

Evolução da Web Semântica

Prof. Dr. José de Jesús Pérez Alcázar EACH - USP

Baseado no curso do Gandon, F. Web of Data Course. Coursera.

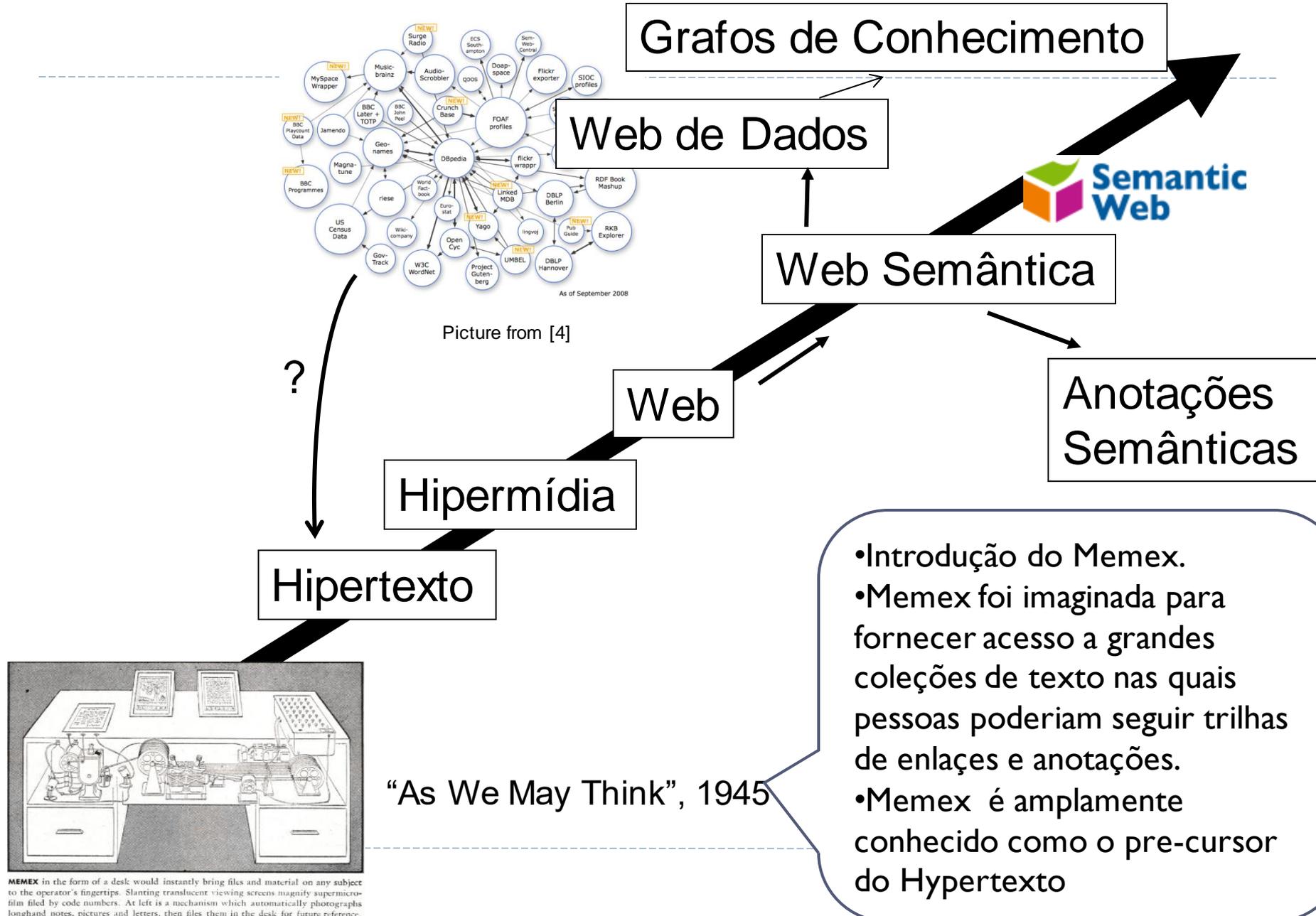
<https://www.coursera.org/learn/web-data/home/welcome>

Agenda

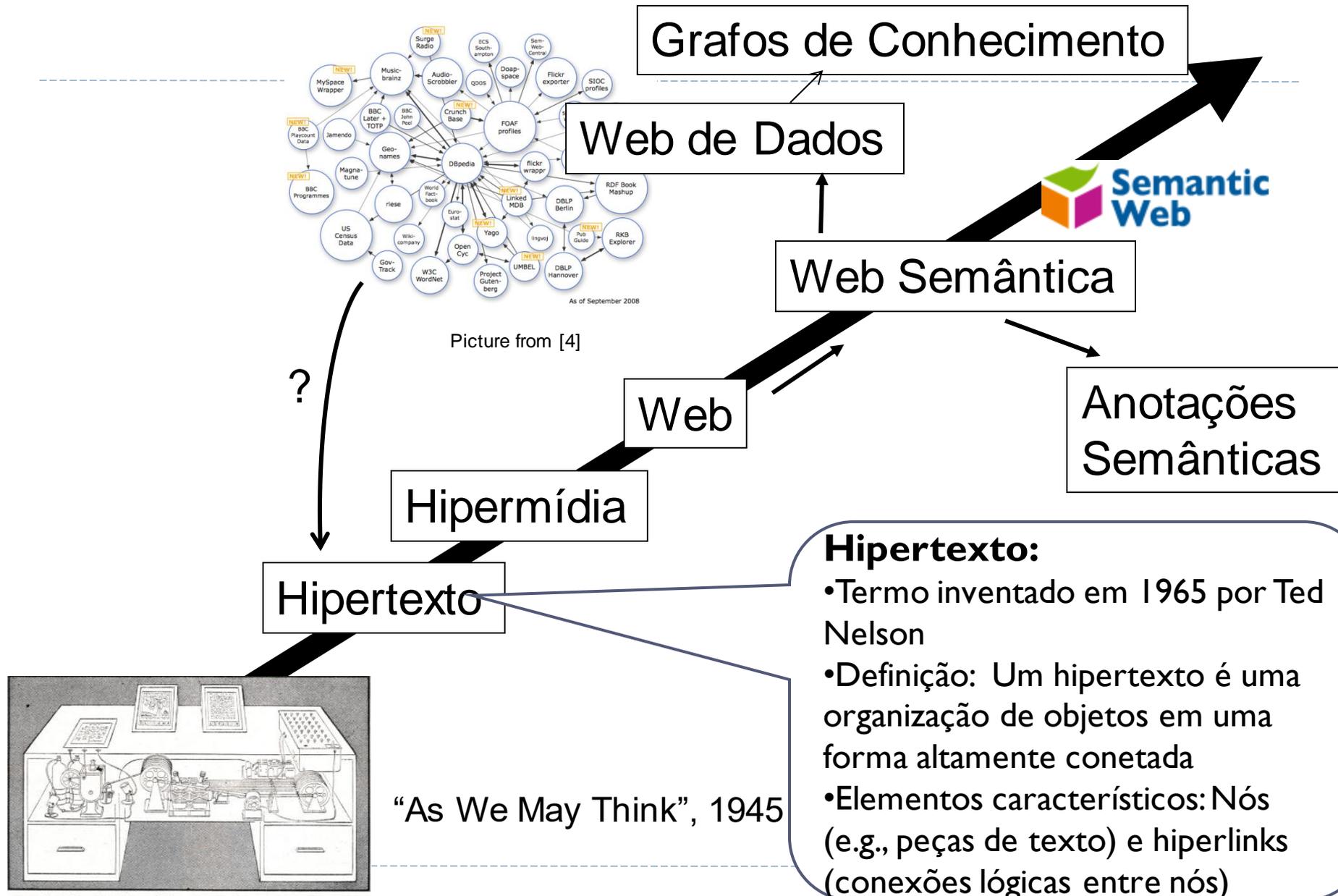
- ▶ Evolução da Web Semântica
- ▶ Web de Dados e dados ligados
- ▶ Tecnologias associadas
- ▶ RDF
- ▶ SPARQL
- ▶ Dados ligados abertos
- ▶ Grafos de Conhecimento
- ▶ Referências



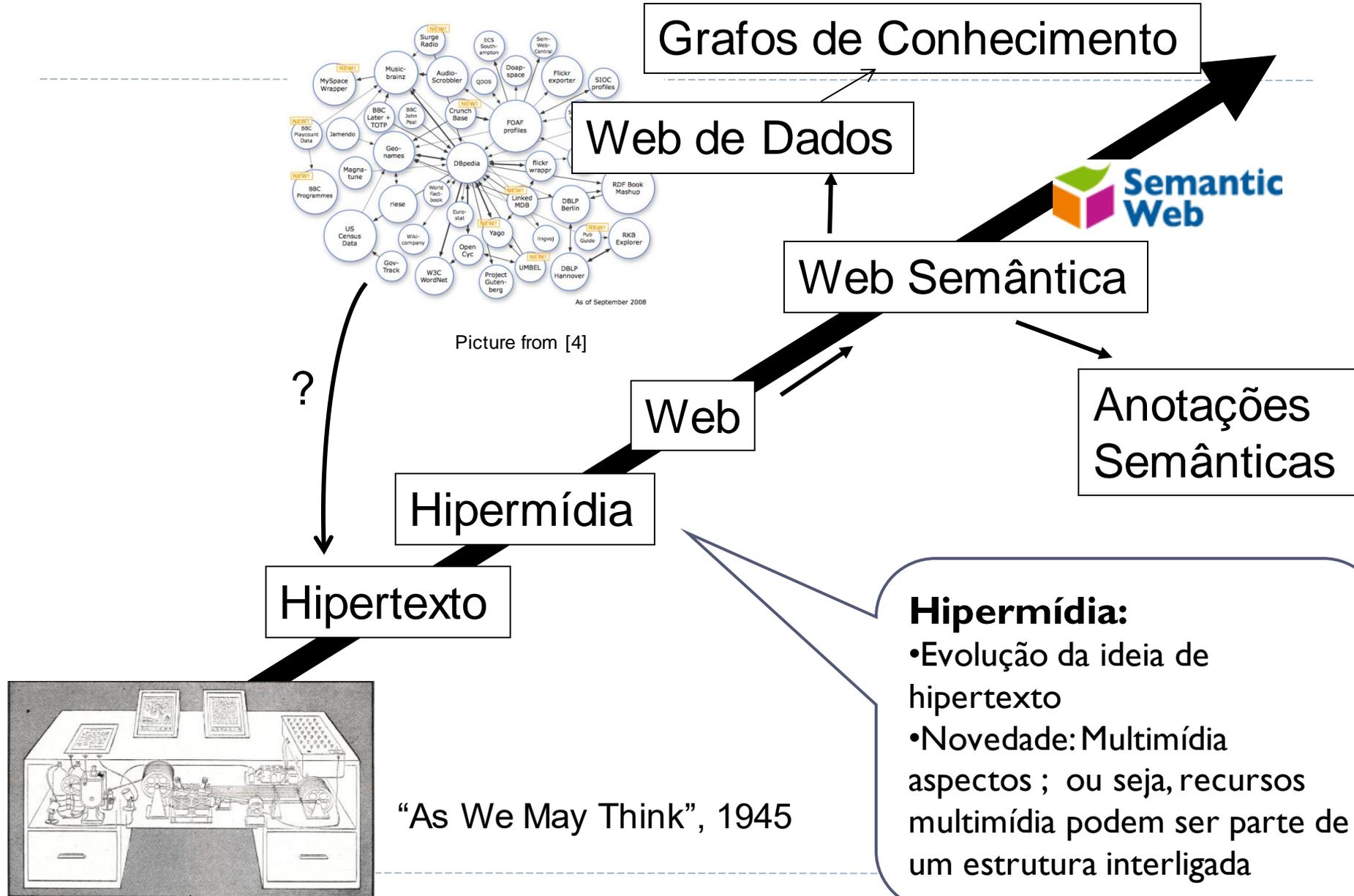
Evolução da Web Semântica



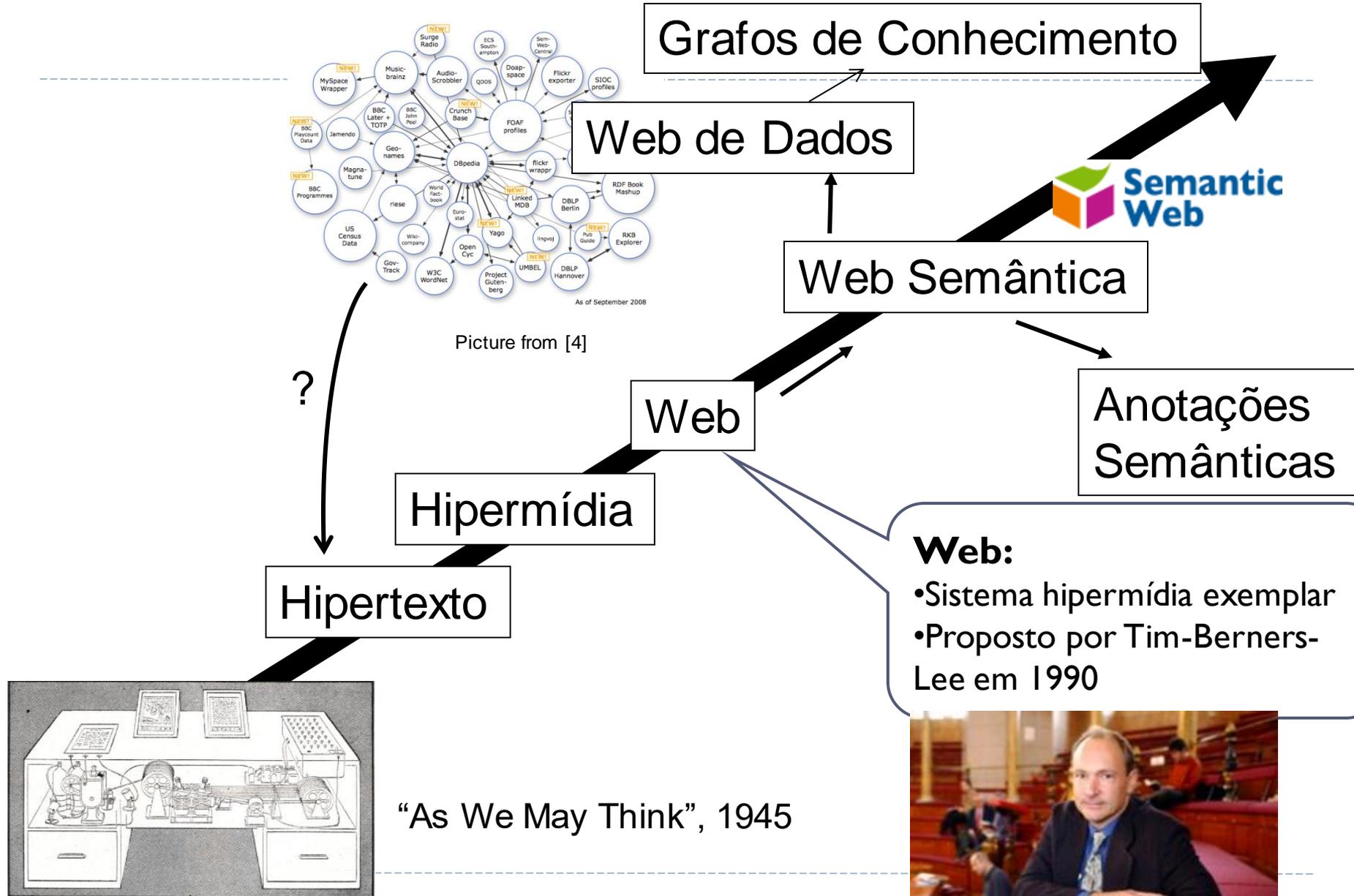
Evolução da Web Semântica



Evolução da Web Semântica

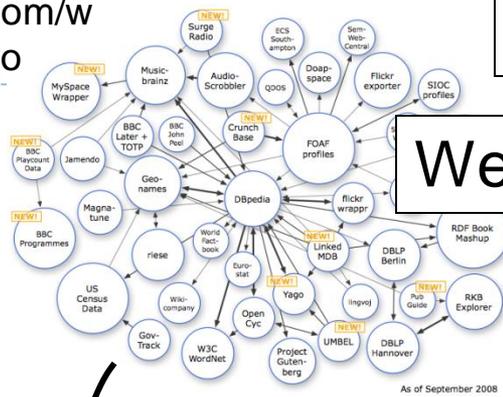


Evolução da Web Semântica



Evolução da Web Semântica

https://www.youtube.com/watch?v=OM6XIIcm_qo



Picture from [4]

Grafos de Conhecimento

Web de Dados

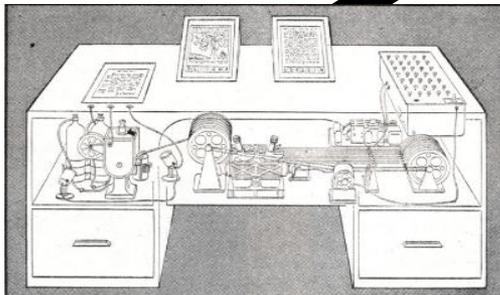
Web Semântica

Web

Hipermídia

Hipertexto

Anotações Semânticas



“As We May Think”, 1945

MEMEX in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference.

