

---

# Grafos de Conhecimento e seu Potencial para Aplicações de Inteligência Artificial

**João Paulo A. Almeida**

jpalmeida@ieee.org

Núcleo de Estudos em Modelagem Conceitual & Ontologias (NEMO)  
Universidade Federal do Espírito Santo

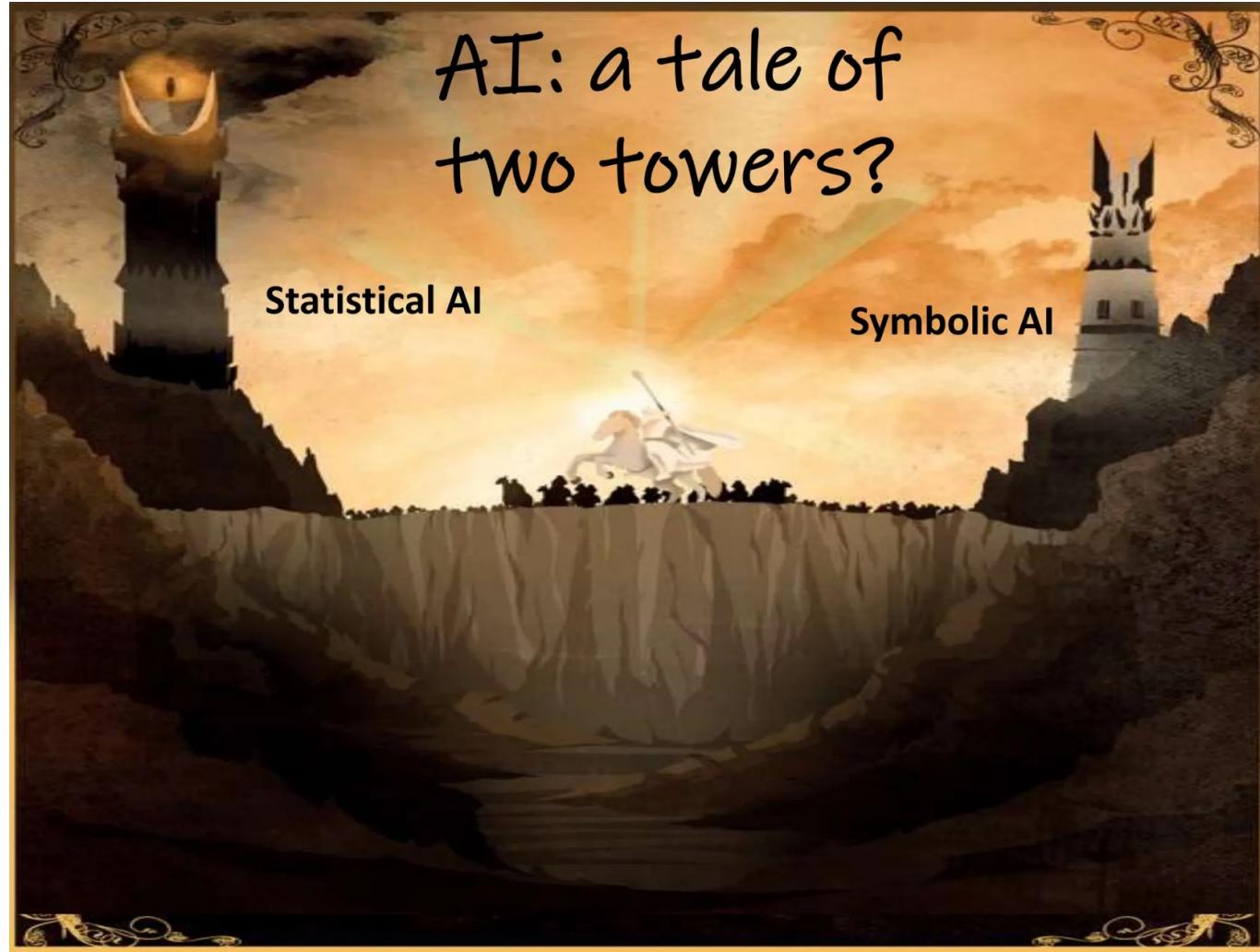


<https://creativecommons.org/licenses/by/4.0/>

# Agenda

- Cenário atual (diagnóstico e requisitos)
- Grafos de conhecimento e ontologias
  - “a little semantics goes a long way”
- Sinergia com Large Language Models (LLMs)
  - Exemplo: interações medicamentosas

# Cenário atual: as duas torres (Frank van Harmelen)



<https://www.slideshare.net/slideshow/systems-that-learn-and-reason-frank-van-harmelen/267008886>

THIS IS YOUR MACHINE LEARNING SYSTEM?

YUP! YOU POUR THE DATA INTO THIS BIG  
PILE OF LINEAR ALGEBRA, THEN COLLECT  
THE ANSWERS ON THE OTHER SIDE.

WHAT IF THE ANSWERS ARE WRONG?

JUST STIR THE PILE UNTIL  
THEY START LOOKING RIGHT.



Randall Munroe  
<https://xkcd.com/1838>

# Cenário atual: a torre de IA estatística

- Ganhos das abordagens de “aprendizado de máquina” com a escala
  - Large Language Models (LLMs) / IA “generativa”
  - Modelos pré-treinados
- Avanços em processamento de linguagem natural sem precedentes
  - Com aplicações para recuperação de informação, verbalização, tradução

# Cenário atual: a torre de IA estatística

- Problemas persistentes de confiabilidade, “explanabilidade”
  - Confiabilidade: “alucinação”
  - Explanabilidade: “caixa preta”
  - Custo computacional do treinamento, inviabiliza retreinamento frente a novas informações
  - Dificuldade de obter escala em aplicações em domínios especializados
    - Dados em organizações
    - “Você não tem dados suficientes” (“Data hunger”)

