



WIKIDATA LAB XXXII

Consultando o Wikidata

17/02/2022



Erika Guetti Suca



Quem sou eu?

Doutoranda em Ciências da Computação pela Universidade de São Paulo.

Atuando nas áreas de representação de conhecimento, privacidade de dados e Web Semântica.

Wikimedista-em-residência no Centro de Pesquisa, Inovação e Difusão em Neuromatemática (CEPID NeuroMat)



NeuroMat

Apoio
FAPESP



Agenda

1. Introdução
2. Conceitos Fundamentais
3. Wikidata e SPARQL
4. Operadores de SPARQL
5. Consultas de Contagem
6. Consultas de Agrupamento
7. Potencialidades gráficas
8. Bibliografía

The background features a series of vertical bars in three colors: red, green, and blue. The bars are arranged in a repeating pattern across the width of the page. The word 'Introdução' is centered in the white space between the bars.

Introdução

Introdução

- A Web Semântica é uma visão sobre uma extensão da Internet existente.
- *Em vez de humanos pesquisando manualmente em uma lista limitada de links, agora os algoritmos navegam em uma vasta quantidade de conjuntos de conteúdo cada vez mais estruturados disponíveis para responder especificamente ou agir sobre uma determinada consulta.*



Web Semântica

- Os dados são enriquecidos com semântica e links significativos e interpretáveis por máquina viabilizando a que os computadores encontrem e manipulem informações com maior precisão.
- As experiências semânticas ajudam a os humanos a encontrar experiências de conteúdo com menos discordância, enquanto uma máquina encontra, filtra, classifica, combina, organiza respostas mais reais.



Web Semântica, Web de Dados ou Web 3.0 .

- O primeiro site criado na Internet por Berners-Lee em 6 de agosto de 1991:

<http://info.cern.ch/hypertext/WWW/TheProject.htm>

- Estatísticas ao vivo da internet:

<https://www.internetlivestats.com/>



World Wide Web

The WorldWideWeb (W3) is a wide-area [hypermedia](#) information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including an [executive summary](#) of the project, [Mailing lists](#), [Policy](#), November's [W3 news](#), [Frequently Asked Questions](#).

What's out there?

Pointers to the world's online information, [subjects](#), [W3 servers](#), etc.

Help

on the browser you are using

Software Products

A list of W3 project components and their current state. (e.g. [Line Mode](#), [X11 Viola](#), [NeXTStep](#), [Servers](#), [Tools](#), [Mail robot](#), [Library](#))

Technical

Details of protocols, formats, program internals etc

Bibliography

Paper documentation on W3 and references.

People

A list of some people involved in the project.

History

A summary of the history of the project.

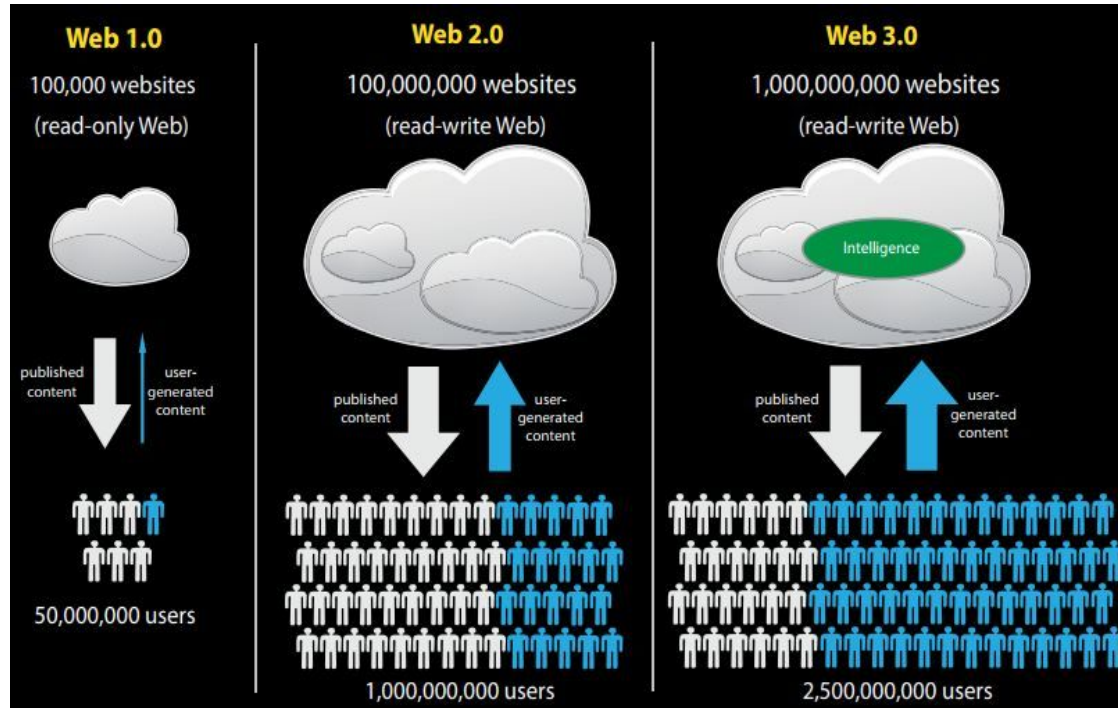
How can I help?