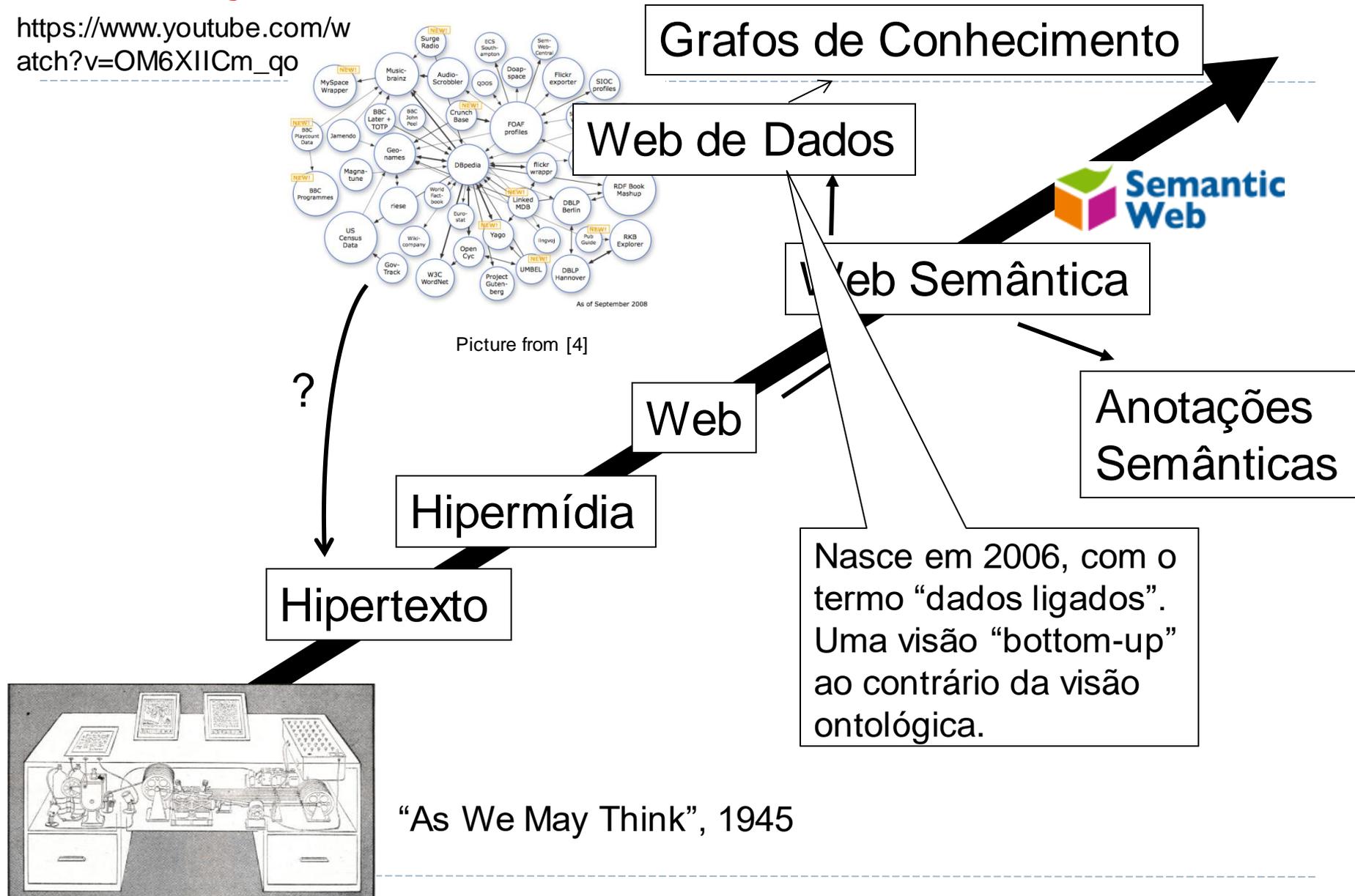
Anotações Semânticas:

- Anotações são geradas para a web existente
- Geração automática, semi-automática, ou baseada manualmente na entrada humana



Evolução da Web Semântica

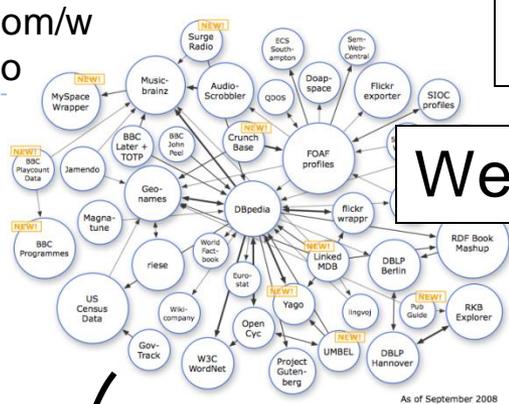
https://www.youtube.com/watch?v=OM6XIICm_qo



MEMEX in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference.

Evolução da Web Semântica

https://www.youtube.com/watch?v=OM6XIICm_qo



Picture from [4]

Grafos de Conhecimento

Web de Dados

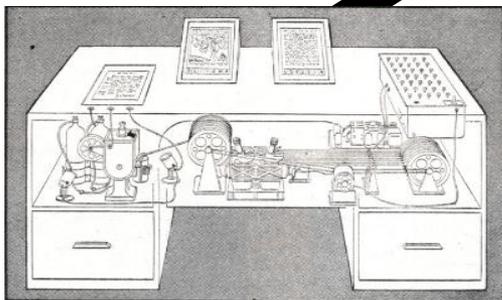
Web Semântica

Web

Hipermídia

Hipertexto

Anotações Semânticas



“As We May Think”, 1945

Em 2012, Google lança o termo Grafo de Conhecimento baseado na ontologia simples schema.org. Conceito bastante usado atualmente na industria: Microsoft, IBM, Facebook, eBay, etc.

MEMEX in the form of a desk would instantly bring files and material on any subject to the operator's fingertips. Slanting translucent viewing screens magnify supermicrofilm filed by code numbers. At left is a mechanism which automatically photographs longhand notes, pictures and letters, then files them in the desk for future reference.

Evolução da Web Semântica

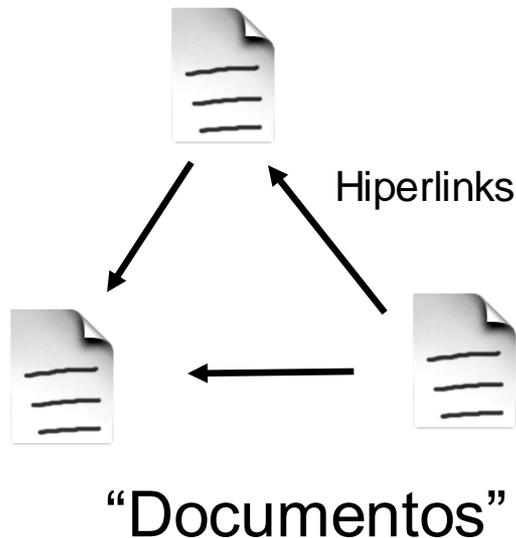


Fonte: Hitzler, P., 2021

-
- Como resultado da primeira fase da web semântica focada em Ontologias e raciocínio, se deram muitos resultados principalmente na área de medicina e biologia.
 - Ontologias a grande escala foram criadas e são amplamente usadas como: Gene Ontology (2008) e a SNOMED CT (www.snomed.org) convertida para OWL.
 - Ferramentas como Protégé (<https://protege.stanford.edu/>), também foram criadas. Veja uma lista completa na W3C (<https://www.w3.org/wiki/SemanticWebTools>).
 - Entretanto, construir Ontologias é todo um trabalho de Engenharia, o que deu passo a uma web semântica mais sóbria: a Web de dados (dados ligados).
-

Da Web de documentos à Web de dados

▶ Web de Documentos



▶ Elementos fundamentais:

1. Nomes (URIs)
2. Documentos (Recursos) descritos em HTML, XML, etc.
3. Interações via HTTP
4. (Hiper)enlaces entre documentos ou âncoras nesses documentos

▶ Deficiências:

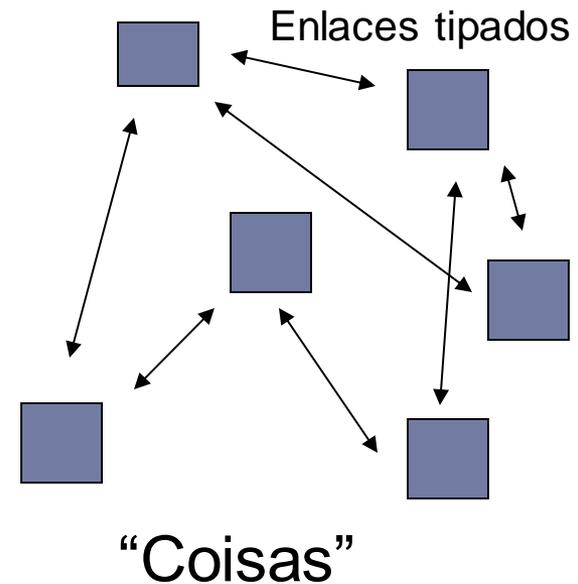
- ▶ Enlaces não tipados
- ▶ Motores de busca na web falham nas consultas complexas

Da Web de documentos à Web de dados

▶ Características:

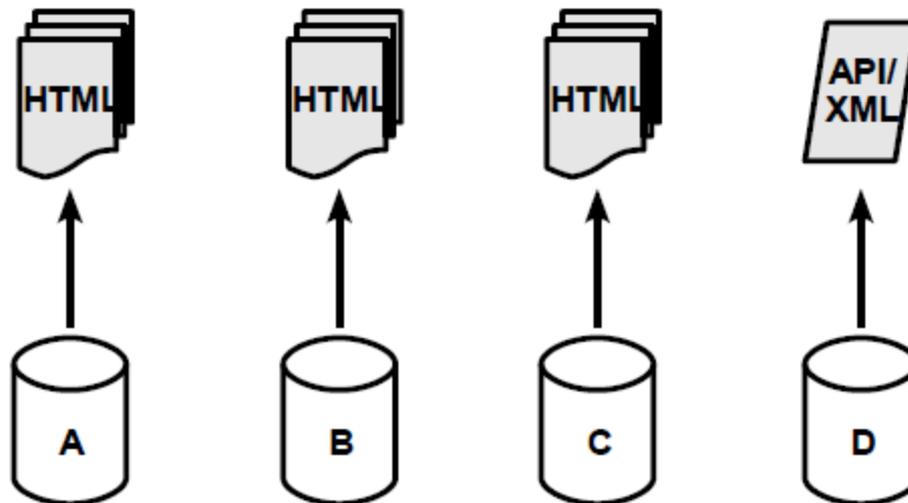
- ▶ Enlaces entre coisas arbitrarias (ex., pessoas, localizações, eventos, prédios)
- ▶ Estrutura dos dados nas páginas web são explicitadas
- ▶ Coisas descritas nas páginas web são nomeadas e têm URIs
- ▶ Enlaces entre coisas são feitas explícitas e são tipadas

▶ Web de Dados

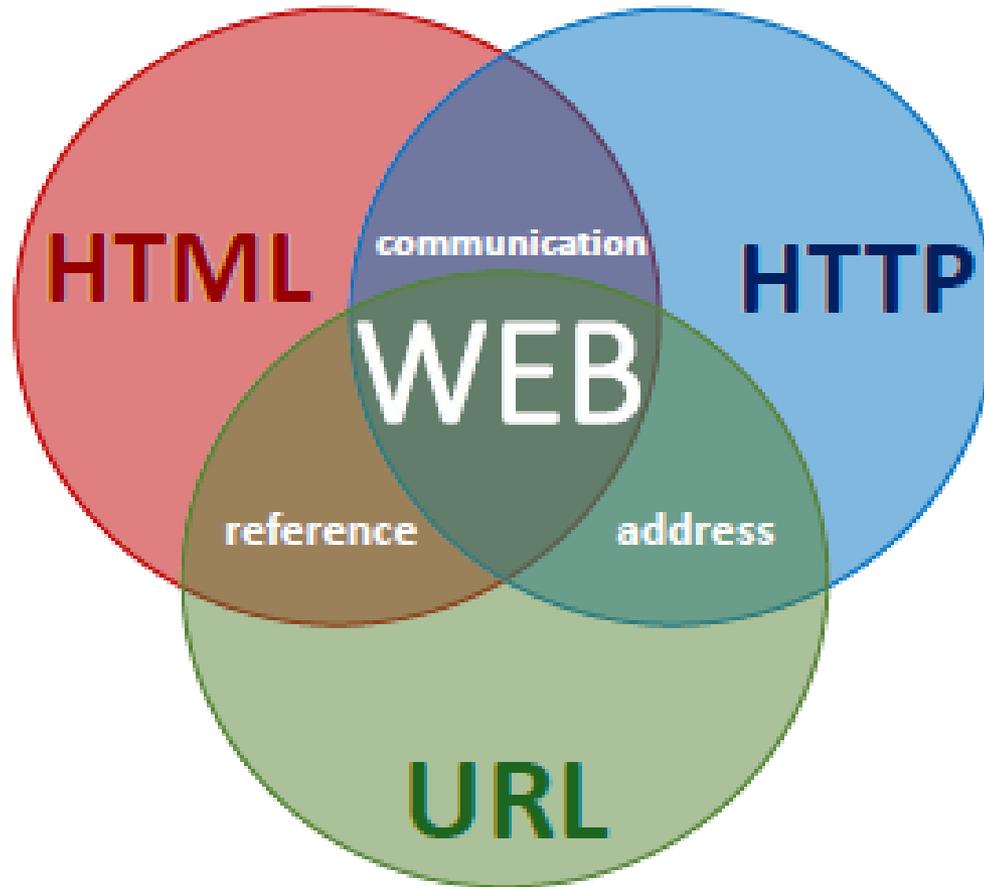


A Web de Dados

- ▶ A Web hoje
 - ▶ Consiste de silos de dados que podem ser acessados via motores de busca especializados de forma isolada.
 - ▶ Um site (silo de dados) tem filmes, outros publicações, outros têm atores.
 - ▶ Várias coisas comuns são representados em múltiplos conjuntos de dados
 - ▶ Identificadores de enlace ligam esses conjuntos de dados

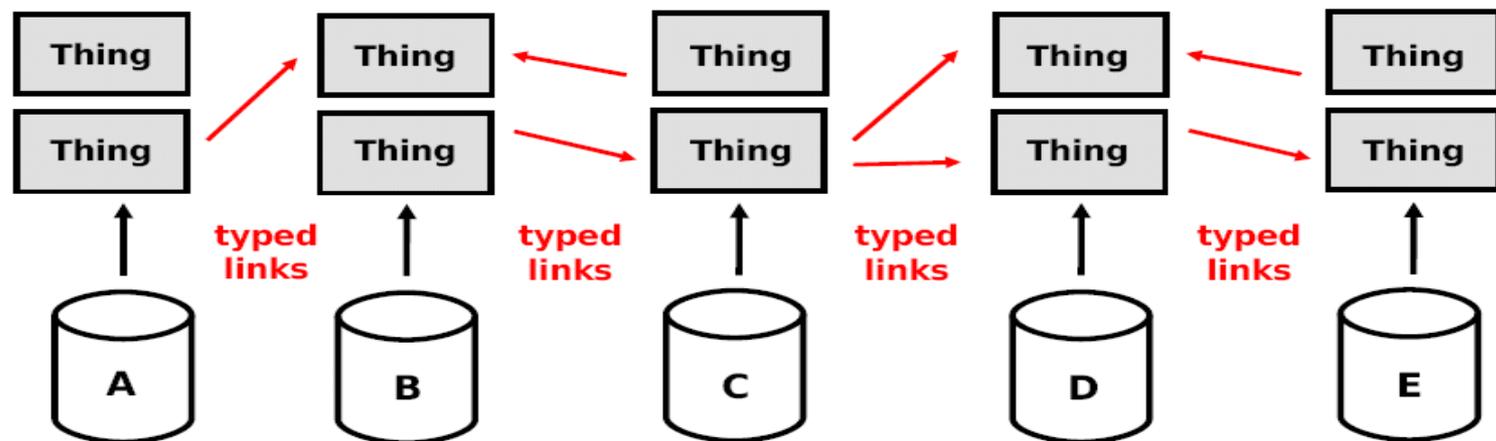


Arquitetura da Web Hoje



A Web de Dados

- ▶ A Web de Dados (Ligados) é imaginada como um banco de dados global
 - ▶ Consiste de objetos e suas descrições
 - ▶ no qual os objetos são ligados entre eles
 - ▶ com um alto grau de estrutura de objeto
 - ▶ com semântica explícita para enlaces e conteúdo
 - ▶ que é projetado para humanos e máquinas



Dados ligados

- ▶ É uma forma de publicar dados na Web que:
 - ▶ Estimula o reuso
 - ▶ reduz redundância
 - ▶ maximiza sua (real e potencial) interconectividade
 - ▶ facilita os efeitos de rede para adicionar valor aos dados



Princípios de dados ligados (Berners-Lee, 2006)

