

УДК 004.89

ФУНКЦИОНАЛЬНОЕ НАПОЛНЕНИЕ И АРХИТЕКТУРА ПРОГРАММНОЙ ЛАБОРАТОРИИ ДЛЯ ОНТОЛОГИЧЕСКОГО АНАЛИЗА ДАННЫХ

В.А. Семенова, С.В. Смирнов*

Самарский федеральный исследовательский центр РАН,
Институт проблем управления сложными системами РАН
Россия, 443020, г. Самара, ул. Садовая, 61

E-mail: queenbfjr@gmail.com

Аннотация. Представлены функциональные возможности и архитектура программной лаборатории для онтологического анализа и через эту «призму» – методология этого анализа. Методический комплекс онтологического анализа и порядок его применения обеспечивают извлечение из данных многомерных наблюдений и измерений предметной области ее семантической модели в виде формальной онтологии – совокупности формальных понятий, каждое из которых определяется объемом и содержанием. Во множестве формальных понятий выявляется частичный порядок (бинарное отношение обобщения) и вскрываются интенциональные отношения, отражающие связи между элементами объемов понятий. От известных инструментальных средств построения формальных онтологий на основе эмпирических объектно-признаковых данных (т. е. на основе данных, представленных в общепризнанной форме протокола регистрации результатов измерений) разработанная программная лаборатория отличается прежде всего учетом реалий накопления информации об исследуемой предметной области. Они в общем случае обуславливают неполноту и противоречивость исходных данных, для обработки которых по необходимости привлечен аппарат многозначной векторной логики. Другим уникальным отличием рассматриваемой программной лаборатории является учет при обработке эмпирических данных априори (до измерений) известных ограничений существования свойств – бинарных отношений обусловленности и несовместимости свойств у объектов исследуемой предметной области. Представляемая программная лаборатория реализована на платформе табличного процессора Excel и языка программирования VBA. Главным мотивом такого выбора было обеспечение доступности и облегчение освоения технологии онтологического анализа данных широким кругом пользователей, использующих Excel в своей профессиональной работе.

Ключевые слова: массовая программная платформа, формальная онтология, онтологический анализ данных, анализ формальных понятий, ограничения существования свойств.

Введение

В [1] был анонсирован проект создания в Институте проблем управления сложными системами РАН прототипа программной лаборатории для онтологического

* Валентина Андреевна Семенова, младший научный сотрудник.

Сергей Викторович Смирнов, главный научный сотрудник, доктор технических наук.

ческого анализа данных, получившей наименование *OntoWorker* (OW). К этому моменту в коллективе потенциальных участников проекта сложилось вполне устойчивое представление об онтологическом анализе данных (ОАД) как о новом методическом комплексе для выявления понятийной структуры, или формальной онтологии, предметной области (ПрО) на основе эмпирической информации, структурированной в виде таблицы «объекты – свойства» (ТОС) [2, 3].

В фундамент ОАД положен метод анализа формальных понятий (АФП), теоретически строго обоснованный в рамках математической теории решеток [4, 5] и вполне отражающий классический логико-философский подход к моделированию фундаментального теоретико-познавательного элемента человеческого сознания – понятия, определяемого объемом и содержанием [6, 7] (формально объем понятия – это множество объектов, каждый из которых характеризуется всяким свойством из набора свойств и только из этого набора, который суть содержание понятия, и наоборот).

Новеллой ОАД стало прежде всего обогащенное методическое обеспечение подготовки контекста для вывода формальных понятий [1]: ввод механизма представления в ТОС эмпирически наблюдаемых отношений между объектами ПрО; учет реалий накопления эмпирической информации, вызывающих неполноту и противоречивость объектно-признаковых данных, интерпретация и консолидация которых возможна лишь в рамках многозначной логики; учет ограничений существования свойств у объектов ПрО. Подобные методические новшества не поддерживались в известных инструментальных средствах анализа данных на базе АФП (см., например, [8–11]), и это определяло актуальность OW-проекта.

Кроме того, в условиях повышенного интереса к семантическому моделированию ПрО актуальной становилась разработка средств онтологического анализа на массовой программной платформе, доступной и освоенной специалистами в самых разных предметных областях. Такой уникальной платформой выделялся табличный процессор Excel, который помимо многих других достоинств характеризуется весьма удобным для разработчиков и пользователей единством среды создания пользовательских приложений и хранения данных [12, 13].

На фоне активного совершенствования методологического базиса ОАД развитие собственно OW-проекта фиксировали публикации [14, 15]. В данной статье очерчивается достигнутый уровень функционального наполнения и архитектурные решения, принятые в апробированной версии программной лаборатории для онтологического анализа данных *OntoWorker*, помещенной в государственный Реестр программ для ЭВМ (номер регистрации 2022669423, дата 19.10.2022, бюллетень № 10).

1. Реализуемая схема онтологического анализа данных

Функциональное наполнение OW достаточно полно демонстрирует реализуемая в программной лаборатории пошаговая схема ОАД (рис. 1). На схеме отражены взаимосвязи основных информационных «переделов» (шагов) и продуктов ОАД, а кроме того, выделены фрагменты анализа, где поддерживаются когнитивные акты субъекта-исследователя ПрО: 1, 4, 8, 17 – осмысление субъектом, находящимся на определенном этапе решения ОАД-задачи, либо результата предшествующего, либо результата любого из последующих шагов анализа (фигуру субъекта, которая занимает в двудольной структуре схемы место «продук-

та), можно интерпретировать как измененное состояние «сознающего ума» [16] (исследователя вследствие полученных результатов анализа).