If you would like to support the web..

Getting code

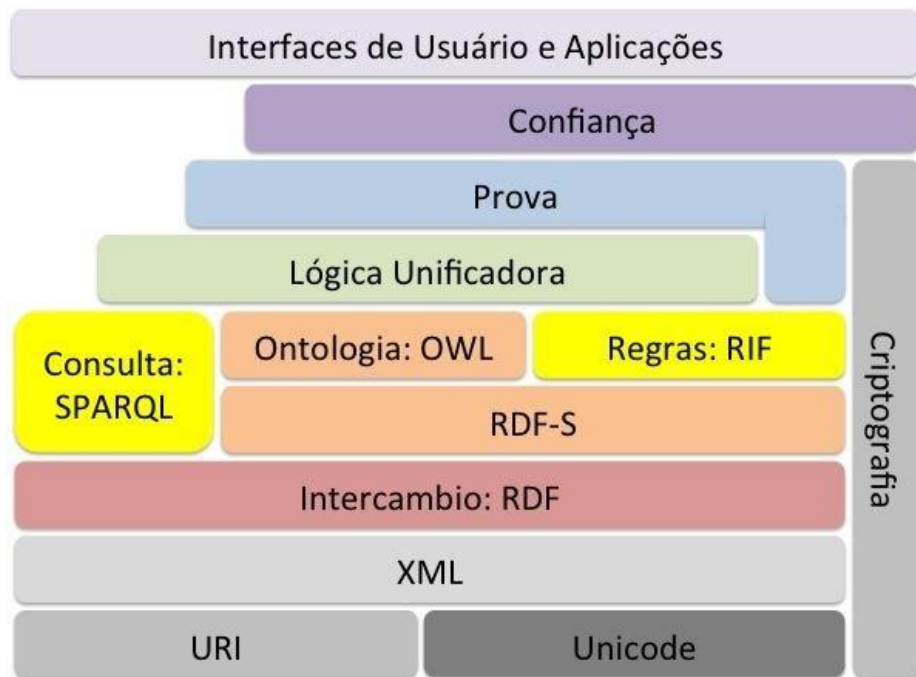
Getting the code by [anonymous FTP](#), etc.

Web Semântica, Web de Dados ou Web 3.0 .



Fonte: [Research Hubs](#)

As camadas da Web Semântica



Principais tecnologias que fundamentam a proposta da Web Semântica (Adaptada por Primo, 2013)



A Web Semântica

- A Web Semântica foi projetada para permitir que computadores, sistemas de software e pessoas trabalhem juntos em uma rede.
- Usa padrões que fornecem uma maneira uniforme e interoperável dados e conteúdo para que possam ser facilmente compartilhados e reutilizados entre aplicativos, empresas e indivíduos.
- Habilita a implementação de regras de inferência que permitem implementar raciocínio automatizado e o compartilhamento, reuso e geração dos dados, tanto entre pessoas quanto entre máquinas.

The background features a series of vertical bars in three colors: red, green, and blue. On the far left, there is a thick red vertical bar. To its right, there are two red bars of medium width, followed by two thin green bars, then a wide blue bar, a thin blue bar, another wide blue bar, and a thin green bar. On the far right, there is a thick green vertical bar. The text "Conceitos Fundamentais" is centered horizontally between the top and bottom rows of these bars.

Conceitos Fundamentais



SPARQL

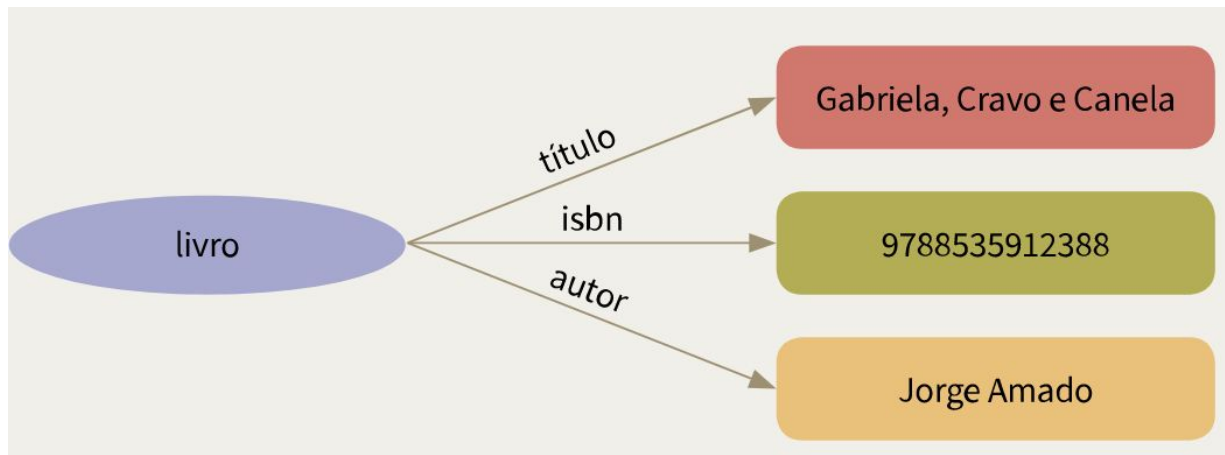
- **SPARQL** é um acrónimo do inglês **S**imple **P**rotocol **a**nd **R**DF **Q**uery **L**anguage.
- SPARQL é a linguagem para consulta, recuperação e manipulação dos dados armazenados no formato **RDF** proposta World Wide Web Consortium (W3C).
- SPARQL possibilita a navegação em diversos bancos de dados de triplas e descobrimento de relacionamentos entre os dados.