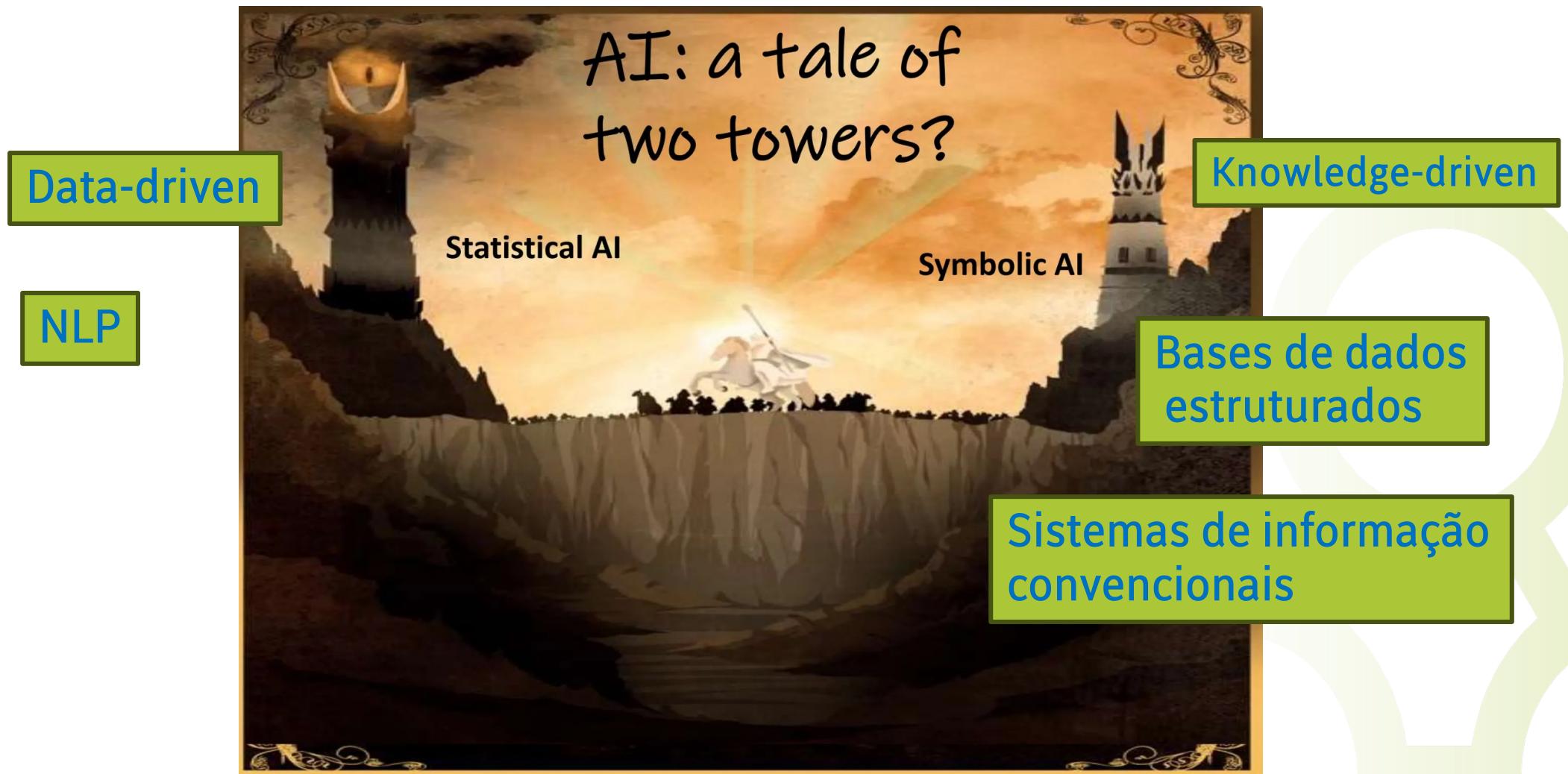
# Cenário atual: a torre de IA simbólica

- Respostas baseada no que está declarado explicitamente (também com raciocínio lógico)
- Custo humano alto
  - Curadoria explícita de regras
  - Curadoria explícita de identificadores de entidades relevantes
  - Curadoria de dados em bases de conhecimento
- Dificuldade com processamento de linguagem natural
  - Rigidez dos identificadores

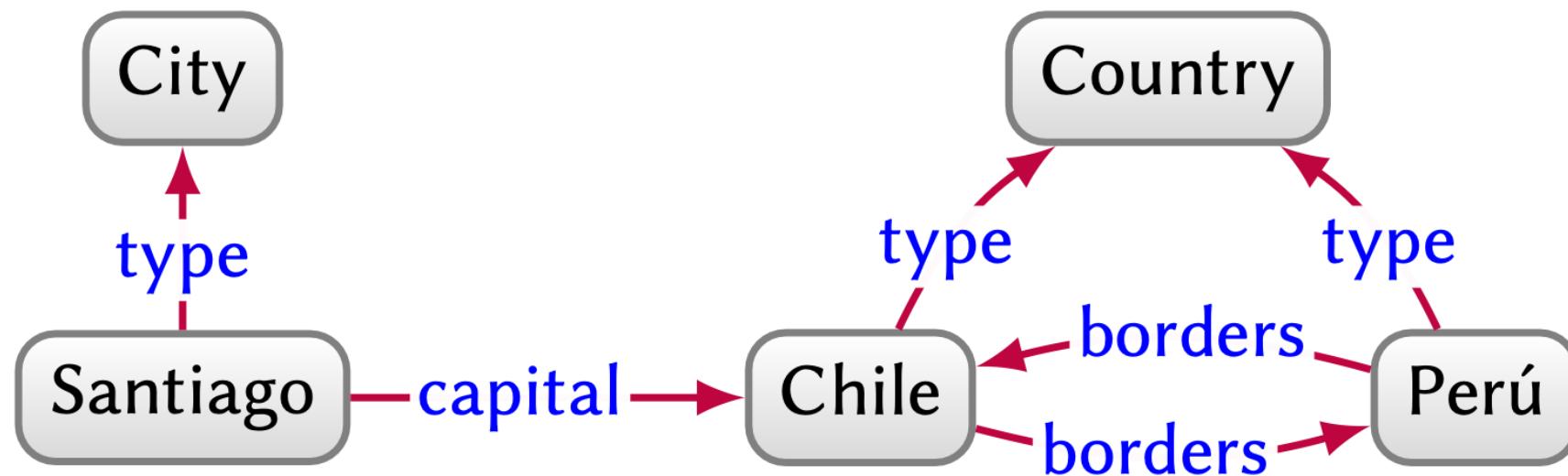
# O que queremos?

- Sistemas confiáveis, que expliquem suas decisões (por exemplo citando fontes científicas), e que apoiem os humanos em suas tarefas
- Capazes de interagir através de linguagem
- Fidedignos à realidade (aos fatos que importam)
- Se eu pergunto: existe interação medicamentosa entre fluconazol e amoxicilina... me importo com o resultado!

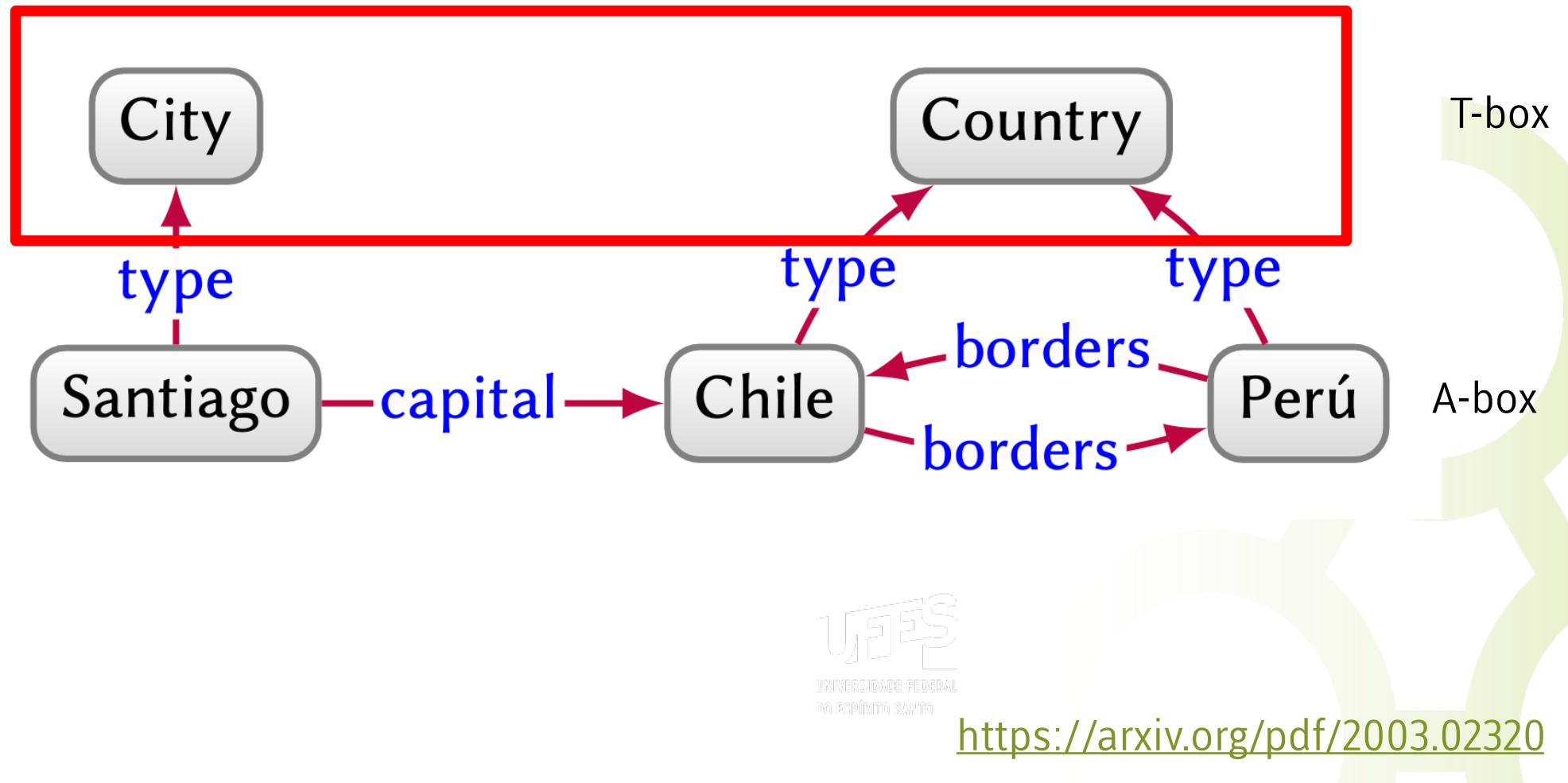
# Combinando as forças das duas torres (Frank van Harmelen)



