

УДК 004.89:004.93

DOI 10.34757/2413-7383.2023.31.4.001

Д. А. Гаркуша

Федеральное государственное бюджетное научное учреждение  
«Институт проблем искусственного интеллекта», г. Донецк  
283048, г. Донецк, ул. Артема, 118 б

## ФУНКЦИОНАЛЬНЫЕ ОСОБЕННОСТИ РЕАЛИЗОВАННЫХ ОНТОЛОГИЧЕСКИХ ПЛАТФОРМ

D. A. Garkusha

Federal State Budgetary Scientific Institution «Institute of Artificial Intelligence Problems»  
283048, Donetsk, Artem str., 118 b

## FUNCTIONAL FEATURES OF IMPLEMENTED ONTOLOGICAL PLATFORMS

В статье приведен анализ инструментов в области построения онтологии. Исследование данной проблемы показало, что на данный момент присутствует малое количество источников, показывающих эффективное использование инструментов в разных задачах, стеках, объем проекта. Результат даст использование данных в комплексном подходе в решении задач онтологического инжиниринга.

**Ключевые слова:** онтология, онтологические платформы.

The article provides an analysis of tools in the field of ontology construction. The study of this problem showed that at the moment there is a small number of sources showing the effective use of tools in different tasks, stacks, and project scope. The result will be the use of data in an integrated approach to solving ontological engineering problems.

**Keywords:** ontology, ontological platforms.

Ориентация на использование онтологий при создании информационных систем становится всё более актуальным вопросом в связи с принятием концепции создания цифровой аналитической платформы предоставления статистических данных, в которой присутствует требование применения онтологии данных, используемой в модели государственных данных [1]. Очевидно, что при разработке программного обеспечения, которое предполагается использовать для хранения и обработки государственных данных, будут требования к использованию инструментальных средств работы со знаниями, разработанных в Российской Федерации.

Существует не меньше десятка зарубежных инструментов онтологического инжиниринга, поддерживающих формализмы для описания знаний и использующих машины вывода из этих знаний.

Рассмотрим кратко некоторые из инструментов работы с онтологиями.

**Protégé** – локальная, свободно распространяемая Java-программа, предназначенная для создания, просмотра и редактирования онтологий явных моделей предметной области с целью включения этих моделей в программный код. Редактор онтологий Protégé позволяет разворачивать иерархическую структуру абстрактных или конкретных классов и слотов. Структура онтологии аналогична иерархической структуре каталога. На основе онтологии Protégé может генерировать формы получения знаний для введения экземпляров классов и подклассов. Имеет графический интерфейс. Поддерживает использование языка OWL. Генерирует HTML-документы, отображающие структуру онтологий. Использует фреймовую модель представления знаний ОКВС – позволяет редактировать модели предметных областей, представленных в форматах UML, XML, SHOE, DAML+OIL, RDF / RDFS и т.п. (а не только в не в OWL).

**Stardog** – это платформа EnterpriseKnowledgeGraph, сочетающая возможности хранения графов и виртуализации для гибкой и экономичной интеграции данных. Stardog масштабируем, безопасен и основан на стандартах W3C Semantic Web.

Stardog обеспечивает, помимо работы в RDF, наличие набора функций и средств для интеграции данных и сложных данных:

- виртуальные коннекторы данных для всех основных SQL-серверов, Cassandra, MongoDB и других для удобного доступа к хранилищам данных;
- конвейер NLP BITES, позволяет включать неструктурированные данные в дополнение к данным SQL и NoSQL в граф знаний;
- виртуализация данных, оптимизирующая ETL (основные процессы управления данными хранилища) и устойчивое добавление данных в граф знаний;
- поддержка BI / SQL Server, который переводит граф знаний обратно в SQL; поддерживаемые платформы включают Tableau, PowerBI, Cognos и др;
- своевременный вывод позволяет исследовать данные с применением различных наборов правил;
- встроенное машинное обучение, включая прогнозную аналитику и поиск по сходству;
- запросы пути позволяют перемещаться по графу, обнаруживая связи в данных.