RDF

- **RDF (Resource Description Framework)** é uma família de especificações do W3C originalmente projetada como um modelo de dados de metadados .
- **RDF** modela afirmações sobre recursos, particularmente recursos da web, em expressões da forma sujeito - predicado - objeto , conhecidas como *triplas* .



RDF





URI

- Um **Uniform Resource Identifier (URI)** é uma sequência única de caracteres que identifica um recurso lógico ou físico usado por tecnologias da web.
- Cada Sujeito ou Propriedade tem um URI permanente.



Wikidata e SPARQL



Wikidata

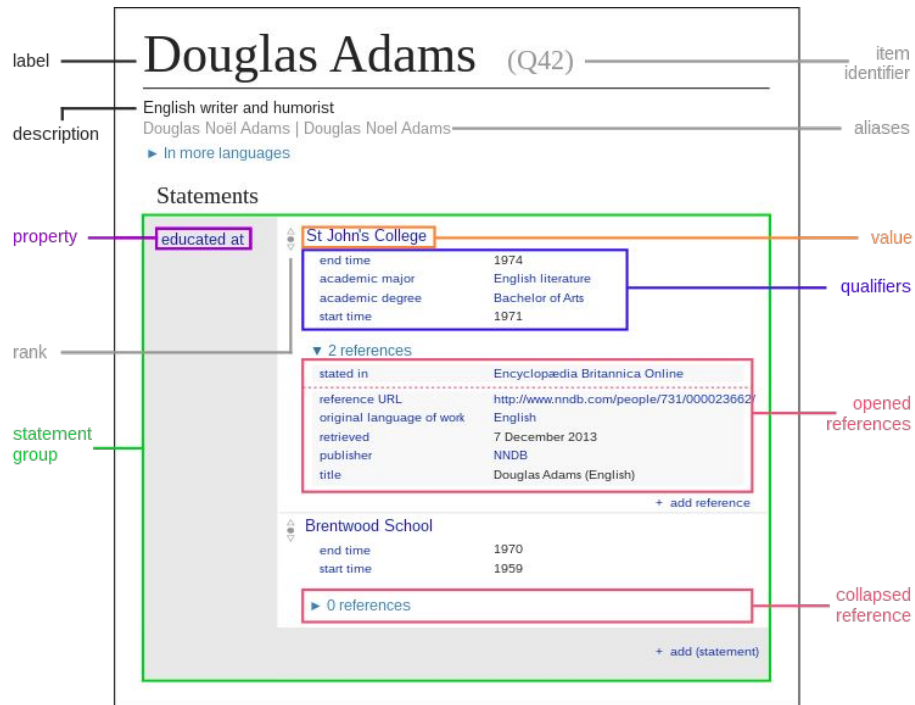
- Wikidata é uma base de conhecimento livre e aberta que pode ser editada tanto por humanos quanto por máquinas.
- É uma plataforma que permite que as informações sejam editáveis por qualquer pessoa com acesso à internet.
- Possibilita a curadoria do conhecimento.
- O Wikidata atua como um repositório central para os dados estruturados de seus projetos irmãos, como a Wikipédia, o Wikivoyage, o Wikcionário, o Wikisource, entre outros.



Porque usar Wikidata?

- Porque queremos preservar ao longo do tempo acesso a nossos dados.
- Porque queremos criar infraestruturas de conteúdo digital inteligentes.
- Porque queremos documentar os relacionamentos entre o software, hardware, formatos dos dados, configurações do entorno, e dados não estruturados (vídeos, imagens, sons, etc), e outros..
- Porque queremos compartilhar conhecimento.
- Porque queremos possibilitar a colaboração.

Modelo de dados do Wikidata





Wikidata e dados ligados

- O Wikidata armazena as informações na forma de **declarações** (ou statements).
- Conhecidas como triplas de Sujeito - Predicado e Objeto, seguem uma estrutura de
 - Item – Propriedade – Valor.
 - Item – Propriedade – Item.



