

Решение задачи сбора и подготовки данных и знаний, поступающих из нескольких информационных систем, для их интеллектуальной обработки (на примере отчетности регистратора)

И.А. Спицина, К.А. Аксенов

Уральский федеральный университет имени первого Президента Б.Н. Ельцина

Аннотация: В работе рассмотрена задача сбора и подготовки данных, поступающих из нескольких информационных систем на примере автоматизации отчетности регистратора. Для описания предметной области могут быть использованы языки OWL, XML, XBRL и семантических сетей. Подготовлен набор критериев для анализа и выбора наиболее подходящего языка представления знаний с целью сбора данных на примере финансовой отчетности. Описаны результаты разработки сервиса и показано применение формата XBRL. При разработке сервиса использовался мультиагентный подход при моделировании и проектировании информационных систем.

Ключевые слова: интеллектуальная обработка данных, модель предметной области, форматы данных, XBRL, бизнес-процесс, сервис, интеграция данных.

Введение

Целью работы является решение задачи сравнительного анализа различных форматов данных с целью использования их для сбора и анализа данных отчетности, поступающих из большого количества источников, и их применения в интеллектуальной обработке данных и знаний. По сравнению с задачей обработки данных процессов промышленного предприятия, описанной в работе [1] (на примере разработки автоматизированной системы выпуска металлургической продукции, в части модуля подготовки данных [2-3], ориентированную на работу с гетерогенными данными [4]), не предполагалось решения задач вывода и поиска на знаниях, а обмен информацией между различными гетерогенными распределенными информационными системами осуществлялся в форматах обмена данными. Вопросам интеграции информационных систем на основе агентно-онтологического подхода посвящена работа [5].

При описании предметной области (ПрО) для последующего сбора различных статистических данных и их машинной обработке необходим

язык, который позволил бы описать ее онтологию и факты (какие-то количественные показатели) в машиночитаемом формате [6, 7].

Для описания семантики предметной области используются различные языки, среди которых можно выделить: семантические сети, OWL (Web Ontology Language) - используется для создания онтологий и описания семантики данных, XBRL (англ. eXtensible Business Reporting Language «расширяемый язык деловой отчетности») и XML (eXtensible Markup Language) - применяется для описания сложных структур данных и их иерархий; используется в системах, где важна семантика данных.

Сравнение OWL, XML, XBRL и семантических сетей

Диалекты DL и Full языка OWL предоставляют формальные правила и структуры для описания понятий (концептов) и их взаимосвязей. Знания о предметной области знания описываются с помощью классов, свойств и отношений между ними. OWL поддерживает автоматизированный вывод, что позволяет системам делать заключения и решать задачи поиска на основе заданных знаний и входных данных, а также проверять согласованность и полноту представленных данных. Свойства, которые устанавливают отношения от индивидов (понятий) к значениям данных, могут рассматриваться как факты [8].

XBRL можно рассматривать как язык представления знаний в определённом контексте, хотя его основной фокус — это представление и обмен финансовой и деловой информацией. XBRL поддерживает ограниченный набор вычислительных методов, необходимых для расчета и обработки передаваемых производственных параметров, параметров бизнес-процессов и/или финансовых данных. Можно выделить несколько аспектов, которые позволяют рассматривать XBRL в контексте представления и обработки знаний:

1. Структурированное представление информации: XBRL использует XML для создания документов, в которых данные описаны с помощью элементов и атрибутов.

2. Таксономии: XBRL включает создание таксономий, которые определяют структуру и семантику данных. Таксономии в XBRL описывают, как финансовые данные должны быть организованы и интерпретированы, что похоже на создание онтологий для описания знаний в других областях.

3. Использование метаданных: в XBRL метаданные (например, определения элементов и их взаимосвязи) помогают в интерпретации данных. Это обеспечивает дополнительный контекст и понимание представленных данных.

4. Анализ и сопоставление данных: XBRL облегчает анализ и сопоставление данных, что позволяет пользователям извлекать знания и выводы из структурированной информации.