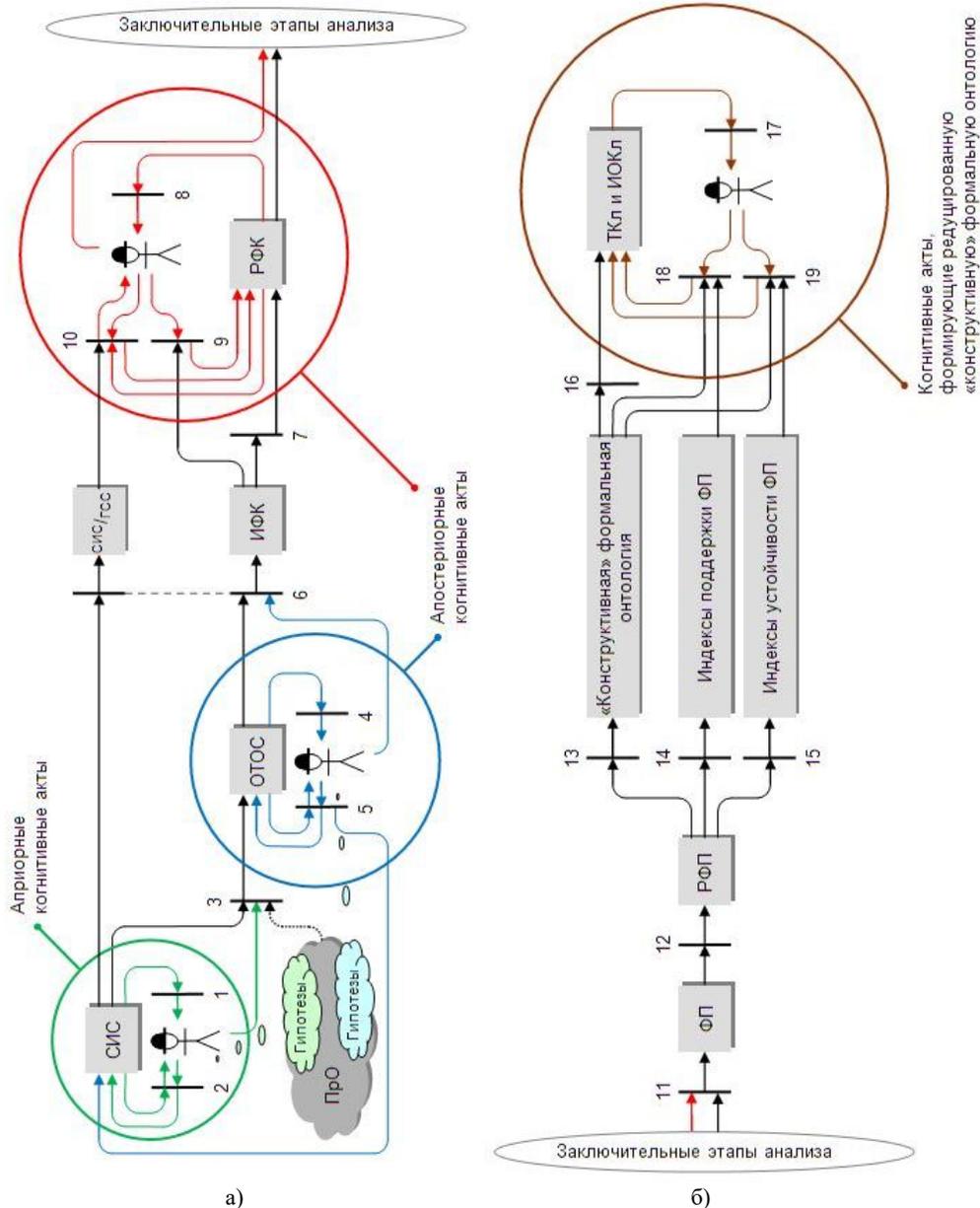


Рис. 1. Пошаговая схема онтологического анализа данных, реализованная в программной лаборатории *OntoWorker*: *a* – подготовка контекста для извлечения формальных понятий и выполняемые при этом когнитивные акты; *б* – заключительные этапы анализа и когнитивные акты, управляющие редукцией его основного продукта – «конструктивной» формальной онтологии

Например, 4-й этап может быть инициирован восприятием субъекта и как ОТОС, и как СИС/ГСС, ИФК, РФК или ФП. Для исключения нагромождения

связей такие переходы изображены лишь в пределах выделенных когнитивных актов.

Здесь и на рис. 1 используются следующие аббревиатуры:

ОТОС – обобщенная в ОАД ТОС [1];

СИС – система измеряемых свойств, понимаемая как множество свойств с априорно заданными на нем ограничениями существования – экзистенциальными бинарными отношениями обусловленности и несовместимости свойств у объектов исследуемой ПрО [17, 18];

СИС/ГСС – редуцированное представление СИС в виде множества пересекающихся групп сопряженных свойств (ГСС), однородных по виду экзистенциального ограничения существования [19];

ИФК – исходный формальный контекст – совокупность объектов измерений, измеряемых свойств и эмпирических оценок истинности базовых семантических суждений (БСС) о ПрО вида «объекту x присуще свойство y », – сформированный непосредственно на основе ОТОС в рамках многозначной векторной логики VTF [20];

РФК – рабочий формальный контекст с однозначными – True или False – оценками БСС, используемый при решении ОАД-задачи;

ФП – формальное понятие;

РФК – решетка формальных понятий – конечное замкнутое множество ФП с заданным на нем бинарным отношением обобщения, являющимся частичным (несовершенным) порядком;

ТКл – таксономия классов объектов – специально определяемая часть РФП – экстенциональная сеть ФП, именуемых классами, используемая для представления формальной онтологии ПрО;

ИОКл – интенциональное отношение классов – воображаемое асимметричное бинарное отношение между ФП, именуемыми классами, которое представляет реально существующее асимметричное бинарное отношение между объектами одного или двух разных классов (множество ИОКл – также часть формальной онтологии ПрО);

2, 5 – формирование и редактирование соответственно СИС и ОТОС, включая концептуальное шкалирование [21] измеряемых свойств. Обратные по отношению к субъекту воздействия связаны с информированием о невозможности выполнить выбранное действие;

3 – измерение (в широком смысле) свойств объектов ПрО;

6 – построение СИС/ГСС и ИФК;

7 – возможная прямая трансформация ИФК в РФК с исключением лишь заведомо ложных БСС [22];

9 – дефазификация ИФК: установка субъективного порога доверия к содержащимся в нем эмпирическим оценкам истинности БСС и обновление РФК путем построения его новой версии как стандартного сечения ИФК с выбранным субъективным порогом доверия;

10 – нормализация РФК – преобразование РФК, в ходе которого снимаются противоречия между эмпирическими данными и априорными ограничениями свойств: множество свойств каждого отдельно взятого объекта измерения нормализуется путем отказа от части эмпирической информации, характеризующейся меньшим значением агрегированного показателя достоверности [22]. Обратное по отношению к субъекту воздействие связано с информированием о невозможности выполнить данное действие;

11, 12 – извлечение из РФК всех ФП и построение РФП;

13 – извлечение формальной онтологии ПрО, «конструктивность» которой означает, что она формируется на основе РФП путем исключения ФП, у которых отсутствуют собственные объем и содержание (то и/или другое лишь наследуется) [1]. В этом смысле «конструктивная» формальная онтология эквивалентна так называемой субиерархии Галуа РФП [23], где множество ФП разделено на непересекающиеся подмножества абстрактных и реальных (data-) классов объектов исследуемой ПрО;

14, 15 – вычисление поддержки и устойчивости ФП [24];

16 – показ извлеченной «конструктивной» формальной онтологии с построением ТКл объектов ПрО и с идентификацией всех ИОКл;

18, 19 – показ «конструктивной» формальной онтологии, осуществляемой построением ТКл при условии, что поддержка/устойчивость включаемых в онтологию классов составляет не менее субъективно указываемого порога, совместно с идентификацией всех ИОКл при указанном пороге поддержки/устойчивости.

