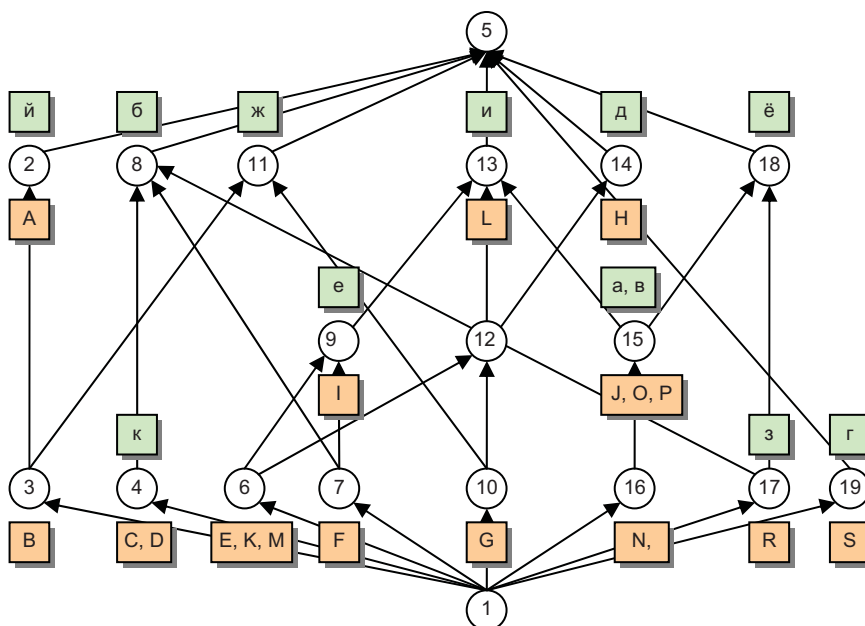


Рисунок 13 – Диаграмма Хассе решётки формальных понятий:

○ - формальное понятие; ↗ - обобщение между понятиями; в прямоугольниках – измеряемые свойства и объёмы обучающей выборки; 1 и 5 – соответственно наименьшее и наибольшее понятия решётки

Наконец, рисунок 14 представляет результат трансформации полученной решётки формальных понятий в формальную онтологию ПрО согласно принципам, изложенным в разделе 4. На рисунке 14 идентификаторы сформированных фундаментальных классов имеют префикс «f». Формальному понятию 12 (см. рисунки 11, 12) класс не сопоставлен, т.к. оно не является ни максимальным для какого-либо свойства, ни минимальным для какого-либо понятия в полученной решётке формальных понятий. Криволинейные связующие стрелки на рисунке 14 показывают нетаксономические отношения выведенных формальных понятий. Исток и сток каждой из них определяет соответствующая ПССВ-связь в схеме ОСС (см. рисунок 7) и принадлежность её аргументов-свойств содержанию определённых формальных понятий (см. рисунок 12).

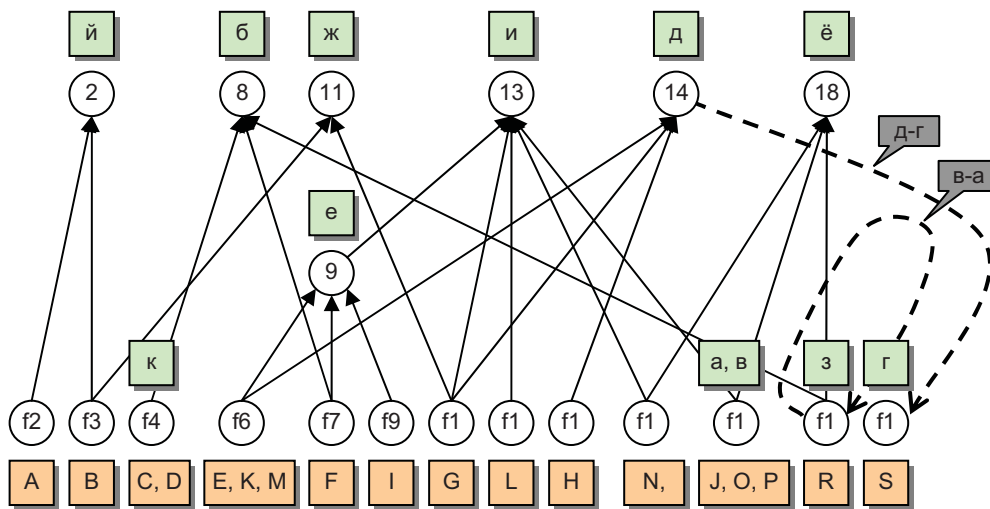


Рисунок 14 – Классы объектов и их связи (формальная онтология):
 -> - нетаксономические отношения между классами