- ▶ Usar URIs como nomes para coisas
 - ▶ Qualquer coisa não somente documentos
 - ▶ Você não é o seu homepage
 - ▶ recursos de informação e não-informação
- ▶ Usar HTTP URIs
 - ▶ Nomes únicos globalmente, propriedade distribuída
 - ▶ Permite às pessoas localizar esses nomes
- ▶ Fornecer informação útil em RDF
 - ▶ Quando alguém procura uma URI
- ▶ Incluir enlaces RDF para outros URIs
 - ▶ Para facilitar a descoberta de informação relacionada



Pilha de Tecnologias dos Dados ligados

- ▶ URIs
- ▶ HTTP
- ▶ RDF
- ▶ (RDFS/OWL)



URIs – Não somente para páginas Web

- ▶ “Um Uniform Resource Identifier (URI) oferece uma forma simples e extensível de identificar um recurso.” -- RFC 3986
- ▶ Vários esquemas diferentes: *http://*, *ftp://*, *tel:*, *urn:*, *mailto:*
- ▶ *Veja Figura.*
- ▶ Alguns URIs para coisas do “mundo real”:
 - ▶ <http://tomheath.com/id/me>
 - ▶ http://dbpedia.org/resource/Talis_Group
 - ▶ <http://sws.geonames.org/4671654/>



URL

identify what exists on the Web

<http://my-site.fr>



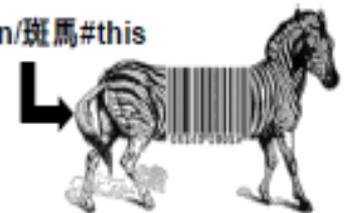
<http://animals.org/zebra#this>



URI

identify, on the Web, what exists

<http://الحيوانات.tn/斑馬#this>



IRI

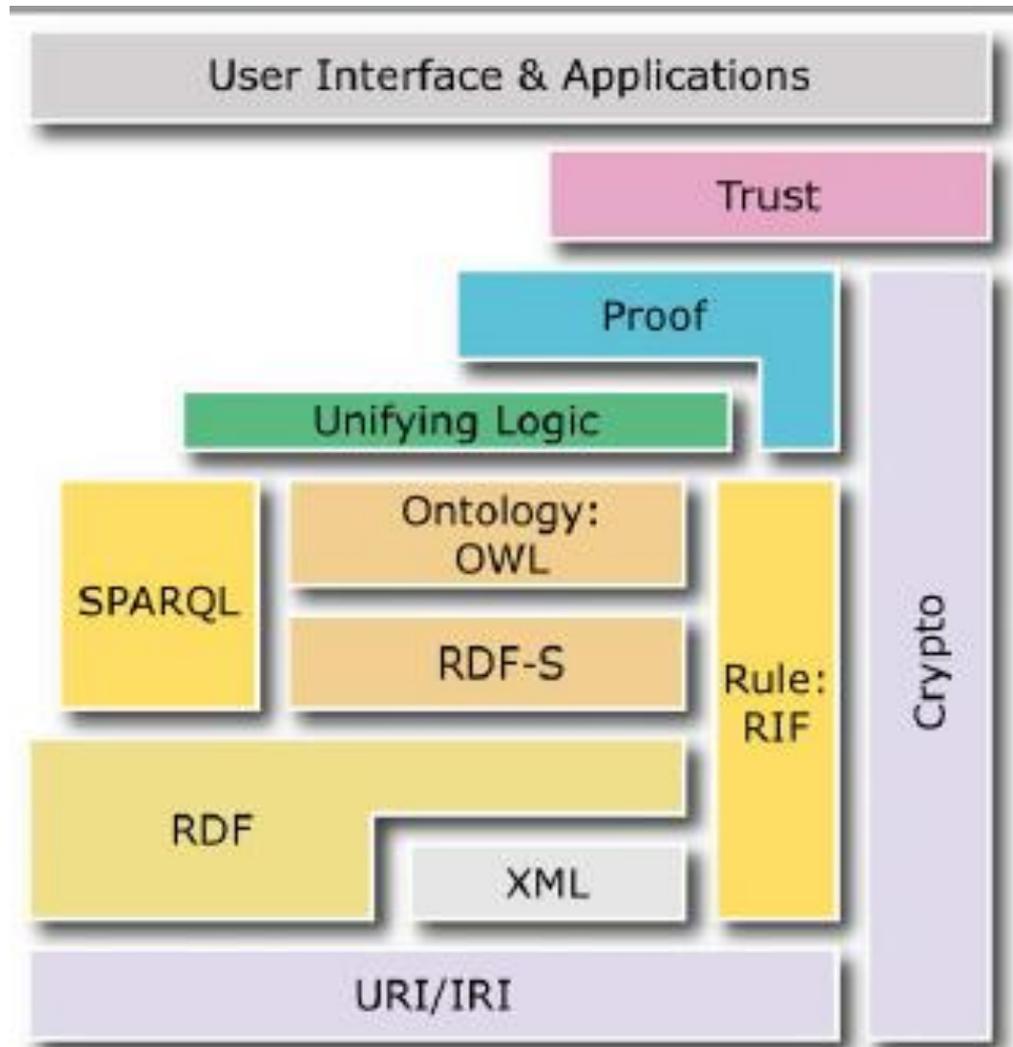
identify, on the Web, in any language, what exists

HTTP

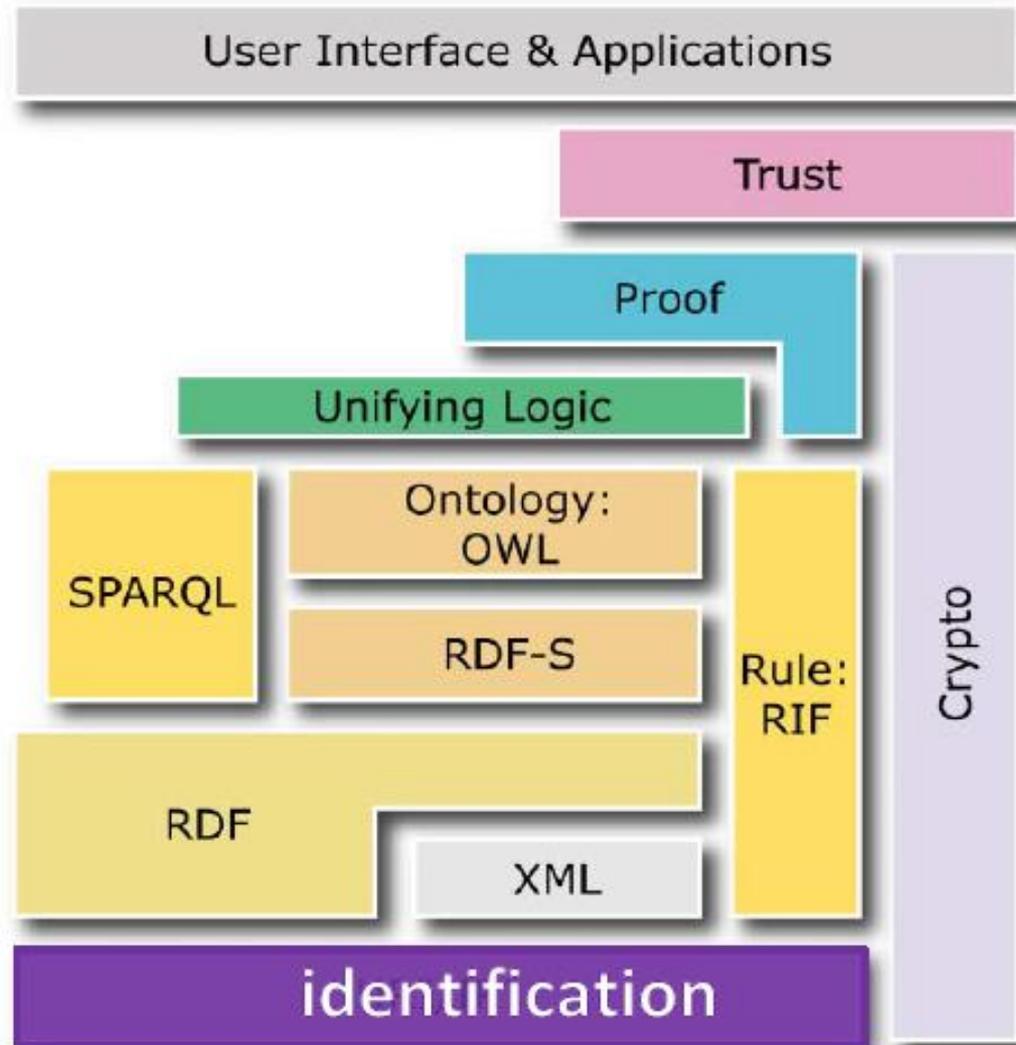
- ▶ Mecanismo de acesso aos Dados
- ▶ Usando *http://*URIs para identificar coisas permite que as pessoas as achem