Схема на рис. 1 дает представление о функциональности разработанной программной системы с позиции методологии ОАД. Иной ракурс при очерке его функциональности – представление о поведении системы с точки зрения пользователя при его взаимодействии с ней – дает диаграмма вариантов использования из арсенала средств визуального моделирования языка UML [25].

Диаграмма вариантов использования OW (рис. 2) показывает, что единственным действующим лицом по отношению к программной системе является пользователь, играющий роль субъекта-исследователя актуальной ПрО и располагающий следующими способами использования системы:

- получение справочной информации об OW;
- взаимодействие с файловой системой компьютера для загрузки и выгрузки ОАД-задачи – Excel-книги с фиксированной структурой;
- подготовка исходных данных для анализа, включая ввод и редактирование СИС и ОТОС. Эти действия помимо прочего расширяются возможностью концептуального шкалирования измеряемых свойств, которое представлено двумя разновидностями: «дизъюнктивное» замещает шкалируемое свойство набором попарно несовместимых свойств, «уточняющее» добавляет в СИС новое измеряемое свойство, которое становится обуславливающим для определенного ряда других свойств в СИС [26];
- генерация исходных данных для анализа со случайно выбираемыми параметрами СИС и ОТОС с возможностью субъективной коррекции таких параметров;
- документирование ОАД-задачи на листе расширенного описания (ЛРО), включенном в Excel-книгу, размещающую эту задачу;
- построение СИС/ГСС и ИФК, включая их показ;
- формирование РФК, включая его показ. Это действие может быть расширено актом обновления РФК путем дефазсификации ИФК с субъективно задаваемым порогом доверия и актом нормализации РФК;
- извлечение множества ФП из РФК, включая показ извлеченных ФП;
- построение РФП, включая ее показ и показ ее транзитивной редукции;
- вычисление поддержки и устойчивости ФП;
- извлечение «конструктивной» формальной онтологии (или ТКл), включая показ ее ТКл, а также показ транзитивной редукции ТКл. Эти действия расширяются за счет возможности формирования и показа «конструктивной» фор-

мальной онтологии, где все data-классы обладают задаваемым уровнем поддержки и/или устойчивости ФП.

В силу ограниченности места на рис. 2 не дана детализация ряда комплексных вариантов использования. Например, редактирование СИС расширяется не только действиями, реализующими концептуальное шкалирование свойства, но и актами добавления, удаления, переименования измеряемых свойств и т. д. [19].

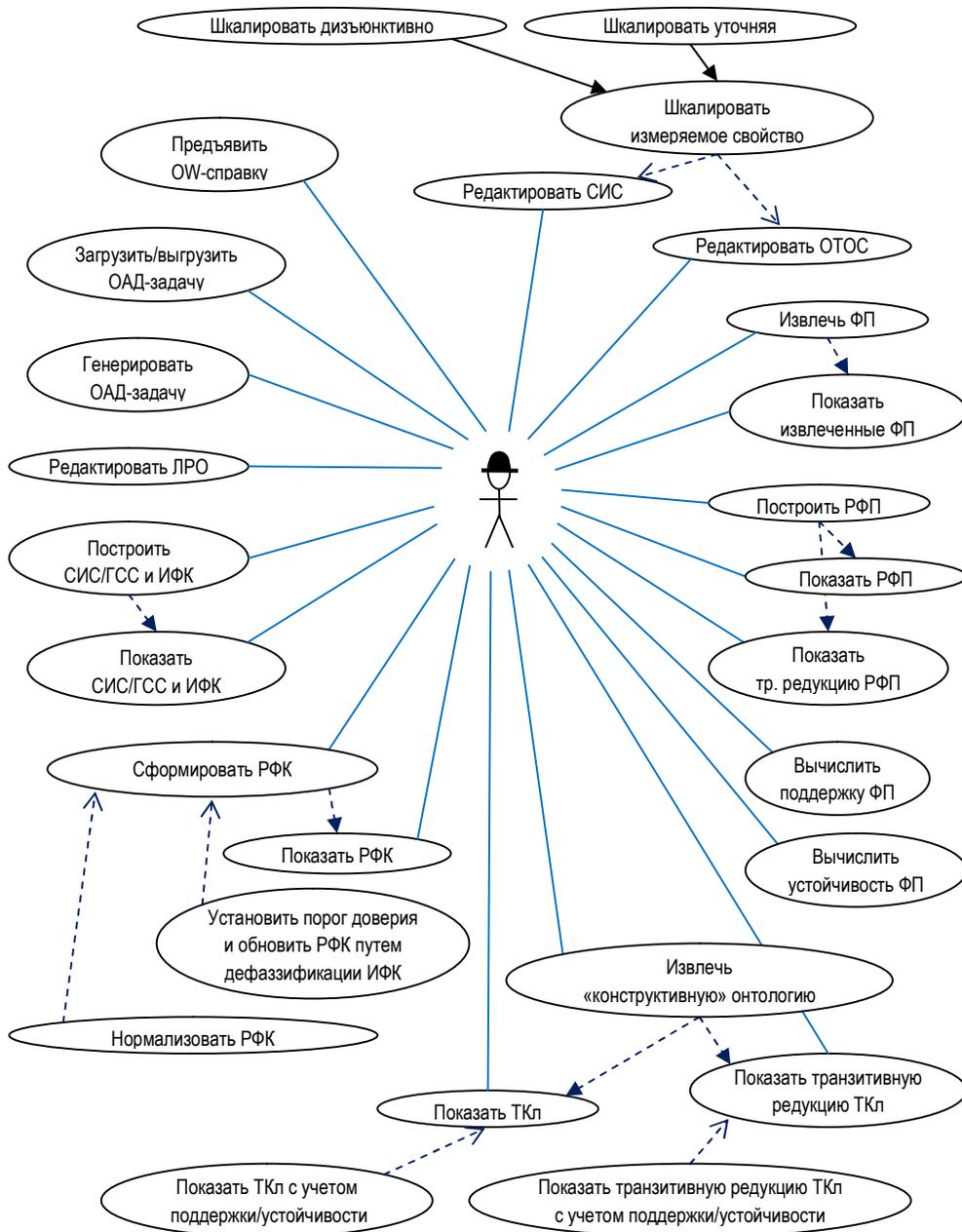


Рис. 2. Диаграмма вариантов использования программной лаборатории для онтологического анализа данных OntoWorker:

— коммуникация; —> обобщение; -> включение (include); -> расширение (extend)

2. Архитектура программной лаборатории

Программная лаборатория OntoWorker представляет собой защищенную Excel-книгу (OW-книгу), включающую набор рабочих листов, часть из которых жестко структурирована, для представления исходных данных, результатов решения и различной сопровождающей информации ОАД-задачи. Между этими OW-листами возможно свободное перемещение.

