

Dados Conectados e Ontologias para o Setor Elétrico: Experimentos com Interoperabilidade Semântica no ONS

Rodrigo Guimarães

Universidade Federal do Rio de Janeiro

rrg@ufrj.br

Resumo. Este relatório, é o resultado da experiência prática com aplicação das técnicas de aquisição de conhecimento, modelagem de base de conhecimento e metodologias discutidas no estudo de caso CONHECIMENTO ESPECIALIZADO NO DOMÍNIO DA ENERGIA ELÉTRICA: ESTUDO DE CASO SOBRE A AQUISIÇÃO DO CONHECIMENTO PARA ONTOLOGIA Teixeira et al. (2021), que utilizou um modelo estruturado de aquisição de conhecimento, para obter termos do domínio utilizado pelo setor elétrico brasileiro, utilizando assistentes de IA para desenvolvimento do trabalho de modelagem de uma base de conhecimento. O trabalho aqui apresentado, com o objetivo de modelar a base de conhecimento de equipamentos do Operador Nacional do Sistema Elétrico (ONS), utiliza técnicas similares, contudo, simplificando o processo através do glossário corporativo, e com a carga de dados reais de alguns equipamentos do ONS através de um processo de ETL. O trabalho discute também os desafios, limitações e a possibilidade de expansão para apoiar a gestão de ativos do setor, assim como operações de troca de informações entre os agentes do setor.

Palavras-chave: Ontologia; Web Semântica; Operador Nacional do Sistema Elétrico; LLM.

1 Introdução

O sistema elétrico brasileiro, é caracterizado por grande complexidade técnica, sistemas heterogêneos e grande carga regulatória. A falta de interoperabilidade entre os agentes do setor, se mostra um desafio recorrente, pois atrasa a inovação, e dificulta um processo transparente de fiscalização pelas agências reguladoras, como ANEEL e o próprio ONS, que é responsável pela apuração de ocorrências nos equipamentos da rede básica do SIN.

O presente trabalho, se inspira no estudo de caso, realizado por Teixeira et al, que, indicou métodos para aquisição de conhecimento na modelagem de ontologias. A proposta consiste, em adaptar parte dessa metodologia ao domínio do ONS, utilizando mais amplamente IA generativa como suporte para a modelagem conceitual. E como ferramenta de apoio, também foi implementado um fluxo de carga para carregar e organizar dados reais em um protótipo de grafo de conhecimento.

O objetivo do experimento é primariamente, validar se o modelo adotado no artigo base, é aderente à realidade de um agente do setor com diferenças base de natureza jurídica e papel no modelo eletroenergético brasileiro, além da geração de consultas SPARQL para validar os dados gerados, a fim de entender se existem inconsistências relacionadas ao negócio e ao processo de geração. Foram realizadas as seguintes etapas:

Com o objetivo de ser um balizador e permitir a avaliação de aderência, foi necessário primeiramente a modelagem manual de um escopo reduzido. Para esta etapa, foram modeladas as entidades Agente, Usina, Linha de Transmissão, Transformador e Subestação, que também é conhecida como instalação. Esta modelagem, foi realizada utilizando a ferramenta Progtegé, e foi utilizado o ontograf para visualização inicial do resultado.