Однако, в отличие от специализированных языков представления знаний, таких как семантические сети, XBRL не предназначен для моделирования сложных семантических отношений или создания обширных онтологий. Рассмотрим, в чем их отличия.

1. Цель и область применения. XBRL применяется для унификации финансовой отчетности. Семантические сети предназначены для создания онтологий, представления сложных отношений между концепциями и поддержки семантического поиска и обработки.

2. Структура и формат. XBRL использует XML. Семантические сети представлены в виде графов, где узлы представляют некоторые сущности предметной области, а связи между узлами — отношения между ними.

3. Гибкость и расширяемость. Семантические сети позволяют легко добавлять новые концепты и отношения, а также изменять существующие.

Для расширения XBRL необходимо изменения или создания новых таксономий.

В таблице № 1 приведено сравнение способов описания информации для сбора и анализа отчетности, и их применения интеллектуальной обработке данных и знаний.

Таблица № 1

Сравнение языков и подходов для сбора и анализа отчетности, и их применения интеллектуальной обработке данных и знаний

	XML	XBRL	OWL DL \ Full	Семантическая сеть
1	2	3	4	5
Тип	Язык разметки	Язык разметки	Язык описания онтологий	Графовая структура
Структура	Элементы, атрибуты.	Таксономии на основе XML	Классы и свойства	Узлы и ребра
Наличие стандарта, формата данных	Да (XML)	Да (XBRL)	Да (OWL)	В зависимости от задач и специфики Про.
Возможность автоматизированной \ автоматической обработки большого объема данных, поступающих из различных источников	Да	Да	Да	Нет (для каждой Про требуется уникальное решение задачи)
Возможность хранения фактов	Да	Да	Да	Нет
Встроенная аналитика (расчет параметров)	Нет	Да, наиболее простые вычисления	Нет	Нет

1	2	3	4	5
Описание знаний	Нет	Для очень узких Про	Да	Да
Области применения	Системы, где важна семантика данных	Финансовая отчетность, производственные параметры, параметры бизнес-процессов	Семантический веб	Системы управления знаниями, искусственный интеллект

Применение XBRL в автоматизации процесса обработки финансовой отчетности

Банк России с 2018 года принимает отчетность профессиональных участников рынка ценных бумаг в формате XBRL. Это открытый стандарт обмена деловой информацией, который получил широкое распространение по всему миру. В основе этого стандарта лежит расширяемый язык разметки XML. В отличие от XML с иерархической структурой данных, XBRL также включает себя следующие составляющие [9]:

1. Схемы таксономии – это xsd-файлы, которые определяют концепты в сфере бизнеса (предметной области), по которым формируется отчетность.
2. База ссылок – xml-файлы, которые содержат различную информацию об элементах, описанных в схемах XBRL, отвечает за семантику.
 - 2.1. Сети ресурсов (ярлыки и ссылки) описывает на понятном для человека языке суть концептов и предоставляет ссылки на нормативные документы.
 - 2.2. Сети взаимосвязей описывают зависимости между концептами.
3. Отчеты.
 - 3.1. Ссылки на таксономию для определения смысла фактов.

- 3.2. Контексты определяют сущность, к которым применим факт, период времени, в течение которого факт имеет значение, и необязательный сценарий – дополнительная информация.
- 3.3. Единицы измерения фактов.
- 3.4. Факты – значения концептов. Каждый из фактов в пределах отчета связан с соответствующим концептом в таксономии.
- 3.5. Комментарии.

Использование XBRL дает следующие преимущества для Банка России:

- ✓ упрощен поиск информации по конкретным источникам;
- ✓ упрощено построение консолидированных отчетов;
- ✓ упрощена проверка полученной от профучастников информации.

В следующем разделе описывается пример применения XBRL для автоматизации процессов обмена данными и знаниями регистратора, взаимодействующего с внешними участниками.