Заключение

Представленные в статье модели и методы показывают новый этап развития онтологического анализа данных - методики выявления понятийной структуры и, в конечном счёте, формальной онтологии экспериментально исследуемой ПрО. Основу методики составляет анализ формальных понятий, в котором сохранено классическое понимание понятия как фундаментального смыслового элемента, определяемого объёмом и содержанием.

В центре внимания выполненного исследования находится проблема отражения реалий накопления эмпирической информации о ПрО, определяющих неполноту исходных данных. Для её решения потребовалось обобщить стандартную модель представления объектно-признаковых данных и применить для её обработки модели и аппарат многозначной векторной логики.

Повышение адекватности моделей исходных данных породило новые задачи онтологического анализа. Потребовалось разработать интеллектуальный метод преобразования исходных данных нового формата в однозначный формальный контекст, для которого известны эффективные алгоритмы вывода формальных понятий. При этом построены модели для учёта априори известных зависимостей между измеряемыми свойствами объектов ПрО.

Наконец, в методику включён ряд прагматических принципов, определивших метод трансформации решетки формальных понятий, выводимой из эмпирических данных, в формальную онтологию ПрО. В полном объёме представленный методический комплекс предполагается внедрить в разрабатываемую в Институте проблем управления сложными систе-

мами РАН систему семантического моделирования и проектирования на массовой программной платформе.