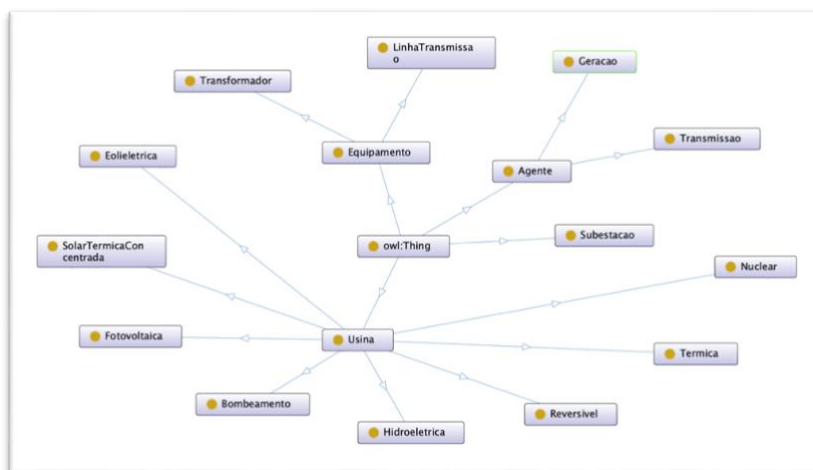


Fig. 1. Resultado da modelagem no Protégé – Ontograf

Foram mapeados também, algumas propriedades básicas de dados e de objetos, estas propriedades serão utilizadas para a montagem de consultas SPARQL com o objetivo de validar os dados de triplas geradas. Esta montagem simples será utilizada para a análise e comparação com os dados gerados pelo processo automático.

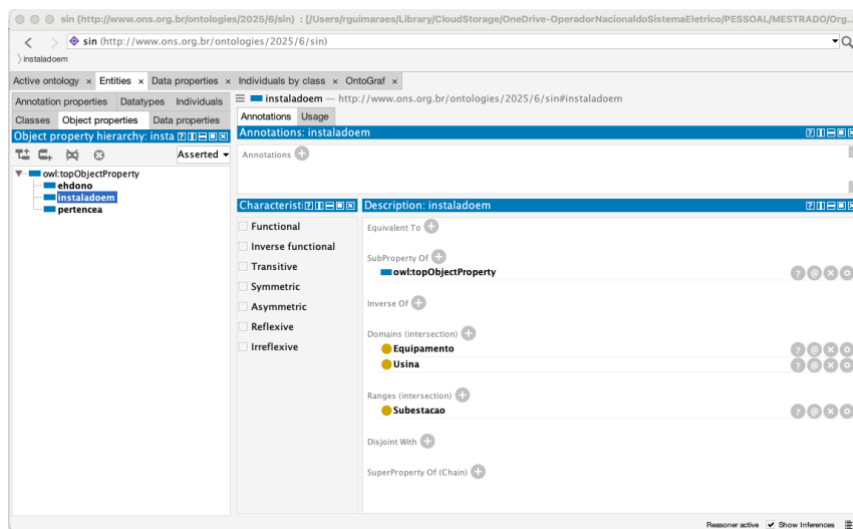
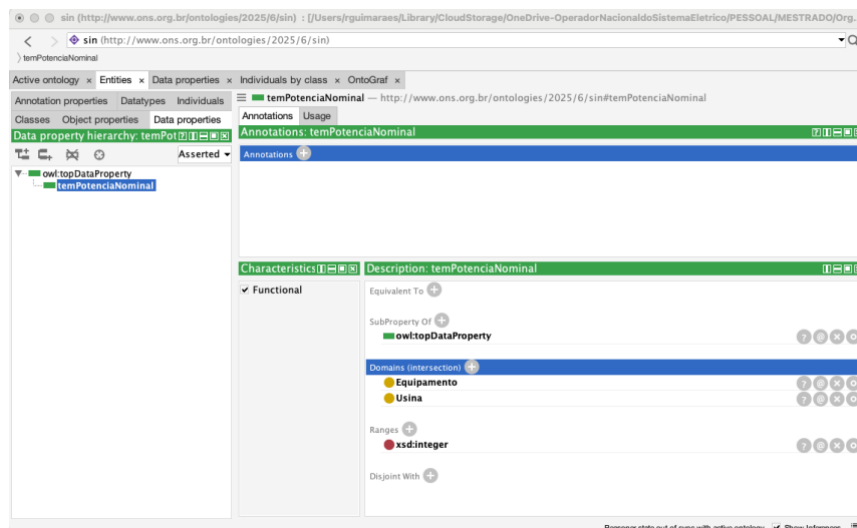
**Fig. 2.** Propriedades de Objetos – Protégé

