

Ontologias ou *Knowledge Graphs?*



NCOR-BR, 2021

O presente material e quaisquer outros recursos que o acompanhem foi produzido pela NCOR-BR para distribuição gratuita e disseminação do tema, sem fins lucrativos.

This work is licensed under



[Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Este trabalho está licenciado sob



[Creative Commons - Atribuição 4.0 Internacional.](https://creativecommons.org/licenses/by-nc-nd/4.0/)

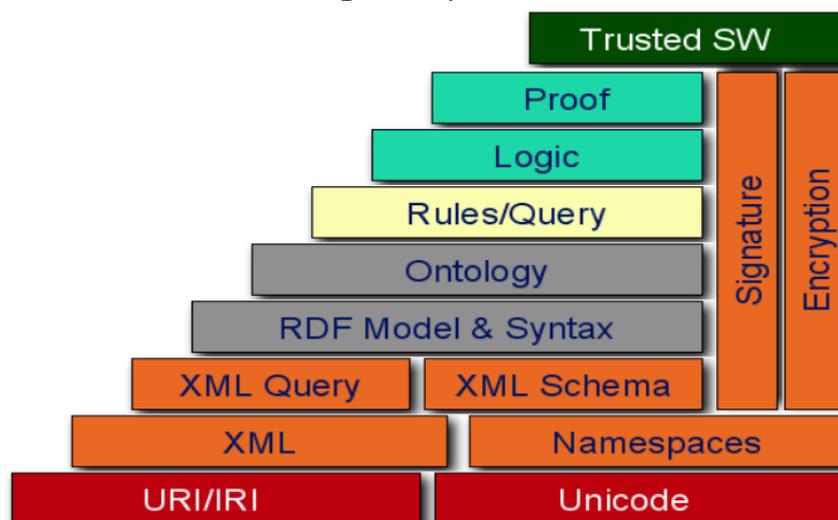
Sumário

A Web Semântica	4
Tecnologias diferentes	6
Linked Data e Knowledge Graphs é o mesmo que ontologia?	7
Referências	10

A Web Semântica

A “Web Semântica” (WS) abrange tecnologias da informação, linguagens, protocolos, padrões, dentre outros recursos que tem por objetivo obter melhorias na recuperação da informação na Web. A “pilha de tecnologias da WS”, publicada pela W3C, exibe esses recursos. São tantos recursos e tantas siglas que as vezes fica complicado entender o que significada cada. Se você está confuso ou esqueceu as linguagens de representação, leia novamente o Caderno 9.

Figura 1: pilha da W3C



Fonte: Bratt, S. Semantic Web Stack. Disponível na Internet em: <https://www.w3.org/2004/Talks/1117-sb-gartnerWS/slide18-0.html>. Acesso em 18-05-2021

O *Linked Data* é muitas vezes confundido com a ontologia e mesmo artigos de científicos relatam o uso da linguagem RDF para construir ontologias. **Isso não é possível!** Ao RDF faltam vários elementos essenciais em uma ontologia. Para citar o mais evidente, o RDF não possui o elemento “classe”, que é um dos blocos mais importantes da ontologia OWL (vide o Caderno 3).