Wikidata e dados ligados

- Um **item** é uma coisa, uma entidade, um conceito. Um objeto, uma pessoa, um evento, um lugar, uma obra de arte, etc. Um conceito abstrato como amor ou socialismo.
- Um item é identificado exclusivamente pela letra Q, seguida de um número. Por exemplo:
- O item <https://www.wikidata.org/wiki/Q17738> representa o filme “Star Wars” de 1977.
- O **URI** de conceito de Douglas Adams é <http://www.wikidata.org/entity/Q42>. Este URI se refere à pessoa no mundo real.



Wikidata e dados ligados

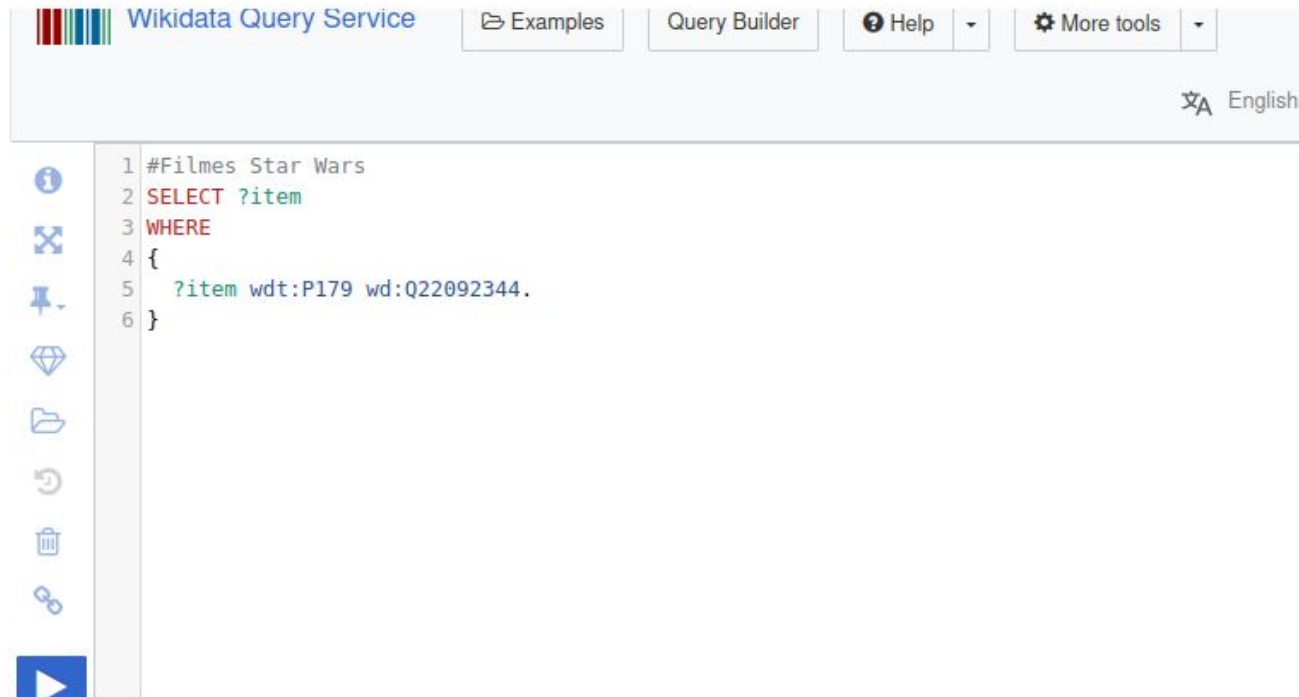
- As propriedades são representadas por um P seguido de um número. Exemplo:
- Duração (P2047): período de tempo de um evento ou processo
- **URI** <https://www.wikidata.org/wiki/Property:P2047>



Wikidata e dados ligados

- O Wikidata possui 17 tipos de dados distintos no total.
- Por exemplo, o valor da propriedade “duração” (P2047) para o item [Q17738](#) é ‘121 minutos’, uma **quantidade**.
- O valor de “data de publicação” (P577) nos Estados Unidos é “25 de maio de 1977”, uma **data**.
- Tipos de dados comumente usados são **strings**, **coordenadas geográficas** e **textos monolínguais** (uma string que não é traduzida para outros idiomas).

Estrutura consulta SPARQL



The image shows the Wikidata Query Service interface. At the top, there is a navigation bar with the Wikidata logo, the text "Wikidata Query Service", and several menu items: "Examples", "Query Builder", "Help", and "More tools". On the right side of the navigation bar, there is a language selector showing "English".

The main area of the interface displays a SPARQL query in a text editor. The query is as follows:

```
1 #Filmes Star Wars
2 SELECT ?item
3 WHERE
4 {
5   ?item wdt:P179 wd:Q22092344.
6 }
```

On the left side of the text editor, there is a vertical toolbar with various icons for editing and executing the query, including an information icon, a refresh icon, a pin icon, a diamond icon, a folder icon, a refresh icon, a trash icon, a link icon, and a play button icon.