ApacheJena – это открытая Java-платформа для создания приложений SemanticWeb и LinkedData. Рассчитана на преобразование и сохранение данных в формате RDF запрос данные RDF с использованием SPARQL.

Демо-версия Ахиграф

Архиграф представляет собой платформу для построения дата-центра архитектуры автоматизированных систем. Компания «ТриниДата» зарегистрирована в Российской Федерации.

Выделяют следующие 3 продукта:

1. АрхиГраф.М – редактор онтологий;
2. АрхиГраф.СУЗ – инструмент редактирования правил и исследования данных;
3. АрхиГраф.МДМ – программный интерфейс доступа к данным.

Назначение продукта АрхиГраф.Мир, совместная работа с онтологическими моделями и данными, представленными в соответствии с ними.

Продукт обещает следующую функциональность:

- создание и редактирование классов, свойств, индивидуальных объектов в соответствии со стандартами RDFS/OWL;
- просмотр ограничений SHACL и результатов их выполнения;
- комментирование и аннотирование элементов онтологии;
- импорт/экспорт элементов онтологии через файлы формата Excel;
- выгрузка OWL-модели в XML или Turtle;
- поиск, фильтрация, сортировка элементов онтологии;
- групповые операции с элементами онтологии.

На веб-ресурсе производителя программного обеспечения (ПО) предоставляется литература по направлениям: онтологическое моделирование; моделирование предприятий; документация по эксплуатации выпускаемого ПО.

**АрхиГраф** предназначается для хранения информационной модели, нормативно-справочной информации, основных и транзакционных данных организации. Обеспечивает доступ прикладным программным компонентам программного обеспечения к данным, размещенным в хранилищах под ее управлением, и к любой информации во внешних хранилищах.

Платформа АрхиГраф позиционируется как ядро сложных прикладных автоматизированных систем, в том числе: ситуационных центров, интегрированных аналитических систем и логических витрин данных, систем построения отчетности и др. Может выступать в роли корпоративного инструмента управления данными (Data Governance) в сложных, интегрированных инфраструктурах, включающих десятки бизнес-приложений. Включает в себя синхронизацию данных между различными приложениями, контроль прав доступа при обращении к информации, версионность модели и данных, извлечение данных из внешних источников в режиме логической витрины, подписку на уведомления об изменении модели и данных, логический контроль информации и многие другие функции.

Производитель обещает кроме основной функциональности – создания и редактирования элементов онтологий – редактор имеет функции, предназначенные для решения дополнительных задач: контроля прав доступа к элементам модели, импорта/экспорта данных, управления версионностью онтологии и отслеживания истории редактирования ее элементов, просмотра правил контроля качества данных и результатов их применения, обсуждения и согласования изменений в онтологии и др.

На веб-ресурсе доступна демо-версия АрхиГраф.Мир после предварительной регистрации. Обнаружены недостатки, демо-версия требует установки на рабочую станцию виртуальной машины, куда и требуется установить ПО АрхиГраф.Мир. Так же система ограничивает пользователя в использовании продукта в корпоративной сети. Возникает проблема с IPv4, если он в вашей локальной сети не начинается на 192.168. Требуется использовать IP домашней сети, которую может предоставить, например роутер. Лицензия на продукт от «ТриниДата» предоставляется на один год, ниже приведена таблица с ценами.

Таблица 1 – Стоимость лицензии на продукты «Тринидата»

Лицензия на один сервер	Цена, руб. без НДС
АрхиГраф.MDM	2 000 000
АрхиГраф.СУЗ	1 200 000
АрхиГраф.Мир	1 000 000
Пульс Дела / Пульс Региона *	1 000 000
Лицензии на один кластер Kubernetes	
АрхиГраф.MDM	4 000 000
Пульс Дела / Пульс Региона *	5 000 000

Высокая цена на лицензию и отсутствие бесплатных решений затрудняет произвести анализ ПО в качестве конкурентно способного продукта в сравнении с бесплатными аналогами, для ознакомления разработчиков с полным спектром возможностей и интеграции в собственные сервисы. Свою очередь ведется подробное изучение документации и возможностей демо-версии продукта. Насколько известно из документации, предоставляемой на официальном сайте «Тринидата», используется свободно ПО «ApacheJena» для хранения данных. Сами программные продукты являются слоем по взаимодействию и управлению данными между пользователем и хранилищем данных.