Fig. 3. Propriedades de dados – Protegé

2.2 Mapeamento do vocabulário utilizando IA

A metodologia de geração da base de IA, consiste primariamente na criação de um prompt padrão, contextualizando todo o negócio envolvido, os tipos de equipamento, relações básicas, incluindo referência para o Glossário ONS e solicitando a geração do vocabulário em um arquivo no formato TTL para visualização, análise e importação no Protegé.

```
prompt - geração vocabulário.txt
1  Dado as informações abaixo, referentes ao operador nacional do sistema elétrico – ONS, cujo o glossário está
2
3  Agente: Representa um agente do setor elétrico. Ex: Furnas, CEMIG.
4  Possui os seguintes atributos:
5      age_id: Identificador do Agente
6      descricao: nome detalhado do agente (descrição)
7      nome: nome comumente utilizado do agente
8      tipo: indica se é um agente da transmissão ou da geração de energia. Deve ser criado na classe Transmissa
9
10
11 Subestação: Representa uma subestação ou também conhecida como Instalação, que é uma area física que contém d
12 Seus dados estão no arquivo subestacao.csv
13 Possui os seguintes atributos:
14     ins_id: Indentificador da subestação
15     nome: nome identificador da subestação
16     descricao: nome detalhado / descrição da subestação
17
18
```

Fig. 4. Exemplo de um trecho do prompt utilizado

Após a geração do prompt e ajustes realizados, os dados foram devidamente importados para o Protegé, e posteriormente para o GraphDB

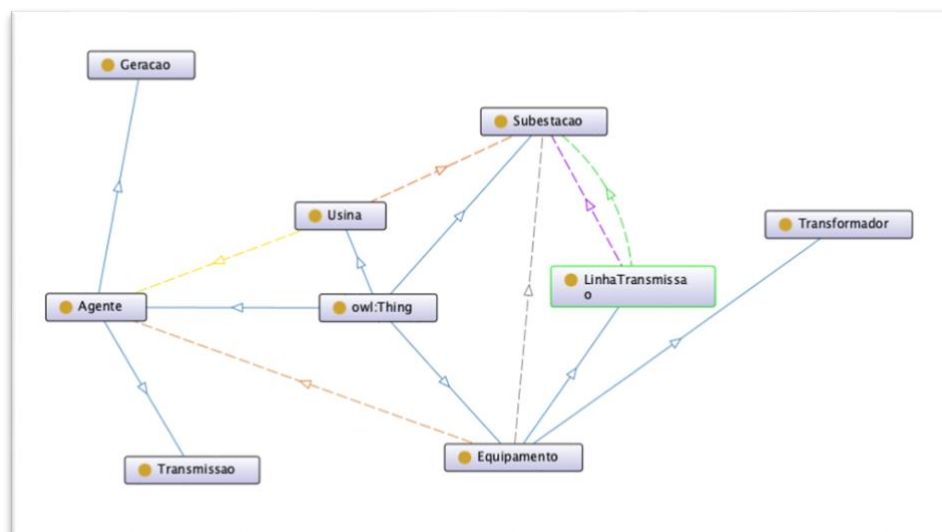


Fig. 5. Vocabulário Gerado pela IA – Protegé Ontogrf

2.3 Massa de dados reais para análise

A geração da base de dados, se deu através da exportação de tabelas do banco de dados do ONS em formato CSV, e a utilização de um script Python, com apoio da biblioteca RDFLib, Biblioteca tal que facilita a carga de maneira estruturada, carregando facilmente cerca de 193 mil triplas.