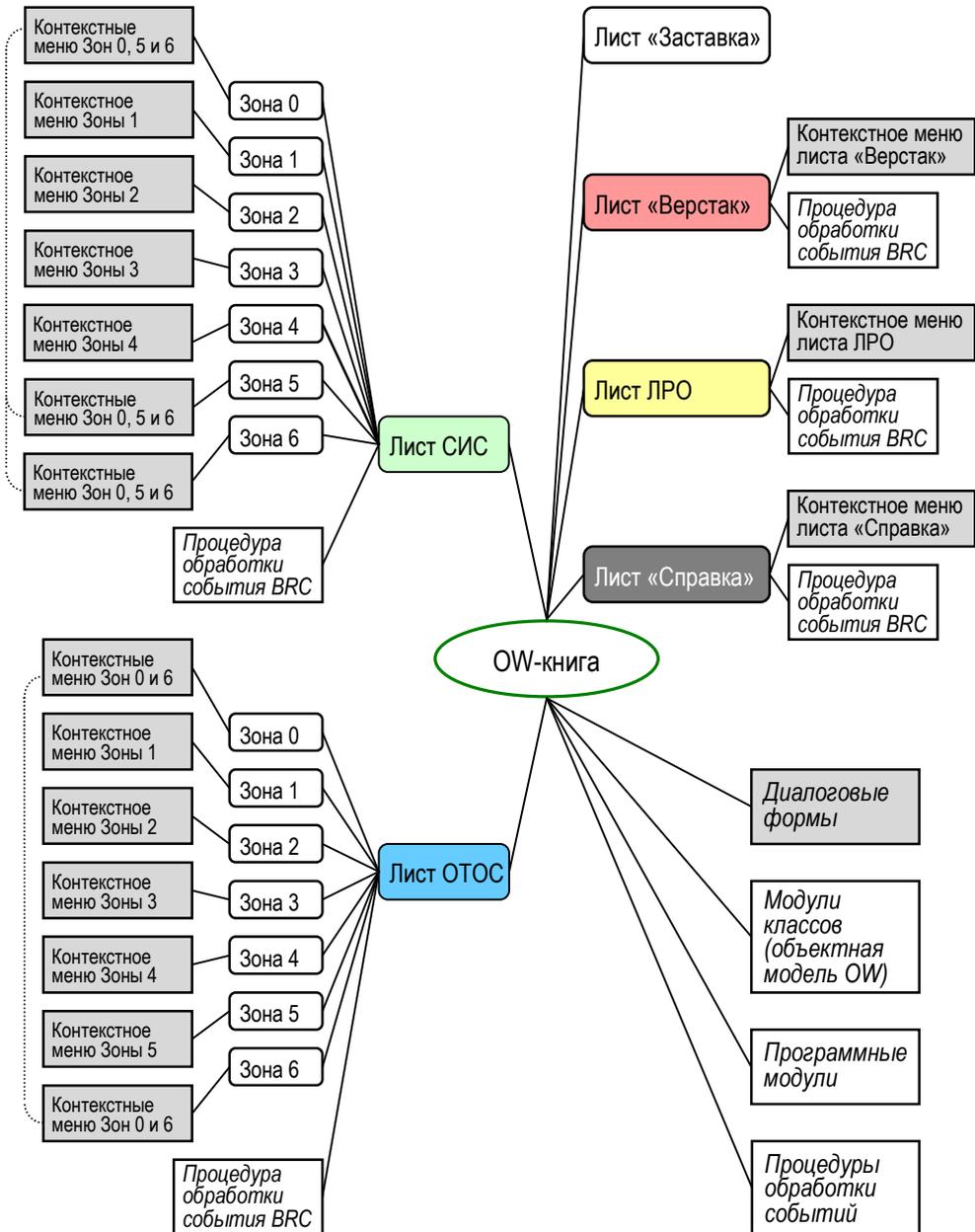


Рис. 3. Архитектура программной лаборатории для онтологического анализа данных OntoWorker

Интеллект-карта [27] на рис. 3 дает представление об архитектуре созданного программного продукта:

- для ввода, редактирования и отображения СИС и ОТОС используются одноименные рабочие листы Excel, которые имеют зонную организацию. Эта организация определяет табличное представление СИС как соответствия «сопряженные пары свойств – свойства» (сопряженная пара свойств – это либо пара несовместимых свойств – Н-пара, либо пара свойств с обусловленностью – О-пара) и табличное представление ОТОС как соответствия «серии измерений – процедуры измерения» [1, 15]. Содержание этих листов защищено, и его изменение возможно только специализированными сервисами, интерфейс к которым осуществляется с помощью различных контекстных меню, локализованных в специализированных зонах этих листов. Наконец, процедура обработки события BRC («Before Right Click») служит здесь и в рамках других OW-листов для вызова контекстных меню, а в случае зонной организации листов – еще и для идентификации зоны, где пользователь кликнул правой кнопкой мыши;

- лист «Справка» содержит справочную информацию о программной лаборатории, включая используемые термины и определения, описание функциональных возможностей и архитектуры OW. Контекстное меню листа ограничивается навигационными задачами;

- лист ЛРО служит для размещения любой информации (тексты, рисунки, мультимедиа и т. п.), сопровождающей загруженную ОАД-задачу. Контекстное меню листа позволяет, в частности, временно снимать его защиту, открывая пользователю возможность редактирования содержания листа ЛРО с помощью всей инструментальной палитры Excel;

- на листе «Верстак» отображаются результаты работы OW. Контекстное меню листа ориентировано прежде всего на предоставление пользователю основных аналитических функций программного приложения, но, в частности, позволяет временно снимать защиту листа для его произвольного редактирования стандартными средствами Excel;

- лист «Заставка» играет исключительно декоративную роль и появляется при открытии OW-книги прежде, чем будет развернут формально основной в программной лаборатории лист СИС;

- наряду с контекстными меню интерактивное взаимодействие пользователя и OW поддерживается с помощью диалоговых форм, а одностороннее – от программных компонентов OW к пользователю – с помощью сообщений;

- объектная модель OW насчитывает почти полтора десятка классов, связанных помимо отношения наследования дюжиной различных интенциональных бинарных отношений (тех, что при описании классов представляют бинарные отношения, в которых состоят экземпляры соответствующих классов, образуя экстенциональную сеть экземпляров);

- функциональные сервисы OW-листов объединены в одноименные программные модули OW-книги. К ним добавлен модуль, содержащий описания разделяемых переменных и сервисов;

- процедуры обработки событий OW-книги служат для реализации надлежащих реакций на ее открытие, закрытие и попытку сохранения (последнее действие некорректно, и OW прекращает работу, прежде предоставляя пользователю возможность сохранить ее результаты).

3. Пример работы программной лаборатории

Для демонстрации работы программной лаборатории OntoWorker использованы объектно-признаковые данные, описывающие инновационные сорта мягкой яровой пшеницы [28]: объектами исследуемой ПрО являются сорта, а также Госреестр, куда включена часть сортов; признаками – характеристики сортов, а также ссылочные свойства (валентности), связывающие сорт и Госреестр.