# O que é um grafo de conhecimento?



# O que é um grafo de conhecimento? (incluindo classes)



# Exemplo de Grafo de Conhecimento: Wikidata

- Independente de domínio, estrutura por trás da Wikipedia

The screenshot shows the Wikidata homepage. On the left, there is a sidebar with various links: Main page, Community portal, Project chat, Create a new item, Recent changes, Random item, Query Service, Nearby, Help, Donate, Lexicographical data, Create a new Lexeme, Recent changes, and Random Lexeme. The main content area is titled "Special page" and displays the "Statistics" page. The statistics are presented in three sections: "Page statistics", "Edit statistics", and "User statistics".

Page statistics	
Content pages	113,988,305
Pages (All pages in the wiki, including talk pages, redirects, etc.)	119,114,068

Edit statistics	
Page edits since Wikidata was set up	2,261,414,291
Average edits per page	18.99

User statistics	
Registered users (list of members)	6,682,966
Active users (list of members) (Users who have performed an action in the last 30 days)	23,854



Item Discussion

Read View history

Search Wikidata



# Earth (Q2)

third planet from the Sun in the Solar System

Planet Earth | World | Gaia | Sol III

► In more languages

## Statements

instance of

terrestrial planet

▼ 0 references

inner planet of the Solar System

follows Venus

followed by Mars

► 1 reference

part of

Earth-Moon system

▼ 0 references

# Princípios básicos

- Representação de entidades através de nós
- Representação de relações através de arcos
- Tipos podem ser usados para construir árvores taxonômicas
- Tipos de arcos restringem possíveis relações
- Identificação única dos nós, tipos de nós, arcos, tipos de arcos
- Consultas por padrões
- Combinado com princípios FAIR:
  - Findable, Accessible, Interoperable, Reusable
  - Incluindo licenças explícitas dos dados

# Taxonomias

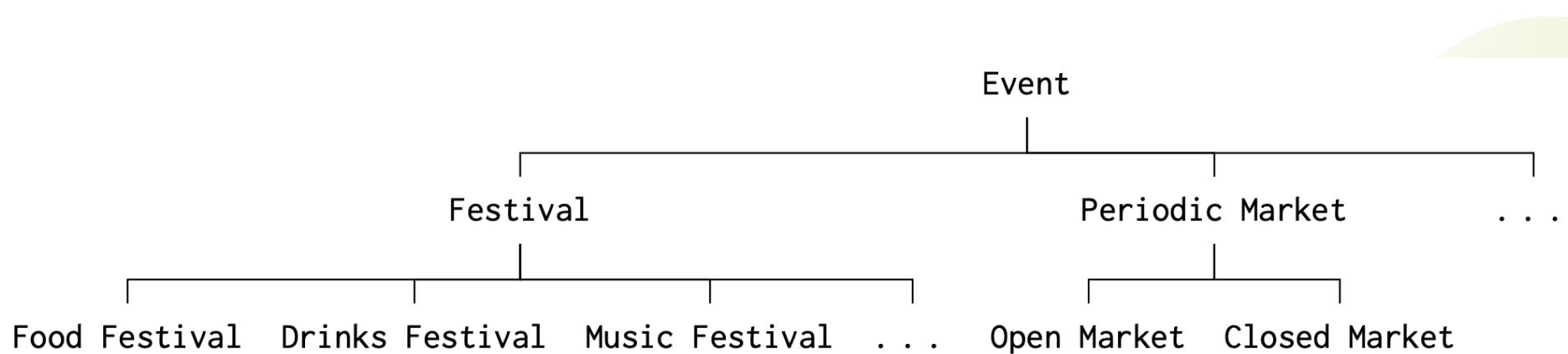
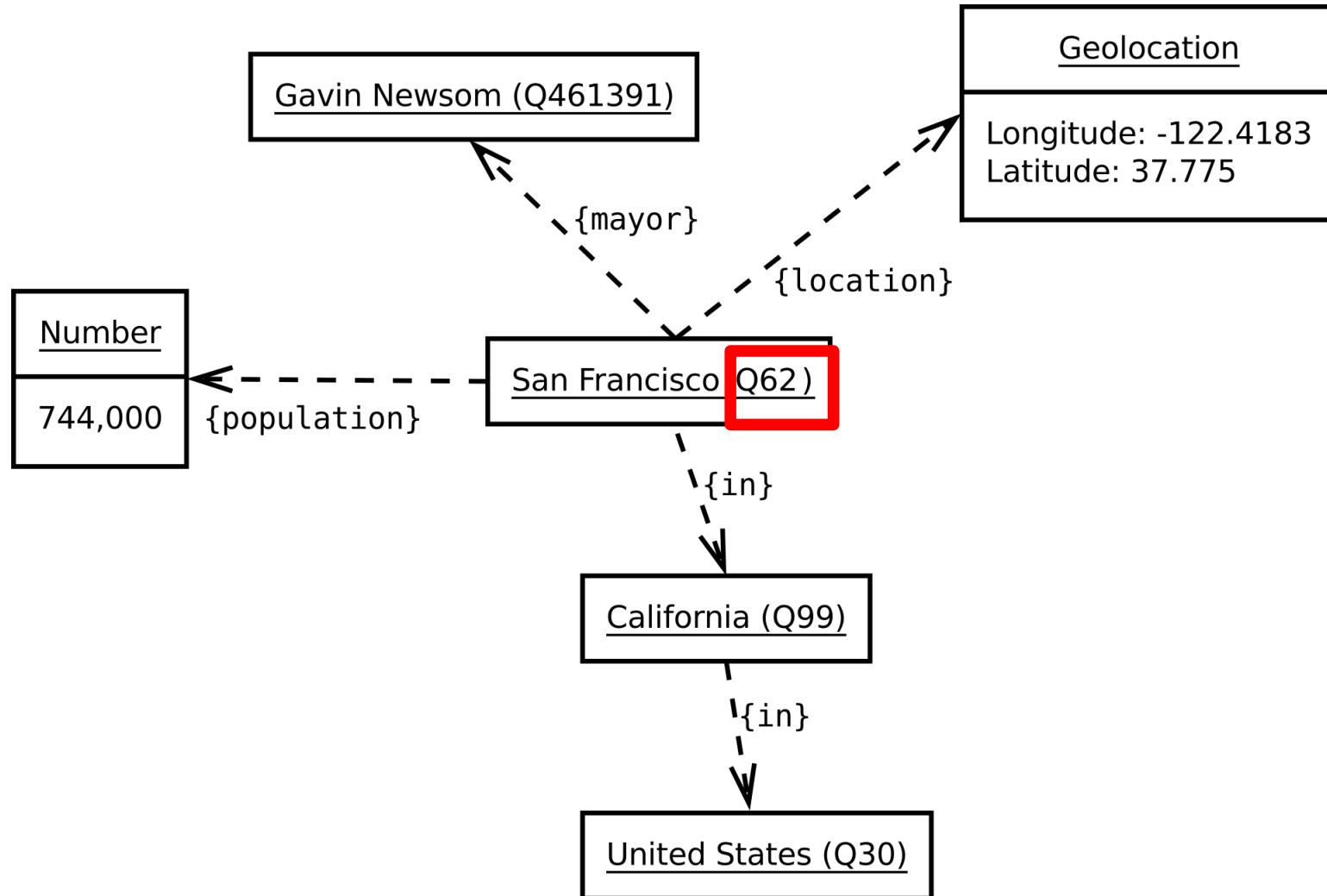