В это же время графовые базы набирают популярность в качестве инструмента для эмуляции графа RDF. Все чаще в сети появляются решения, которые предлагают не погружаться в веб-семантику, что сопровождается трудоемким процессом изучения новой технологии и вынужденной адаптации под данный существующий набор инструментов. Вместо этого предлагается изучить только один новый инструмент и базовые принципы использования. Данные производители программного обеспечения уже как правило предоставляют документацию как моделировать те или иные графы. В этот список входит Neo4J, OrientDB, MarkLogic и др. инструменты.

**Neo4J** – графовая система управления базами данных с открытым исходным кодом, реализованная на Java. Данные хранит в собственном формате, приспособленном для представления графовой информации, такой подход в сравнении с моделированием графовой базы данных средствами реляционной СУБД позволяет применять дополнительную оптимизацию в случае данных с более сложной структурой.

Характеризуется высокой производительностью и масштабируемостью. Распределенная кластерная архитектура Neo4j позволяет клиентам выполнять сложные рабочие нагрузки онлайн-обработки транзакций (OLTP) и обработки данных, сохраняя при этом соответствие ACID (атомарность, согласованность, изолированность, прочность) и целостность данных.

Neo4j может быть развернута на собственной, гибридной или мультиоблачной платформе. Обеспечивает 1000-кратную производительность при неограниченном масштабе.

Производительность обеспечивается за счет декларативного языка, имеющего самое большое в мире сообщество опытных разработчиков графов. Считается первым в отрасли Graph ML для предприятий.

**MarkLogic** – это современная мультимодельная база данных. Имеет все функции NoSQL баз данных.

Хранилище документов: нативная XML СУБД; RDF хранилище; поисковая система;

Рассмотрим особенности, делающие MarkLogic удобной для интеграции данных:

1. Многомодельная база данных. MarkLogic хранит и запрашивает данные в документах, графических данных или реляционных данных, обеспечивая невероятную гибкость.

2. Встроенный поиск. В отличие от других баз данных, MarkLogic имеет встроенную поисковую систему. Это приводит к меньшим затратам времени и усилий на создание и настройку индексов для стандартных запросов и не требует привязанной поисковой системы для полнотекстового поиска, как в других базах данных.

3. Эластичная масштабируемость. MarkLogic является масштабируемой базой данных, которая масштабируется по горизонтали в кластерах на стандартном оборудовании до сотен узлов, петабайт данных и миллиардов документов, и при этом обрабатывает десятки тысяч транзакций в секунду.

Вывод «ТриниДата» предоставляет ПО отечественного решения в качестве аналогов зарубежных систем онтологического инжиниринга. Оно может использоваться для работы с государственными данными и других структурах на территории Российской Федерации. Но сейчас отечественное решение является труднодоступным, поскольку нацелено на корпоративный сектор, такие как АО "НИПИГАЗ", Газпром недра, Россети Урал, ПАО "Россети", Универмаг Au Pont Rouge, ПАО "РусГидро" и т.д. предоставляя услуги следующего спектра:

- проектирование и создание архитектур ИТ-систем;
- создание онтологических моделей;
- разработка и внедрение ИТ-систем;
- интеграция автоматизированных систем;
- создание каталогов мастер-данных и НСИ;

Следовательно, сотрудники организаций небольшого уровня не могут полноценно изучить данную технологию под отечественным ПО из-за отсутствия решений для малого сектора что является минусом. Проведем параллель, Microsoft SQL Server, есть бесплатное решение, которое подходит для разработчиков и малого бизнеса, если у вас возникает потребность перейти на более «высокий уровень» то, поставщик предлагает уже платное решение. Тем самым стимулирует интеграцию своей среды и переманивая разработчиков. Не малую роль играет и поддержка сообщества так как наиболее популярные решения удобны широкой поддержкой. MarkLogic предоставляет учебную программу с менторами и отдельно для самоизучения, а также доступ к своим серверам бесплатно, что делает ее привлекательной. Возможно, Тринидата в скором времени предложит решение, которое смогут использовать разработчики в небольших командах, но, на данный момент все вышеизложенное заставляет выбирать зарубежное ПО, которое может использоваться для построения онтологии.