В адекватной постановке опыт решения задачи онтологического анализа рассматриваемой ПрО представлен в [29]. Здесь же ввиду того, что преследуется цель лишь продемонстрировать работу OW при решении ОАД-задачи, данные о ПрО в части признаков объектов радикально урезаны – во внимание приняты сведения только о засухоустойчивости сортов, их классности (определяется главным образом содержанием белка в зерне) согласно ГОСТ [30] и включении в Госреестр.

На рис. 4 дан фрагмент скриншота листа СИС, где отражены состав и сопряженность учитываемых характеристик сортов пшеницы. О-пары свойств отмечают естественные отношения обусловленности между данными в [28] оценками засухоустойчивости сортов: если сорт «высоко засухоустойчив» (взу), то он и «засухоустойчив» (зу); если он «экстремально засухоустойчив» (эзу), то он и «высоко засухоустойчив», и «засухоустойчив». Н-пары закрепляют установленное ГОСТ [30] разделение пшеницы мягких сортов на пять непересекающихся классов. Представленные в [28] сорта характеризуются в рамках 1–3-го классов, и в СИС фиксируется их несовместимость. Наконец, В-пара описывает сопряженные, или «встречные» валентности объектов (в ОАД валентности описывают потенциальные способности объектов ПрО вступать в связи, реализующие различные несимметричные бинарные отношения): сорт может быть «включен» в Госреестр, а по отношению к некоторому сорту состояние Госреестра может характеризоваться тем, что он «включает» этот сорт.

	1	2	3	4	5	6	7	8	9
1			зу	взу	эзу	1 класс	2 класс	3 класс	включает
2	О-пара	взу - зу	О	Х					
3	О-пара	эзу - взу		О	Х				
4	О-пара	эзу - зу	О	Х	Х				
5	Н-пара	1 класс - 2 класс				Х	Х		
6	Н-пара	1 класс - 3 класс				Х	Х	Х	
7	Н-пара	2 класс - 3 класс					Х	Х	

Рис. 4. Фрагмент скриншота листа СИС программной лаборатории OntoWorker, описывающий задаваемую в задаче онтологического анализа данных систему измеряемых свойств

Фрагмент скриншота листа ОТОС, где зафиксирована фактическая информация о сортах мягкой яровой пшеницы, приведен на рис. 5 [28]. В [28] классность сортов нередко характеризуется отнесением сорта сразу к двум смежным классам ГОСТ с указанием процента зерна того и другого класса. Это отражено в рассматриваемой ОТОС путем задания надлежащих элементов соответствия «объекты – свойства» в виде оценок истинности V^{TF} -логики, в которой вектор $\langle b^+, b^- \rangle$, $b^+, b^- \in [0, 1]$ свидетельствует «за» принадлежность свойства объекту со степенью b^+ и «против» со степенью b^- (в OW степени указываются в процен-

тах). При этом в рассматриваемой ОТОС согласно фактам из [28] выполняется логический принцип исключенного третьего. В общем случае в результате логической интерпретации и консолидации неполных и противоречивых эмпирических данных этот принцип нарушается, но логика V^{TF} вполне адекватна для описания таких ситуаций [1, 20].

	1	2	3	4	5	6	7	8	9	10
1				p_1	p_2	p_3	p_4	p_5	p_6	p_7
2				100	100	100	100	100	100	100
3				зу	взу	эзу	1 класс	2 класс	3 класс	включает
4	s_1	100	Жигулевская	X	X		<80; 20>	<20; 80>		
5	s_2	100	Самсар	X			<50; 50>	<50; 50>		
6	s_3	100	Тулайковская 5	X			X			
7	s_4	100	Тулайковская 10	X	X		X			
8	s_5	100	Тулайковская 100	X	X		<80; 20>	<20; 80>		
9	s_6	100	Тулайковская золотистая	X	X	X	X			
10	s_7	100	Экада 70	X			X			
11	s_8	100	Экада 66	X			X			
12	s_9	100	Тулайковская 110	X	X		X			
13	s_10	100	Тулайковская победа	X	X		<75; 25>	<25; 75>		
14	s_11	100	Тулайковская надежда	X	X			<25; 75>	<75; 25>	
15	s_12	100	Тулайковская 116	X	X	X	<83; 17>	<17; 83>		

Рис. 5. Фрагмент скриншота листа ОТОС программной лаборатории OntoWorker, описывающий задаваемую в задаче онтологического анализа данных эмпирическую информацию о предметной области в виде обобщенной таблицы «объекты – свойства»: s_i , p_j – имена серий измерений и процедур измерения соответственно; цифры (здесь везде 100) – достоверности серий измерений и степени доверия к процедурам измерения

Рис. 6 демонстрирует фрагмент скриншота листа «Верстак» со сформированным и нормализованным РФК: все непустые элементы соответствия «объекты – свойства» отвечают БСС, принятым как однозначно истинные (**True**), а пустые указывают БСС, признанные ложными (**False**). При формировании этого РФК специального порогового сечения ИФК не проводилось; РФК сформирован из ИФК лишь путем исключения заведомо ложных БСС [22]. Нормализация в данном случае свелась к устранению исходной двусмысленности в определении классности сортов мягкой пшеницы.

	зу	взу	эзу	1 класс	2 класс	3 класс	включает	включён
Жигулевская	<100; 0>	<100; 0>		<80; 20>				<100; 0>
Самсар	<100; 0>			<50; 50>				<100; 0>
Тулайковская 5	<100; 0>			<100; 0>				<100; 0>
Тулайковская 10	<100; 0>	<100; 0>		<100; 0>				<100; 0>
Тулайковская 100	<100; 0>	<100; 0>		<80; 20>				<100; 0>
Тулайковская золотистая	<100; 0>	<100; 0>	<100; 0>	<100; 0>				<100; 0>
Экада 70	<100; 0>			<100; 0>				<100; 0>
Экада 66	<100; 0>			<100; 0>				<100; 0>
Тулайковская 110	<100; 0>	<100; 0>		<100; 0>				<100; 0>
Тулайковская победа	<100; 0>	<100; 0>		<75; 25>				<100; 0>
Тулайковская надежда	<100; 0>	<100; 0>				<75; 25>		<100; 0>
Тулайковская 116	<100; 0>	<100; 0>	<100; 0>	<83; 17>				<100; 0>
Экада 214	<100; 0>					<55; 45>		<100; 0>
Зауральская волна	<100; 0>			<100; 0>				<100; 0>
Госреестр							<100; 0>	

Рис. 6. Фрагмент скриншота листа «Верстак» программной лаборатории OntoWorker, описывающий рабочий формальный контекст задачи онтологического анализа данных

Фрагменты скриншота листа «Верстак», приводимые на рис. 7 и рис. 8, *а*, представляют соответственно извлеченное из РФК множество ФП и построенную на этой основе «конструктивную» онтологию. Онтология представлена на рис. 8, *а* двумя таблицами: первая показывает ее в виде транзитивной редукции таксономии классов, вторая описывает интенциональные отношения на классах. Рис. 8, *б* дает содержанию этих таблиц графическую интерпретацию.