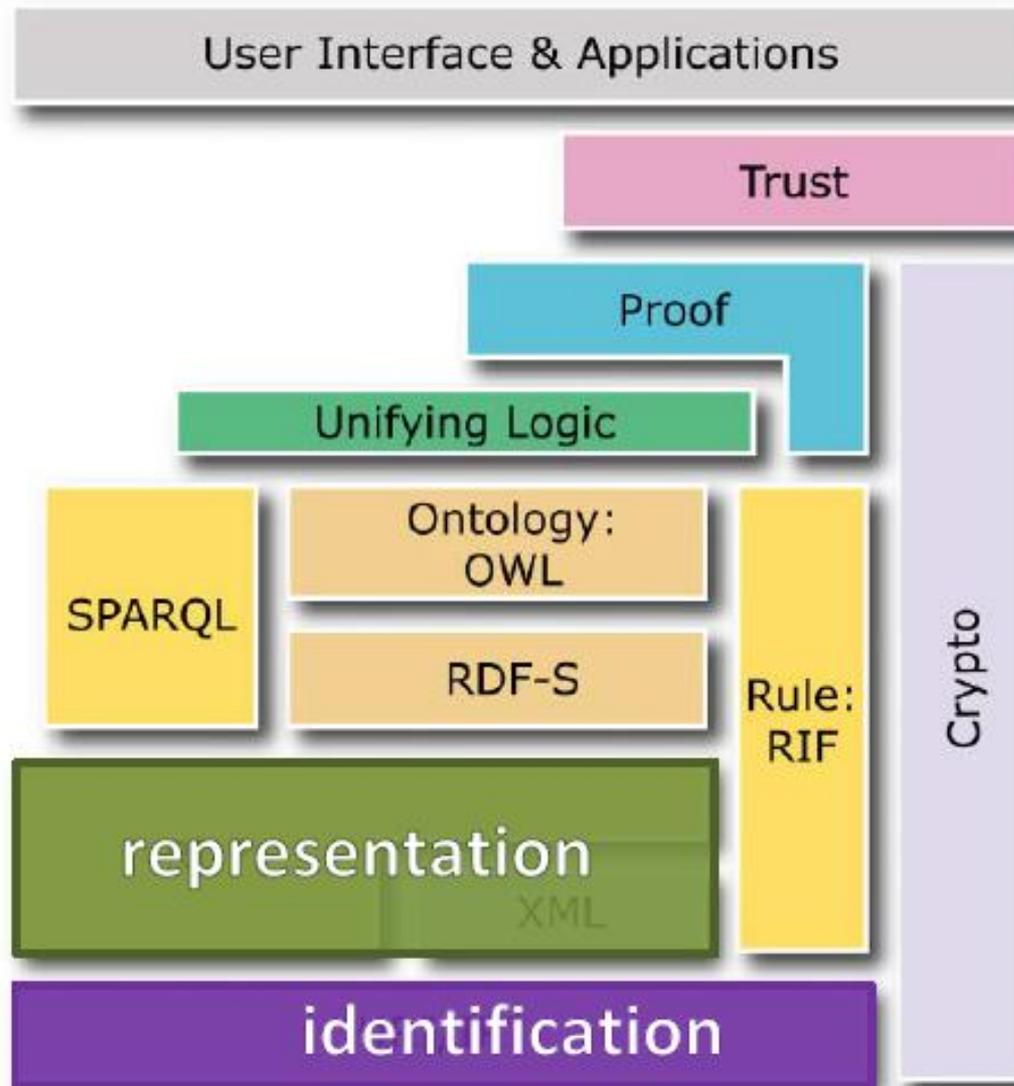
Padrões W3C – Web Semântica



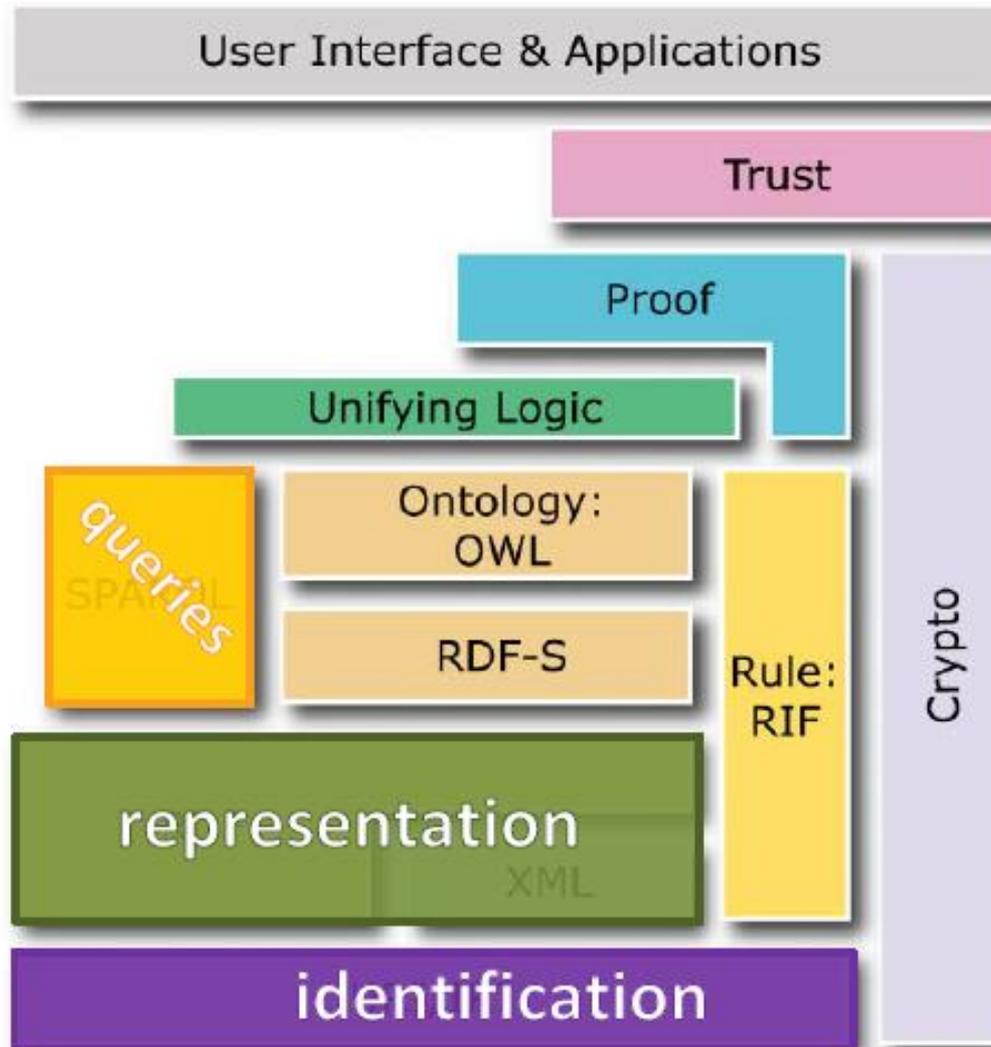
Padrões W3C – Web Semântica



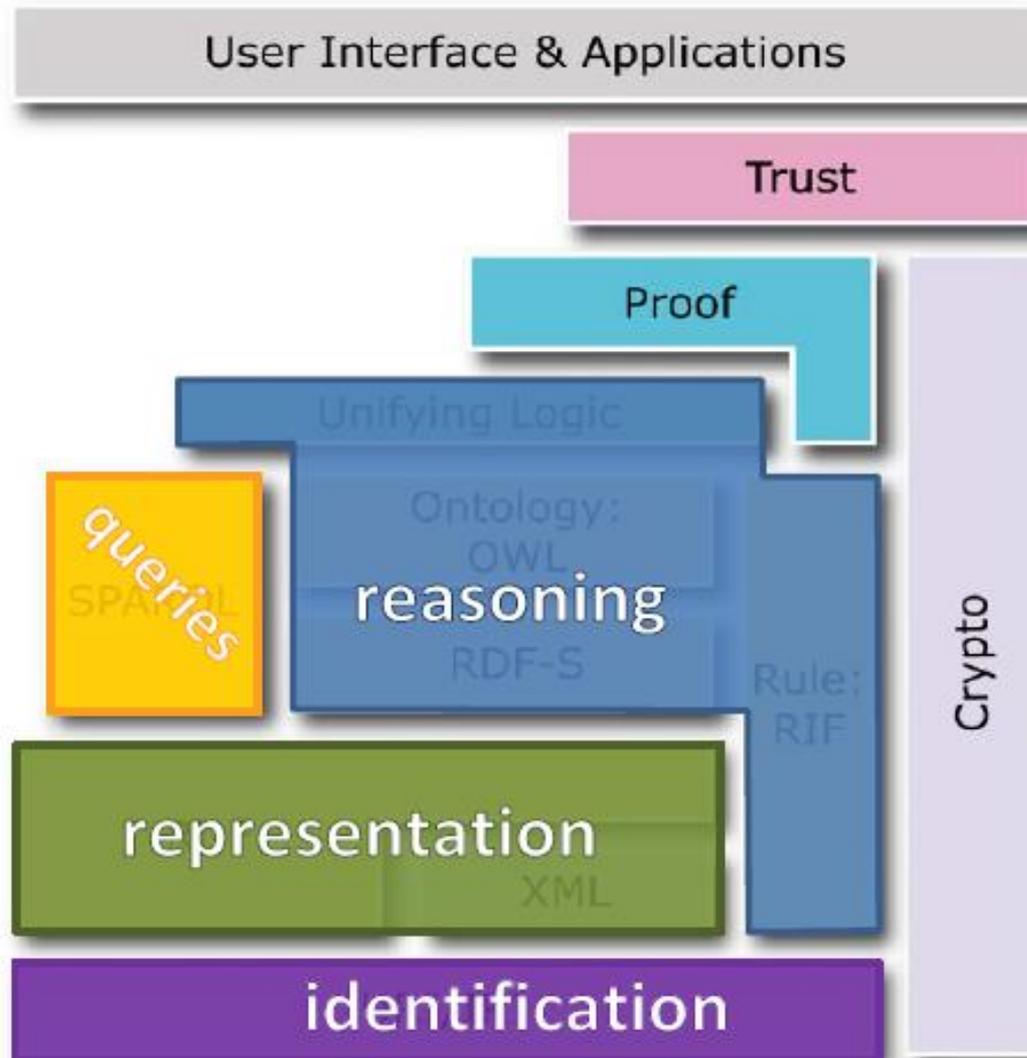
Padrões W3C – Web Semântica



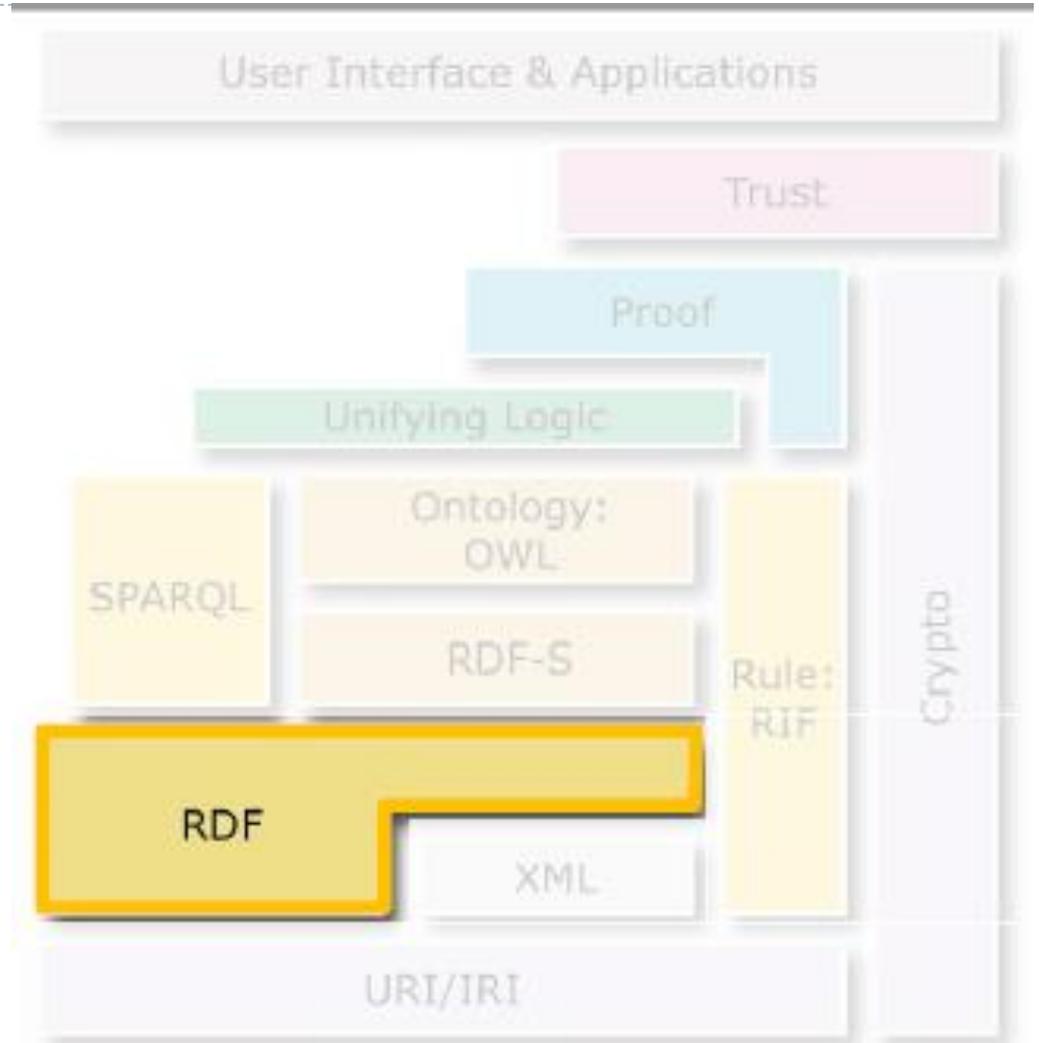
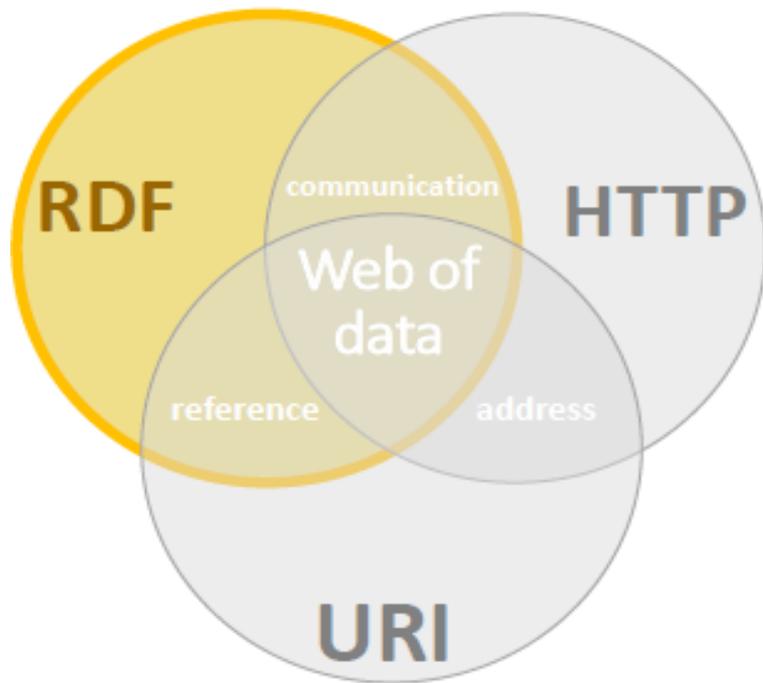
Padrões W3C – Web Semântica



Padrões W3C – Web Semântica



Padrões W3C



RDF: Resource Description Framework

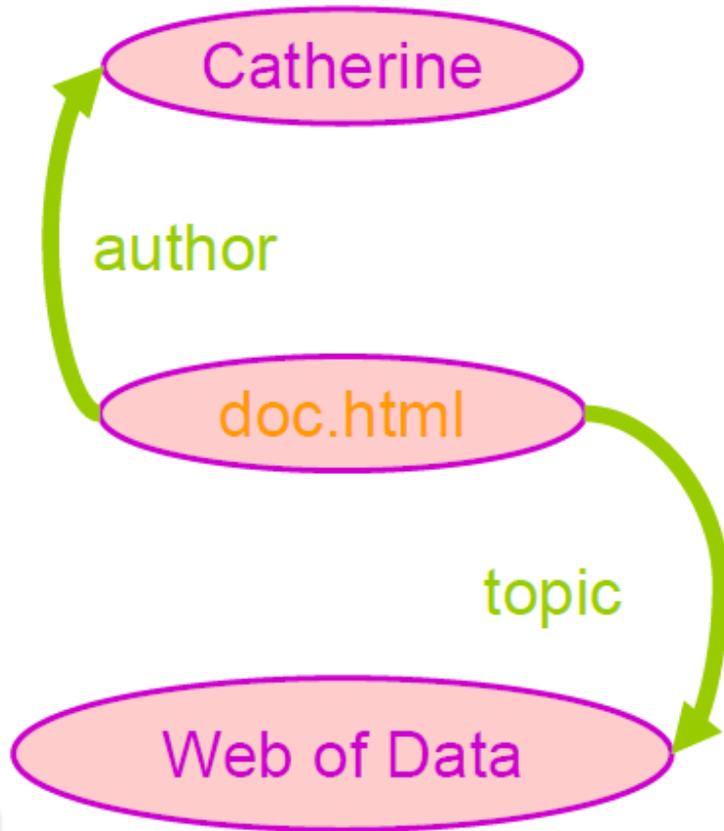
- Recursos: páginas, cadeiras, pessoas, ideias...
tudo o que pode ter uma URI
- Descrição: atributos, características, e relações entre recursos
- Framework: modelo, linguagem e sintaxe para essas descrições



RDF: Resource Description Framework

- ▶ Formato de dados para descrever coisas e suas inter-relações
- ▶ O modelo de dados RDF:
 - ▶ *Triplas*
 - ▶ sujeito → predicado → objeto
 - ▶ *Jose* → *worksFor* → *EACH-USP*
 - ▶ *EACH-USP* → *basedIn* → *São Paulo*
 - ▶ `<uri>` → `<uri>` → `<uri>` ou "literal"
 - ▶ “Talis Group is Based Near Birmingham”
 - ▶ `<http://dbpedia.org/resource/Talis_Group >`
 - ▶ `<http://xmlns.com/foaf/0.1/based_near >`
 - ▶ `<http://sws.geonames.org/3333125/>`

RDF é um modelo de grafos



(doc.html , author , Catherine)
(doc.html , topic , Web of Data)

RDF é um modelo **multigrafo** rotulado orientado



Várias arestas podem ligar os mesmos dois nós



RDF é um modelo multigrafo rotulado **orientado**



Arestas são orientadas:
a cabeça é o objeto e a
cauda é o sujeito.



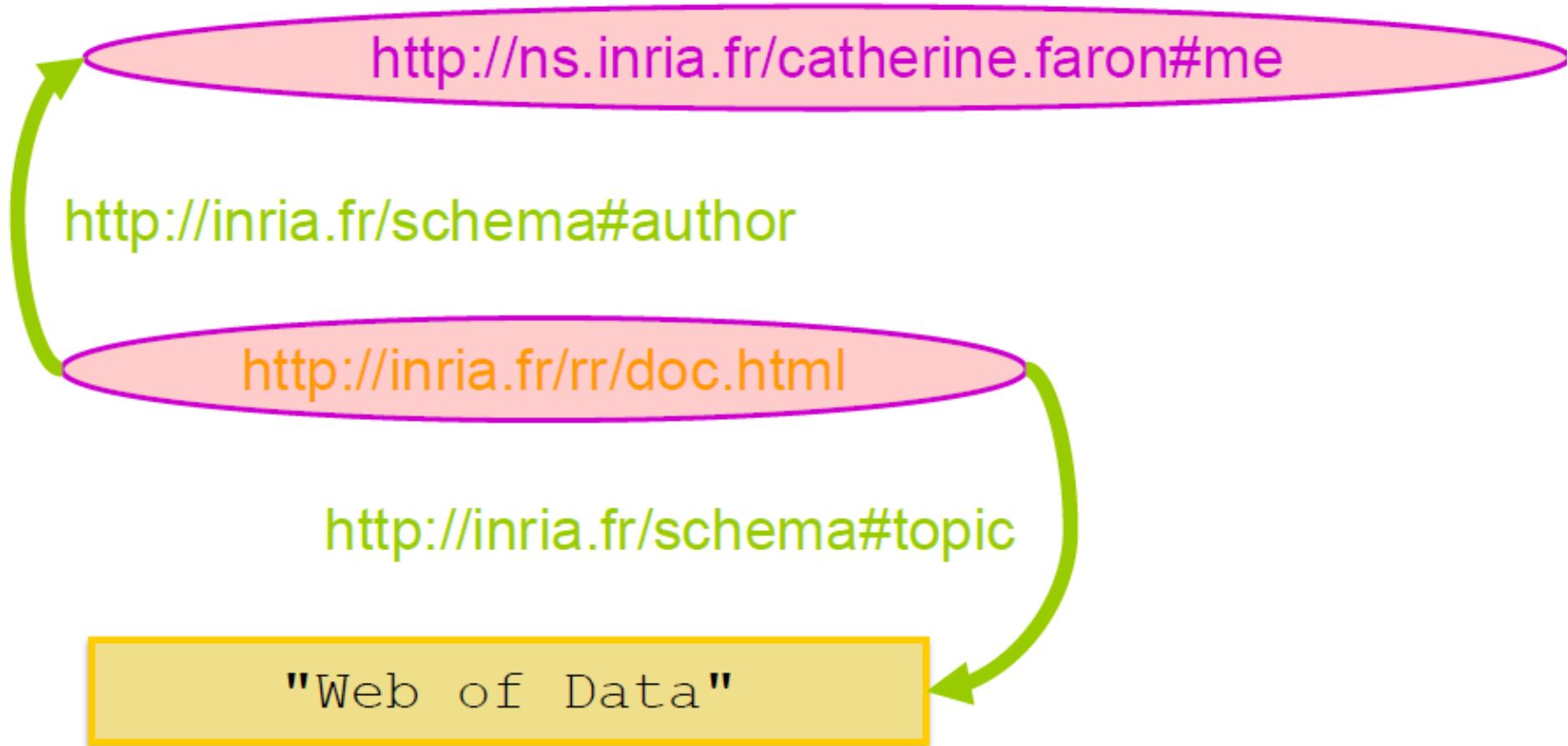
RDF é um modelo multigrafo **rotulado** orientado