## Список литературы

1. Дорохина Г.В. Формализованные онтологии и задачи построения компьютерной информационной технологии цифрового сбора, обработки и анализа данных. *Искусственный интеллект: теоретические аспекты, практическое применение : материалы Донецкого международного научного круглого стола*. Донецк: ГУ ИПИИ, 2020. С. 229–234
2. Дорохина Г. В. Требования к информационной технологии цифрового сбора, обработки и анализа данных. *Проблемы искусственного интеллекта*. 2020. № 4 (19). С. 4–9.
3. Муромцев Д. Индустриальные графы знаний – интеллектуальное ядро цифровой экономики [Текст] / Дмитрий Муромцев, Алексей Романов, Дмитрий Волчек. *Control Engineering Россия*. № 5 (83). октябрь 2019. С. 23–39.

4. Охтилев П.А. Алгоритмы и онтологические модели информационно-аналитической поддержки процессов создания и применения космических средств [Текст]: Диссертация на соискание ученой степени кандидата технических наук. Специальность 05.13.01 – Системный анализ, управление и обработка информации (технические системы) Дата защиты: 19.12.2019 г./ Петр Алексеевич Охтилев. Санкт-Петербург, 2019. 408 с.
5. Bruno Borlini Duarte An Ontology-based Reference Model for the Software Systems Domain with a focus on Requirements Traceability/ Bruno Borlini Duarte. – Vitória, ES, 2022- 149 p.: il.; 30 cm. (дата обращения: 10.10.2023)
6. ГОСТ Р 59798-2021. Информационные технологии. Онтологии высшего уровня (TLO). Часть 2. Базисная формальная онтология (BFO): национальный стандарт Российской Федерации: дата введения в действие: 30.04.2022 / Федеральное агентство по техническому регулированию и метрологии. Изд. официальное. Москва Российский институт стандартизации 2021. 36 с.
7. Литейное производство открытых биологических и биомедицинских онтологий. Литейный завод ОБО [просмотрено 2019-11-1], доступно по адресу <http://obofoundry.org> (OPEN BIOLOGICAL AND BIOMEDICAL ONTOLOGIES FOUNDRY. The OBO Foundry)
8. Гаркуша, Д. А. Анализ онтологических платформ. *Искусственный интеллект: теоретические аспекты, практическое применение : материалы Донецкого международного научного круглого стола*. Донецк : ФГБНУ «ИПИИ», 2023. 252 с. С. 37–40.
9. Криводубский, О. А. Математические модели в системах управления. *Искусственный интеллект: теоретические аспекты, практическое применение : материалы Донецкого международного научного круглого стола*. Донецк : ФГБНУ «ИПИИ», 2023. 252 с. С. 111–113.
10. Gruber T.R. The role of common ontology in achieving sharable, reusable knowledge bases. *Principles of Knowledge Representation and Reasoning. Proceedings of the Second International Conference*. J.A. Allen, R. Fikes, E. Sandewell – eds. Morgan Kaufmann, 1991, 601-602.
11. Загоруйко Ю.А. Современные средства формализации семантики областей знаний на основе онтологий. *Информационные и математические технологии в науке и управлении*. 2018. № 3 (11). С. 27–36. DOI:10.25729/2413-0133-2018-3-03.
12. JenaApiDocumentation [Электронный ресурс]. Режим доступа: <https://jena.apache.org/documentation/ontology/> (Дата обращения: 18.06.2023)
13. Родионцев, Н.Н. Анализ применения онтологий при разработке информационно-коммуникационных систем нефтегазовой отрасли. *Московский экономический журнал*. 2019. № 2. С. 695–699.
14. Семерханов, И.А. Методы и алгоритмы автоматизированной интеграции информационных ресурсов на основе онтологического подхода. Диссертация на соискание ученой степени кандидата технических наук. СПб.: Питер, 2014. С. 137.
15. Горшков, С.В. Введение в онтологическое моделирование. ООО «ТриниДата», 2014-2018.– 144 с.
16. Трассировка сервисов в мобильной транспортной сети. Как мы пришли к графовой БД Neo4j / Сураев Д.В. [Режим доступа] <https://habr.com/ru/post/517254/> (Дата обращения: 18.06.2023).
17. Корпоративные автоматизированные системы на основе онтологических моделей: книга рецептов – изд. ООО «ТриниДата» 2020 81с.
18. Онтологическое моделирование предприятий: методы и технологии : монография ; [отв. ред. С. В. Горшков] ; предисл. С. В. Горшкова. Екатеринбург : Изд-во Урал. ун-та, 2019. 234 с.
19. DBengines [Электронный ресурс]. Режим доступа: <https://db-engines.com/en/system/Stardog> (Дата обращения: 20.06.2023).
20. DBengines [Электронный ресурс]. Режим доступа: <https://db-engines.com/en/system/MarkLogic> (Дата обращения: 20.06.2023).
21. Stardog [Электронный ресурс]. Режим доступа: <https://docs.stardog.com/> (Дата обращения: 21.06.2023).