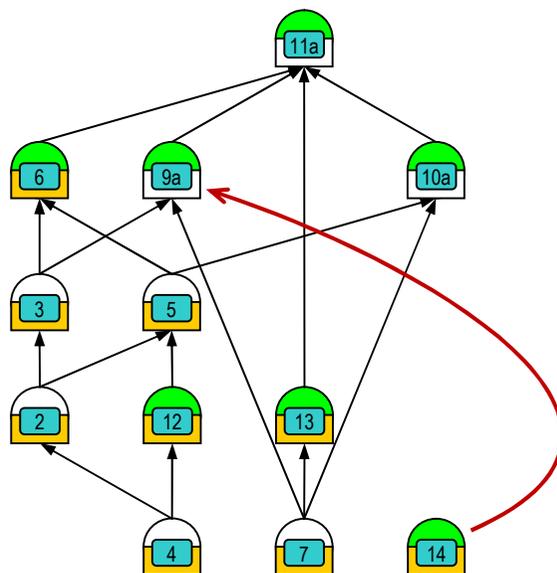
Имя	Объем	Содержание
1		зу; взу; эзу; 1 класс; 2 класс; 3 класс; включает; включён
2	Жигулевская; Тулайковская 10; Тулайковская 100; Тулайковская золотистая; Тулайковская 110	зу; взу; 1 класс; включён
3	Жигулевская; Самсар; Тулайковская 5; Тулайковская 10; Тулайковская 100; Тулайковская золотистая; Экада 70; Экада 66; Тулайковская 110	зу; 1 класс; включён
4	Тулайковская золотистая	зу; взу; эзу; 1 класс; включён
5	Жигулевская; Тулайковская 10; Тулайковская 100; Тулайковская золотистая; Тулайковская 110; Тулайковская победа; Тулайковская 116	зу; взу; 1 класс
6	Жигулевская; Самсар; Тулайковская 5; Тулайковская 10; Тулайковская 100; Тулайковская золотистая; Экада 70; Экада 66; Тулайковская 110; Тулайковская победа; Тулайковская 116; Зауральская волна	зу; 1 класс
7	Тулайковская надежда	зу; взу; 3 класс; включён
8	Жигулевская; Тулайковская 10; Тулайковская 100; Тулайковская золотистая; Тулайковская 110; Тулайковская надежда	зу; взу; включён
9	Жигулевская; Самсар; Тулайковская 5; Тулайковская 10; Тулайковская 100; Тулайковская золотистая; Экада 70; Экада 66; Тулайковская 110; Тулайковская надежда	зу; включён
10	Жигулевская; Тулайковская 10; Тулайковская 100; Тулайковская золотистая; Тулайковская 110; Тулайковская победа; Тулайковская надежда; Тулайковская 116	зу; взу
11	Жигулевская; Самсар; Тулайковская 5; Тулайковская 10; Тулайковская 100; Тулайковская золотистая; Экада 70; Экада 66; Тулайковская 110; Тулайковская победа; Тулайковская надежда; Тулайковская 116; Экада 214; Зауральская волна	зу
12	Тулайковская золотистая; Тулайковская 116	зу; взу; эзу; 1 класс
13	Тулайковская надежда; Экада 214	зу; 3 класс
14	Госреестр	включает
15	Жигулевская; Самсар; Тулайковская 5; Тулайковская 10; Тулайковская 100; Тулайковская золотистая; Экада 70; Экада 66; Тулайковская 110; Тулайковская победа; Тулайковская надежда; Тулайковская 116; Экада 214; Зауральская волна; Госреестр	

Рис. 7. Фрагмент скриншота листа «Верстак» программной лаборатории OntoWorker, описывающий формальные понятия, извлеченные в ходе решения задачи онтологического анализа данных

Класс	Собственный объём	Отличительное содержание	Ярус 0	Ярус 1	Ярус 2
2	Жигулевская; Тулайковская 10; Тулайковская 100; Тулайковская 110			2	3; 5
3	Самсар; Тулайковская 5; Экада 70; Экада 66				3
4	Тулайковская золотистая		4	2; 12	
5	Тулайковская победа				5
6	Зауральская волна	1 класс			
7	Тулайковская надежда		7	13	
9a		включён			
10a		взу			
11a		зв			

Отношение	max-X-класс	max-O-класс	data-X-классы
включает - включён	14	9a	14

a



б

Рис. 8. Транзитивная редукция таксономии классов «конструктивной» онтологии и их интенциональные отношения: *a* – фрагменты скриншотов листа «Верстак»; *б* – графическая интерпретация

Заключение

Создание программной лаборатории OntoWorker как конкурентоспособного программного продукта в области инженерии знаний стало возможным благодаря успешному развитию в ИПУСС РАН – СамНЦ РАН наукоемкой и практически значимой методологии онтологического анализа данных. Совмещая известные и новые технологии семантической обработки многомерной информации, программная лаборатория имеет значительный потенциал развития прежде всего

в части функционального наполнения. Например, здесь можно обозначить следующие задачи:

– построить содержательные процедуры концептуального шкалирования измеряемых свойств при формировании СИС, вводя описание доменов значений свойств. Сейчас это шкалирование осуществляется умозрительно, формируя тем не менее гипотетическую понятийную структуру исследуемой ПрО и пополняя ограничения существования свойств;

– добавить стандартные возможности извлечения из РФК импликаций и ассоциативных правил на признаках [4, 5];

– расширить палитру критериев и методов оценки качества выводимых понятий, включая возможность агрегирования различных критериев, и поддержать перспективные методики построения «конструктивных» онтологий при редуцировании множества формальных понятий на основе учета их качества по различным критериям;

– интегрировать в программную лабораторию возможности графического представления результатов анализа.