Operadores de SPARQL

Operador “UNION”

```
SELECT ?item ?itemLabel
```

```
WHERE {
```

```
{?item wdt:P179 wd:Q22092344.} # item filmes Star Wars
```

```
UNION
```

```
{?item wdt:P179 wd:Q1102217.} # item filmes Star Trek
```

```
SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]". }
```

```
}
```

Operador “OR”

```
SELECT ?item ?itemLabel
```

```
WHERE {
```

```
  ?item wdt:P179 ?series.
```

```
  FILTER (?series = wd:Q22092344 || ?series = wd:Q1102217)
```

```
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]". }
```

```
}
```



Consultas de Contagem



Consultas de Contagem

- DISTINCT: elimina resultados duplicados.
- COUNT, SUM, MAX, MIN: Determinam o número de recursos distintos que atendem a determinados critérios.
- AVG: Calculando a média dos recursos que atendem determinados critérios.

Consultas de Contagem

```
SELECT DISTINCT ?diretor ?diretorLabel ?data
WHERE {
  ?item wdt:P179 wd:Q22092344. # item faz parte da saga de filmes Star Wars
  # o valor da propriedade "diretor" de ?item é armazenado na variável ?diretor
  ?item wdt:P57 ?diretor .

  # a data de nascimento do diretor é armazenado em ?data
  ?diretor wdt:P569 ?data.

  SERVICE wikibase:label { bd:serviceParam wikibase:language [AUTO_LANGUAGE]". }
}
```

Consultas de Contagem

```
SELECT (COUNT (?item) as ?contadorAtrizes)
```

```
WHERE {
```

```
?item wdt:P31 wd:Q5.      # Item é uma instância de ser humano.
```

```
?item wdt:P21 wd:Q6581072.  # Item tem sexo / gênero feminino.
```

```
?item wdt:P106 wd:Q33999.    # Item tem ocupação: atriz.
```

```
SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]" . }
```

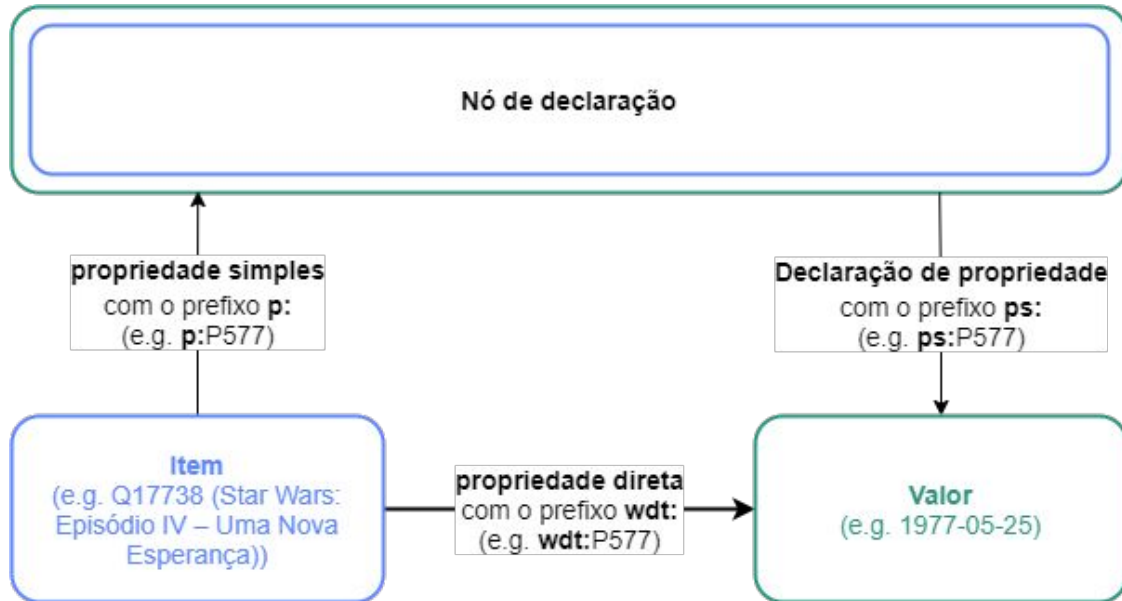
```
}
```



Consultas de Agrupamento

Consultas de Agrupamentos

- Qualificadores: permitem expandir o que pode ser expresso com um simples par de propriedades-valor.