Arestas e nós são rotulados

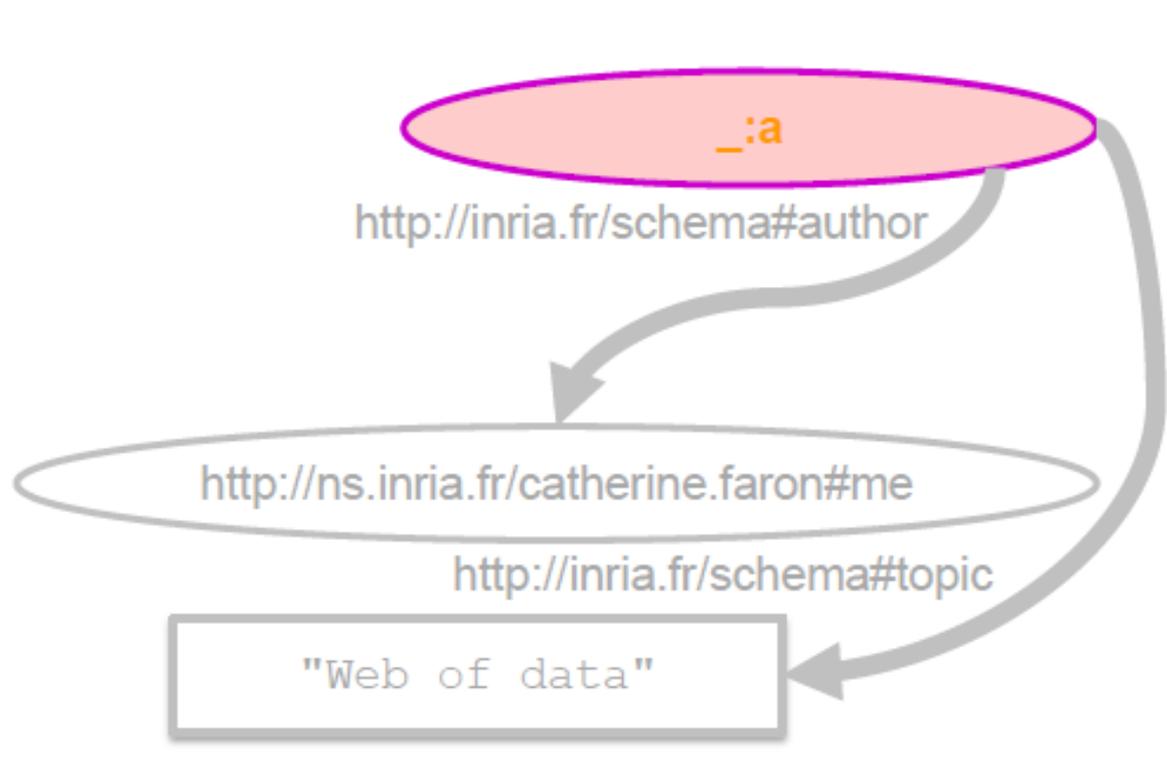


Grafos RDF são rotulados por recursos e literais

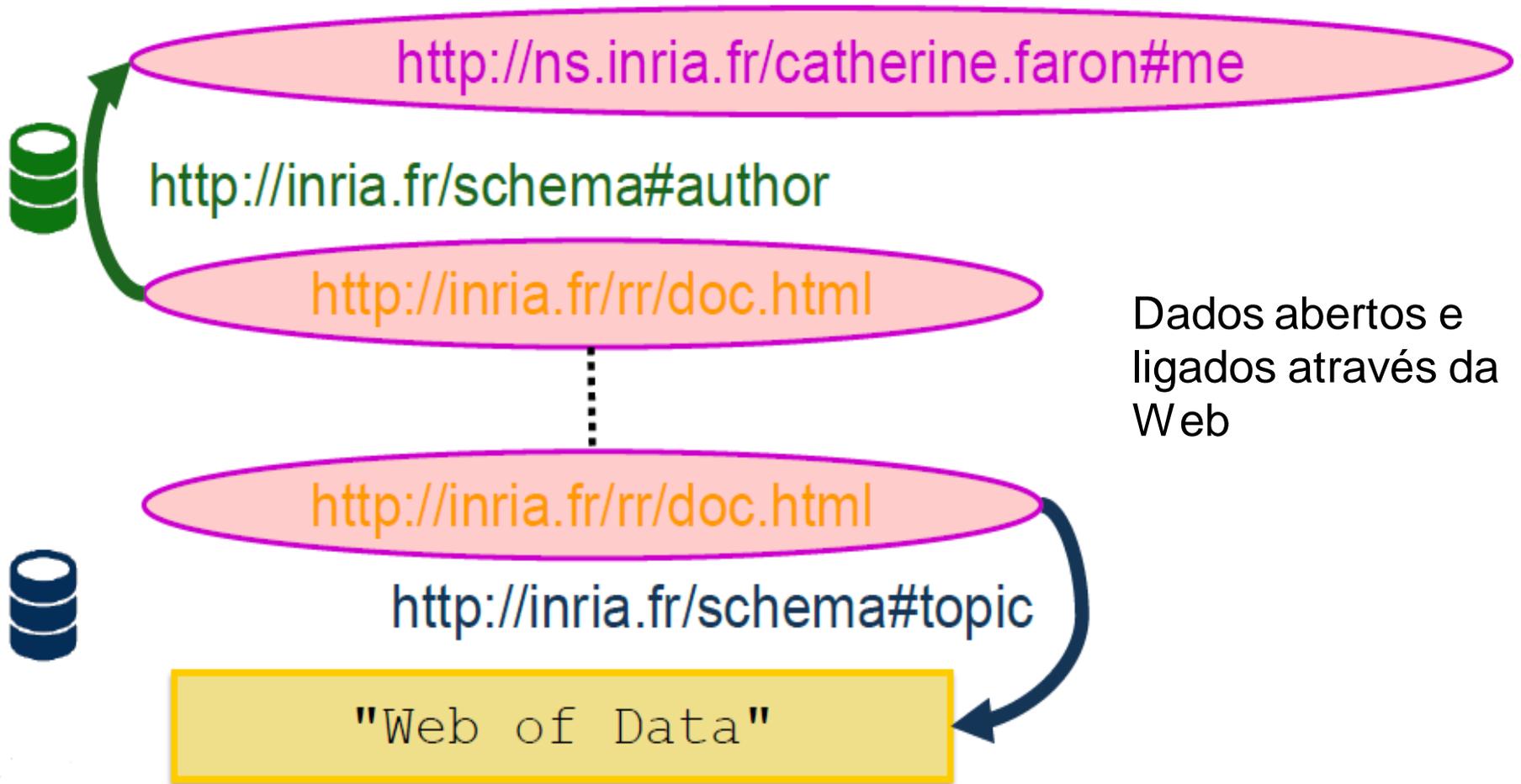


Grafos RDF são rotulados por recursos e literais

- Um recurso é ou uma URI ou um recurso anônimo (“blank node”)
- Quantificação existencial: existe um recurso tal que ...



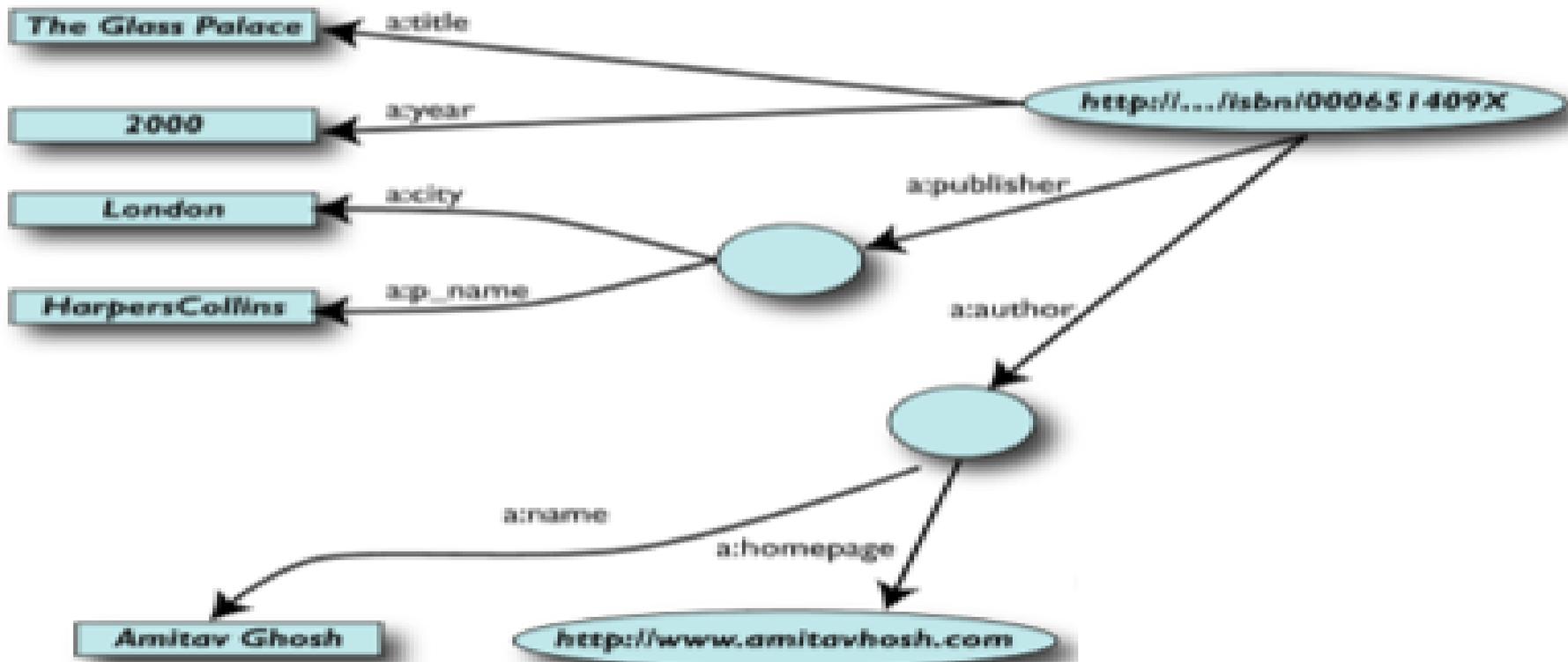
RDF Grafo gigante global



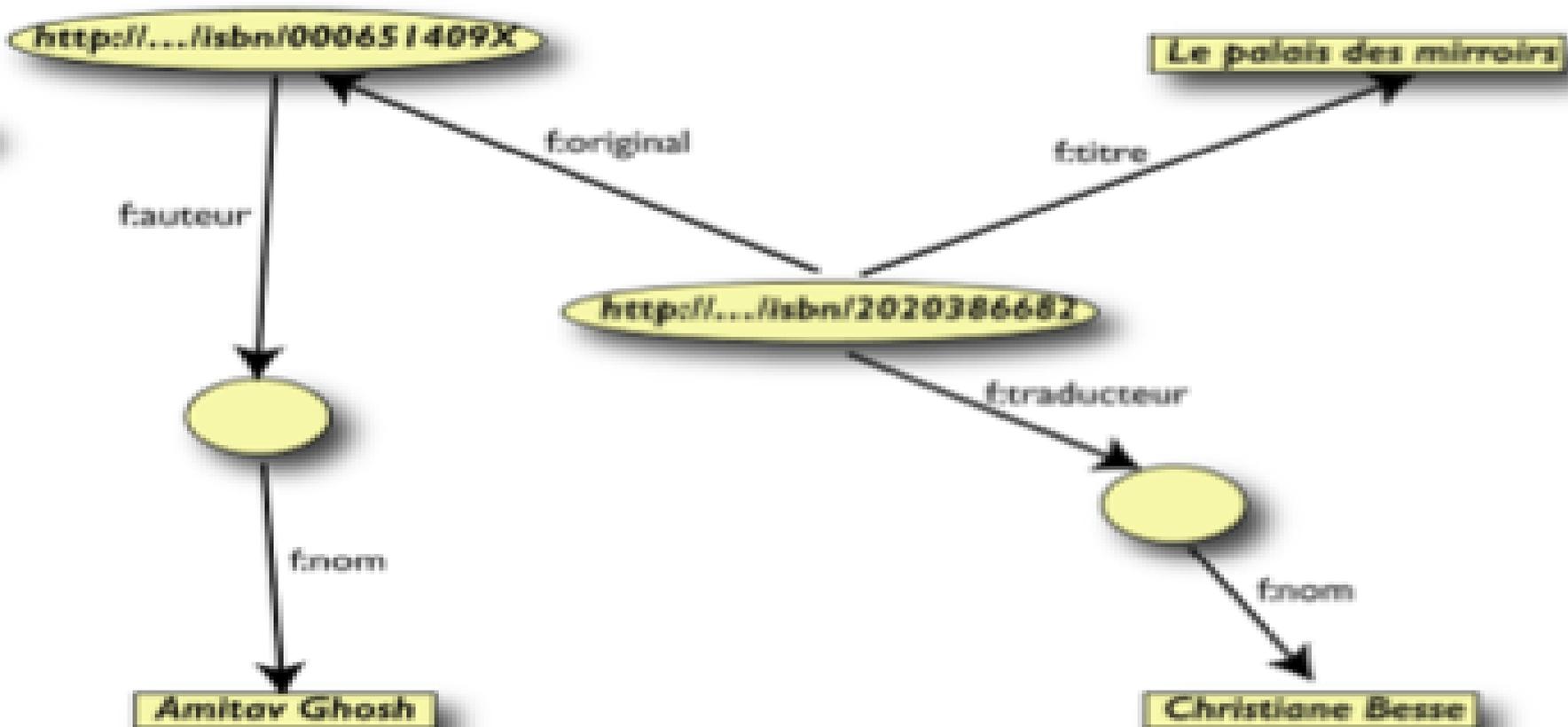
Misturando dados com RDF

1. Exporta o primeiro conjunto de dados como um grafo RDF

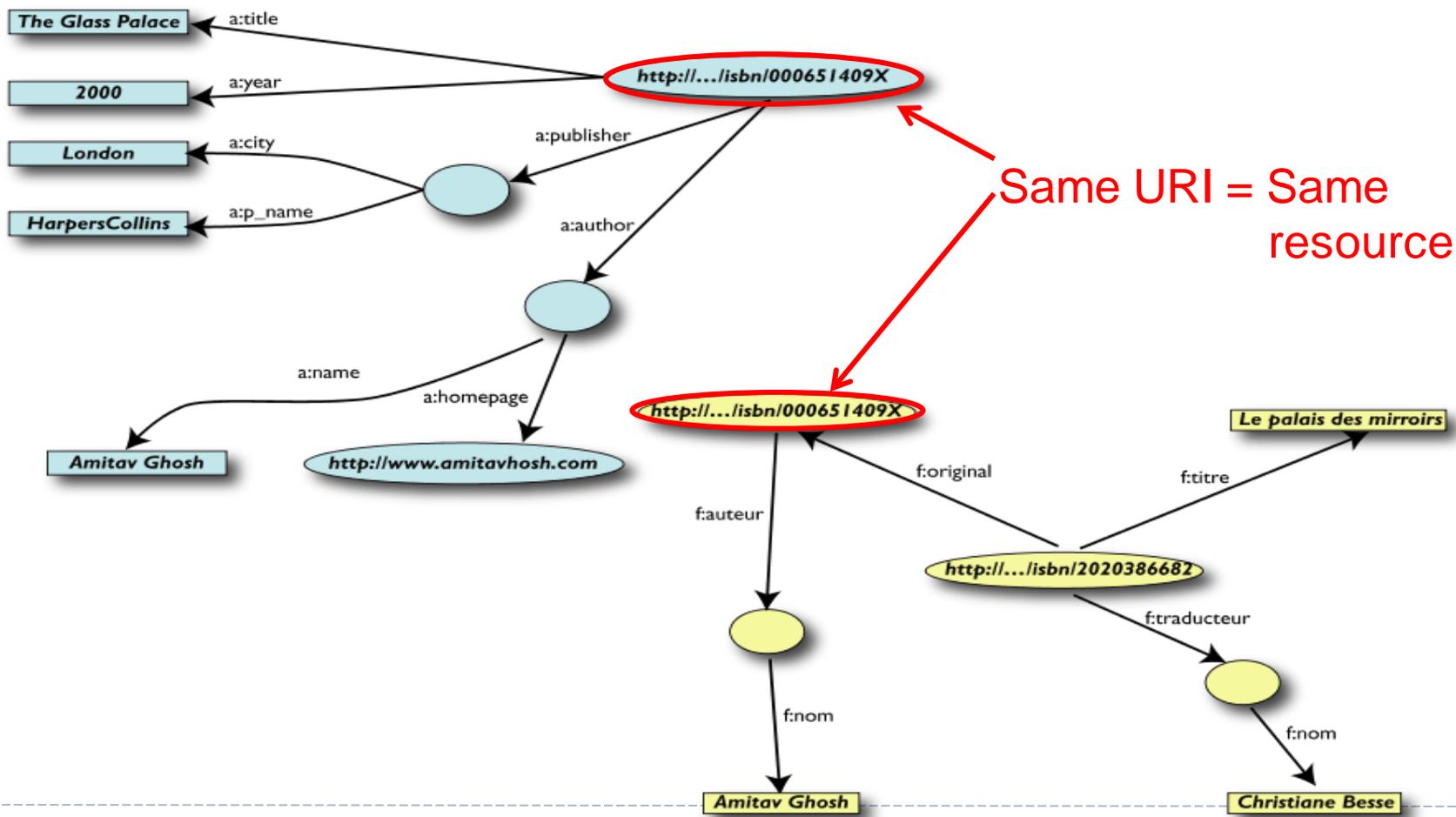
Por exemplo o seguinte grafo RDF contem o livro “The Glass Palace” do Amitav Ghosh

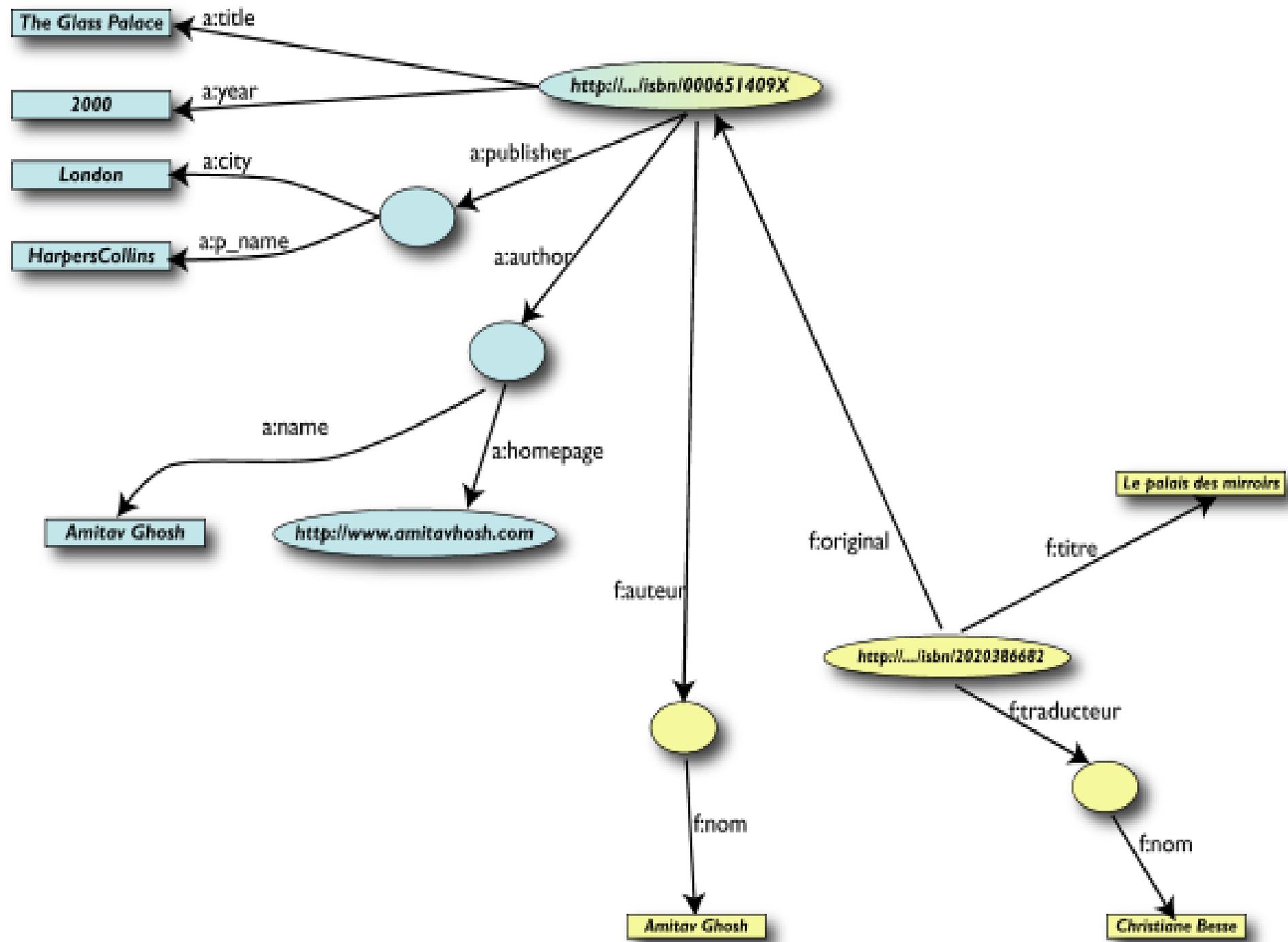


2. **Exporte o segundo conjunto de dados como um grafo RDF:**
Informa sobre o mesmo livro em Francês.