Наконец, учитывая экспоненциальную сложность ряда алгоритмов онтологического анализа и ограниченную производительность интерпретирующей программной платформы, каковой является связка Excel-VBA, оправданно обозначить вопрос о реинжиниринге OntoWorker. Наследуя предложенные архитектурные решения, разумно рассмотреть разнообразные современные возможности в диапазоне от разработки настольного приложения промышленной базы данных на основе мощного средства автоматизации программирования до создания широкодоступного «облачного» сервиса.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Семенова В.А., Смирнов В.С., Смирнов С.В. OntoWorker: программная лаборатория для онтологического анализа данных // Проблемы управления и моделирования в сложных системах: Тр. XVII Международной конференции. Самара: СамНЦ РАН, 2015. С. 382–393.
2. Загоруйко Н.Г. Когнитивный анализ данных. Новосибирск: Гео, 2013. 186 с.
3. Барсегян А.А., Куприянов М.С., Холод И.И., Тесс М.Д., Елизаров С.И. Анализ данных и процессов. СПб.: БХВ-Петербург, 2009. 512 с.
4. Ganter B., Wille R. Formal Concept Analysis. Mathematical foundations. Berlin-Heidelberg: Springer-Verlag, 1999. 290 p.
5. Ferré S., Huchard M., Kaytoue M., Kuznetsov S.O., Napoli A. Formal Concept Analysis: From Knowledge Discovery to Knowledge Processing // In: Marquis P., Papini O., Prade H. A Guided Tour of Artificial Intelligence Research. Vol. II: AI Algorithms Springer Int. Publishing, 2020. Pp. 411–445.
6. Философский энциклопедический словарь. М.: ИНФРА-М, 1997. 576 с.
7. Гладкий А.В. Введение в современную логику. М.: URSS, 2022. 240 с.
8. The Concept Explorer. <http://conexp.sourceforge.net> (обращение 07.04.2023).
9. Toscanaj. <http://toscanaj.sourceforge.net> (обращение 07.04.2023).
10. Galicia Lattice Builder. <http://www.iro.umontreal.ca/~galicia> (обращение 07.04.2023).
11. Python FCA Tool. <https://github.com/ae-hse/fca> (обращение 07.04.2023).
12. Гарнаев А.Ю., Рудикова Л.В. Microsoft Office Excel 2010: разработка приложений. СПб.: БХВ-Петербург, 2011. 528 с.
13. Уокенбах Дж. Excel 2010: профессиональное программирование на VBA. М.: И.Д. Вильямс, 2011. 994 с.
14. Зубцов Р.О., Семенова В.А., Смирнов С.В. Алгоритмическое и программное обеспечение онтологического анализа данных // Open Semantic Technologies for Intelligent Systems: Материалы VI международной науч.-техн. конференции. Мн: Изд-во БГУИР, 2016. С. 83–88.

15. *Лецева Д.В., Семенова В.А., Смирнов С.В.* Пользовательский интерфейс программной лаборатории для онтологического анализа данных // Перспективные информационные технологии: Тр. международной науч.-техн. конференции. Самара: СамНЦ РАН, 2018. С. 479–482.
16. *Чалмерс Д.* Сознательный ум: в поисках фундаментальной теории. М.: URSS, 2019. 512 с.
17. *Lammari N., Metais E.* Building and maintaining ontologies: a set of algorithms // Data & Knowledge Engineering. 2004. Vol. 48(2). Pp. 155–176.
18. *Пронина В.А., Шипилина Л.Б.* Использование отношений между атрибутами для построения онтологии предметной области // Проблемы управления. 2009. № 1. С. 27–32.
19. *Семенова В.А., Смирнов С.В.* Алгоритмизация формирования и прагматической трансформации ограничений существования свойств предметной области // Онтология проектирования. 2020. Т. 10, № 3. С. 361–379.
20. *Аршинский Л.В.* Векторные логики: основания, концепции, модели. Иркутск: Изд-во Иркутского государственного ун-та, 2007. 228 с.
21. *Ganter B., Wille R.* Conceptual scaling // Applications of Combinatorics and Graph Theory to the Biological and Social Sciences. New York Springer-Verlag, 1989. P. 139–167.
22. *Semenova V., Smirnov S.* Revealing attributes of an object based on incomplete and inconsistent empirical data // VIII Int. Conf. on Information Technology and Nanotechnology. IEEE Xplore. 2022. Pp. 1–5.
23. *Arévalo G., Berry A., Huchard M., Perrot G., Sigayret A.* Performances of Galois Subhierarchy-building Algorithms // Formal Concept Analysis. ICFCA 2007. Lecture Notes in Computer Science. Vol 4390. Berlin, Heidelberg: Springer, 2007. Pp. 166–180.
24. *Kuznetsov S.O.* On stability of a formal concept // Annals of Mathematics and Artificial Intelligence. 2007. Vol. 49(1–4). Pp. 101–115.
25. *Иванов Д.Ю., Новиков Ф.А.* Моделирование на UML. СПб.: Наука и техника, 2010. 640 с.
26. *Лецева Д.В., Семенова В.А.* Управление данными при кластеризации объектов многомерных наблюдений и экспериментов // Перспективные информационные технологии: Тр. международной науч.-техн. конференции. Самара: СамНЦ РАН, 2020. С. 127–132.
27. *Бьюзен Т., Бьюзен Б.* Интеллект-карты. Практическое руководство. Минск: Попурри, 2010. 368 с.
28. Каталог инновационных разработок Самарского НИИ сельского хозяйства имени Н.М. Тулайкова на 2018 год. Самара: СамНЦ РАН, 2018. 92 с.
29. *Семенова В.А.* Опыт бикластеризации данных о сортах сельскохозяйственных культур // Известия Самарского научного центра Российской академии наук. 2020. Т. 2, № 1. С. 86–92.
30. ГОСТ 9353-2016 Пшеница. Технические условия. <https://pdf.standartgost.ru/catalog/Data2/1/4293751/4293751950.pdf> (дата обращения: 07.04.2023).

Статья поступила в редакцию 17 апреля 2023 г.

FUNCTIONAL CONTENT AND ARCHITECTURE OF SOFTWARE LABORATORY FOR ONTOLOGICAL DATA ANALYSIS

*V.A. Semenova, S.V. Smirnov**

Samara Federal Research Scientific Center RAS,
Institute for the Control of Complex Systems RAS
61, Sadovaya st., Samara, 443020, Russian Federation

E-mail: queenbfjr@gmail.com

Abstract. *The article presents the functionality and architecture of the software laboratory for ontological analysis and through this “prism” – the methodology of this analysis. The methodological complex of ontological analysis and the procedure for its application provide extraction from the data of multidimensional observations and measurements of the knowledge domain of its semantic model in the form of a formal ontology – a set of formal concepts, each of which is determined by extent and intent. In the set of formal concepts, a partial order is revealed (a binary relation of generalization) and intensional relations are revealed that reflect the connections between the elements of the concepts extents. The developed software laboratory differs from the well-known tools for constructing formal ontologies based on empirical object-feature data (i.e., based on data presented in the generally recognized form of a measurement results registration protocol), first of all, by taking into account the realities of accumulating information about the knowledge domain under study. In the general case, they cause the incompleteness and inconsistency of the initial data, for the processing of which the apparatus of multi-valued vector logic is involved. Another unique difference of the software laboratory is the consideration of a priori known (i.e., known before measurements) constraints on the properties existence during the primary processing of empirical data – binary relations of conditionality and incompatibility of objects properties of the studied knowledge domain. The presented software laboratory is implemented on the Excel table processor platform and the programming language Visual Basic for Application. The main motive for this choice was to ensure the availability and facilitate the familiarization of ontological data analysis technology by a very wide range of users who use Excel in their professional work.*