Описание сервиса подготовки отчетности регистратора в формате XBRL

Информация для отчетов нефинансовых институтов фондового рынка, к которым относится регистратор, включает в себя показатели бухгалтерской, надзорной и статистической отчетности. В Центральный Банк России необходимо отправлять пакет отчетности в виде единого файла. Таким образом, требовалось объединять данные из бухгалтерской и системы ведения реестра (СВР), причем объем данных не позволял формировать отчеты вручную. На момент внедрения новой формы отчетности на рынке было программное обеспечение «Фабрика XBRL», в которой была реализована возможность интеграции с бухгалтерскими системами и некоторая валидация. Также была реализована возможность загрузки файлов с надзорной отчетностью в формате XBRL. Поэтому было принято решение интеграция с СВР на уровне файлов. На рис. 1 представлена диаграмма

контекста С4. Она показывает взаимодействие сервиса с пользователями и внешним приложением «Фабрика XBRL».



Рис 1. – Диаграмма контекста С4

Сервис XBRL состоит из клиентского настольного и серверного приложений, которое взаимодействует с базой данных. Диаграмма контейнеров С4 представлена на рис. 2. Она показывает архитектуру сервиса XBRL и основные технологические решения. Поскольку с ним работает один сотрудник, занимающийся отчетностью в офисе компании на основе конфиденциальной информации, то была выбрана архитектура настольного приложения.

При разработке сервиса использовался мультиагентный подход при моделировании и проектировании информационных систем [10]. Сервис XBRL подготовки отчетности можно рассматривать как агента концептуальной модели мультиагентных процессов преобразования ресурсов (МППР) [11] в контексте предметной области рынка ЦБ, взаимодействующего с сотрудником регистратора.