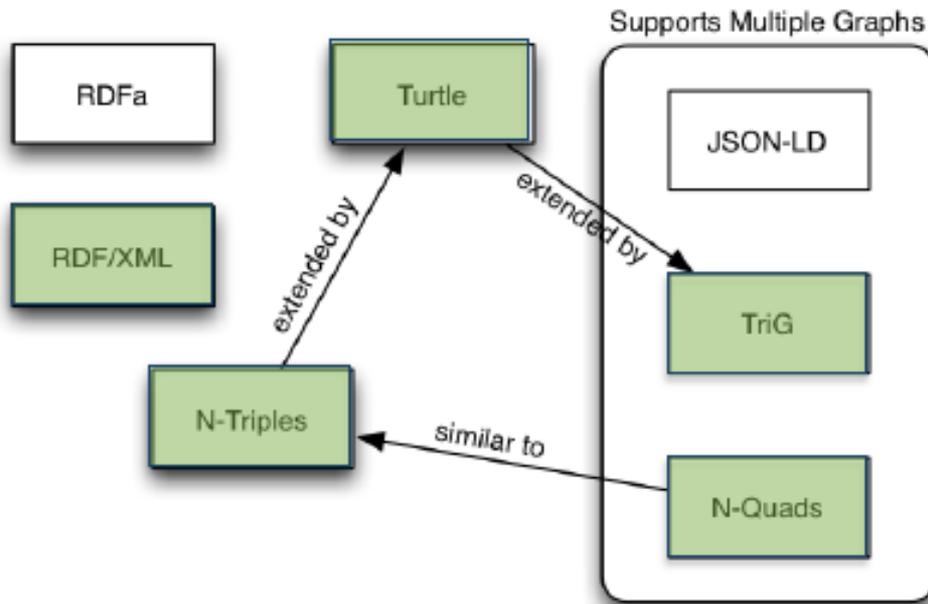


- ▶ **Misture recursos idênticos (ex: recursos que têm o mesmo URI) de conjuntos de dados diferentes**





RDF tem várias sintaxes



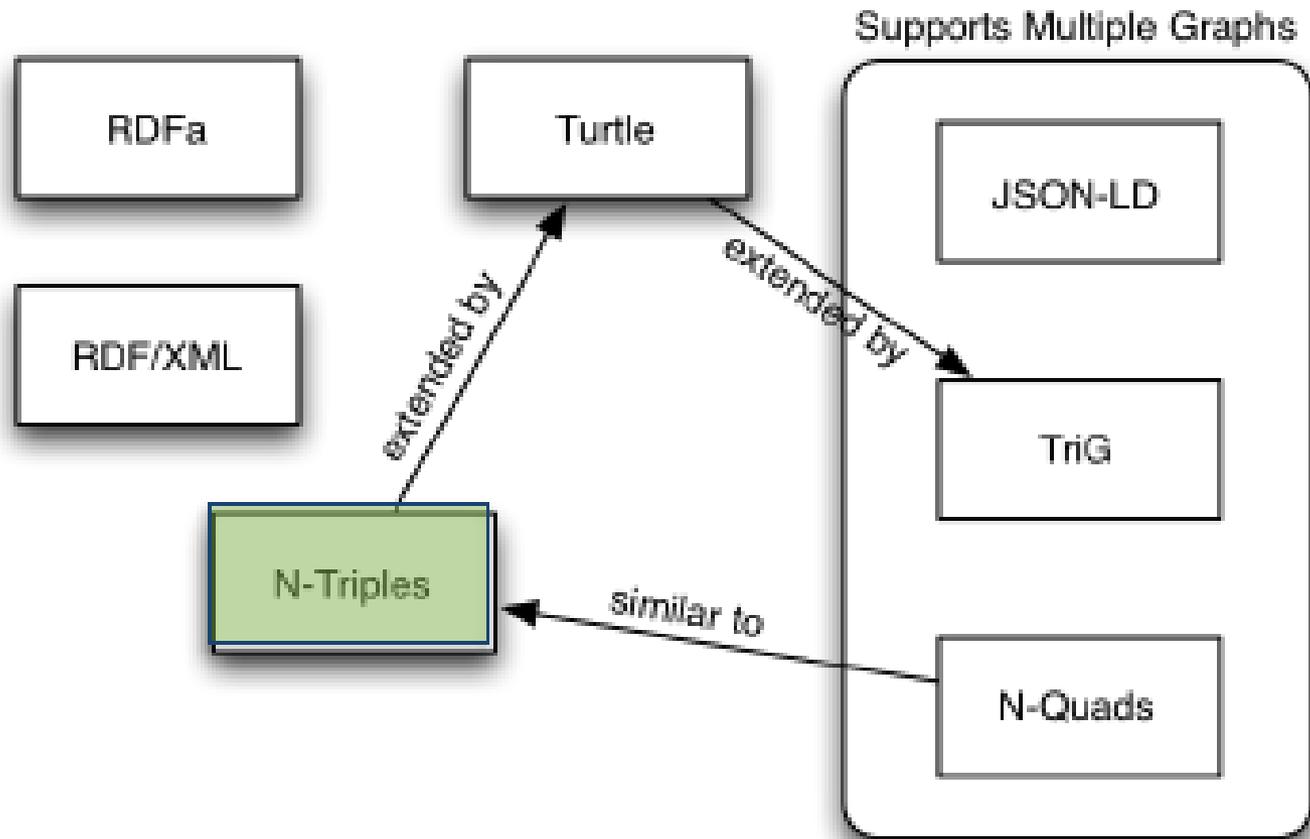
Sintaxe XML é a notação clássica

N-Triples é a sintaxe minimalista



RDF

N-Triples: uma sintaxe minimalista



W3C ©

RDF N-Triples

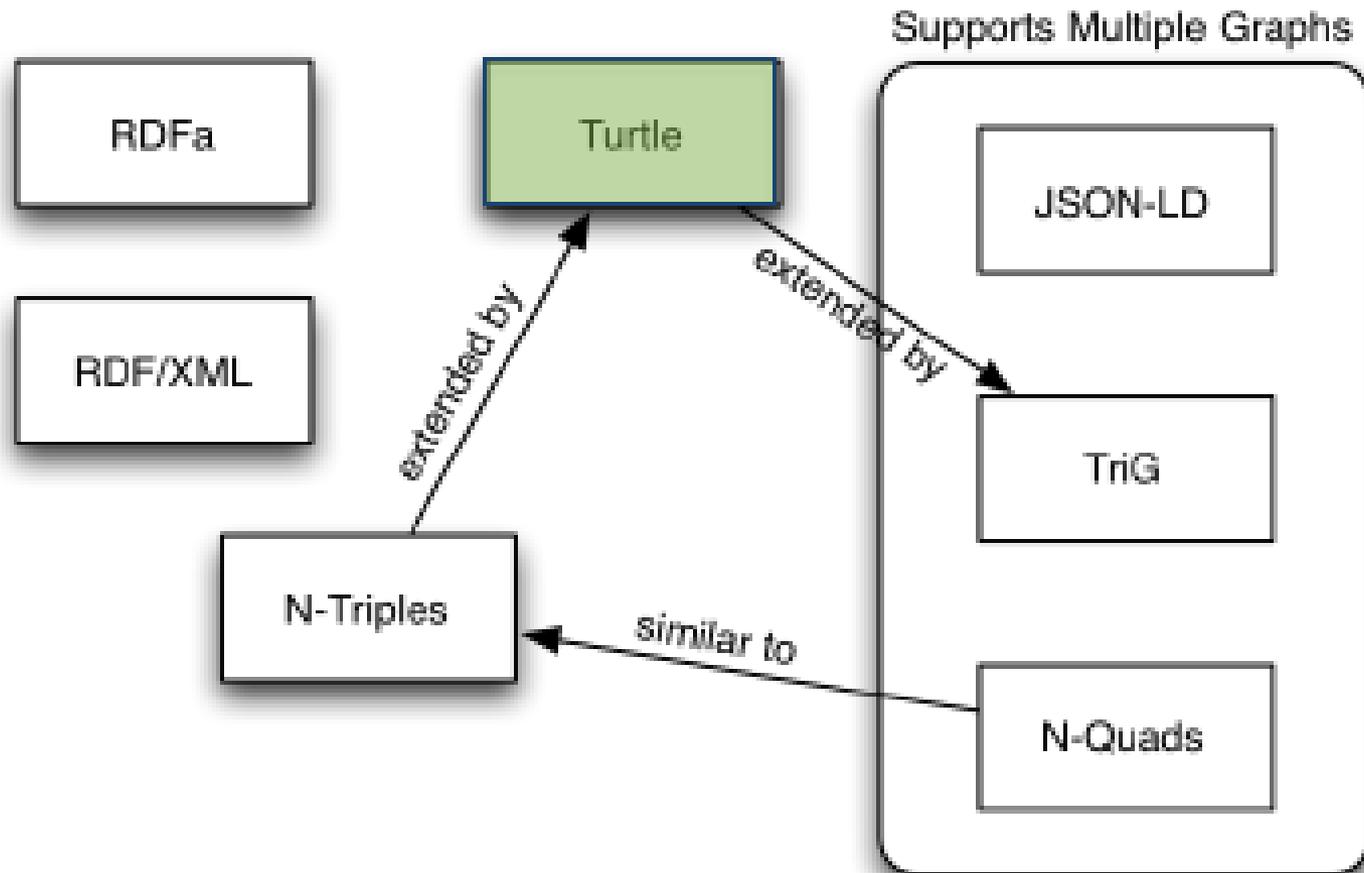
- URIs entre símbolos de menor e maior que.
- Valores literais entre haspas
- Triplas separadas por ponto final

```
<http://inria.fr/rr/doc.html>  
  <http://inria.fr/schema#author>  
    <http://ns.inria.fr/catherine.faron#me> .
```

```
<http://inria.fr/rr/doc.html>  
  <http://inria.fr/schema#topic> "Web" .
```



Turtle: a syntaxe RDF mais popular



W3C ©

RDF Turtle: uma sintaxe muito concisa

Declaração de prefixos

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .  
@prefix inria: <http://inria.fr/schema#> .
```

Várias propriedades

```
<http://inria.fr/rr/doc.html>  
  inria:author <http://ns.inria.fr/catherine.faron#me> ;  
  inria:topic "Web of data", "Semantic Web".
```

Nome qualificado

Vários valores



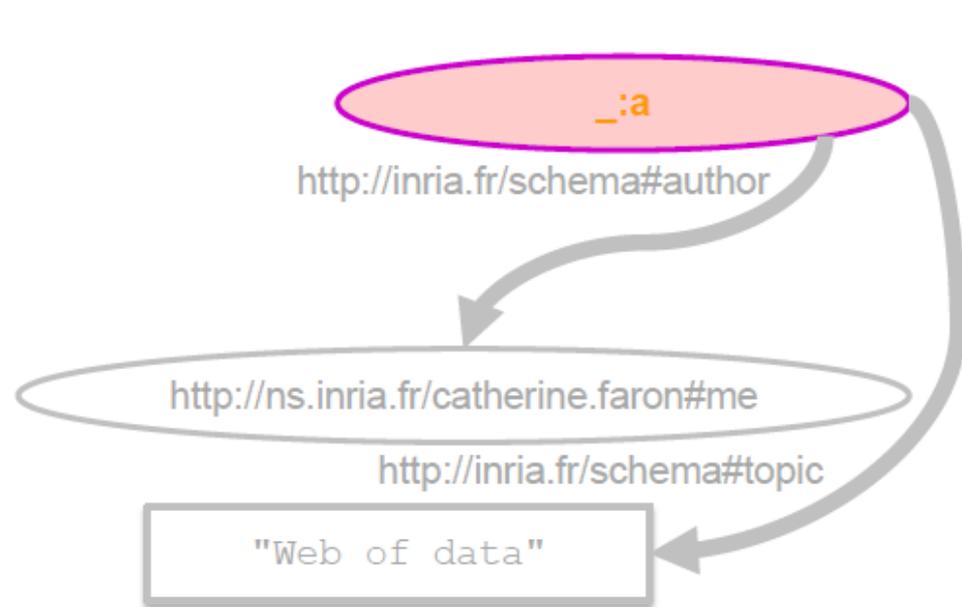
RDF Turtle: [recursos anônimos]

@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

@prefix inria: <http://inria.fr/schema#> .

[inria:author <http://ns.inria.fr/catherine.faron#me> ;

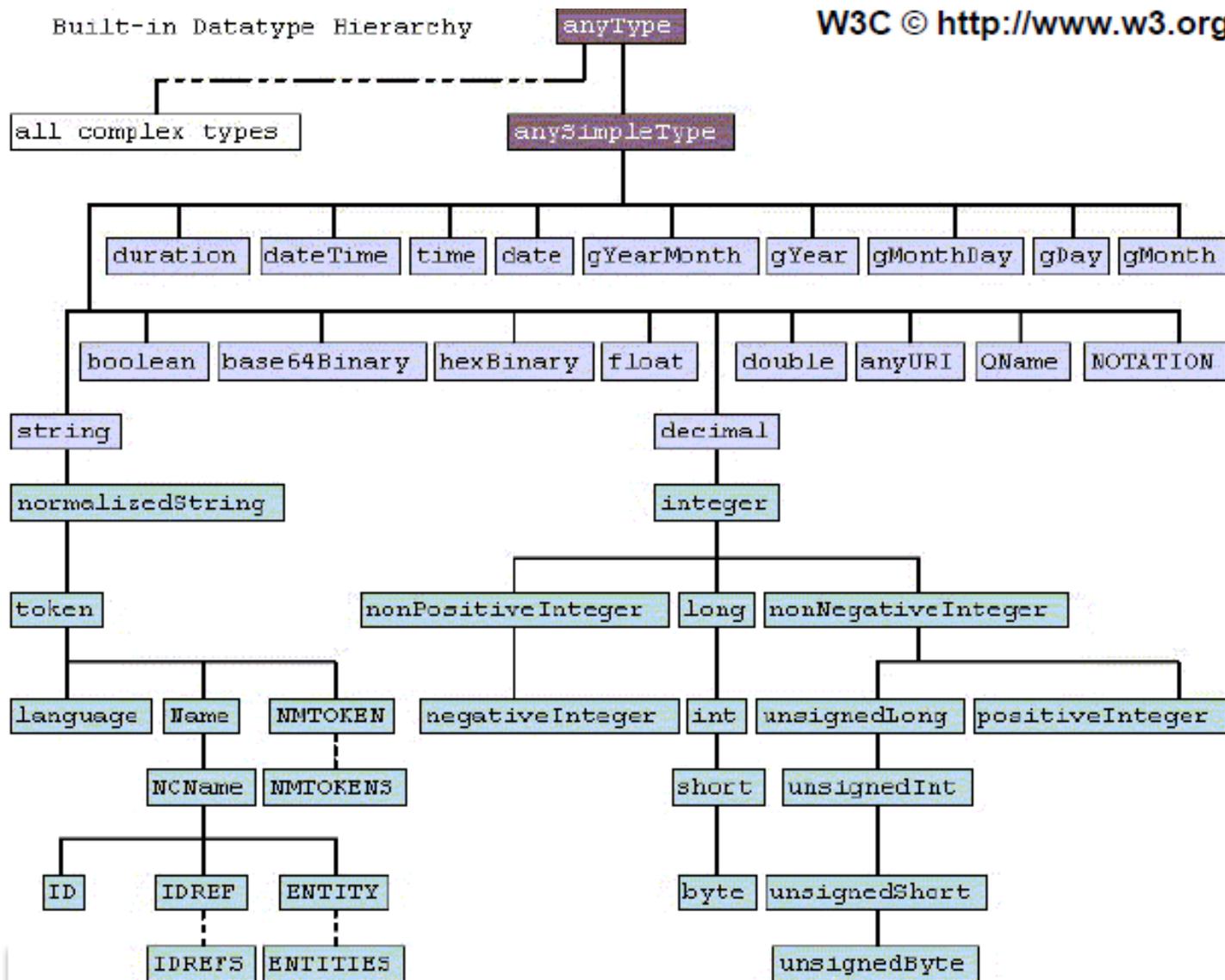
inria:topic "Web of data" .]



Tipos de dados XML Schema para tipificar literais

Built-in Datatype Hierarchy

W3C © <http://www.w3.org/TR/xmlschema-2/>



Indicando o idioma de um valor de texto

- Literais podem ser associados a um idioma