Keywords: *mass software platform, formal ontology, ontological data analysis, formal concept analysis, property existence constraints.*

REFERENCES

1. *Semenova V.A., Smirnov V.S., Smirnov S.V.* OntoWorker: Program Laboratory for Ontological [Data Analysis] // Proc. of the XVII Int. Conf. Complex Systems: Control and Modeling Problems. Samara: Samara Scientific Center of RAS, 2015. Pp. 382–393. (In Russian).
2. *Zagoruyko N.G.* Kognitivnyy analiz dannykh [Cognitive data analysis]. Novosibirsk: Geo Publisher, 2013. 186 p. (In Russian).
3. *Barsegyan A.A., Kupriyanov M.S., Holod I.I., Tess M.D., Elizarov S.I.* Analiz dannykh i protsessov [Data and Process Analysis]. St. Petersburg: BHV-Petersburg. 2009. 512 p. (In Russian).
4. *Ganter B., Wille R.* Formal Concept Analysis. Mathematical foundations. Berlin-Heidelberg: Springer-Verlag, 1999. 290 p.
5. *Ferré S., Huchard M., Kaytoue M., Kuznetsov S.O., Napoli A.* Formal Concept Analysis: From Knowledge Discovery to Knowledge Processing // In: Marquis P., Papini O., Prade H.

* *Valentina A. Semenova, Junior Researcher.
Sergey V. Smirnov (Dr. Sci. (Techn.)), Chief Researcher.*

- A Guided Tour of Artificial Intelligence Research. Vol. II: AI Algorithms Springer Int. Publishing, 2020. P. 411–445.
6. Filosofskiy entsiklopedicheskiy slovar' [Philosophical Encyclopedic Dictionary]. M.: INFRA-M, 1997. 576 p. (In Russian).
 7. *Gladkiy A.V.* Vvedeniye v sovremennuyu logiku [Introduction to Modern Logic]. M.: URSS, 2022. 240 p. (In Russian).
 8. The Concept Explorer. <http://conexp.sourceforge.net> (accessed April 07, 2023).
 9. ToscanaJ. <http://toscanaj.sourceforge.net> (accessed April 07, 2023).
 10. Galicia Lattice Builder. <http://www.iro.umontreal.ca/~galicia> (accessed April 07, 2023).
 11. Python FCA Tool. <https://github.com/ae-hse/fca> (accessed April 07, 2023).
 12. *Garnaev A.Yu., Rudikova L.V.* Microsoft Office Excel 2010: razrabotka prilozheniy [Microsoft Office Excel 2010: Application Development]. St. Petersburg: BHV-Peterburg, 2011. 528 p. (In Russian).
 13. *Walkenbach J.* Excel 2010: professional'noye programmirovaniye na VBA [Excel 2010: power programming with VBA]. Moscow: Williams, 2011. 994 p. (In Russian).
 14. *Zybtsov R.O., Semenova V.A., Smirnov S.V.* Algorithmic and software for ontological data analysis // Proc. of the VI Int. Scientific and Technical Conf. "Open Semantic Technologies for Intelligent Systems". Minsk: BGUIR. 2016. Pp. 83–88. (In Russian).
 15. *Leschyova D.V., Semenova V.A., Smirnov S.V.* User interface of a software laboratory for ontological data analysis // Proc. Int. Conf. "Perspective information technology". Samara Scientific Center of RAS, 2018. Pp. 479–482. (In Russian).
 16. *Chalmers D.* Soznayuschiy um: v poiskakh fundamental'noy teorii [The Conscious Mind: In Search of a Fundamental Theory]. Moscow: URSS, 2019. 512 p. (In Russian).
 17. *Lammari N., Metais E.* Building and maintaining ontologies: a set of algorithms // Data & Knowledge Engineering. 2004. Vol. 48(2). Pp. 155–176.
 18. *Pronina V.A., Shipilina L.B.* Using the relationships between attributes to build domain ontology // Problemy upravleniya. 2009. No 1. Pp. 27–32. (In Russian).
 19. *Semenova V.A., Smirnov S.V.* Algorithms for the formation and pragmatic transformation of Existence Constraints // Ontologiya proyektirovaniya. 2020. Vol. 10, no. 3. Pp. 361–379. (In Russian).
 20. *Arshinskiy L.V.* Vectornyye logiki: osnovaniya, kontseptsii, modeli [Vector logics: foundations, concepts, models]. Irkutsk: Irkutsk State University Publishing House. 2007. 228 p. (In Russian).
 21. *Ganter B., Wille R.* Conceptual scaling. In: Applications of Combinatorics and Graph Theory to the Biological and Social Sciences. New York Springer-Verlag, 1989. Pp. 139–167.
 22. *Semenova V., Smirnov S.* Revealing attributes of an object based on incomplete and inconsistent empirical data. VIII Int. Conf. on Information Technology and Nanotechnology. IEEE Xplore. 2022. Pp. 1–5.
 23. *Arévalo G., Berry A., Huchard M., Perrot G., Sigayret A.* Performances of Galois Subhierarchy-building Algorithms. In: Formal Concept Analysis. ICFA 2007. Lecture Notes in Computer Science. Vol 4390. Berlin, Heidelberg: Springer, 2007. Pp. 166–180.
 24. *Kuznetsov S.O.* On stability of a formal concept // Annals of Mathematics and Artificial Intelligence. 2007. Vol. 49 (1–4). Pp. 101–115.
 25. *Ivanov D.Yu., Novikov F.A.* Modelirovaniye na UML [Modeling in UML]. St. Petersburg: Nauka i tekhnika. 2010. 640 p.
 26. *Leschyova D.V., Semenova V.A.* Data Management in Clustering Objects of Multidimensional Observations and Experiments // Proc. Int. Conf. "Perspective information technology". Samara Scientific Center of RAS, 2020. Pp. 127–132. (In Russian).
 27. *Buzan T., Buzan B.* Intellekt-karty. Prakticheskoye rukovodstvo [Mind cards. Practical guide]. Minsk: Poppuri. 2010. 368 p. (In Russian).
 28. Katalog innovatsionnykh razrabotok Samarskogo NII sel'skogo khozyaystva imeni N.M. Tulaykova na 2018 god [Catalog of innovative developments of the Samara Research Institute of Agriculture named after N.M. Tulaykov for 2018]. Samara: Samara Scientific Center of RAS, 2018. 92 p. (In Russian).
 29. *Semenova V.A.* Experience of Biclustering Data on Crop Varieties // Izvestiya Samarskogo nauchnogo tsentra Rossiiskoy Akademiyi Nauk. 2020. T. 2, no. 1. Pp. 86–92. (In Russian).
 30. GOST 9353-2016 Pshenitsa. Tekhnicheskoye usloviya [GOST 9353-2016 Wheat. Specifications]. <https://pdf.standartgost.ru/catalog/Data2/1/4293751/4293751950.pdf> (accessed: April 7, 2023).