

УДК 004.891.3

doi:10.18720/SPBPU/2/id24-162

*Бурячек Ирина Юрьевна*<sup>1</sup>,

инженер;

*Хохловский Владимир Николаевич*<sup>2</sup>,

доцент, канд. техн. наук, доцент

## **ТЕХНОЛОГИЯ ФОРМИРОВАНИЯ БАЗЫ ЗНАНИЙ НА ОСНОВЕ КОНЦЕПЦИИ СЕМАНТИЧЕСКОЙ ИНТЕРОПЕРАБЕЛЬНОСТИ**

<sup>1,2</sup> Россия, Санкт-Петербург, Санкт-Петербургский политехнический  
университет Петра Великого;

<sup>1</sup>buryachek\_iyu@spbstu.ru, <sup>2</sup>hohlovskij\_vn@spbstu.ru

**Аннотация.** В работе исследуются концепция и способы реализации семантической интероперабельности в области автоматизации технологических процессов в энергетике. Проведен анализ технологий формирования баз знаний, показаны преимущества и перспективы использования данной технологии. Описывается практический подход к созданию системы, взаимодействующей со SCADA-приложениями разных вендоров, рассматривается реализация модуля семантической интероперабельности.

**Ключевые слова:** семантическая интероперабельность, нотация RDF, онтология SAREF4ENERGY, базы знаний, SCADA-системы, программирование C#.

*Irina Yu. Buryachek*<sup>1</sup>,

Engineer;

*Vladimir N. Khokhlovskiy*<sup>2</sup>,

Candidate of Technical Sciences (PhD), Associate Professor

## **SEMANTIC-INTEROPERABILITY DRIVEN CONCEPT FOR FORMING A KNOWLEDGE BASE**

<sup>1,2</sup> Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia;

<sup>1</sup>buryachek\_iyu@spbstu.ru, <sup>2</sup>hohlovskij\_vn@spbstu.ru

**Abstract.** The work explores the concept and methods of implementing semantic interoperability in the field of automation of processes in energy industry. An analysis of technologies for forming knowledge bases is shown, as well as the advantages and prospects for using this technology. A practical approach is considered to build the system exchanging data with SCADA-applications of different vendors. Implementation the semantic interoperability module is described.

**Keywords:** semantic interoperability, RDF notation, SAREF4ENERGY ontology, knowledge bases, SCADA systems, C# programming.

## **Введение**

В современном информационном обществе, характеризующемся огромными объемами данных и широким разнообразием информационных ресурсов, создание и эффективное управление базами знаний являются одними из ключевых задач. Эффективное использование знаний становится неотъемлемой частью успешного функционирования организаций и обеспечения качественных информационных сервисов для конечных пользователей [1]. В данном контексте концепция семантической интероперабельности выделяется как фундаментальный подход, направленный на совместное использование, интеграцию и обмен знаний между различными информационными системами. Семантическая интероперабельность, опираясь на стандартизированные семантические модели и онтологии, позволяет преодолевать барьеры, возникающие из-за различий в терминологии, структуре и семантике данных.

Цель данной статьи заключается в рассмотрении и анализе технологии формирования базы знаний на основе концепции семантической интероперабельности применительно к одной предметной области — автоматизации технологических процессов в энергетике. Рассматриваются теоретические аспекты и практические реализации данной концепции для организации взаимодействия SCADA-приложений разных вендоров, а также особенности, преимущества и перспективы, связанные с использованием данной технологии. В работе представлена структура и описано взаимодействие компонентов системы, предназначенной для создания базы знаний на основе концепции семантической интероперабельности. Приложение реализовано на языке C# с применением библиотеки dotNetRDF [2] для обработки семантических аспектов системы.

### **1. Назначение и структура системы**

Семантическая интероперабельность представляет собой важный аспект современных информационных технологий, который призван обеспечить эффективное взаимодействие и обмен данными между различными информационными системами, обладающими различной структурой данных и предметной терминологией. В контексте систем управления технологическими процессами семантическая интероперабельность становится важным компонентом для обеспечения надежной связи и интеграции между двумя или более SCADA-системами.

Основная цель семантической интероперабельности в данном контексте заключается в унификации семантики данных, которые собираются, обрабатываются и передаются между SCADA-системами. Это позволяет устранить различия в терминологии, структуре и форматах данных, которые могут возникнуть из-за использования разными вендорами различных правил использования и стандартов SCADA. Применение семантической интероперабельности обеспечивает следующие преимущества:

1. Единая семантика данных: семантическая интероперабельность позволяет обеспечить единое понимание данных между системами. Например, если в одной SCADA-системе используется термин «теплопроизводительность котельного оборудования», а в другой — «генерация тепла котлоагрегатами», то при реализации семантической интероперабельности можно связать эти термины и задать, что они означают одно и то же.