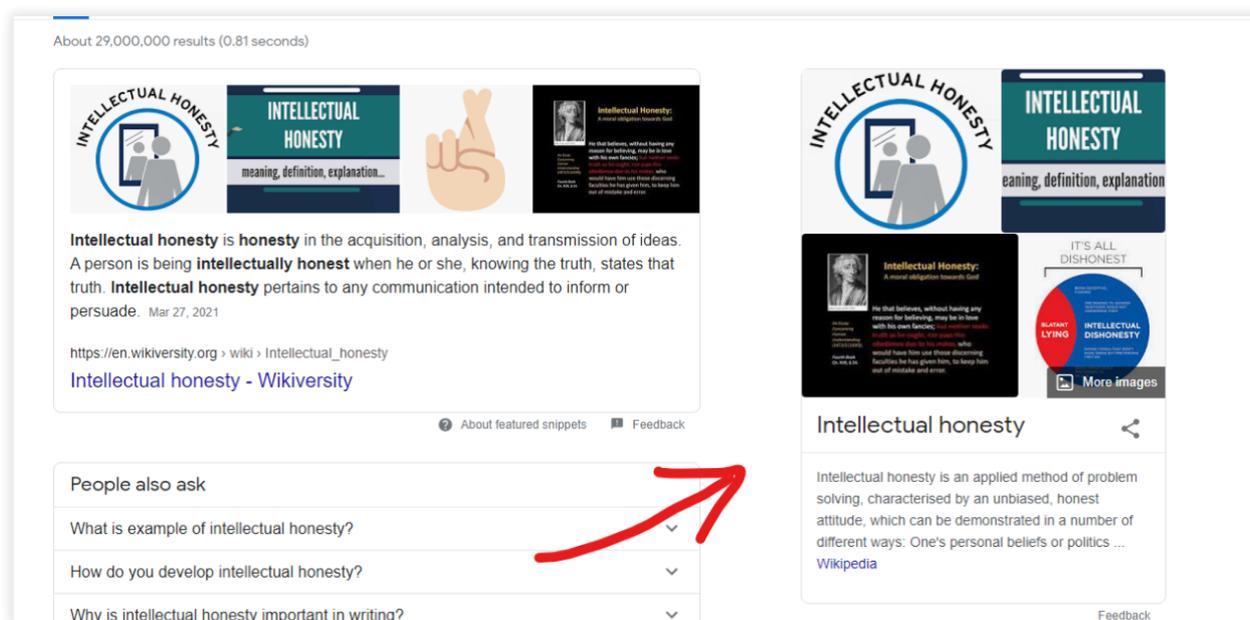
Pela variedade de recursos, nomes, siglas, notações, tipos, além de uma história de mais de 20 anos, muitas vezes é complicado entender como tudo isso se relaciona. Vamos explicar um pouco disso aqui.

Tecnologias diferentes

Em meados da década de 2000, um conjunto de tecnologias passou a ser identificado pelo termo “*Linked Data*”. Trata-se de uma iniciativa aberta, pública e gratuita que representou grande impulso para a Web Semântica.

O *Linked Data* se mantém em uso até hoje em aplicações diversas na Web. Os usuários do Google, por exemplo, têm acesso a algo parecido ao fazer buscas, na verdade, uma variação da tecnologia original. Aquelas informações que aparecem do lado direito dos resultados de busca, na forma de um quadro com dados básicos sobre o tema da busca e *links* para mais dados, são *Linked Data* levemente modificado pelo Google.

Figura 2 – *Knowledge Graph* da Google (seta)

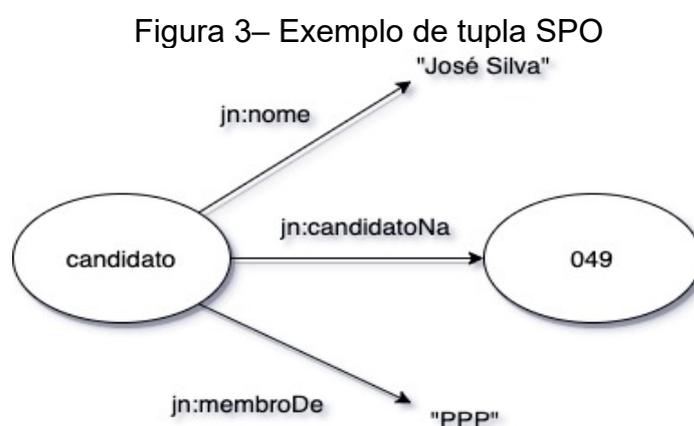


Fonte: busca no site Google pelo termo “*honestidade intelectual*”

Entretanto, sempre fica a dúvida sobre *Linked Data* – que pode ser traduzido para “dados ligados” – e as ontologias são a mesma coisa. Isso ainda fica mais complicado com a pluralidade de nomes encontrados na Web.

Linked Data e Knowledge Graphs é o mesmo que ontologia?

O *Linked Data* se refere ao conjunto de grafos escritos em RDF e conectados pelo compartilhamento de URIs. As URIs representam triplas sujeito-predicado-objeto (SPO) de uma descrição RDF, a unidade básica dos grafos (veja caderno 9). A Figura 3 exibe um exemplo como Sujeito (S) = *Candidato*, Predicado (P) = *candidatoNa* e Objeto (O) = *Zona Eleitoral 049*.