Список источников

- [1] Искусственный интеллект. – В 3-х кн. Кн. 2. Модели и методы: Справочник / Под ред. Д.А. Поспелова. – М.: Радио и связь, 1990. – 304 с.
- [2] *Лукашевич, Н.В.* Тезаурусы в задачах информационного поиска / Н.В. Лукашевич. – М.: Изд-во Московского университета, 2011. – 512 с.
- [3] *Мальцев, А.И.* Алгебраические системы. – М.: Наука, 1970. – 392 с.
- [4] *Смирнов, С.В.* Онтологическое моделирование в ситуационном управлении // Онтология проектирования. – 2012. – №2(4). – С. 16-24.
- [5] *Гаврилова, Т.А.* Интеллектуальные технологии в менеджменте: инструменты и системы / Т.А. Гаврилова, Д.И. Муромцев. – СПб.: Изд-во «Высшая школа менеджмента»; Изд. дом СПбГУ, 2008. – 488 с.
- [6] *Смирнов, А.В.* Онтологии в системах искусственного интеллекта: способы построения и организации / А.В. Смирнов, М.П. Пашкин, Н.Г. Шилов, Т.В. Левашова // Новости искусственного интеллекта. – 2002. – №1. – С. 3-13 (Часть 1); №2. – С. 3-9 (Часть 2).
- [7] *Suarez-Figueroa, M.S.* The NeOn Methodology for Ontology Engineering / M.S. Suarez-Figueroa, A. Gomez-Perez, M. Fernandez-Lopez // In: Ontology Engineering in a Networked World. – Berlin-Heidelberg: Springer, 2012. – P. 9-34.
- [8] *Ломов, П.А.* Автоматизация синтеза составных онтологических паттернов содержания // Онтология проектирования. – 2016. – Т. 6, №2(20). – С. 162-172.
- [9] *Загоруйко, Н.Г.* Прикладные методы анализа данных и знаний / Н.Г. Загоруйко. – Новосибирск: Институт математики СО РАН, 1999. – 270 с.
- [10] *Барсегян, А.А.* Анализ данных и процессов / А.А. Барсегян, М.С. Куприянов, И.И. Холод, М.Д. Тесс, С.И. Елизаров. – 3-е изд., перераб. и доп. – СПб.: БХВ-Петербург, 2009. – 512 с.
- [11] *Ganter, B.* Formal Concept Analysis. Mathematical foundations / B. Ganter, R. Wille. – Berlin-Heidelberg: Springer-Verlag, 1999. – 290 p.
- [12] *Carpineto, C.* Concept Data Analysis: Theory and Applications / C. Carpineto, G. Romano. – Wiley, 2004.
- [13] *Ganter, B.* Conceptual Exploration / B. Ganter, S. Obiedkov. – Springer, 2016.
- [14] *Ignatov, D.I.* Introduction to Formal Concept Analysis and Its Applications in Information Retrieval and Related Fields / D.I. Ignatov // In: P. Braslavski, N. Karpov, M. Worring, Y. Volkovich, D.I. Ignatov (Eds.): Information Retrieval. Revised Selected Papers 8th Russian Summer School, RuSSIR 2014 (Nizhniy Novgorod, Russia, August 18-22, 2014). – Springer International Publishing, 2015. – P. 42-141.
- [15] Formal Concept Analysis Homepage - <http://www.upriss.org.uk/fca/fca.html> (дата обращения 20.08.2016).
- [16] *Пронина, В.А.* Использование отношений между атрибутами для построения онтологии предметной области / В.А. Пронина, Л.Б. Шипилина // Проблемы управления. – 2009. – №1. – С. 27-32.
- [17] *Смирнов, С.В.* Построение онтологий предметных областей со структурными отношениями на основе анализа формальных понятий / С.В. Смирнов // Знания – Онтология – Теория: Материалы Всероссийской конф. с международным участием (3-5 октября 2011 г., Новосибирск, Россия). Т. 2. – Новосибирск: Институт математики СО РАН, 2011. – С. 103-112.
- [18] *Смирнов, С.В.* Онтологический анализ предметных областей моделирования / С.В. Смирнов // Известия Самарского научного центра РАН. – 2001. – Т. 3, № 1. – С. 62-70.
- [19] *Ganter, B.* Conceptual scaling / B. Ganter, R. Wille // In: F. Roberts (Ed.): Applications of Combinatorics and Graph Theory to the Biological and Social Sciences. – New York Springer-Verlag, 1989. – P. 139-167.
- [20] *Аршинский, Л.В.* Векторные логики: основания, концепции, модели / Л.В. Аршинский. – Иркутск: Иркутский гос. ун-т, 2007. – 228 с.
- [21] *Зверев, Г.Н.* Неклассические логики в задачах идентификации / Г.Н. Зверев // Идентификация систем и задачи управления: Труды международной конф. SICHRO'2000 (26-28 сентября 2000 г., Москва, Россия). – М.: ИПУ РАН, 2000. – С. 1607-1616.
- [22] *Belohlavek, R.* Computing the lattice of all fixpoints of a fuzzy closure operator / R. Belohlavek, B. De Baets, V. Ouatra, J. Vychodil // IEEE Trans. on Fuzzy systems. – 2010. – Issue 3, Vol. 18. – P. 546-557.
- [23] *Tho, Q.T.* Automatic Fuzzy Ontology Generation for the Semantic Web / Q.T. Tho, S.C. Hui, A.C.M. Fong, T.H. Cao // IEEE Transactions on Knowledge and Data Engineering. – 2006. – Issue 6, Vol. 18. – P. 842-856.
- [24] *Yang, K.M.* Fuzzy Concept Mining based on Formal Concept Analysis / K.M. Yang, E.H. Kim, S.H. Hwang, S.H. Choi // Int. J. of Computers. – 2008. – Issue 3, Vol. 2. – P. 279-290.