2. Интеграция разнородных источников данных: SCADA-системы могут собирать данные из разных источников, таких как датчики, устройства, базы данных и другие системы. Средства реализации семантической интероперабельности позволяют объединить их в единое хранилище знаний, где данные могут быть связаны и проанализированы совместно.

3. Улучшение принятия решений: механизмы исполнения семантической интероперабельности упрощают поиск, фильтрацию и анализ данных, что способствует более точному и быстрому принятию решений в реальном времени.

Семантическая интероперабельность в контексте интеграции данных из нескольких SCADA-систем достигается через разработку специального приложения, которое выполняет несколько ключевых функций. Приложение собирает данные из различных SCADA-систем, которые могут иметь разнообразные форматы, терминологию и структуру данных. Собранные данные при помощи выбранной онтологии переводятся в унифицированный семантический формат, который обеспечивает единое понимание терминов и отношений между данными. Затем данные интегрируются в единое хранилище знаний, где они становятся доступными для анализа и дополнительной обработки. Это позволяет объединить информацию из разных SCADA-систем в единое пространство, где данные связываются и могут быть проанализированы с учетом их семантической значимости (см. рис. 1).

База данных формируется с использованием приложения, разработанного на языке программирования C#. Приложение реализует структуру базы на модели данных, применяя стандарт RDF. В этой модели данные представлены в формате триплетов: субъект, предикат и объект.

В данном приложении используется онтология SAREF4ENERGY [3], которая представляет собой формально описанную семантическую модель, ориентированную для использования в области энергетики. Она определяет структуру и семантику данных, связанных с управлением и мониторингом энергетических систем, включающую информацию об энергопотреблении, электропотребителях, устройствах и других аспектах энергетической инфраструктуры (см. рис. 2 [4]).