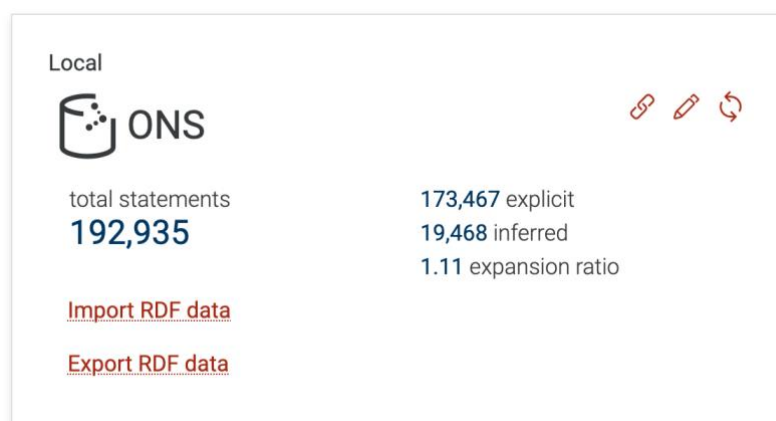


Fig. 6. Repositório carregado – GRAPHDB

Durante o processo, a fim de enriquecer a análise, houve a tentativa de utilização de outras ferramentas, como o cellfie, e foram feitos testes com R2RML, mas devido a características dos dados utilizados, este mapeamento se tornou menos eficiente do que o script Pyhon.

3 Resultados

Foram modelados conceitos relacionados a ativos e processos do ONS, utilizando inicialmente o GPT-4, mas as etapas iniciais se mostraram desafiadoras, pois o modelo, não reconheceram nuances de negócio, demandando vários ajustes manuais nos dados gerados inicialmente, como por exemplo, a primeira versão gerada do vocabulário com o GPT-4, que identificou erradamente o domínio do ONS

```
1 @prefix : <http://www.ons.gov.br/ontologia/energia#> .
2 @prefix owl: <http://www.w3.org/2002/07/owl#> .
3 @prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
4 @prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
5
6 <http://www.ons.gov.br/ontologia/energia#> rdf:type owl:Ontology .
7
8 ### Classes
9
10 :Agente rdf:type owl:Class ;
11     rdfs:label "Agente" ;
12     rdfs:comment "Entidade que representa os proprietários ou operadores de ativos (geradores, transmissores, distribuidores, etc.)." .
13
14 :Instalacao rdf:type owl:Class ;
15     rdfs:label "Instalação" ;
16     rdfs:comment "Instalação física do sistema elétrico, como usinas ou subestações." .
17
18 :Usina rdf:type owl:Class ;
19     rdfs:subClassOf :Instalacao ;
20     rdfs:label "Usina" ;
```

Fig. 7. Exemplo de problema com RDF Gerado pelo GPT-4

3.1 Vocabulário

Ao analisar o vocabulário gerado, foi constatado que a base foi corretamente representada, contudo, alguns pontos demandam atenção, em especial as questões de nomenclaturas dos predicados, como por exemplo, ao informar que um equipamento está instalado em uma subestação, o modelo definiu o nome do relacionamento com o código identificador, ao invés de criar algo mais semanticamente correto, como “instaladoEm”. Isto aconteceu em todos os tipos de relação mapeados. Além disso, foram encontrados diversos erros no processo de importação dos dados de indivíduos para o GraphDB, o que demandou muitos ajustes manuais nos arquivos, por este motivo, modificamos a IA utilizada para Q, produzido pela AMAZON, que se mostrou mais preparada para atender as solicitações relacionadas à RDF, o que permitiu uma quantidade menor de ajustes e gerou algo mais “próximo” ao real, e gerou dados que foram importados corretamente para o GraphDB.

Comparando com o vocabulário gerado manualmente, foi observado que de maneira geral, existe uma grande similaridade entre eles, e que com poucos detalhes que

precisaram ser tratados manualmente, ele atende perfeitamente as necessidades do trabalho proposto.

```
78 # Propriedades para Equipamentos
79 ons:eqp_id a owl:DatatypeProperty ;
80     rdfs:domain ons:Equipamento ;
81     rdfs:range xsd:string ;
82     rdfs:label "Identificador do Equipamento" .
83
84 ons:age_id_prop a owl:ObjectProperty ;
85     rdfs:domain ons:Equipamento ;
86     rdfs:range ons:Agente ;
87     rdfs:label "Agente Proprietário" .
88
89 # Propriedades para Linha de Transmissão
90 ons:instalacao_de a owl:ObjectProperty ;
91     rdfs:domain ons:LinhaTransmissao ;
92     rdfs:range ons:Subestacao ;
93     rdfs:label "Subestação Terminal DE" .
94
95 ons:instalacao_para a owl:ObjectProperty ;
```