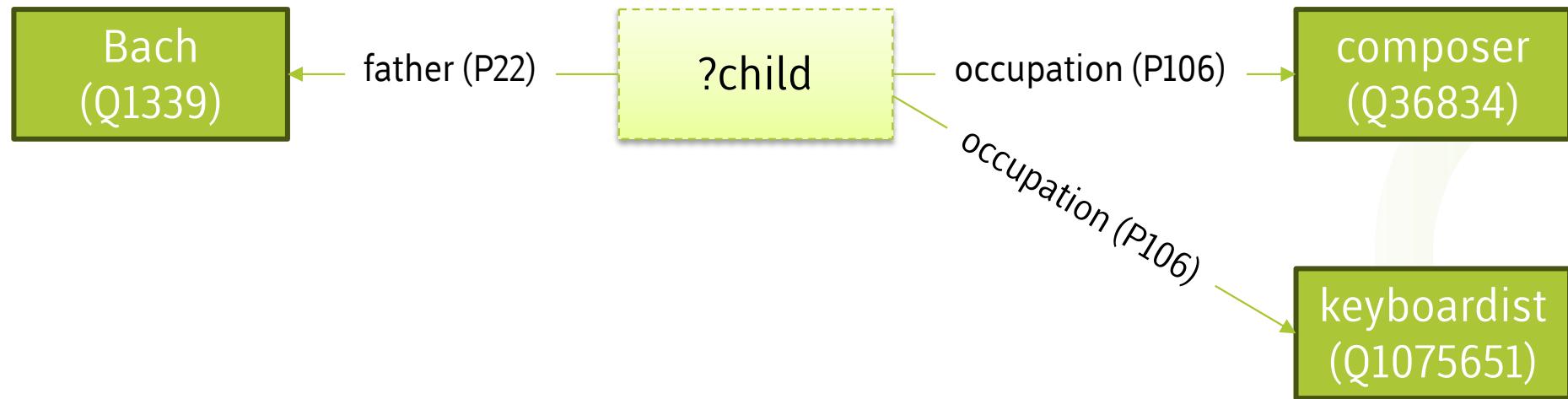
Fig. 11. Example class hierarchy for Event

# Identificação única de nós



<https://www.wikidata.org/wiki/Wikidata:Introduction#>

# Consulta por padrões





```
1 SELECT ?child ?childLabel
2 WHERE
3 {
4     ?child wdt:P22 wd:Q1339;    # child whose father is Bach
5         wdt:P106 wd:Q36834; # occupation composer
6         wdt:P106/wdt:P279+ wd:Q1075651. # occupation keyboardist or specialization
7     SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]" . }
8 }
```

child	childLabel
<a href="#">wd:Q106641</a>	Johann Christian Bach
<a href="#">wd:Q107277</a>	Wilhelm Friedemann Bach
<a href="#">wd:Q710996</a>	Gottfried Heinrich Bach
<a href="#">wd:Q57225</a>	Johann Christoph Friedrich Bach
<a href="#">wd:Q107277</a>	Wilhelm Friedemann Bach
<a href="#">wd:Q470198</a>	Johann Gottfried Bernhard Bach