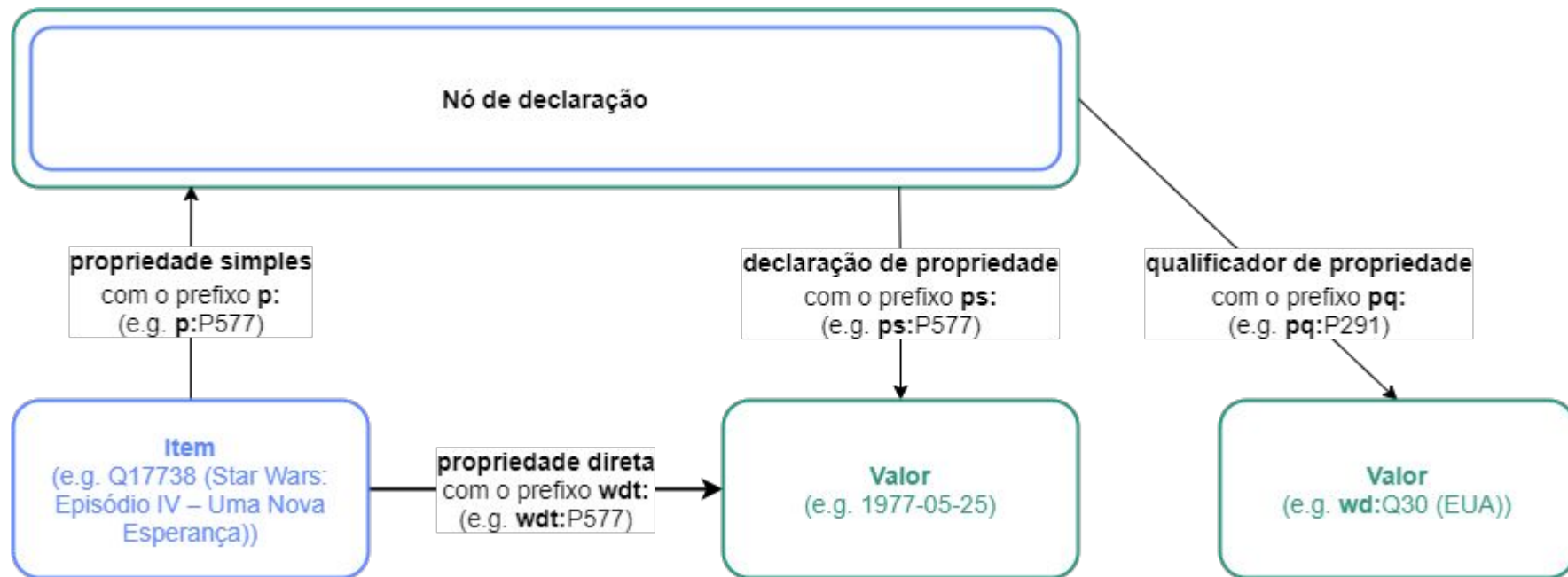


Consultas de Agrupamentos

Qualificadores: uma forma de descrever as informações é por meio das seguintes declarações:

Item	Propriedade	Valor
Q17738	wdt:P577	1977-05-25
Q17738	p:P577	Q17738 – nó de declaração
Q17738 – nó de declaração	ps:P577	1977-05-25

Consultas de Agrupamentos





Consultas de Agrupamentos

- ORDER BY (DESC ou ASC): classifica os resultados por algum critério. A ordem de classificação pode ser crescente ou decrescente.
- HAVING: permite afinar um critério de agrupamento.
- LIMIT: limita a lista de resultados em a um número natural.

Consultas de Agrupamentos

```
SELECT ?publisher ?publisherLabel (AVG(?pages) AS ?avgPages)
```

```
WHERE {
```

```
  ?book wdt:P123 ?publisher;    wdt:P1104 ?pages.
```

```
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE]". }
```

```
}
```

```
GROUP BY ?publisher ?publisherLabel
```

```
HAVING(COUNT(?book) > 1)
```

```
ORDER BY DESC(?avgPages)
```

Consultas de Agrupamentos

publisher	publisherLabel	avgPages
Q wd:Q4002388	UTET	7841.3333333333333333333333333333
Q wd:Q40380767	Q40380767	5702.5
Q wd:Q2077283	Gale	3577.5
Q wd:Q12297010	Trud	2621
Q wd:Q3375355	Pergamon Press	2581
Q wd:Q4080157	Bashkir Encyclopedia, publisher	2500
Q wd:Q108460866	Wood Egg LLC	2418
Q wd:Q5953914	Shanghai Lexicographical Publishing House	1743



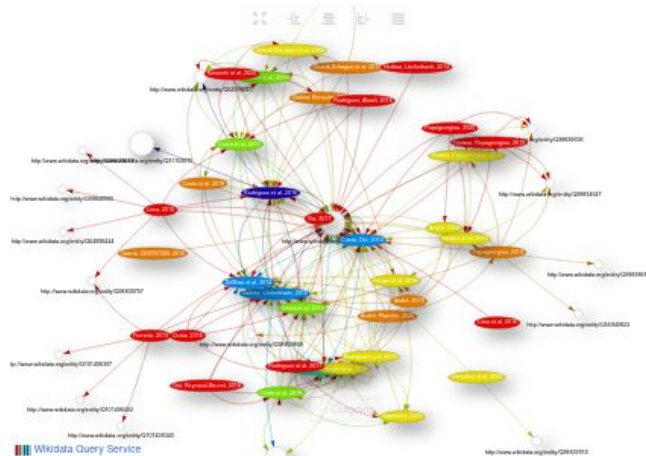
Potencialidades Gráficas

Caso de exemplo: Vitrine do NeuroMAT

NeuroMat (Centro de Pesquisa, Inovação e Difusão em Neuromatemática)