Fig. 8. Exemplo de mapeamento

3.2 Carga de dados

A carga de dados foi um processo desafiador, pois os dados de origem possuíam caracteres especiais e acentuação não compatíveis com algumas ferramentas para automação do processo, sendo assim, o que se mostrou mais prático foi a geração do arquivo TTL com os indivíduos através de um script Python com a biblioteca RDFLib, que atendeu de maneira simples e modular as necessidades com o mínimo de esforço. Inclusive na questão da performance, pois o script tratou cerca de 190.000 registros em 3,5 segundos, o que se entende como uma performance excelente, visto que em um teste inicial utilizando um ETL este tempo foi 10 vezes maior.

```
def adicionar_linhatransmissao(row): #eqp_id,age_id,descricao,instalacao_de,instalacao_para
    linha_uri = ONS[f'LinhaTransmissao_{row["eqp_id"]}']
    individuals.add((linha_uri, RDF.type, ONS.LinhaTransmissao))
    individuals.add((linha_uri, ONS.eqp_id, Literal(row['eqp_id'])))
    individuals.add((linha_uri, ONS.age_id, Literal(row['age_id'])))
    individuals.add((linha_uri, ONS.descricao, Literal(remover_acentos(row['descricao']))))
    individuals.add((linha_uri, ONS.instalacao_de, ONS[f'Subestacao_{row["instalacao_de"]}']))
    individuals.add((linha_uri, ONS.instalacao_para, ONS[f'Subestacao_{row["instalacao_para"]}']))
```

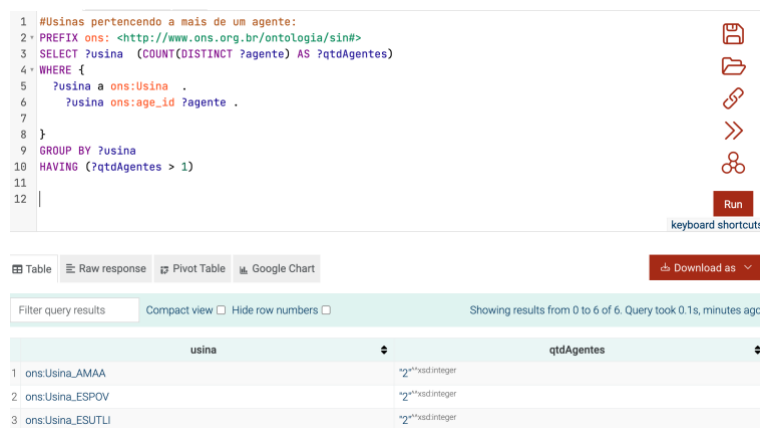
Fig. 9. Exemplo do Código para criação dos indivíduos – Linha de Transmissão

Também foram realizados testes, solicitando que a IA gerasse diretamente as entidades, fornecendo os arquivos de dados em formato CSV, o que se demonstrou inviável, primeiramente pelo tamanho dos arquivos, que foram rejeitados pela API da IA, mesmo em arquivos reduzidos, não houve a possibilidade de carga sem erros de referências e sintaxe.

Uma terceira tentativa foi solicitar que a IA, através da base regulatória oficial, gerasse a base, o que se demonstrou promissor, o que será indicado como uma possibilidade para trabalho futuro, já que seria uma base atualizada automaticamente sempre que um empreendimento receber as licenças regulatórias.

3.3 Consultas para validação

Com o objetivo de validar o experimento e avaliar a qualidade geral dos dados gerados, foram desenvolvidas consultas SPARQL, estas consultas, ao serem executadas no GraphDB, geraram análises interessantes, como por exemplo, foi possível identificar um erro no banco de dados original, em que havia usinas pertencendo a mais de um agente, o que é um erro causado pelo conflito do código no sistema que gerou os dados de origem, e também, questões como, agentes somente cadastrados como geradores de energia, serem proprietários de ativo da transmissão, o que também é uma inconsistência de negócio.



```

1 #Usinas pertencendo a mais de um agente:
2 PREFIX ons: <http://www.ons.org.br/ontologia/sin#>
3 SELECT ?usina (COUNT(DISTINCT ?agente) AS ?qtdAgentes)
4 WHERE {
5   ?usina a ons:Usina .
6   ?usina ons:age_id ?agente .
7 }
8
9 GROUP BY ?usina
10 HAVING (?qtdAgentes > 1)
11
12