- [25] *Lammari, N.* Building and maintaining ontologies: a set of algorithms / N. Lammari, E. Metais // *Data & Knowledge Engineering*. – 2004. - Vol. 48(2). - P. 155-176.
- [26] *Игнатов, Д.И.* Решётки формальных понятий для анализа данных социологических опросов / Д.И. Игнатов, О.Н. Кононыхина // *Интегрированные модели и мягкие вычисления в искусственном интеллекте: Труды V-й Международной научно-технической конференции (20-30 мая 2009 г., Коломна, Россия)*. Т 1. – М.: Физматлит, 2009. – С. 230-240.
- [27] *Офицеров, В.П.* Метод альфа-сечения нестрогих формальных контекстов в анализе формальных понятий / В.П. Офицеров, В.С. Смирнов, С.В. Смирнов // *Проблемы управления и моделирования в сложных системах: Труды XVI международной конф. (30 июня - 03 июля 2014 г., Самара, Россия)*. – Самара: СамНЦ РАН, 2014. - С. 228-244.
- [28] *Зубцов, Р.О.* Алгоритмическое и программное обеспечение онтологического анализа данных / Р.О. Зубцов, В.А. Семенова, С.В. Смирнов // *Открытые семантические технологии проектирования интеллектуальных систем = Open Semantic Technologies for Intelligent Systems (OSTIS-2016): Материалы VI международной научно-технической конф. (18-20 февраля 2016 г. Минск, Беларусь) / Редкол.: В.В. Голенков (отв. ред.) [и др.]*. – Минск: БГУИР, 2016. – С. 83-88.
- [29] *Kovartsev, A.N.* Intelligent Design of Class Structure Model based on Ontological Data Analysis / A.N. Kovartsev, V.S. Smirnov, S.V. Smirnov // *Труды ИСП РАН*. 2015. Т. 27. Выпуск 3. – с.73-86.
- [30] *Виноградов, И.Д.* Изучение автомобильного рынка на основе онтологий потребительских предпочтений / И.Д. Виноградов, В.А. Виттих, В.М. Карпов, А.Н. Москалюк, С.В. Смирнов // *Вестник компьютерных и информационных технологий*. - 2005. - №1. - С. 2-7.

INCOMPLETE DATA ANALYSIS FOR BUILDING FORMAL ONTOLOGIES

D.E. Samoilo^{1,2}, V.A. Semenova^{1,2}, S.V. Smirnov²

¹ *Samara National Research University named after academician S.P. Korolev, Samara, Russia*
dmitrysam3@gmail.com, queenbfjr@gmail.com

² *Institute for the Control of Complex Systems of Russian Academy of Sciences, Samara, Russia*
smirnov@iccs.ru

Abstract

The article considers the problem of automating the formation of ontological specifications subject domains on the basis of measurements. This problem is the pivotal issue of ontological analysis. The article presents models and methods, aimed at identifying conceptual structure and, ultimately, detecting the formal ontology of the considered subject domain. Fundamental realities of accumulation of empirical data (multiple independent measurements for each training sample properties of the object; congruence of the measurement procedures; differentiation of trust to different data sources) are reflected in the "objects-properties" summary table model. Imperfection (inaccuracy, inconsistency, uncertainty) of this information implies the need to use many-valued logic models for its primary processing. The result of this treatment - fuzzy formal context - should be approximated as a unique context, from which the possible conclusion of formal concepts in the framework of applied branch of the lattice theory, known as the "formal concept analysis". The genesis of the "properties' limits of existence" that affect the correctness of the approximation of fuzzy formal context is studied. The models and the method of accounting for this additional information are proposed. Guidelines for conversion of lattice formal concepts into a formal ontology are formulated. A model example of the developed models and ontological data analysis methods is presented.

Key words: *ontological data analysis, formal ontology, formal concept analysis, multi-valued vector logic, properties existence constraints.*

Citation: Samoilo DE., Semenova VA, Smirnov SV. Incomplete data analysis of for building formal ontologies [In Russian]. *Ontology of designing*. 2016; 6(3): 317-339. DOI: 10.18287/2223-9537-2016-6-3-317-339.

References

- [1] *Artificial intelligence Handbook. Vol. 2. Models and methods*. Ed.: DA Pospelov [In Russian]. – Moscow: "Radio i svyaz" Publisher, 1990.