```
@prefix (...)          Turtle
<http://inria.fr/rr/doc.html>
  inria:topic "Web of Data"@en ;
  inria:topic "Web de données"@fr .
```

```
<rdf:RDF (...)>      RDF/XML
<rdf:Description rdf:about="http://inria.fr/rr/doc.html">
  <inria:topic xml:lang='en'>Web of Data</inria:topic>
  <inria:topic xml:lang='fr'>Web de données</inria:topic>
</rdf:Description>
</rdf:RDF>
```



Tipificando recursos

- A propriedade `rdf:type` liga os URIs de recursos aos URIs da suas classes

@prefix (...) Turtle

```
<http://ns.inria.fr/catherine.faron#me>  
  a inria:Woman, inria:Researcher.
```

<http://inria.fr/schema#Woman>

<http://ns.inria.fr/catherine.faron#me>

rdf:type

rdf:type

<http://inria.fr/schema#Researcher>

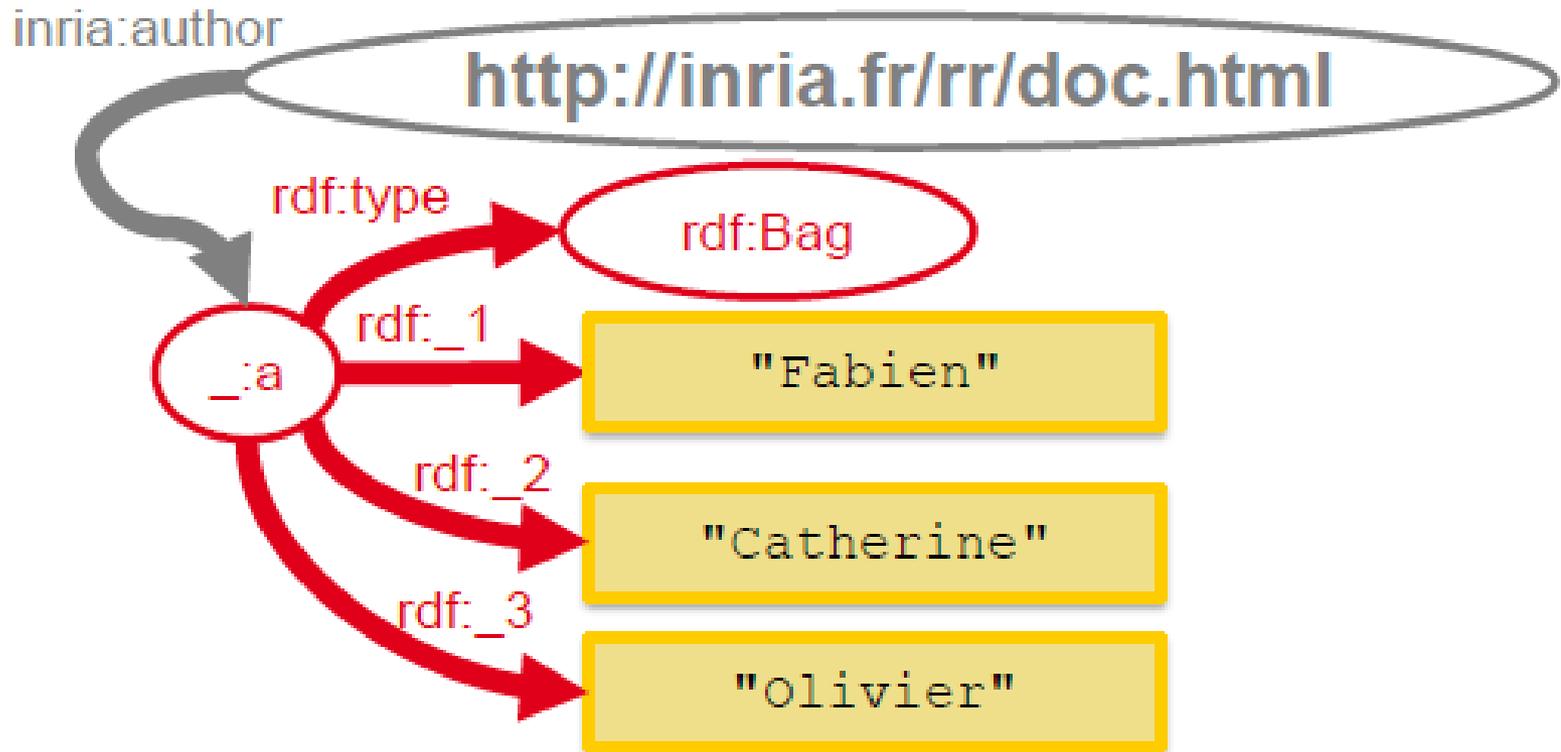
<rdf:RDF (...)> RDF/XML

```
<inria:Researcher rdf:about="http://ns.inria.fr/catherine.faron#me">  
  <rdf:type rdf:resource="http://www.inria.fr/schema#Woman"/>  
</inria:Researcher>  
</rdf:RDF>
```



Bags (rdf:Bag) de Recursos ou literais

Grupos simples sem ordem



Bags (rdf:Bag) de Recursos ou literais

```
@prefix (...)          Turtle
<http://inria.fr/rr/doc.html> inria:author [ a rdf:Bag ;
  rdf:li "Fabien", "Catherine", "Olivier" . ] .
```

```
<rdf:RDF (...) >      RDF/XML
<rdf:Description rdf:about="http://inria.fr/rr/doc.html">
  <inria:author>
    <rdf:Bag>
      <rdf:li>Fabien</rdf:li> <rdf:li>Catherine</rdf:li>
      <rdf:li>Olivier</rdf:li>
    </rdf:Bag>
  </inria:author>
</rdf:Description>
</rdf:RDF>
```

Também existem sequências (ordenadas) e alternativas.

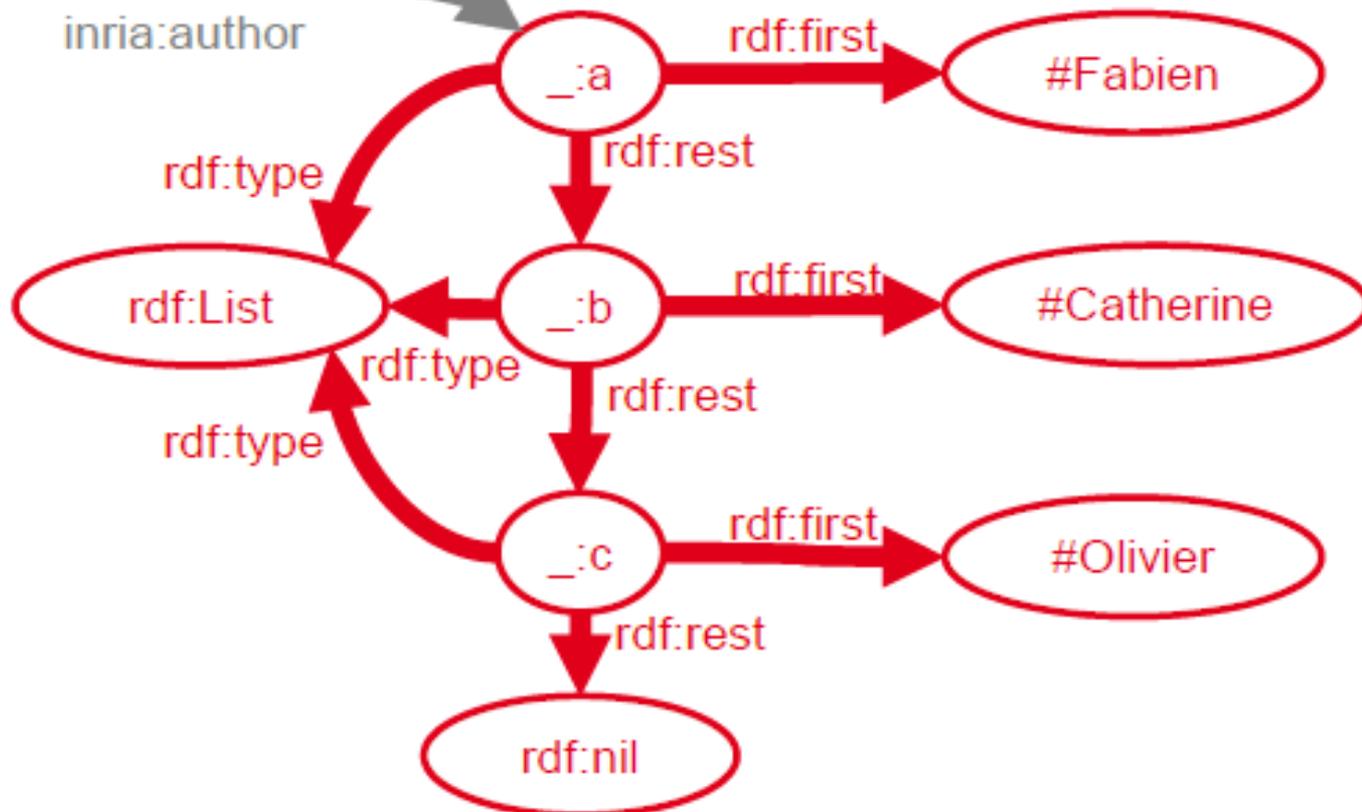


Coleções

- Uma limitação destes contenedores é que não existe nenhuma forma de fechar eles.
Estruturas abertas.
 - “estes são **todos** os membros do contenedor”
 - RDF fornece suporte para a descrição de grupos que contem somente os membros especificados, na forma de **coleções RDF.**
Fechadas.
 - estrutura **list** no grafo RDF
 - construído usando uma coleção predefinida
vocabulário: **rdf:List**, **rdf:first**, **rdf:rest** e **rdf:nil**
-



<http://inria.fr/rr/doc.html>



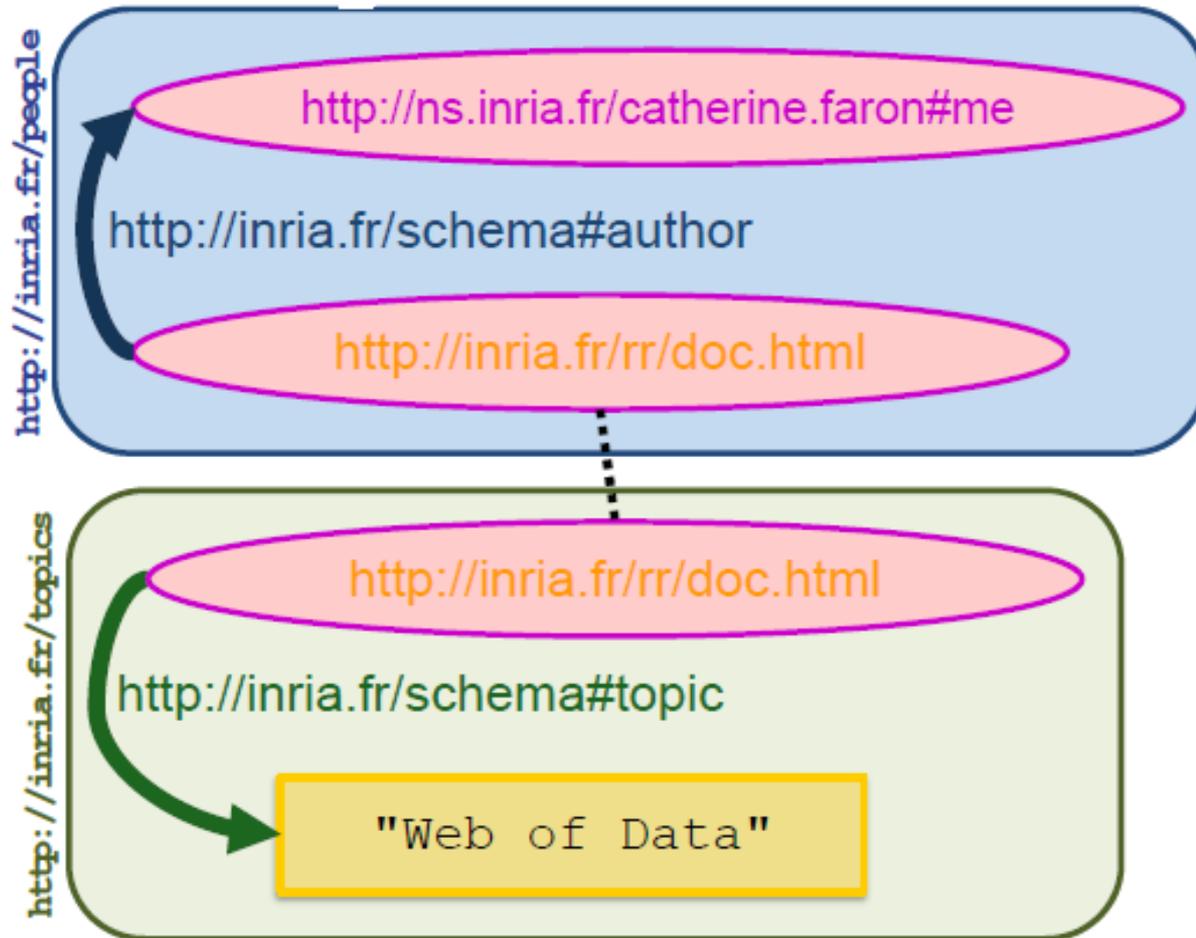
```
@prefix (...)          Turtle
<http://inria.fr/rr/doc.html> inria:author
  ( <#Fabien> <#Catherine> <#Olivier> ).
```

```
<rdf:RDF (...)>          RDF/XML
<rdf:Description rdf:about="http://inria.fr/rr/doc.html">
  <inria:author rdf:parseType="Collection">
    <rdf:Description rdf:about="#Fabien"/>
    <rdf:Description rdf:about="#Catherine"/>
    <rdf:Description rdf:about="#Olivier"/>
  </inria:author>
</rdf:Description>
</rdf:RDF>
```



Grafos Nomeados

- Agrupando triplas em sub-grafos identificados por URIs



RDFS – RDF Schema

- O RDFS oferece um vocabulário padrão para ser usado em descrições RDF
- RDFS reusa o vocabulário de RDF e introduz construções adicionais
- Um vocabulário RDF é um conjunto de declarações de propriedades e declarações de classes



Associando um Namespace a um Vocabulário

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>  
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>  
@base <http://inria.fr/2005/humans.rdfs>  
(...)
```



```
<rdf:RDF xml:base="http://inria.fr/2005/humans.rdfs"  
  xmlns:rdf ="http://www.w3.org/1999/02/22-rdf-syntax-ns#"  
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">  
(...)  
</rdf:RDF>
```



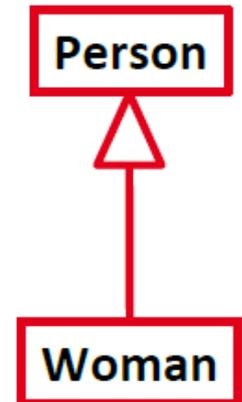
Declarando Classes de Recursos

- Nomeando classes e organizando-as em hierarquias.

```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
@base <http://inria.fr/2005/humans.rdfs>
<Woman> a rdfs:Class ;
    rdfs:subClassOf <Person>, <Female> .
```



```
<rdf:RDF xml:base="http://inria.fr/2005/humans.rdfs"
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">
  <rdfs:Class rdf:ID="Woman">
    <rdfs:subClassOf rdf:resource="#Person"/>
    <rdfs:subClassOf rdf:resource="#Female"/>
  </rdfs:Class>
</rdf:RDF>
```



Declarando tipos de propriedades

- Nomeando tipos de propriedades e organizando-os em hierarquias. A propriedade Class está no namespace de RDF já que propriedades são a chave das triplas RDF.