O artigo [Infinite Systems of Interacting Chains with Memory of Variable Length—A Stochastic Model for Biological Neural Nets](#), escrito por [Antonio Galves](#) e [Eva Löcherbach](#) em 2013, constitui a pedra angular do [Centro de Pesquisa, Inovação e Difusão em Neuromatemática \(CEPID NeuroMat\)](#). Este trabalho introduziu um novo modelo para redes neurais e lançou uma nova linha de investigação.

No contexto do CEPID NeuroMat, participaram desta linha [105](#) pesquisadores que produziram [37](#) artigos com base no modelo de 2013 ou contribuições posteriores. O modelo fundador do NeuroMat, chamado [Modelo Galves-Löcherbach](#), foi também o objeto de análise de publicações não filiadas ao NeuroMat, como: [Replica-Mean-Field Limits for Intensity-Based Neural Networks](#).



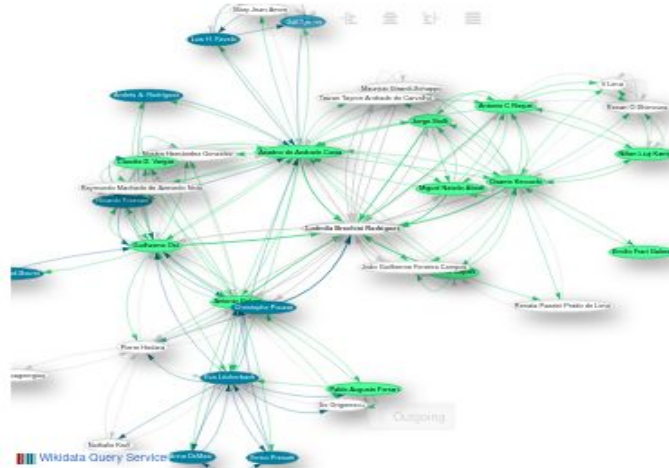
[Experimente a consulta completa!](#)

O artigo que lançou o [Modelo Galves-Löcherbach](#) teve impacto em várias áreas do conhecimento. As publicações associadas ao novo modelo deram-se em [11](#) revistas académicas, em áreas como probabilidade, estatística, computação e neurobiologia.

Caso de exemplo: Vitrine do NeuroMAT

Publicações

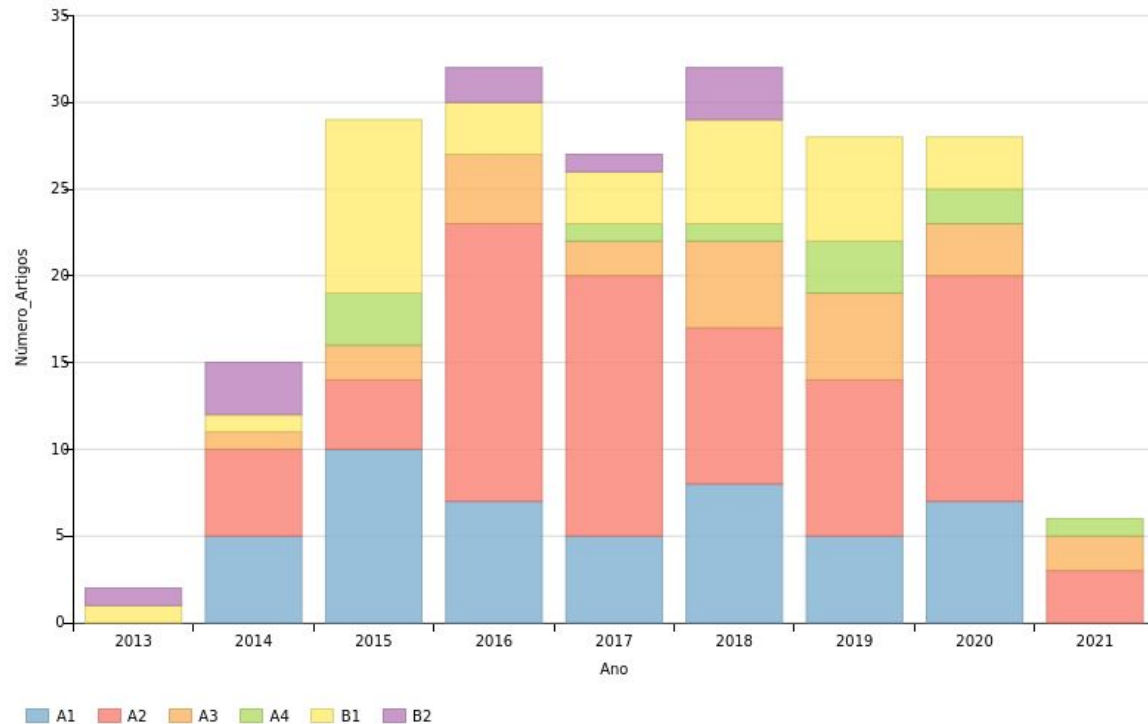
Desde a sua criação em 2013, a equipe de pesquisadores do CEPID NeuroMat produziu 361 artigos e foram citados 776 vezes. Como ilustrado abaixo, a produção dependeu de uma rede ampla de co-autores, com 1157 pessoas. [Experimente a consulta completa!](#)