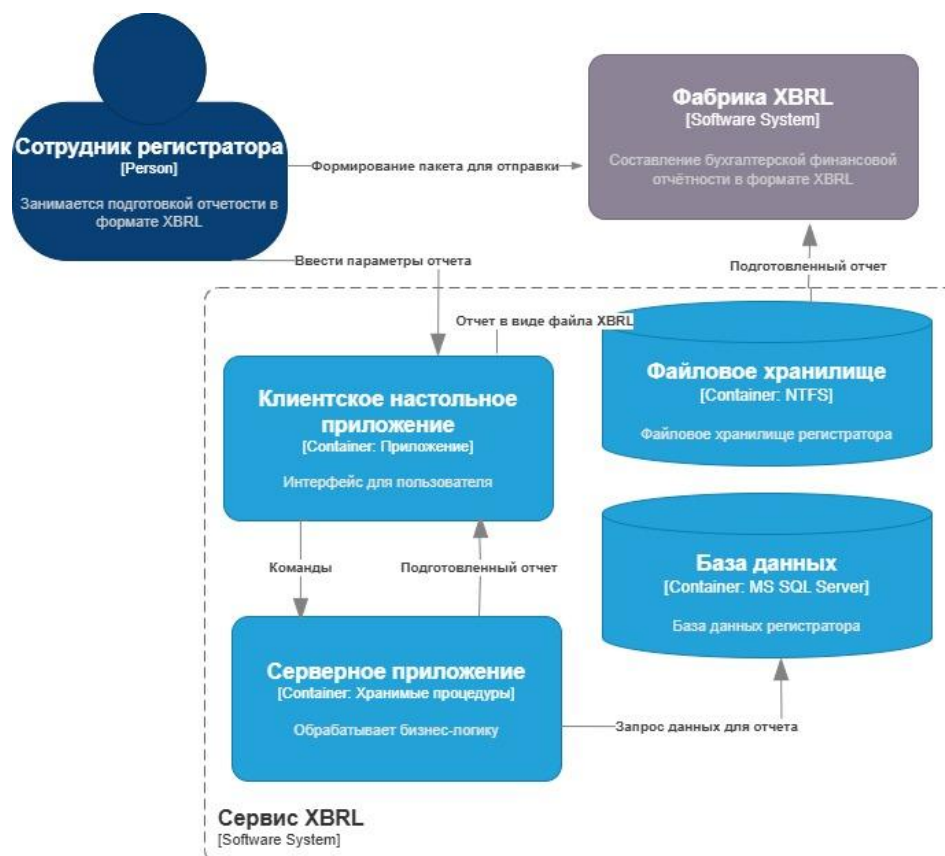


Рис. 2. – Диаграмма контейнеров С4

Пример описания контекста в формате XBRL:

```
<xbrli:context id="C1-01-00177-A">
  <xbrli:entity>
    <xbrli:identifier
      scheme="http://www.cbr.ru">1026605227923</xbrli:identifier>
    </xbrli:entity>
    <xbrli:period>
      <xbrli:instant>2024-08-31</xbrli:instant>
    </xbrli:period>
    <xbrli:scenario>
      <xbrldi:typedMember dimension="dim-int:ID_CzennojBumagiTaxis">
        <dim-int:ID_CzennojBumagi_TypedName>1-01-00177-
A_RU0009191967_ESVXFR_УРАЛПРЕДПРИЯТИЕ</dim-
int:ID_CzennojBumagi_TypedName>
```


</xbrldi:typedMember>

</xbrli:scenario>

</xbrli:context>

Пример описания факта (описание типа ценных бумаг):

```
<nfo_dic:TipCZenBumVidFIEnumerator contextRef="C1-01-00177-A">mem-  
int:SHS3_akcziiprochixrezidentov_obyknovennyeMember</nfo_dic:TipCZenBu  
mVidFIEnumerator>
```

Разработанный сервис успешно применяется регистратором для сдачи отчетности в Центральный банк.

Заключение

В работе выделены критерии сравнения языков описания предметных областей и проведен их сравнительный анализ. Для описания предметной области финансовой отчетности профессиональных участников рынка ценных бумаг используется язык XBRL. Для решения задачи сбора и подготовки данных и знаний, поступающих из нескольких информационных систем с целью подготовки отчетности регистратора, был разработан сервис сбора надзорной и статистической информации.

Литература

1. Сысолетин Е.Г., Аксенов К.А., Круглов А.В. Интеграция гетерогенных информационных систем современного промышленного предприятия // Современные проблемы науки и образования, 2015. № 1. URL: science-education.ru/121-19030.

2. Ponomareva, O., Porshnev, S., Borodin, A., & Mirvoda, S. (2021). Date preparation module of automated metallurgical products production system. IOP Conference Series: Materials Science and Engineering, 1047(1), [012003]. URL: doi.org/10.1088/1757-899X/1047/1/012003.



3. Бородин, А. М., Мирвода, С. Г., & Поршневу, С. В. Метод управления нормативно справочной информацией (НСИ) в автономных информационных системах // Инженерный вестник Дона, 2013, №4. URL: ivdon.ru/ru/magazine/archive/n4y2013/2036.

4. Porshnev, S., Borodin, A., Ponomareva, O., Mirvoda, S., & Chernova, O. (2021). The development of a heterogeneous MP data model based on the ontological approach. *Symmetry*, 13(5), [813]. URL: doi.org/10.3390/sym13050813.

5. Skobelev, P., Zhilyaev, A., Larukhin, V., Grachev, S., Simonova, E. Ontology-based open multi-agent systems for adaptive resource management // In Proceedings of the 12th International Conference on Agents and Artificial Intelligence, Valletta, Malta, 2020. P. 127–135. URL: doi:10.5220/0008896301270135.

6. Колесникова Д.С., Верещагина Е.А., Гуляев В.Е. Построение онтологической модели для предметной области «Информационная безопасность» // Инженерный вестник Дона, 2023, №7. URL: ivdon.ru/uploads/article/pdf/IVD_33__6y23_kolesnikova_vereshhagina_guljaev.pdf_63ff16e0d3.pdf.

7. Колесникова Д.С., Верещагин Е.А. Применение онтологий в обучающих системах // Инженерный вестник Дона, 2023, №6. URL: ivdon.ru/uploads/article/pdf/IVD_10__6y23_kolesnikova_vereshhagina.pdf_e3b1a8c17d.pdf.