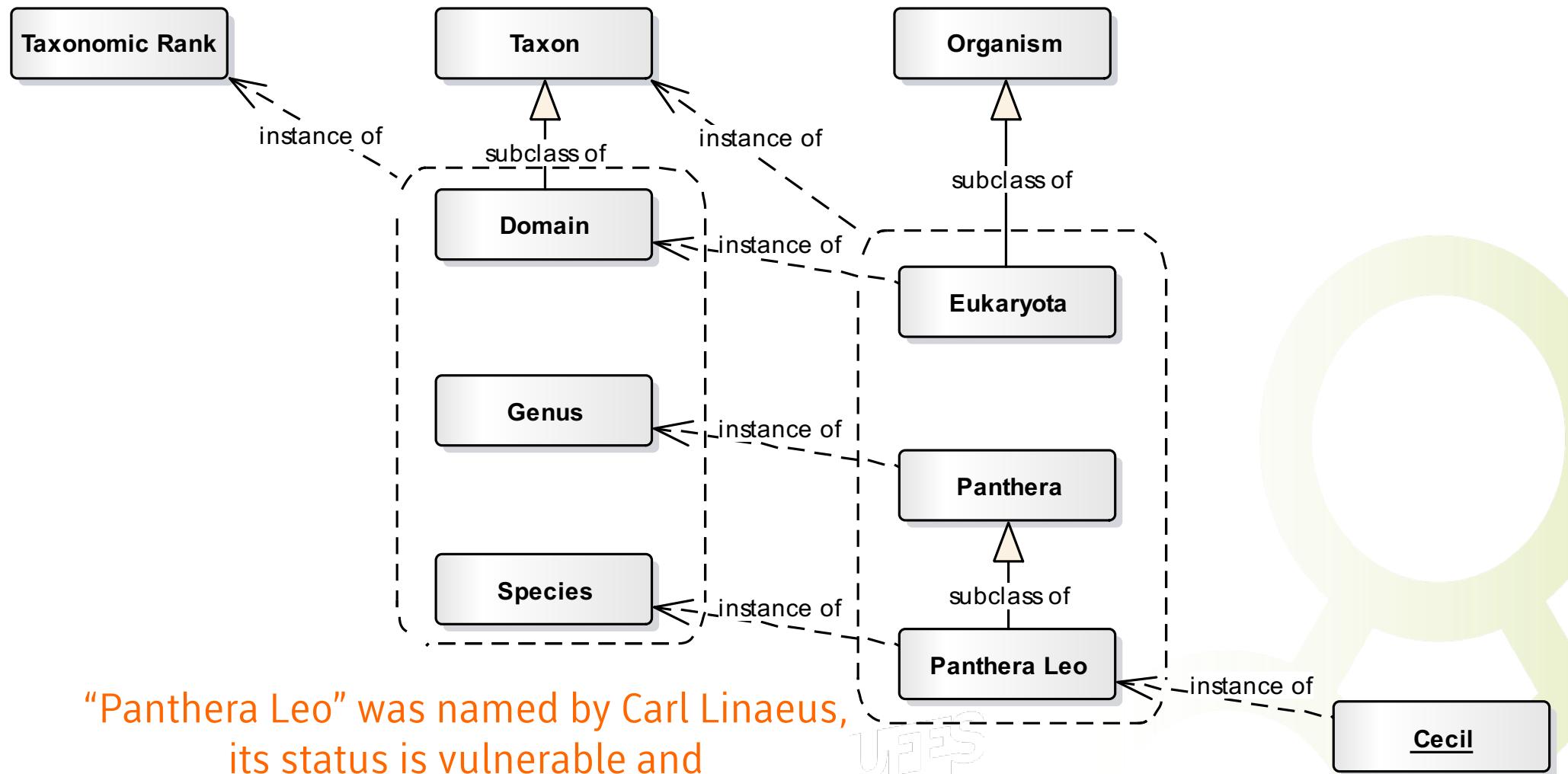
DIRETORIA DE CULTURA

DO ESPÍRITO SANTO

<https://w.wiki/BaHD>

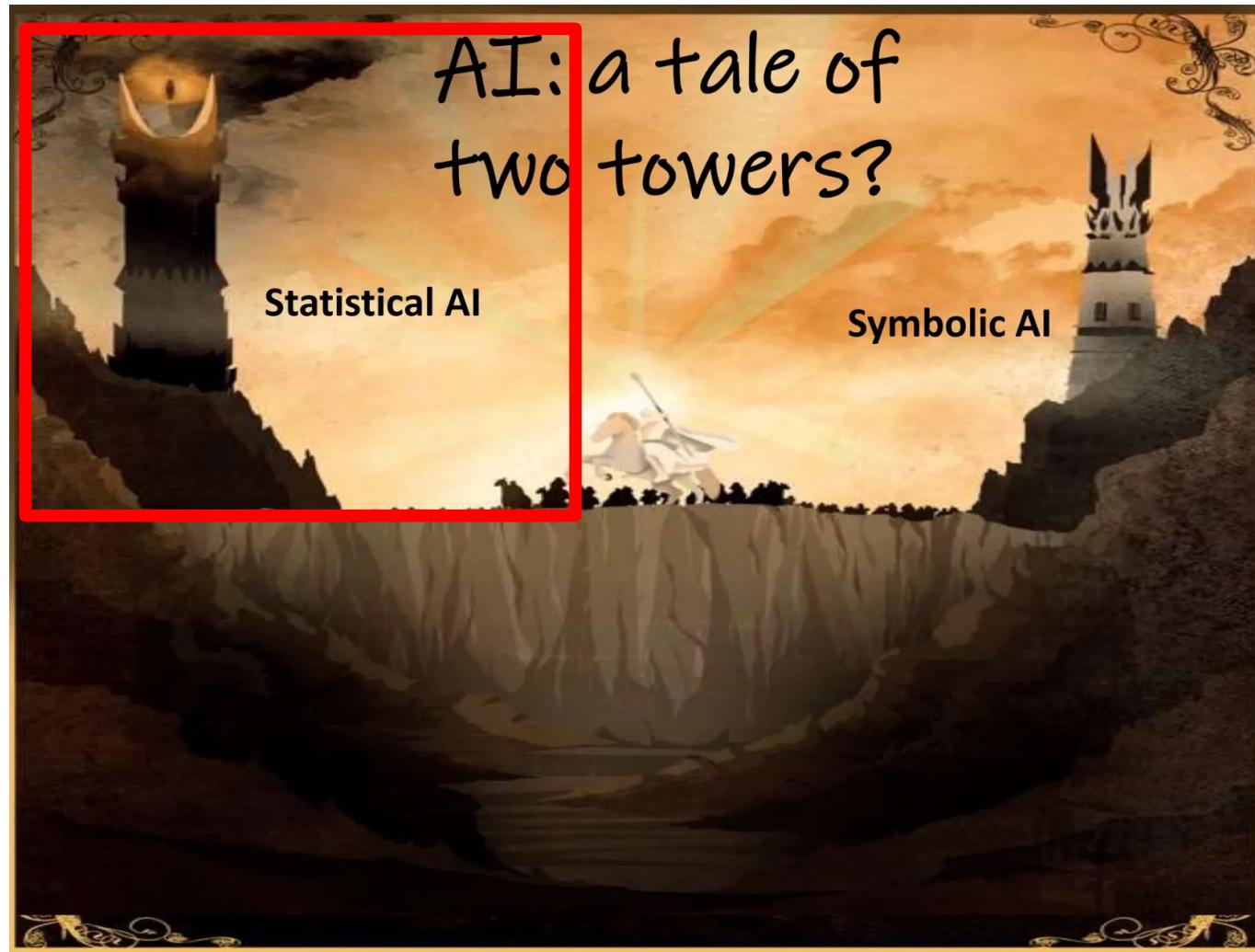
# Construtos

- Classificação
  - Tipos, classes, categorias
- Estruturas taxonômicas
  - é-um
- Instanciação
  - é-ístance-de
- Atributos
  - qualidades
- Relacionamentos
- Relações todo-parte
- Tipos de tipos



“Panthera Leo” was named by Carl Linnaeus,  
its status is vulnerable and  
all its instances are warmblooded

## Cenário atual: as duas torres (Frank van Harmelen)



# Exemplo: interações medicamentosas

ChatGPT 4o >



Any drug interactions between Pentobarbital and Acetaminophen?



There are no significant drug interactions between pentobarbital and acetaminophen.

Pentobarbital is a barbiturate, primarily used as a sedative or anticonvulsant, while acetaminophen is a non-opioid analgesic and antipyretic. They work through different mechanisms and do not generally affect each other's metabolism.

However, combining pentobarbital with other medications that depress the central nervous system (CNS) can enhance sedative effects, so caution is advised. Always consult with a healthcare professional before using these medications together, especially if other medications or conditions are involved.



**About:** [DDI between Pentobarbital and Acetaminophen - Barbiturates may increase the metabolism of Acetaminophen. This may 1\) diminish the effect of acetaminophen; and 2\) increase the risk of liver damage.](#) [drugbank\_resource:DB00312\_DB00316] [Goto](#) [Sponge](#) [NotDistinct](#) [Permalink](#)An Entity of Type : [http://bio2rdf.org/drugbank\\_vocabulary:Resource](http://bio2rdf.org/drugbank_vocabulary:Resource), within Data Space : [bio2rdf.org](http://bio2rdf.org) associated with source [document\(s\)](#).

Type: drugbank resource [drugbank\_vocabulary:Resource]

 New Facet based on Instances of this Class**Attributes****Values**[rdf:type](#)[drug-drug\\_interaction](#) [[drugbank\\_vocabulary:Drug-Drug-Interaction](#)]  
[drugbank\\_resource](#) [[drugbank\\_vocabulary:Resource](#)][rdfs:label](#)

DDI between Pentobarbital and Acetaminophen - Barbiturates may increase the metabolism of Acetaminophen. This may 1) diminish the effect of acetaminophen; and 2) increase the risk of liver damage. [drugbank\_resource:DB00312\_DB00316]

[dcterms:title](#)

DDI between Pentobarbital and Acetaminophen - Barbiturates may increase the metabolism of Acetaminophen. This may 1) diminish the effect of acetaminophen; and 2) increase the risk of liver damage.