## References

1. Dorokhina G.V. Formalized ontologies and tasks for constructing computer information technology for digital collection, processing and analysis of data. [Artificial intelligence: theoretical aspects, practical application: materials of the Donetsk international scientific round table. – Donetsk: GU IPII, 2020. – pp. 229–234
2. Dorokhina G. V. Requirements for information technology of digital data collection, processing and analysis. [Artificial intelligence: theoretical aspects, practical application: materials of the Donetsk international scientific round table. – Donetsk: GU IPII, 2020. – №4 (19). – Pp. 4-9.
3. Muromtsev D. Industrial graphs of knowledge – the intellectual core of the digital economy [Text] / Dmitry Muromtsev, Alexey Romanov, Dmitry Volchek // Control Engineering Russia. – No. 5 (83). – October 2019. – pp. 23-39.

4. Okhtilev P.A. Algorithms and ontological models of information and analytical support for the creation and application of space facilities [Text]: Dissertation for the degree of Candidate of Technical Sciences. Specialty 05.13.01 – System analysis, management and information processing (technical systems) Date of protection: 19.12.2019/ Peter Alekseevich Okhtilev. – St. Petersburg, 2019. – 408 p.
5. Bruno Borlini Duarte An Ontology-based Reference Model for the Software Systems Domain with a focus on Requirements Traceability/ Bruno Borlini Duarte. – Vitória, ES, 2022- 149 p.: il.; 30 cm. (дата обращения: 10.10.2023)
6. SS R 59798-2021. Information technology. Higher-level Ontologies (TLO). Part 2. Basic Formal Ontology (BFO): National Standard of the Russian Federation: effective date: 30.04.2022 / Federal Agency for Technical Regulation and Metrology. – Official edition. – Moscow Russian Institute of Standardization 2021 – 36 p.
7. FOUNDRY PRODUCTION OF OPEN BIOLOGICAL AND BIOMEDICAL ONTOLOGIES. Foundry ABOUT [viewed 2019-11-1], available at <http://obofoundry.org> (OPEN BIOLOGICAL AND BIOMEDICAL ONTOLOGIES FOUNDRY. The OBO Foundry)
8. Garkusha, D. A. Analysis of ontological platforms [Artificial intelligence: theoretical aspects, practical application: materials of the Donetsk international scientific round table. – Donetsk : FGBNU "IPII", 2023. – 252 p. – pp. 37-40.
9. Krivodubsky, O. A. Mathematical models in control systems [Artificial intelligence: theoretical aspects, practical application: materials of the Donetsk international scientific round table. – Donetsk: GU IPII, 2023. – 252 p. – pp. 111-113.
10. Gruber T.R. The role of common ontology in achieving sharable, reusable knowledge bases // Principles of Knowledge Representation and Reasoning. Proceedings of the Second International Conference. J.A. Allen, R. Fikes, E. Sandewell – eds. Morgan Kaufmann, 1991, 601-602.
11. Zagorulko Yu.A. Modern means of formalizing the semantics of knowledge fields based on ontologies // Information and mathematical technologies in science and management. – 2018. № 3 (11). – P. 27 36. – DOI:10.25729/2413-0133-2018-3-03.
12. JenaApiDocumetion [Электронный ресурс]. – Режим доступа: <https://jena.apache.org/documentation/ontology/> (Дата обращения: 18.06.2023)
13. Rodiontsev, N.N. Analysis of the application of ontologies in the development of information and communication systems in the oil and gas industry / N.N. Rodiontsev. // Moscow Economic Journal. – 2019. – No. 2.– С. 695-699.
14. Semerkhanov, I.A. Methods and algorithms of automated integration of information resources based on an ontological approach / I.A. Semerkhanov // Dissertation for the degree of Candidate of Technical Sciences. – St. Petersburg,: Peter, 2014. – p. 137.