As três principais revistas em que a equipe do NeuroMat publicou são: *Journal of Statistical Physics*, *Scientific Reports* e *Physical Review E*, respectivamente com 15, 11 e 9 artigos. Essas revistas têm classificação [Qualis Capes](#), em ordem, B1, A2 e A2.

No gráfico abaixo é possível visualizar a quantidade de artigos publicados em revistas difusoras das pesquisas por ano, desde o surgimento do CEPID NeuroMat, a partir da classificação B2 do Qualis Capes.

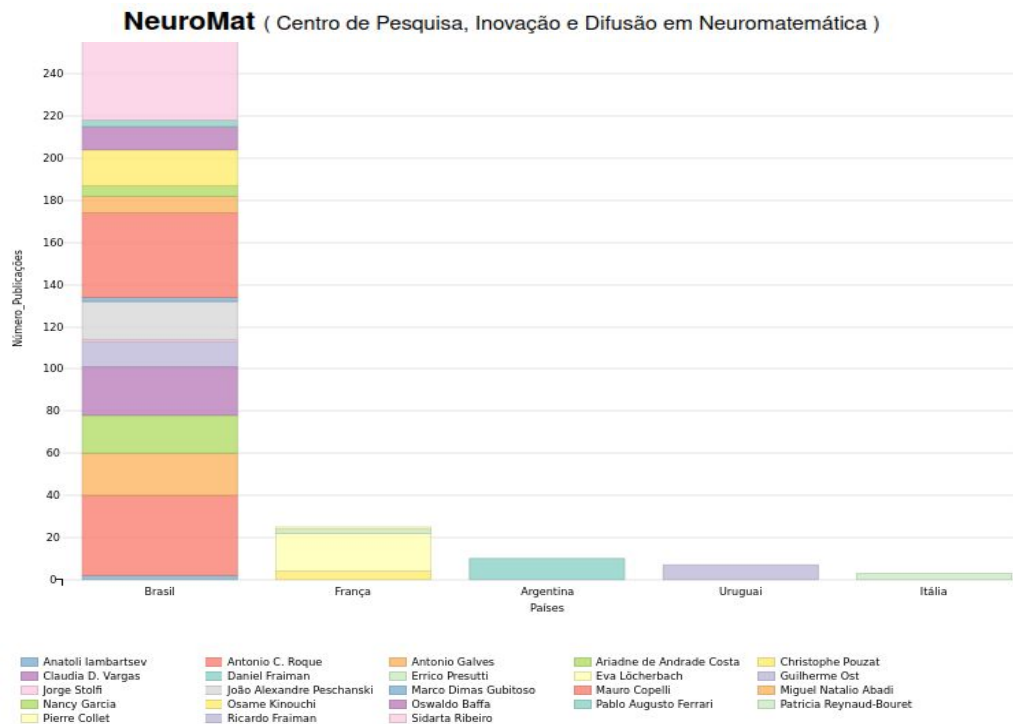
Caso de exemplo: Vitrine do NeuroMAT



Caso de exemplo: Vitrine do NeuroMAT



Caso de exemplo: Vitrine do NeuroMAT



Bibliografia

1. <https://wdqs-tutorial.toolforge.org/?lang=pt-br>
2. https://www.wikidata.org/wiki/Wikidata:SPARQL_tutorial
3. <https://www.w3c.br/Padroes/WebSemantica>
4. https://stringfixer.com/pt/Dereferenceable_Uniform_Resource_Identifier
5. <https://wmnobraasil.org/as-potencialidades-do-wikidata/>
6. <https://ceweb.br/guias/web-semantica/>
7. Tiago Primo Thompsen. Método de representação de conhecimento baseado em Ontologias para apoiar Sistemas de Recomendação Educacionais. 2013. Disponível em: <<http://www.lume.ufrgs.br/handle/10183/83654>>. Acesso em: 02 de Setembro de 2017.
8. <https://simplea.com/Articles/what-is-the-semantic-web>
9. Wikidata as a semantic framework for the Gene Wiki initiative. Sebastian Burgstaller-Muehlbacher, et al. 2016.
10. <https://vitrine.numec.prp.usp.br/>