

# PowerBI Ontology Extractor

Инструмент для извлечения и управления  
онтологиями из моделей данных Power BI

Автор: vladspace\_ubuntu24

Дата: 04 February 2026

## Аннотация

---

В данной статье представлен инструмент PowerBI Ontology Extractor — программное решение для автоматического извлечения семантических метаданных из моделей данных Power BI и их преобразования в формат OWL-онтологий. Инструмент решает задачу формализации бизнес-знаний, заложенных в аналитических моделях, делая их доступными для семантического анализа, валидации и интеграции с системами искусственного интеллекта.

## 1. Введение

---

### 1.1 Проблематика

Современные организации накапливают значительный объём бизнес-знаний в своих аналитических платформах. Power BI, будучи одной из ведущих платформ бизнес-аналитики, содержит в своих моделях данных ценную семантическую информацию: структуру сущностей, связи между ними, бизнес-метрики и правила расчёта показателей.

Однако эта информация остаётся «замкнутой» внутри инструмента и недоступна для:

- Семантического анализа и рассуждений
- Интеграции с системами управления знаниями
- Валидации действий AI-агентов
- Автоматизированной проверки соответствия бизнес-правилам

### 1.2 Предлагаемое решение

PowerBI Ontology Extractor автоматически извлекает метаданные из различных форматов Power BI моделей (PBIP/TMDL, BIM, VPAX) и генерирует формальные OWL-онтологии, которые могут быть использованы в системах семантического веба, knowledge graphs и AI-приложениях.

## 2. Архитектура системы

---

### 2.1 Ключевые компоненты

Система состоит из четырёх основных компонентов:

### **Power BI Model Parser**

Модуль парсинга поддерживает три формата моделей Power BI: PBIP/TMDL (новый формат с разделением на файлы), BIM (JSON-представление табличной модели) и VPAX (XML-экспорт из DAX Studio). Парсер извлекает таблицы, колонки, меры, связи и иерархии из модели.

### **OWL Ontology Generator**

Генератор преобразует извлечённые метаданные в OWL-онтологию с использованием библиотеки RDFLib. Сущности моделируются как OWL-классы с аннотациями типов данных, связи — как ObjectProperty, меры — как DatatypeProperty с формулами в rdfs:comment.

### **Business Rules Engine**

Модуль позволяет определять бизнес-правила в формате OWL-ограничений и SWRL-правил. Поддерживаются ограничения кардинальности, допустимых значений, а также ролевые ограничения доступа для интеграции с системами безопасности.

### **Ontology Chat (AI Q&A)**

Интеллектуальный чат-интерфейс на базе OpenAI GPT-4o-mini, позволяющий задавать вопросы об онтологии на естественном языке. Поддерживает русский и английский языки, учитывает роль пользователя (Admin, Analyst, Viewer) при формировании ответов.

## **2.2 Технологический стек**

- Python 3.10+ — основной язык разработки
- Streamlit — веб-интерфейс с 8 вкладками
- RDFLib — работа с RDF/OWL онтологиями
- OpenAI API — AI-функциональность чата
- Pydantic — валидация данных

## 3. Функциональные возможности

---

### 3.1 Извлечение онтологий

Система поддерживает автоматическое извлечение следующих элементов:

- Entities (сущности) — таблицы модели с типами колонок и описаниями
- Relationships (связи) — foreign key связи между таблицами
- Measures (меры) — DAX-формулы с метаданными форматирования
- Hierarchies (иерархии) — аналитические иерархии для drill-down

### 3.2 Управление онтологиями

Веб-интерфейс предоставляет полный набор инструментов:

- Создание и редактирование сущностей с атрибутами
- Визуальное управление связями между сущностями
- Настройка ролевых разрешений (CRUD per entity)
- Редактор бизнес-правил с SWRL-синтаксисом
- Предпросмотр OWL в форматах RDF/XML, Turtle, N-Triples
- Сравнение и слияние версий онтологий (Diff & Merge)

### 3.3 AI-функциональность

Ontology Chat позволяет пользователям взаимодействовать с онтологией через естественный язык. Примеры запросов:

- «Какие сущности есть в онтологии?» — список всех entities
- «Покажи связи между таблицами» — таблица relationships
- «Опиши структуру модели данных» — развёрнутое описание архитектуры
- «Какие меры связаны с продажами?» — фильтрация по контексту

## 4. Интеграции

---

### 4.1 OntoGuard AI

Проект интегрируется с OntoGuard AI — семантическим файрволом для AI-агентов. Сгенерированные онтологии могут использоваться для валидации действий агентов на основе OWL-правил. Например, правило «только Admin может удалять записи» будет автоматически проверяться при каждом запросе агента.

### 4.2 Universal Agent Connector

Онтологии могут экспортироваться для использования в Universal Agent Connector — платформе для подключения AI-агентов к корпоративным базам данных. Это обеспечивает семантическую валидацию SQL-запросов и естественно-языковых команд на уровне бизнес-правил.

### **4.3 Knowledge Graphs**

Экспортированные OWL-онтологии совместимы со стандартами семантического веба и могут загружаться в графовые базы данных (Neo4j, Amazon Neptune) для построения корпоративных knowledge graphs.

## 5. Результаты и тестирование

---

### 5.1 Статистика проекта

- 14 реализованных задач (Tasks) — 100% completion
- ~3000 строк Python кода
- 10 ключевых функций
- 8 вкладок пользовательского интерфейса
- Поддержка 3 форматов Power BI моделей

### 5.2 Тестирование

Система протестирована на реальных моделях Power BI:

#### Sales Returns Sample

- 15 сущностей (таблиц)
- 9 связей между таблицами
- Успешная генерация OWL-онтологии
- AI-чат отвечает на вопросы корректно

#### Adventure Works DW 2020

- 11 сущностей
- 13 связей
- Сложная star schema корректно преобразована
- Описание модели данных на английском языке

### 5.3 Тестирование с Playwright MCP

Функциональное тестирование выполнено с использованием Playwright MCP — инструмента браузерной автоматизации от Microsoft. Все 8 вкладок интерфейса проверены, AI-чат протестирован на различных типах запросов (списки, таблицы, развёрнутые описания) на русском и английском языках.

## 6. Заключение

---

PowerBI Ontology Extractor решает актуальную задачу извлечения семантических знаний из аналитических моделей и их формализации в виде OWL-онтологий. Инструмент обеспечивает мост между миром бизнес-аналитики и семантическими технологиями, открывая возможности для:

- Построения корпоративных knowledge graphs на основе BI-моделей
- Семантической валидации действий AI-агентов

- Автоматизированной проверки соответствия бизнес-правилам
- Интеграции аналитических метаданных в системы управления знаниями

Проект имеет открытый исходный код и доступен на GitHub. Дальнейшее развитие предполагает расширение поддержки форматов (Azure Analysis Services, SSAS), улучшение AI-функциональности и интеграцию с большим количеством семантических платформ.

## Ссылки

---

- GitHub: <https://github.com/vpakspace/powerbi-ontology-extractor>
- OntoGuard AI: <https://github.com/vpakspace/ontoguard-ai>
- Universal Agent Connector: <https://github.com/vpakspace/universal-agent-connector>
- OWL Web Ontology Language: <https://www.w3.org/OWL/>
- RDFLib Documentation: <https://rdflib.readthedocs.io/>