[dcterms:identifier](#)

drugbank\_resource:DB00312\_DB00316

[void:inDataset](#)[http://bio2rdf.org/drugbank\\_resource:bio2rdf.dataset.drugbank.R3](http://bio2rdf.org/drugbank_resource:bio2rdf.dataset.drugbank.R3)[Bio2RDF identifier](#)

DB00312\_DB00316

[Bio2RDF namespace](#)

drugbank\_resource

[Bio2RDF uri](#)[http://bio2rdf.org/drugbank\\_resource:DB00312\\_DB00316](http://bio2rdf.org/drugbank_resource:DB00312_DB00316)is [ddi interactor in ...di-interactor-in](#) of [Pentobarbital](#) [[drugbank:DB00312](#)]  
[Acetaminophen](#) [[drugbank:DB00316](#)]

Barbiturates may increase the metabolism of Acetaminophen. This may 1) diminish the effect of acetaminophen; and 2) increase the risk of liver damage.

[https://bio2rdf.org/drugbank\\_resource:DB00312\\_DB00316](https://bio2rdf.org/drugbank_resource:DB00312_DB00316)

# Exemplo: usando os nomes comerciais...

ChatGPT 4o >

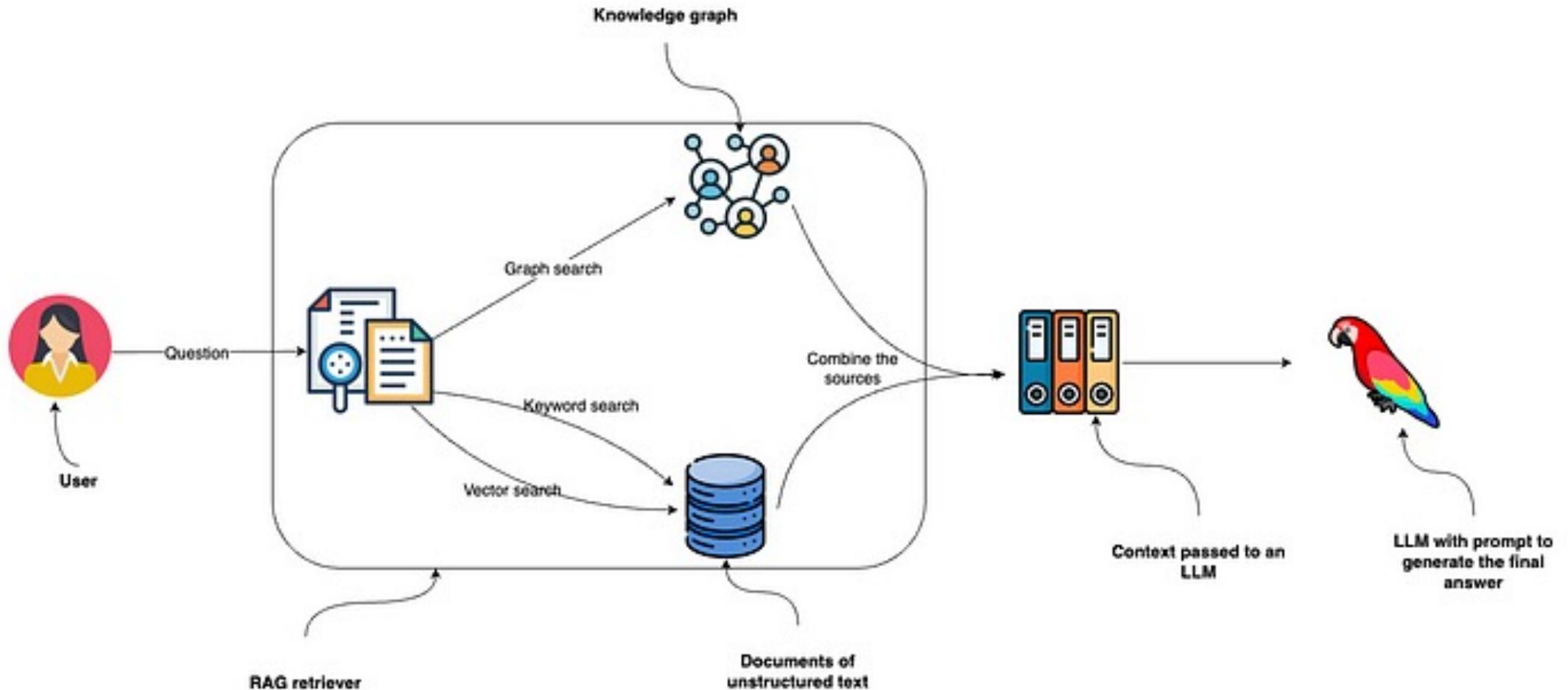


Any drug interactions between Nembutal and Tylenol?

 Nembutal (pentobarbital) and Tylenol (acetaminophen) do not have a direct drug interaction. However, caution is still needed when using them together:

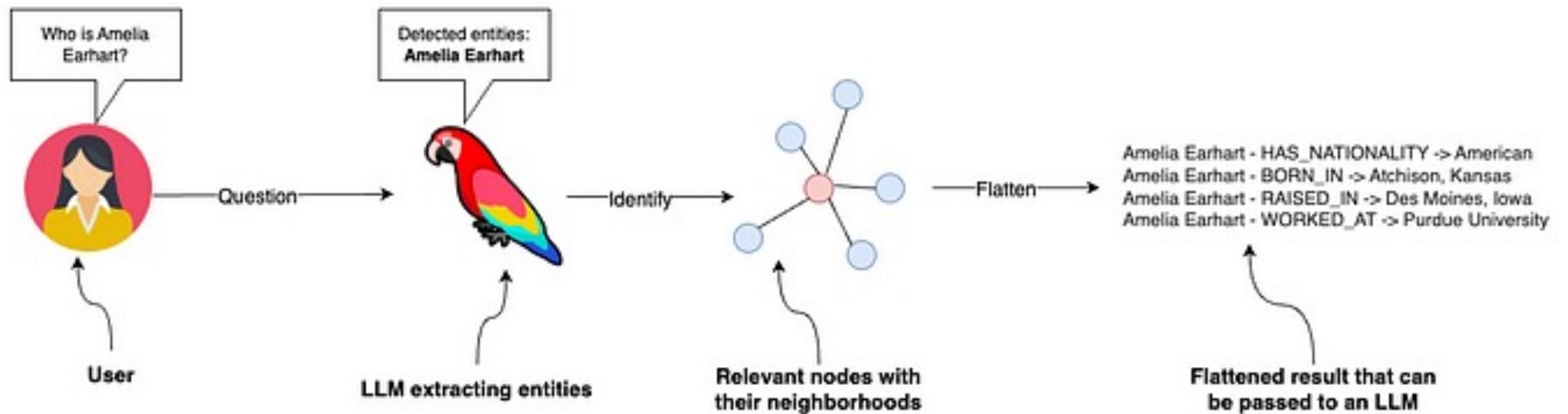
1. **Central Nervous System (CNS) Effects:** Nembutal is a barbiturate that acts as a sedative and can cause CNS depression. When taken with other medications that may have sedative properties or affect the liver, like acetaminophen, there could be an increased risk of drowsiness or sedation.
2. **Liver Considerations:** Tylenol is metabolized by the liver, and combining it with barbiturates like Nembutal, which can also affect liver enzymes, may increase the risk of liver toxicity, especially at higher doses or with chronic use.