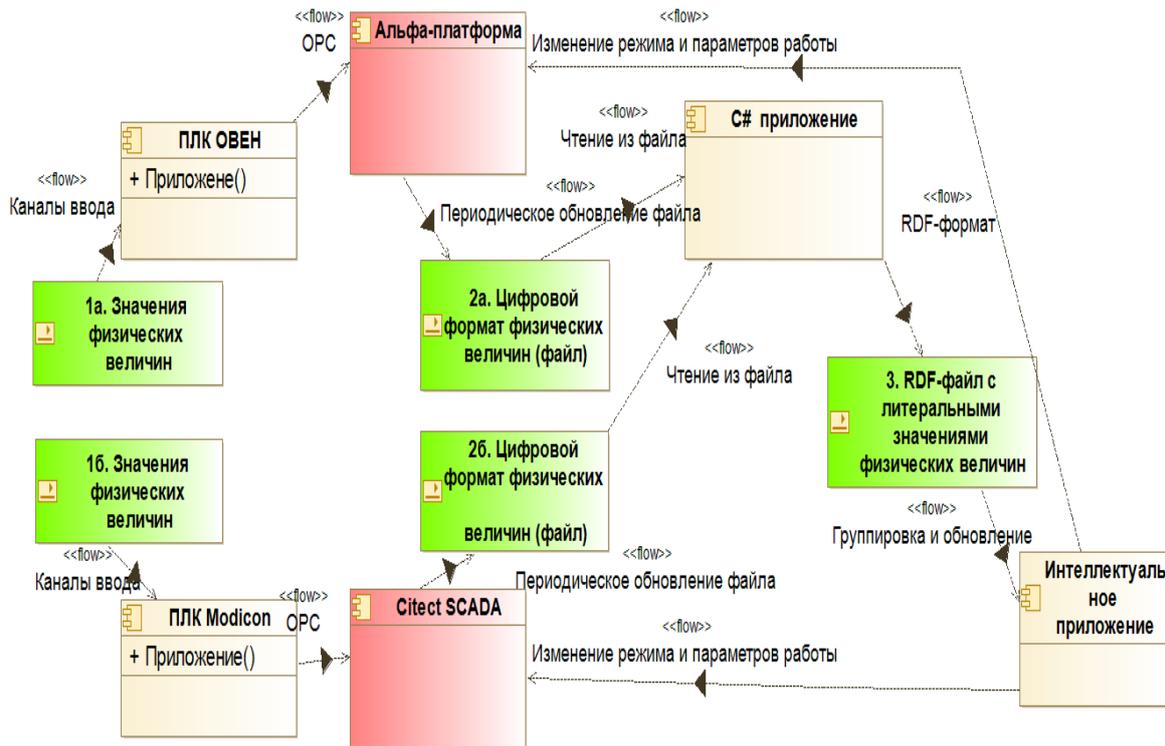


Рис. 1. Взаимодействие компонентов системы

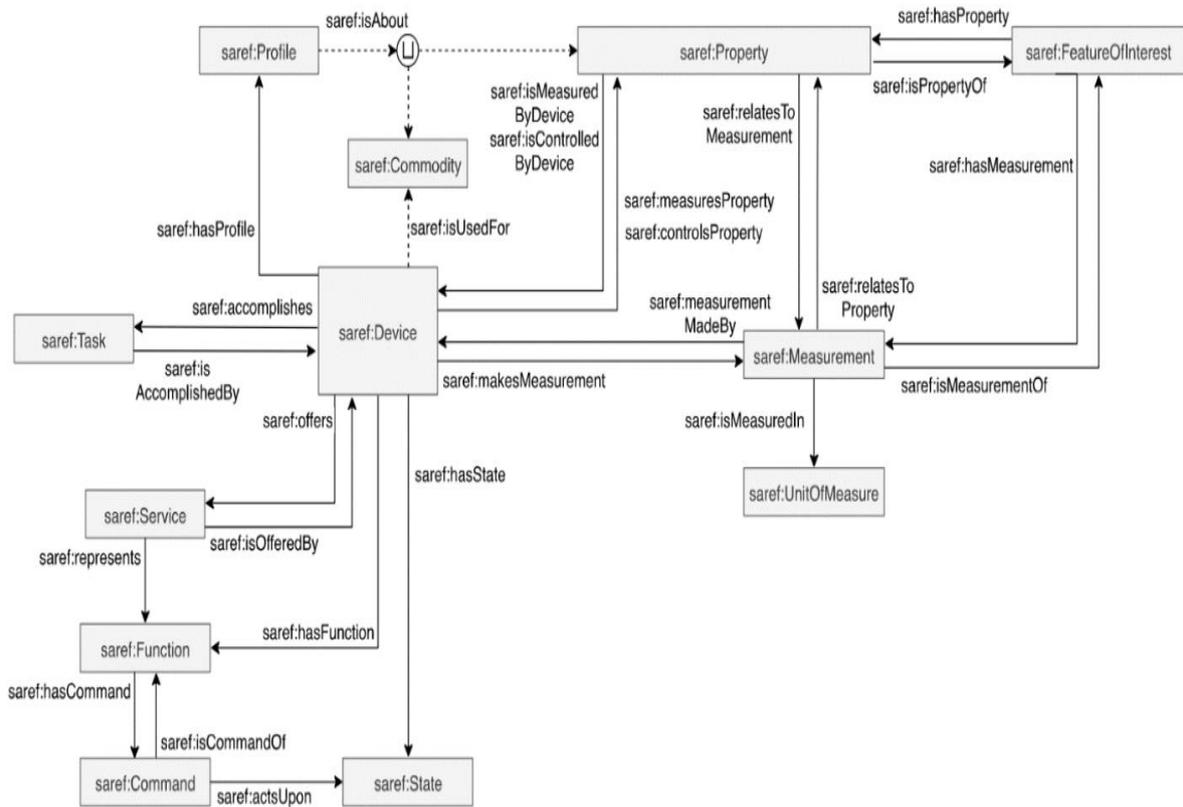


Рис. 2. Основные классы SAREF и их взаимосвязи

Приложение осуществляет организацию данных в базе в соответствии с определенными онтологическими классами и свойствами. Это позволяет создать унифицированное хранилище данных — базу знаний, где каждый триплет представляет собой семантически связанные сущности, описанные в соответствии с онтологией (на рисунке 1 база знаний представлена в формате RDF как объект данных с номером «3»). Пояснить сказанное можно с помощью отображения текстового описания, представленного в таблице 1, в виде графа знаний, показанного на рисунке 3.

Таблица 1

Триплеты базы знаний в виде текста

Субъект: "счетчик1" Предикат: "имеет_параметр" Объект: "параметр1"	Субъект: "параметр1" Предикат: "имеет_название" Объект: "Сила_тока"	Субъект: "параметр1" Предикат: "имеет_описание" Объект: "Измерение силы тока"
--	---	---



Рис. 3. Триплеты в виде графа знаний

## 2. Последовательность шагов по формированию базы знаний

Созданное приложение выполняет процесс обработки входных данных и их последующее преобразование в формат RDF с применением онтологии SAREF4ENERGY. Входные данные представлены в виде текстового файла, содержащего информацию, полученную от SCADA-систем. Каждая строка в данном текстовом файле представляет собой отдельную запись данных, в которой значения разделены запятыми. Формат входного файла был выбран в виде текстового файла (формат txt) из-за его удобства использования и универсальности.

Процесс функционирования приложения может быть изложен следующим образом:

1. Инициализация программы.
2. Установка путей к исходному и целевому файлам.