- [2] *Lukashevch NV*. Thesauri in Information Retrieval [In Russian]. - Moscow: Moscow University Publisher, 2011.
- [3] *Mal'cev AI*. Algebraic Systems. - Berlin-Heidelberg New York: Springer-Verlag, 1973.
- [4] *Smirnov SV*. Ontological modeling in situational management [In Russian]. *Ontology of Designing*, 2012; 2(4): 16-24.
- [5] *Gavrilova TA, Muromtsev DI*. Intelligent technologies in management: tools and systems [In Russian]. – St. Petersburg: Management Higher school Publisher, 2008.
- [6] *Smirnov AV, Pashkin MP, Shilov NG, Levashova TV*. Ontologies in Artificial intelligence systems: building and management methods [In Russian]. *News of Artificial intelligence*, 2002; 1: 3-13 (Part 1); 2: 3-9 (Part 2).
- [7] *Suárez-Figueroa MC, Gómez-Pérez A, Fernández-López M*. The NeOn Methodology for Ontology Engineering. In: *Ontology Engineering in a Networked World*. Springer Berlin Heidelberg; 2012: 9-34.
- [8] *Lomov PA*. Automation of synthesis of composite content ontology design pattern [In Russian]. *Ontology of Designing*, 2016; 6(2): 162-172.
- [9] *Zagoruyko NG*. Applied methods of data and knowledge analysis [In Russian]. Novosibirsk: Sobolev Institute of Mathematics, SB RAS, 1999.
- [10] *Barsegyan AA, Kupriyanov MS, Holod II, Tess MD, Elizarov SI*. Data and Process Analysis [In Russian]. – St. Petersburg: BHV- Petersburg, 2009.
- [11] *Ganter B, Wille R*. Formal Concept Analysis. Mathematical foundations. Springer Berlin-Heidelberg, 1999.
- [12] *Carpineto C, Romano G*. Concept Data Analysis: Theory and Applications. Wiley, 2004.
- [13] *Ganter B, Obiedkov S*. Conceptual Exploration. Springer, 2016.
- [14] *Ignatov DI*. Introduction to Formal Concept Analysis and Its Applications in Information Retrieval and Related Fields. In: P. Braslavski, N Karpov, M. Worring, Y. Volkovich, D.I. Ignatov (Eds.): *Information Retrieval (Revised Selected Papers 8th Russian Summer School, RuSSIR 2014, Nizhniy Novgorod, Russia, 2014, August 18-22,)*. Springer International Publishing, 2015: 42-141.
- [15] Formal Concept Analysis Homepage. Source: <http://www.upriss.org.uk/fca/fca.html>
- [16] *Pronina VA, Shipilina LB*. Using the relationships between attributes to build domain ontology [In Russian]. *Control Science*, 2009; 1: 27-32.
- [17] *Smirnov SV*. Building knowledge domain ontologies with structural relationships based on Formal Concept Analysis [In Russian]. *Proc. 3rd All-Russian Conf. "Knowledge-Ontologies-Theories"* (Novosibirsk, Russia, October 3-5, 2011). Vol. 2. Novosibirsk: Sobolev Institute of Mathematics, SB of RAS; 2011: 103-112.
- [18] *Smirnov SV*. Ontological analysis of modeling domain [In Russian]. *Bulletin of the Samara Scientific Center of RAS*, 2001. 3(1): 62-70.
- [19] *Ganter B, Wille R*. Conceptual scaling. In: F. Roberts (Ed.): *Applications of Combinatorics and Graph Theory to the Biological and Social Sciences*. Springer-Verlag New York; 1989: 139-167.
- [20] *Arshinskii LV*. Substantial and formal deductions in logics with vector semantics. *Automation and Remote Control*, 2007; 68(1): 139-148.
- [21] *Zverev GN*. Non-classical logics in identification problems [In Russian]. *Proc. of Int. Conf. SICHRO'2000* (Moscow, Russia, 2000, 26-28 September). Moscow: Institute of Control Sciences of RAS; 2000: 1607-1616.
- [22] *Belohlavek R, De Baets B, Outrata B, Vychodil J*. Computing the lattice of all fixpoints of a fuzzy closure operator. *IEEE Trans. on Fuzzy systems*, 2010; 18(3): 546-557.
- [23] *Tho QT, Hui SC, Fong ACM, Cao TH*. Automatic Fuzzy Ontology Generation for the Semantic Web. *IEEE Trans. on Knowledge and Data Engineering*, 2006; 18(6): 842-856.
- [24] *Yang KM, Kim EH, Hwang SH, Choi SH*. Fuzzy Concept Mining based on Formal Concept Analysis. *Int. J. of Computers*, 2008; 2(3): 279-290.
- [25] *Lammari N, Metais E*. Building and maintaining ontologies: a set of algorithms. *Data & Knowledge Engineering*, 2004; 48(2): 155-176.
- [26] *Ignatov DI, Kononychina ON*. Formal concept lattices for data analysis in sociological interrogations [In Russian]. *Integrated models and soft computation in Artificial Intelligence: Proc. of 5th Int. Conf. (Kolomna, Russia, 2009, May 20-30)*. Vol. 1. Moscow: "Fizmathlit" Publisher; 2009: 230-240.
- [27] *Ofitserov VP, Smirnov VS, Smirnov SV*. Alpha-section of non-strict formal contexts in Formal Concept Analysis [In Russian]. *Proc. of XVI Int. Conf. "Complex systems: Control and Modeling Problems"* (Samara, Russia, 2014, June 30 – July 03). Samara: Samara Scientific Center of RAS; 2014: 228-244.
- [28] *Zubtcov RO, Semenova VA, Smirnov SV*. Algorithms and software for ontological data analysis [In Russian]. *Proc. of VI Int. Conf. "Open Semantic Technologies for Intelligent Systems (OSTIS-2016)"* (Minsk, Belarus, 2016, February 18-20). Minsk: Byelorussian state university for informatics and radio electronics Publisher; 2016: 83-88.
- [29] *Kovartsev AN, Smirnov VS, Smirnov SV*. Intelligent Design of Class Structure Model based on Ontological Data Analysis. *Proceedings of the Institute for System Programming of the RAS*, 2015; 27(3): 73-86.
- [30] *Vinogradov ID, Vittikh VA, Karpov VM, Moscalyuk AN, Smirnov SV*. Automobile marketing based on ontologies of user's needs. *Advances in Concurrent Engineering: Proc. of the 9-th ISPE Int. Conf. on Concurrent Engineering:*