# Graph Retrieval Augmented Generation



UNIVERSIDADE FEDERAL  
DO RIO DE JANEIRO  
<https://blog.langchain.dev/enhancing-rag-based-applications-accuracy-by-constructing-and-leveraging-knowledge-graphs/>

# Usando LLMs para consultas a grafos de conhecimento



<https://blog.langchain.dev/enhancing-rag-based-applications-accuracy-by-constructing-and-leveraging-knowledge-graphs/>

# Exemplo: OpenAI embeddings + DrugBank + ChatGPT

 DRUGBANK online      Explore ▾      For Drug Discovery ▾      For Clinical Software ▾      For Academic Research      

 OPEN DATA



The DrugBank Open Data datasets are public domain datasets that can be used freely in your application or project (including commercial use). It is released under a [Creative Commons CC0 International License](#). To the extent possible under law, the person who associated CC0 with the DrugBank Open Data has waived all copyright and related or neighboring rights to the DrugBank Open Data. This work is published from: Canada.

Dataset	Description	Released On	Version	Size	Format	Command	Download
DrugBank Vocabulary	DrugBank identifiers, names, and synonyms to permit easy linking and integration into any type of project.	2024-03-14	5.1.12	926 KB	CSV	<a href="#">Example</a>	<a href="#">Download</a>
DrugBank Structures	DrugBank structures, names, and synonyms to permit easy linking and integration into any type of project.	2024-03-14	5.1.12	4.29 MB	SDF	<a href="#">Example</a>	<a href="#">Download</a>

Agradecimentos a Renzo Leão (IC) e Profa. Veruska Zamborlini

# Exemplo: OpenAI embeddings + DrugBank + ChatGPT

DrugBank ID,Accession Numbers,Common name,CAS,UNII,Synonyms,Standard InChI Key  
DB00001,BTD00024 | BIOD00024,Lepirudin,138068-37-8,Y43GF64R34,"[Leu1, Thr2]-63-desulfohirudin | Desulfatohirudin | Hirudin variant-1 | Lepirudin | Lepirudin recombinant | R-hirudin",  
DB00002,BTD00071 | BIOD00071,Cetuximab,205923-56-4,PQX0D8J21J,Cetuximab | Cétuximab | Cetuximabum,  
DB00003,BTD00001 | BIOD00001,Dornase alfa,143831-71-4,953A260A1Y,"Deoxyribonuclease (human clone 18-1 protein moiety) | Dornasa alfa | Dornase alfa | Dornase alfa, recombinant | Dornase alpha | Recombinant deoxyribonuclease (DNase)",  
DB00004,BTD00084 | BIOD00084,Denileukin diftitox,173146-27-5,25E79B5CTM,Denileukin | Denileukin diftitox | Interleukin-2/diphtheria toxin fusion protein,  
DB00005,BTD00052 | BIOD00052,Etanercept,185243-69-0,OP401G70JC,Etanercept | etanercept-szzs | etanercept-ykro | Recombinant human TNF | rhu TNFR:Fc | rhu-TNFR:Fc | TNFR-Immunoadhesin,  
DB00006,BTD00076 | EXPT03302 | BIOD00076 | DB02351,Bivalirudin,128270-60-0,TN9BEX005G,Bivalirudin | Bivalirudina | Bivalirudinum,OIRCOABEOLEUMC-GEJPAHFPSA-N  
...  
DB04817,"",Metamizole,50567-35-6,934T64RMN(),Dipyrone | Metamizol, VWZTYCIRDMEY-UHFFFAOYSA-N



# Exemplo: OpenAI embeddings + DrugBank + ChatGPT

DrugBank ID,Accession Numbers,Common name,CAS,UNII,Synonyms,Standard InChI Key  
DB00001,BTD00024 | BIOD00024,Lepirudin,138068-37-8,Y43GF64R34,"[Leu1, Thr2]-63-des  
Desulfatohirudin | Hirudin variant-1 | Lepirudin | Lepirudin recombinant | R-hirud  
DB00002,BTD00071 | BIOD00071,Cetuximab,205923-56-4,PQX0D8J21J,Cetuximab | Cétuxima  
DB00003,BTD00001 | BIOD00001,Dornase alfa,143831-71-4,953A260A1Y,"Deoxyribonucleas  
moiety) | Dornasa alfa | Dornase alfa | Dornase alfa, recombinant | Dornase alpha  
deoxyribonuclease (DNase)",  
DB00004,BTD00084 | BIOD00084,Denileukin diftitox,173146-27-5,25E79B5CTM,Denileukin  
Interleukin-2/diphtheria toxin fusion protein,  
DB00005,BTD00052 | BIOD00052,Etanercept,185243-69-0,OP401G70JC,Etanercept | etaner  
| Recombinant human TNF | rhu TNFR:Fc | rhu-TNFR:Fc | TNFR-Immunoadhesin,  
DB00006,BTD00076 | EXPT03302 | BIOD00076 | DB02351,Bivalirudin,128270-60-0,TN9BEX0  
Bivalirudina | Bivalirudinum,OIRCOABEOLEUMC-GEJPAHFPSA-N

... DB04817, "",Metamizole,50567-35-6,934T64RMN(),Dipyrone | Metamizol, VWZTYCIRDMEY-UH

Adicionamos coluna  
de embedding  
usando "text-  
embedding-3-large"



# Passo 1: Gerar embeddings para os medicamentos

```
def get_embedding(text, model="text-embedding-3-large"):
    return client.embeddings.create(input = [text], model=model).data[0].embedding

df = pd.read_csv('drugbank vocabulary.csv')

# Process the 'Synonyms' column to replace '||' with 'or'
df['Synonyms'] = df['Synonyms'].fillna('').apply(lambda x: x.replace('||', 'or'))

# Combine 'Common name' and 'Synonyms' columns
df['Combined'] = df['Common name'].fillna('') + ' or ' + df['Synonyms'].fillna('')

# Calculate embeddings for each row and store them in a new column
df['Embedding'] = df['Combined'].apply(lambda x: get_embedding(x))

# Save the DataFrame to a CSV file
output_file_path = 'indexed-vocab.csv'
df.to_csv(output_file_path, index=False)
```



## Passo 2: Dados nomes de medicamentos, busca

```
term1 = get_embedding("Pentobarbital")
term2 = get_embedding("tylenol")
drug1=find_closest_embedding(df, term1)[0]
drug2=find_closest_embedding(df, term2)[0]
```

Importante para o exemplo: “tylenol” não ocorre no DrugBank que não tem nomes comerciais

Resposta do Passo 2:

```
drug1['DrugBank ID']=DB00312
drug1['DrugBank ID']=DB00316
```

## Passo 3: Busca no grafo de conhecimento

```
# Define the SPARQL query
query = """
PREFIX biovoc: <http://bio2rdf.org/bio2rdf_vocabulary:>
PREFIX dbvoc: <http://bio2rdf.org/drugbank_vocabulary:>
PREFIX dcterms: <http://purl.org/dc/terms/>
SELECT DISTINCT *
WHERE {
    <http://bio2rdf.org/drugbank:"""+drug1['DrugBank ID']+""">
        dbvoc:ddi-interactor-in ?interactionLink;
        biovoc:identifier ?drugIdentifier1.

    <http://bio2rdf.org/drugbank:"""+drug2['DrugBank ID']+""">
        dbvoc:ddi-interactor-in ?interactionLink;
        biovoc:identifier ?drugIdentifier2.

    ?interactionLink dcterms:title ?interactionDescription;
        biovoc:identifier ?interactionIDs.

}
"""
"""


```

## Resultado do Passo 3:

- [http://bio2rdf.org/drugbank\\_resource:DB00312\\_DB00316](http://bio2rdf.org/drugbank_resource:DB00312_DB00316)
- DDI between Pentobarbital and Acetaminophen - Barbiturates may increase the metabolism of Acetaminophen. This may 1) diminish the effect of acetaminophen; and 2) increase the risk of liver damage.