3. Формирование графа для взаимодействия с данными в формате RDF.

4. Установка префиксов и пространств имен для модели SAREF4ENERGY (граф).

5. Открытие исходного файла для чтения.

6. Построчное считывание данных из исходного файла.

7. Разбиение каждой строки на элементы, включающие значение параметра электроэнергии, имя параметра, единицу измерения, устройство и временную метку.

8. Формирование имен (QName) для класса устройства и параметра.

9. Создание узлов-устройств и узлов-параметров в графе.

10. Генерация триплетов для хранения информации о значении параметра, его единицах измерения, ассоциации с конкретным устройством и временной меткой.

11. Занесение сформированных триплетов в граф (БЗ).

12. Повторение операций с 6 по 11 до окончания исходного файла.

13. Сохранение полученного графа в целевом файле с помощью Notation3Writer.

14. Вывод уведомления об успешном завершении процесса преобразования данных.

15. Чтение содержимого целевого файла и загрузка данных в новый граф (БЗ).

16. Проведение SPARQL-запросов для извлечения информации из графа (БЗ).

17. Вывод заголовков таблицы, включая атрибуты «Устройство», «Параметр», «Значение», «Единицы измерения» и «Временная метка», в консоли.

18. Отображение разделительной линии в консоли.

19. Вывод информации для каждой строки запроса, включая «Устройство», «Параметр», «Значение», «Единицы измерения» и «Временную метку» в консоли.

20. Повторение операций с 16 по 19 для каждого нового запроса.

21. Завершение выполнения программы.

Было разработано приложение, использующее язык программирования C# с библиотекой dotNetRDF, с целью преобразования данных, полученных от SCADA-систем Альфа-платформа и Citect SCADA, в базу данных на основании правил и формата RDF. Приложение обеспечивает средство автоматизированного переноса и структурирования информа-

ции из системы мониторинга SCADA в базу данных, предназначенную для агрегирования и последующего анализа показателей качества электроэнергии.

### **Заключение**

В работе были рассмотрены преимущества применения семантической интероперабельности в контексте интеграции данных из нескольких SCADA-систем. База данных формируется с использованием приложения, разработанного на языке программирования C#, используется онтология SAREF4ENERGY. Был разработан алгоритм построения приложения. На основании вышеизложенного можно сделать вывод, что приложение обеспечивает семантическую совместимость SCADA-систем и, как результат — интеграцию разнородных и разноформатных данных. База знаний, полученная в результате работы приложения, может использоваться для прогнозной аналитики, в системах поддержки принятия решений и в других интеллектуальных приложениях.

### **Благодарности**

Работа выполнена в рамках гранта РФФИ № 23-29-00551 от 13.01.2023 г.

### **Список источников**

1. Daniele L.M., Punter M., Brewster C., García Castro R., Poveda M., Fernández A. A SAREF extension for semantic interoperability in the industry and manufacturing domain // In: M. Zelm, F.-W. Jaekel, G. Doumeingts, M. Wollschlaeger (eds.), *Enterprise Interoperability: Smart Services and Business Impact of Enterprise Interoperability*. – John Wiley & Sons Inc., 2018. – Pp. 201–207. – URL: <https://doi.org/10.1002/9781119564034.ch25>.

2. Шкодырев В.П., Хохловский В.Н., Бурячек И.Ю. Семантическая интероперабельность SCADA-приложений / В сборнике: XVI Всероссийская мультиконференция по проблемам управления (МКПУ-2023). Материалы мультиконференции. В 4-х томах / Редколлегия: И.А. Каляев, В.Г. Пешехонов, С.Ю. Желтов [и др.]. – Волгоград, 2023. – С. 96–99.

3. Bauer M., Baqa H., Bilbao S., Corchero A., Daniele L., Esnaola-Gonzalez I., Fernandez I., Frånberg Ö., García Castro R., Girod-Genet M., Guillemain P., Gyrard A., El Kaed Ch., Kung A., Lee J., Lefrançois M., Li W., Raggett D., Wetterwald M. Semantic IoT solutions – a developer perspective [Electronic Source]. – October 2019. – URL: [https://www.researchgate.net/publication/336679022\\_Semantic\\_IoT\\_Solutions\\_-\\_A\\_Developer\\_Perspective](https://www.researchgate.net/publication/336679022_Semantic_IoT_Solutions_-_A_Developer_Perspective). – DOI: 10.13140/RG.2.2.16339.53286.

4. da Rocha H., Espirito-Santo A., Abrishambaf R. Semantic interoperability in the Industry 4.0 using the IEEE 1451 standard // IECON 2020 The 46th Annual Conference of the IEEE Industrial Electronics Society. – IEEE, 2020. – Pp. 5243–5248.