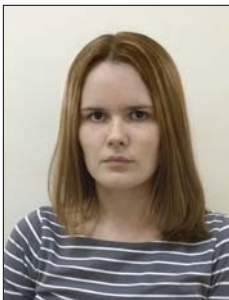
Сведения об авторах



Самойлов Дмитрий Евгеньевич, 1997 г. рождения. Окончил с золотой медалью МБОУ «Лицей авиационного профиля №135» г. Самары (2015). Многократный призёр областных и всероссийских олимпиад по физике, астрономии, математике. Студент второго курса бакалавриата Самарского национального исследовательского университета им. С.П. Королёва по направлению «Прикладная математика и информатика», лаборант Института проблем управления сложными системами РАН. Соавтор двух научных работ по тематике анализа формальных понятий.

Samoilov Dmitry Evgenyevich (b. 1997). He graduated with a gold medal from "Lyceum Profile Aviation №135» (2015). Multiple medalist of regional and All-Russian Olympiads in physics, astronomy and mathematics. Second-year student of the Samara National Research University in "Applied Mathematics and Computer Science", lab assistant of the Institute for the Control of Complex Systems of Russian Academy of Sciences. Co-author of two scientific papers on the subject of formal

concept analysis.



Семёнова Валентина Андреевна, 1994 г. рождения. Окончила бакалавриат Самарского государственного аэрокосмического университета им. С.П. Королёва по направлению «Прикладная математика и информатика» (2015). Студентка второго курса магистратуры Самарского национального исследовательского университета им. С.П. Королёва по направлению «Механика и математическое моделирование», инженер Института проблем управления сложными системами РАН. Соавтор пяти научных работ в области интеллектуального анализа данных.

Semyonovna Valentina Andreevna (b. 1994). She graduated from the Bachelor of Samara State Aerospace University, "Applied Mathematics and Computer Science" (2015). A student of the second year master Samara National Research University in the direction of "Mechanics and

Mathematical Modeling", engineer of the Institute for the Control of Complex Systems of Russian Academy of Sciences. Co-author of five scientific papers in the field of data mining.



Смирнов Сергей Викторович, 1952 г. рождения. Окончил Куйбышевский авиационный институт им. С.П. Королёва в 1975 г., д.т.н. (2002). Заместитель директора Института проблем управления сложными системами РАН, профессор кафедры «Инженерия знаний» Поволжского государственного университета телекоммуникаций и информатики. Член Российской ассоциации искусственного интеллекта. В списке научных трудов более 150 работ в области прикладной математики, компьютерного моделирования сложных систем, создания интеллектуальных систем поддержки принятия решений в технологических и организационных сферах.

Smirnov Sergei Victorovich (b. 1952) graduated from the Korolyov aviation Institute (Kuibyshev-city) in 1975, D. Sc. Eng. (2002). Deputy director at Institute for the Control of Complex Systems of Russian Academy of Sciences, professor at Povolzhskiy State University of Tele-

communication and Informatics (Knowledge engineering sub-department). He is Russian Association of Artificial Intelligence member. He is co-author of more than 150 publications in the field of applied mathematic, complex systems simulation and development of knowledge based decision support systems in control and management.