8. Буракова Е. Е., Боргест Н. М., Коровин М. Д. Языки описания онтологий для технических предметных областей // Вестник Самарского государственного аэрокосмического университета, 2014. №3(45). С. 144-154.

9. Fourny Ghislain The Xbrl Book: Simple, Precise, Technical // Createspace Independently published, 2023. 453 p.

10. Aksyonov K, Antonova A, Aksyonova O. A Method for Planning the Supply of Petroleum Products to Filling Stations Based on Multi-agent Resource Conversion Processes. Lecture Notes in Electrical Engineering: Book Series. Vol. 986. Springer. 2023. pp. 176-185. Chapter 18. (Advances in Automation IV). doi: 10.1007/978-3-031-22311-2_18. URL: link.springer.com/chapter/10.1007/978-3-031-22311-2_18.

11. Спицина И. А., Аксенов К.А. Метод поддержки принятия решений при разработке информационных систем на основе мультиагентного подхода: Екатеринбург: УрГПУ, 2018. 156 с.

References

1. Sysoletin E.G., Aksyonov K.A., Kruglov A.V. Sovremennyye problemy nauki i obrazovaniya, 2015. № 1. URL: science-education.ru/121-19030.

2. Ponomareva, O., Porshnev, S., Borodin, A., & Mirvoda, S. (2021). Data preparation module of automated metallurgical products production system. IOP Conference Series: Materials Science and Engineering, 1047(1), [012003]. URL: doi.org/10.1088/1757-899X/1047/1/012003.

3. Borodin, A. M., Mirvoda, S. G., & Porshnev, S. V. Inzhenernyj vestnik Dona, 2013, №4. URL: ivdon.ru/ru/magazine/archive/n4y2013/2036.

4. Porshnev, S., Borodin, A., Ponomareva, O., Mirvoda, S., & Chernova, O. (2021). The development of a heterogeneous MP data model based on the ontological approach. Symmetry, 13(5), [813]. URL: doi.org/10.3390/sym13050813.

5. Skobelev, P., Zhilyaev, A., Larukhin, V., Grachev, S., Simonova, E. Proceedings of the 12th International Conference on Agents and Artificial Intelligence, Valletta, Malta, 2020. pp. 127–135. URL: [doi:10.5220/0008896301270135](https://doi.org/10.5220/0008896301270135).

6. Kolesnikova D.S., Vereshchagina E.A., Gulyayev V.E. Inzhenernyj vestnik Dona, 2023, №7. URL:



ivdon.ru/uploads/article/pdf/IVD_33__6y23_kolesnikova_vereshhagina_guljaev.pdf_63ff16e0d3.pdf.

7. Kolesnikova D.S., Vereshchagina E.A. Inzhenernyj vestnik Dona, 2023, №6. URL:

ivdon.ru/uploads/article/pdf/IVD_10__6y23_kolesnikova_vereshhagina.pdf_e3b1a8c17d.pdf.

8. Burakova E. E., Borgest N. M., Korovin M. D. Vestnik Samarskogo gosudarstvennogo aerokosmicheskogo universiteta, 2014. №3(45). pp. 144-154.

9. Fourny Ghislain The Xbrl Book: Simple, Precise, Technical. Createspace Independently published, 2023. 453 p.

10. Aksyonov K, Antonova A, Aksyonova O. A Method for Planning the Supply of Petroleum Products to Filling Stations Based on Multi-agent Resource Conversion Processes. Lecture Notes in Electrical Engineering: Book Series. Vol. 986. Springer. 2023. P. 176-185. Chapter 18. (Advances in Automation IV). doi: 10.1007/978-3-031-22311-2_18. URL: link.springer.com/chapter/10.1007/978-3-031-22311-2_18.

11. Spitsina I.A, Aksyonov K.A. Metod podderzhki prinyatiya resheniy pri razrabotke informatsionnykh sistem na osnove mul'tiagentnogo podkhoda [Method of decision support for software development based on multi agent approach]: Yekaterinburg: UrGPU, 2018. 156 p.

Дата поступления: 14.09.2024

Дата публикации: 29.10.2024