O número de grafos RDF disponíveis publicamente continua a crescer, cobrindo uma variedade de temas, como governo, Medicina, Geografia, Ciências da Vida, Biologia Linguística, Mídia, publicações científicas, museus, dentre outros. Esse crescimento nem sempre foi notado, pois os usuários acessam apenas a interface de consulta ou páginas HTML com informações administrativas.

Um dos conjuntos de *Linked Data* mais populares é o *DBpedia*¹, contendo dados extraídos da *Wikipédia* disponibilizados via uma interface de consulta. Por essa interface – chamada *end-point* – um usuário pode obter dados de forma simples, sem precisar abrir várias páginas Wiki e reunir dados específicos. Em 2008 anunciava-se um número superior a 100 milhões de triplas armazenado nesse repositório², tornando a *DBpedia* um recurso importante na Web Semântica e nas iniciativas de dados abertos.

No início da década de 2010, o entusiasmo no *Linked Data* deu lugar a uma perspectiva mais objetiva porque as dificuldades de integração de dados exigiam mais esforço do que esperado. Um dos principais motivos para isso era, e ainda é, a abundância de esquemas superficiais e pouco significativos. A partir daí, intensificou-se a busca por abordagens mais eficientes para compartilhamento, integração e reutilização.

Nesse período, o *Google* passou a adotar o termo “*knowledge graphs*”. Porém, a noção de grafo do conhecimento não é criação do *Google* e remonta a técnicas dos anos 1990, em particular, as “redes semânticas”. Outras empresas como *BBC*, *Microsoft*, *IBM*, *FaceBook* seguiram o *Google*, cada uma tornando acessíveis dados em RDF ou alguma variação. O *Knowledge Graphs* da *Google* (Figura 2), por exemplo, não está disponível para *download*, mas pode ser usado com alguns poucos conhecimentos de programação.

Apesar do que de todas essas iniciativas, ***Knowledge Graphs e Linked Data não são ontologias***, mesmo que muitas vezes os termos sejam considerados sinônimos. Dentre as razões para isso, a principal é que ***Linked Data usa RDF e ontologia usa OWL***. O RDF não possui a expressividade do OWL e as ontologias OWL possuem três fundamentos principais:

- *Axiomas*: declarações expressas por um tipo de lógica
- *Classes, propriedades e indivíduos*: elementos para se referir a entidades
- *Expressões*: combinações de entidades, formando descrições complexas

Observe que o RDF não possui a forma de declarar classes como a ontologia OWL requer, o segundo elemento da lista acima. Essa possibilidade só vai aparecer no RDFS, precursor do OWL.

A OWL é linguagem de propósito geral que objetiva formular o conhecimento de forma explícita e o faz através de unidades elementares denominadas *declarações* ou *proposições*. Nesse aspecto, reside outra diferença importante entre ontologias e *Linked*

Data: o OWL tem capacidade de gerar conhecimento novo, de forma automática, obtendo consequências, conclusões ou *inferências*.

O fundamental pode ser resumido na expressividade possibilitada pela existência de semântica no OWL, ausente em RDF. A semântica da Web Semântica é uma modalidade de *semântica formal* que possibilita o que tem sido chamado “enriquecimento” de dados. Por todas essas características, *Linked Data* e ontologias não são a mesma tecnologia, apesar de terem a mesma origem. Se você não se lembra bem da origem dessas linguagens dê uma olhada no Caderno 9.

Referências

ALMEIDA, M.B. Ontologia em Ciência da Informação: Tecnologia e Aplicações. Curitiba: CRV, 2021, 340 p. Coleção Representação do Conhecimento em Ciência da Informação: Volume 2.

Esse material é gratuito, mas você pode ajudar na divulgação citando a fonte:

NCOR-BR. (2021). Ontologias ou *Knowledge Graphs*? Caderno 10. Disponível na internet em <<http://ncor-br.org>>

¹ *DBPedia*. Disponível na Internet em: <<https://www.dbpedia.org/>>. Acesso em 11-02-2021.

² Disponível na internet em <<https://blog.dbpedia.org/?s=triples>>. Acesso 11/02/2021