```

Run

keyboard shortcut

Table Raw response Pivot Table Google Chart Download as

Filter query results Compact view Hide row numbers Showing results from 0 to 6 of 6. Query took 0.1s, minutes ago

	usina	qtdAgentes
1	ons:Usina_AMAA	2**xsd:integer
2	ons:Usina_ESPOV	2**xsd:integer
3	ons:Usina_ESUTLI	2**xsd:integer

Fig. 10. Consulta para identificar usinas que pertençam a mais de um agente


```

1 PREFIX ons: <http://www.ons.org.br/ontologia/sin#>
2
3 SELECT DISTINCT ?agente ?linha
4 WHERE {
5   ?agente a ons:Geracao .
6   FILTER NOT EXISTS {
7     ?agente a ons:Transmissao .
8   }
9   ?linha ons:age_id ?agente .
10 }
11

```

Press Alt+Enter to autocomplete keyboard shortcuts

Table Raw response Pivot Table Google Chart Download as 1 2 3 4

Filter query results Compact view Hide row numbers Showing results from 0 to 1000 of 3846. Query took 0.1s, moments ago.

	agente	linha
1	ons:Agente_EPE	ons:LinhaTransmissao_AMAA-0AD-7
2	ons:Agente_EPE	ons:LinhaTransmissao_AMAA-0AD-8

Fig. 11. Consulta para identificar agentes proprietários em contradição com seu cadastro

4 Considerações finais

O uso de IA generativa se mostrou um forte mecanismo de apoio na aquisição de conhecimento, principalmente no que diz respeito a obtenção de dados em lote, o que gera grande produtividade no processo como um todo. Entretanto, verificou-se a necessidade de revisão humana para corrigir ambiguidades ou imprecisões.

A tentativa de se realizar o processo de cargas com um ETL se mostrou promissora, mas devido a dificuldade de padronização nos dados, o script Python aparece como a melhor alternativa, pois a facilidade para a geração da base, somado ao amplo engajamento da comunidade que mantém as atualizações do componente, somente corroboram com esta decisão.

Todo o processamento, dados carregados e consultas, indicam que há total viabilidade da utilização deste modelo no contexto do ONS, indicando a possibilidade implícita de melhoria na interoperabilidade entre sistema, e seu uso, certamente será fomentado internamente na empresa.

Todos os arquivos relacionados se encontram disponíveis no repositório GITHUB abaixo:

5 Trabalhos Futuros

O trabalho aqui apresentado, abre a possibilidade de novos projetos e atividades no ONS, para futuras iniciativas, será considerado a expansão do grafo para outros ativos do setor elétrico, outros processos e áreas de negócio, além da integração com estes dados nos sistemas de supervisão e controle da operação e sua possível disponibilização através de endpoint publico de SPARQL no portal de dados abertos do ONS

Referencias

1. Teixeira, L.M.D., et al. “Conhecimento Especializado no Domínio da Energia Elétrica: Estudo de Caso sobre a Aquisição do Conhecimento para Ontologia.” *Encontros Bibli*, 2021.
2. RDFLib. RDFLib is a Python library for working with RDF, a simple yet powerful language for representing information. Disponível em: <https://github.com/RDFLib/rdfliib> Acesso em: 08 jul. 2025
3. OPERADOR NACIONAL DO SISTEMA ELÉTRICO (ONS). Glossário de termos do setor elétrico. Disponível em: <https://www.ons.org.br/paginas/conhecimento/glossario>. Acesso em: 10 jul. 2025.
4. BERNERS-LEE, T.; HENDLER, J.; LASSILA, O. The Semantic Web. *Scientific American*, v. 284, n. 5, p. 34–43, 2001.