## Passo 4: LLM verbaliza com o contexto da consulta ao grafo

```
messages = [
    {"role": "system", "content": "You are a helpful assistant."
     +" Take into account that: "+interactionDescription},
    {"role": "user", "content": "Any drug interactions between
     "+label1+" and "+label2+"?"}
]

# Generate the chat completion
completion = client.chat.completions.create(
    model="gpt-4o",
    messages=messages
)
# Print the response
print(completion.choices[0].message)
```

## Resposta final enriquecida

Yes, there is a known drug interaction between pentobarbital and Tylenol (acetaminophen). Pentobarbital, a barbiturate, may increase the metabolism of acetaminophen. This interaction can potentially lead to two main effects:

- 1. Diminished Effect of Acetaminophen:** The increased metabolism may reduce the efficacy of acetaminophen, making it less effective for relieving pain or reducing fever.
- 2. Increased Risk of Liver Damage:** The enhanced metabolic activity can increase the production of toxic metabolites from acetaminophen, potentially leading to an increased risk of liver damage.

It's important to exercise caution when using these medications together and consult with a healthcare provider to ensure safe use. They might adjust the dosages or suggest alternative therapies to mitigate these risks.



# Mais possibilidades

**DRUGBANK Online** Explore ▾ For Drug Discovery ▾ For Clinical Software ▾ For Academic Research 

Type your search... 

## Interactions Found

	SEVERITY 	DESCRIPTION
<b>Pentobarbital [Nembutal]</b> 	<b>Acetaminophen [Tylenol]</b>	<b>MAJOR</b>
<b>EXTENDED DESCRIPTION</b>	The subject drug is a strong CYP3A4 inducer and the affected drug is metabolized by CYP3A4. Concomitant administration will increase the metabolism of the affected drug, leading to decreased serum concentrations and reduced therapeutic effect. <a href="#">READ LESS</a>	
<b>REFERENCES</b>	<ol style="list-style-type: none"><li>Zhou SF: Drugs behave as substrates, inhibitors and inducers of human cytochrome P450 3A4. <i>Curr Drug Metab.</i> 2008 May;9(4):310-22. <a href="#">[Article]</a></li><li>Lynch T, Price A: The effect of cytochrome P450 metabolism on drug response, interactions, and adverse effects. <i>Am Fam Physician.</i> 2007 Aug 1;76(3):391-6. <a href="#">[Article]</a></li><li>Klein K, Zanger UM: Pharmacogenomics of Cytochrome P450 3A4: Recent Progress Toward the "Missing Heritability" Problem. <i>Front Genet.</i> 2013 Feb 25;4:12. doi: 10.3389/fgene.2013.00012. eCollection 2013. <a href="#">[Article]</a></li><li>Drugs related to CYP3A4 <a href="#">[File]</a></li></ol> <a href="#">READ LESS</a>	



Main page  
Community portal  
Project chat  
Create a new Item  
Recent changes  
Random Item  
Query Service  
Nearby  
Help  
Donate

Lexicographical data

Create a new Lexeme  
Recent changes  
Random Lexeme

Tools

What links here  
Related changes  
Special pages  
Permanent link  
Page information  
Concept URI  
Cite this page  
Get shortened URL

Item

Discussion

Read

View history

Search Wikidata



▼

# paracetamol (Q57055)

chemical compound with antipyretic and pain relieving properties

edit

p-acetylaminophenol | acetaminofen4-(acetylamino)phenol

| N-acetyl-p-aminophenol4-hydroxyacetanilide | N-(4-hydroxyphenyl)acetamide

| p-hydroxyacetanilide | p-acetaminophenol | p-acetamidophenol | 4'-hydroxyacetanilide

| p-hydroxy-acetanilid | p-hydroxyphenolacetamide | 4-acetamidophenol | acetaminophen

| p-(acetylamino)phenol | APAP | L484

► In more languages

## Statements

instance of



type of chemical entity



▼ 0 references

+ add reference

+ add value

subclass of



acetamides



▼ 0 references

CosIng number

31569

 edit

 1 reference

 + add value

PesticideInfo chemical ID

PRI776

 edit

 0 references

 + add reference

 + add value

Nikkaji ID

J4.025H

 edit

 0 references

 + add reference

 + add value

UM-BBD compound ID

c1405

 edit

 0 references

 + add reference

 + add value

ATC code

▼ N02BE01

edit

▼ 1 reference

stated in DrugBank

DrugBank ID 00316



language of work or name English

title Acetamin  
ophen  
(English)

publication date 17  
Novembe  
r 2015

+ add reference

+ add value

World Health Organisation  
international non-proprietary  
name numeric ID

▼ 626

edit

▼ 0 references

+ add reference

+ add value

# Mainstream



## Introduction:

In Q1 2025, SAP is planning to release the [knowledge graph engine](#) as a new capability for SAP HANA Cloud platform. This new multi-model engine allows businesses to connect and analyze information more effectively. SAP HANA Cloud will enable smarter applications and improve user experiences by using data stored in a graph to answer complex questions.

Knowledge graphs connect corporate knowledge, which makes them especially beneficial for powering large language models (LLMs) and generative AI capabilities. Developers can use business context to create applications that provide more precise and relevant answers, enhancing user interactions and improving business processes.

For many, the topic of knowledge graphs is new and exciting. I share below a brief introduction on the major concepts and how SAP plans for those concepts to materialize in SAP HANA Cloud.

## Ontologies

An ontology is a guiding framework used to define at a higher level (i.e. classes, categories, groupings, etc.) the facts that occur in a domain and how those higher-level descriptions should relate to each other. Instead of focusing on the facts, an ontology regulates the resulting knowledge graph. For example, in the context of movies, an ontology could define the concept of "movies" and its attributes like title, director, genre, and release date, making the framework flexible and scalable for future data. Ontologies act as the blueprint for knowledge graphs, structuring and connecting data to allow deeper insights and reasoning.

<https://community.sap.com/t5/technology-blogs-by-sap/connecting-the-facts-sap-hana-cloud-s-knowledge-graph-engine-for-business/ba-p/13888597>

# Conclusões

- Sinergia entre as torres **nas duas direções**
- Várias arquiteturas possíveis de combinação
  - LLMs como componente
  - Grafo de conhecimento como componente
- Melhora significativa de resultados especialmente em domínios especializados
- Requer grafos de conhecimento de alta qualidade
- Outros tipos de sinergia:
  - Uso de LLMs para auxiliar na manutenção de grafos de conhecimento
  - LLMs para escrever e explicar queries

# NEMO

## Núcleo de Estudos em Modelagem Conceitual e Ontologias



Camila



João Paulo



### MEMBROS SÊNIOR

→ 18 anos de existência

→ +600 publicações nacionais e internacionais

→ +70 alunos formados (graduação e pós)

→ +20 orientações em andamento



Ricardo (in memoriam)



Veruska



Vítor



Giancarlo



Renata

### MEMBROS EXTERNOS



<https://nemo.inf.ufes.br>