15. Gorshkov, S.V. Introduction to ontological modeling/ S.V. Gorshkov – LLC "TriniData", 2014-2018.– 144 p.
16. . Tracing services in the mobile transport network. How we came to the Neo4j graph database / Suraev D.V. [Access mode] <https://habr.com/ru/post/517254/> (Date of access: 06/18/2023).
17. Corporate automated systems based on ontological models: recipe book – ed. LLC "TriniData" – 2020 – 81с.
18. . Ontological modeling of enterprises: methods and technologies: Monograph ; [ed. by S. V. Gorshkov]; preface by S. V. Gorshkov. Yekaterinburg : Ural Publishing House. unita, 2019. - 234 p.
19. Dbengines [Electronic resource]. - Access mode: <https://db-engines.com/en/system/Stardog> (date of application: 06/20/2023).
20. Dbengine [Electronic resource]. – Access mode: <https://db-engines.com/en/system/MarkLogic> (Date of application: 06/20/2023).
21. Stardog [Electronic resource]. – Access mode: <https://docs.stardog.com/> (Date of access: 06/21/2023).

## RESUME

*D. A. Garkusha*

### *Functional Features Of Implemented Ontological Platforms*

At the moment, the existing description of ontology implementation approaches does not contain concise, understandable and comprehensive information about the choice of using an ontology platform tool and their compatibility with various stacks. Each task, the purpose of which is the creation and implementation of an ontology in information systems, software systems and other software products, necessitates an analysis and highlighting the purpose of constructing an ontology to understand what task must be solved.

To work effectively, it is necessary to analyze the selected platforms and evaluate them according to various criteria. The criteria should be based on the software's accessibility, ease of use, and support for the developer to work across multiple platforms. The work involves improving the assessment of the suitability of using a particular ontological platform in a variety of tasks.

## РЕЗЮМЕ

*Д. А. Гаркуша*

### *Функциональные особенности реализованных онтологических платформ*

На данный момент существующее описание подходов реализации онтологии не несет в себе краткую, понятную и исчерпывающую информацию о выборе использования инструмента онтологической платформ и совместимость их с различными стеками. Каждая задача целью которой является создание и внедрение онтологии в информационные системы, программные комплексы и другие программные продукты, вызывает необходимость проводить анализ и выделять цель построения онтологии для понимания какая задача должна быть решена.

Для эффективной работы необходимо провести анализ выбранных платформ и оценить по различным критериям. Критерии должны основаться на доступности программного обеспечения, удобства ее использования и поддержки работы в различных платформах для разработчика. Работа предполагает улучшение оценки пригодности в использовании той или иной онтологической платформы в самых различных задачах.

**Гаркуша Дмитрий Алексеевич** – аспирант отдела системного анализа и интеллектуальных систем Федерального государственного бюджетного научного учреждения «Институт проблем искусственного интеллекта», *Область научных интересов:* онтологии, системы искусственного интеллекта; эл. почта *Dmitrii.garkusha.97@mail.ru*, адрес: 283087, г. Донецк, Гвардейский переулок, 30, телефон +7949381045.

Статья поступила в редакцию 14.11.2023.