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>  
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>  
@base <http://inria.fr/2005/humans.rdfs>
```

```
<hasMother> a rdf:Property ;  
  rdfs:subPropertyOf <hasParent> .
```



```
<rdf:RDF xml:base="http://inria.fr/2005/humans.rdfs"  
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"  
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#">  
  <rdf:Property rdf:ID="hasMother">  
    <rdfs:subPropertyOf rdf:resource="#hasParent"/>  
  </rdf:Property>  
</rdf:RDF>
```

hasParent



hasMother



Declarando assinaturas de propriedades

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
```

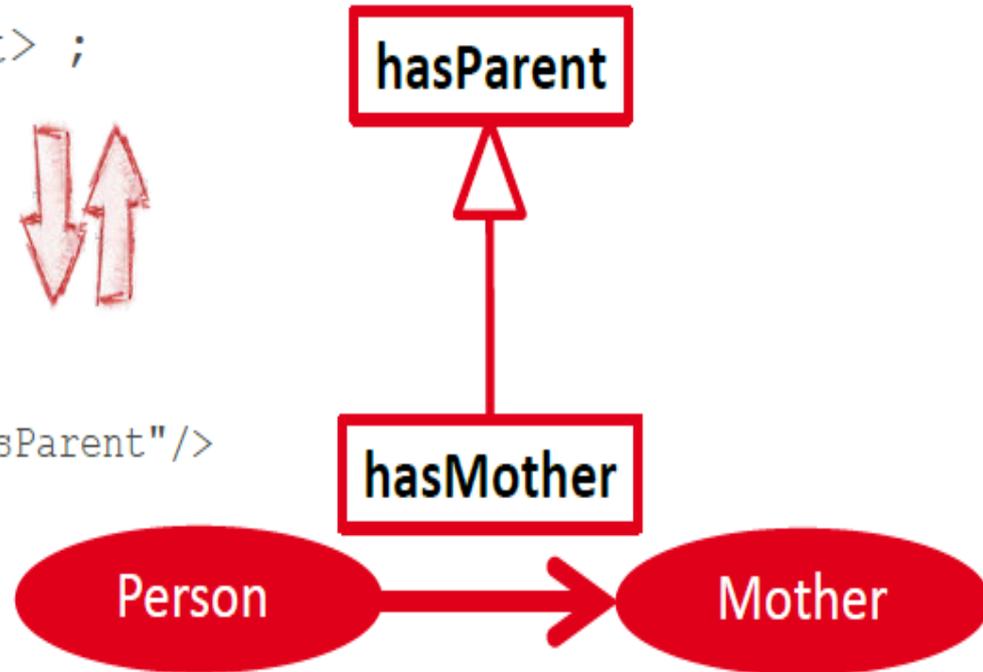
```
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
```

```
@base <http://inria.fr/2005/humans.rdfs>
```

```
<hasMother> a rdf:Property ;  
  rdfs:subPropertyOf <hasParent> ;  
  rdfs:domain <Person> ;  
  rdfs:range <Woman> .
```



```
<rdf:RDF ... >  
  <rdf:Property rdf:ID="hasMother">  
    <rdfs:subPropertyOf rdf:resource="#hasParent"/>  
    <rdfs:domain rdf:resource="#Person"/>  
    <rdfs:range rdf:resource="#Woman"/>  
  </rdf:Property>  
</rdf:RDF>
```



Documentando as declarações de classes e propriedades

```
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>
@base <http://inria.fr/2005/humans.rdfs>

<Woman> a rdfs:Class ;
  rdfs:label "woman"@en ;
  rdfs:comment "an adult female person"@en .

<hasMother> a rdf:Property ;
  rdfs:label "has for mother"@en ;
  rdfs:comment "to have a woman for mother"@en .
```



Referenciando e usando esquemas

- Na descrição de um recurso

@prefix **h**: <http://inria.fr/2005/humans.rdfs#>

@base <http://inria.fr/2005/humans.rdfs-instances>

<Alice> a **h:Woman**; **h:hasMother** <Laura> .



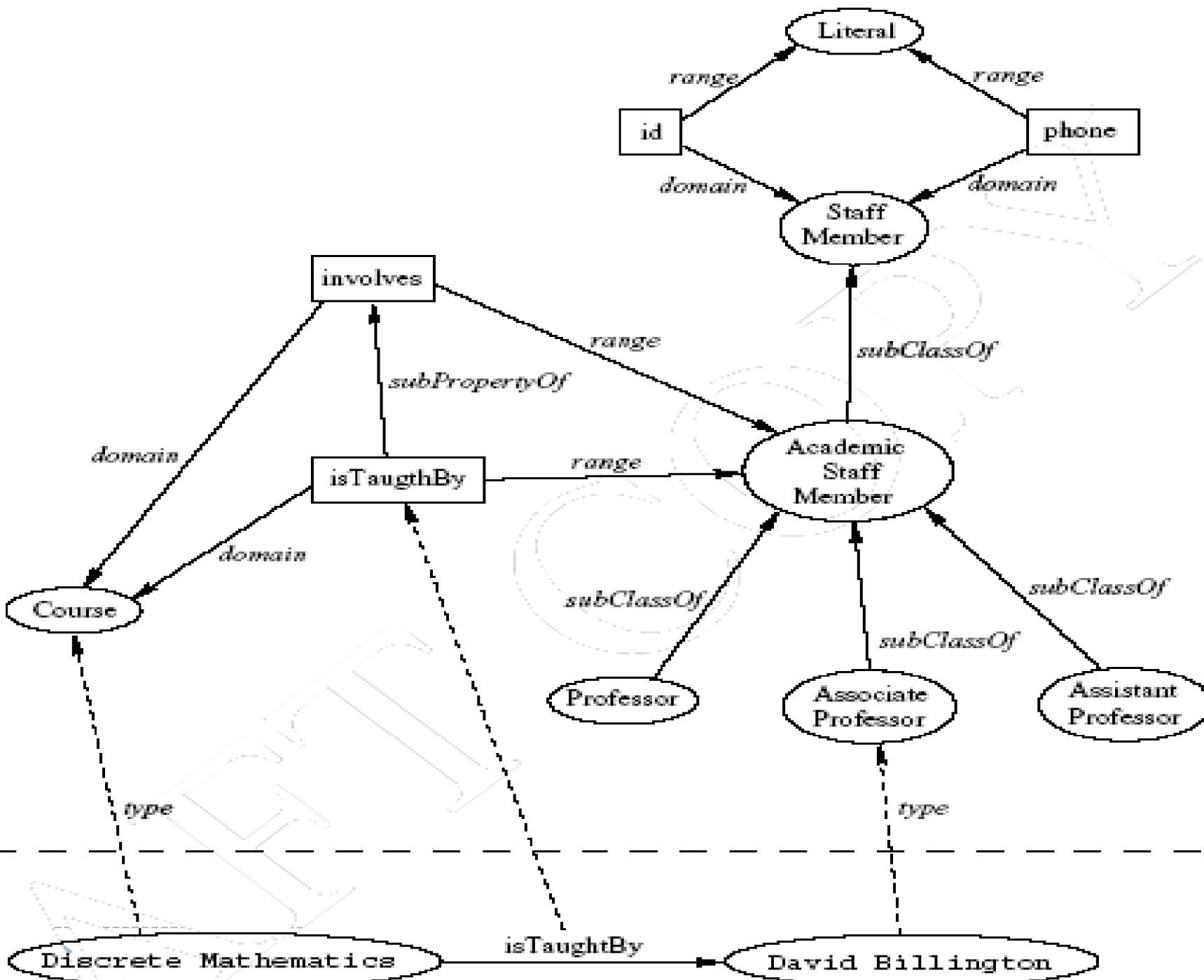
```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:h="http://inria.fr/2005/humans.rdfs#" xml:base="http://inria.fr/2005/humans.rdfs-instances" >
```

```
<h:Woman rdf:ID="Alice">
```

```
<h:hasMother rdf:resource="#Laura"/>
```

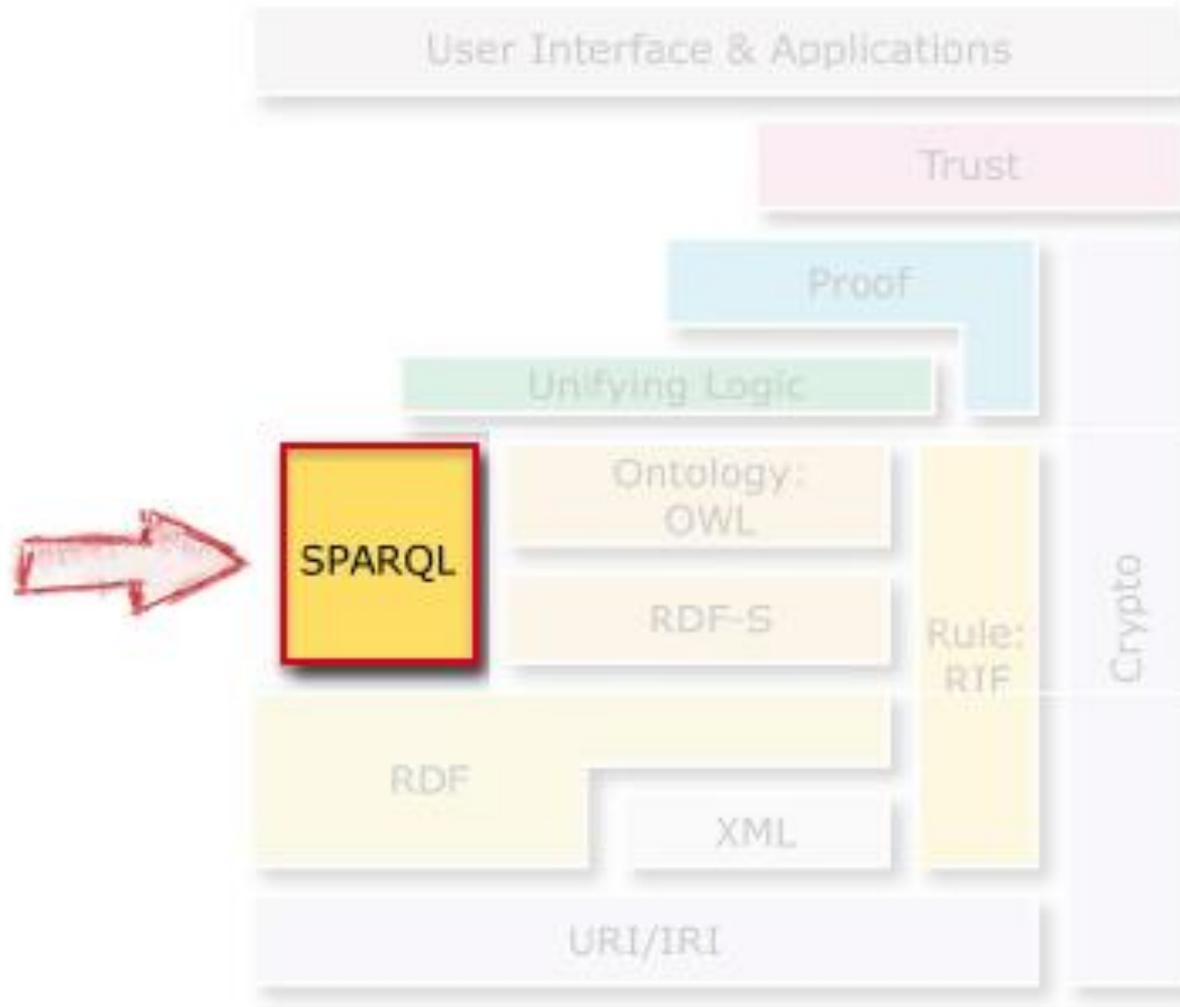
```
</h:Woman>
```

```
</rdf:RDF>
```



RDFS
RDF

Linguagem de consulta SPARQL



SPARQL Protocol e RDF Query Language

1. Linguagem de Consulta (Sintaxe Turtle)

- SPARQL 1.1 Query Language - W3C REC 21 Mar. 2013
- SPARQL 1.1 Update - W3C REC 21 Mar. 2013

2. Formato dos resultados

- SPARQL Query Results XML Format - W3C REC 21 Mar. 2013



Consultas em SPARQL

SELECT o *que* você quer

FROM de *onde* você quer

WHERE {*como* você quiser}

FROM é opcional



Triplas SPARQL

- Usa sintaxe Turtle com interrogações para variáveis:

```
?x rdf:type ex:Person
```

- Especifica um padrão de grafo a ser achado:

```
SELECT ?subject ?property ?value
```

```
WHERE { ?subject ?property ?value }
```

- Um padrão de grafo básico é uma conjunção de triplas

```
SELECT ?x
```

```
WHERE { ?x rdf:type ex:Person .
```

```
      ?x ex:name ?name . }
```



Mesmas abreviações que Turtle

- Triplas com sujeito comum

```
SELECT ?name ?fname  
WHERE {?x a ex:Person;  
       ex:name ?name ;  
       ex:firstname ?fname ;  
       ex:author ?y . }
```



```
SELECT ?name ?fname  
WHERE{?x rdf:type ex:Person.  
      ?x ex:name ?name .  
      ?x ex:firstname ?fname .  
      ?x ex:author ?y . }
```

- Vários valores:

?x ex:firstname "Fabien", "Lucien" .

- Variáveis anônimas (blank nodes)

[ex:firstname "Fabien"]

[] ex:firstname "Fabien" .



Declarando prefixos e Namespaces

- Declaração de **prefixos** para vocabulários usados na consulta.

```
PREFIX mit: <http://www.mit.edu#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?student
WHERE {
  ?student mit:registeredAt ?x .
  ?x foaf:homepage <http://www.mit.edu> .
}
```

- Declaração de namespace para URIs relativos

```
BASE <http://ns.inria.ft/>
```



Especificando Idioma e tipo de dados de literais

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?x ?f WHERE {
  ?x foaf:name "Fabien"@fr ; foaf:knows ?f .
}
```

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?x WHERE {
  ?x foaf:name "Fabien"@fr ;
     foaf:age "21"^^xsd:integer .
}
```



Alguns comandos : Padrão opcional

- Quando parte do padrão de grafo não é obrigatório,

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?person ?name
WHERE {
  ?person foaf:homepage <http://fabien.info> .
  OPTIONAL { ?person foaf:name ?name . }
}
```

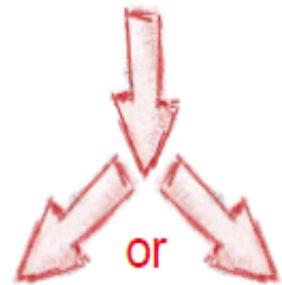
?name variable may be « unbound » in result



Alguns comandos : Padrões alternativos

- União de resultados de padrões de grafos

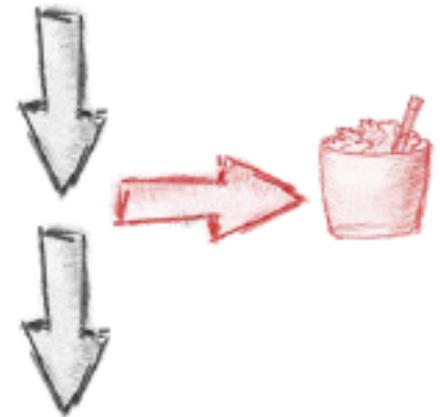
```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?person ?name
WHERE {
  ?person foaf:name ?name .
  {
    ?person foaf:homepage <http://fabien.info> .
  }
  UNION
  {
    ?person foaf:homepage <http://bafien.org> .
  }
}
```



Negação

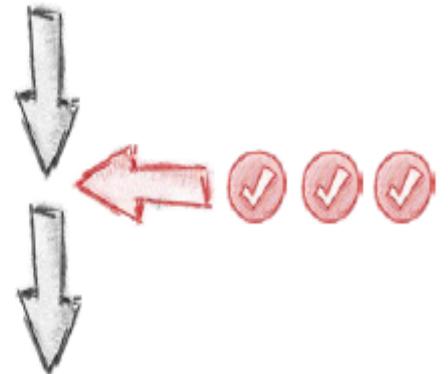
- Remove resultados que casam com um padrão

```
PREFIX ex: <http://www.example.abc#>  
SELECT ?x  
WHERE {  
    ?x a ex:Person  
    MINUS { ?x a ex:Man }  
}
```



Variable Binding

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?person ?name
WHERE {
  ?person ex:fname ?fname ;
          ex:lname ?lname .
  BIND (concat(?fname, ?lname) AS ?name)
}
```



Propriedades de caminho

- Expressões regulares sobre propriedades de caminho entre recursos

/ : seqüência

| : alternativo

+ : um ou vários

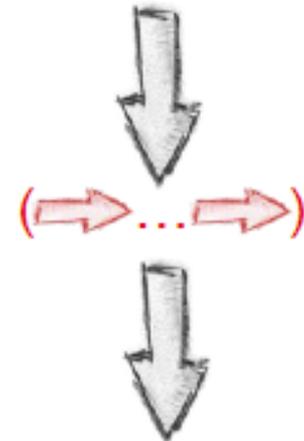
* : zero ou vários

? : opcional

^ : reverso

! : negação

```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
SELECT ?friend WHERE {
  ?x foaf:name "Fabien Gandon" ;
  foaf:knows+ ?friend .
}
```



Mantenha resultados diferentes

- Mantenha uma ocorrência de resultados semelhantes com os mesmos valores para as mesmas variáveis

PREFIX foaf: <http://xmlns.com/foaf/0.1/>

SELECT DISTINCT ?name

WHERE { ?person foaf:name ?name . }



Filtragem de resultados

- Declara restrições sobre valores de variáveis
- **Select** : seleciona os valores a serem retornados
- **Where**: padrão de grafo
- **Filter** : restringe os valores retornados no where com cláusulas com expressões e funções



Pessoas maiores iguais de 18 anos

```
PREFIX ex: <http://inria.fr/schema#>
SELECT ?person ?name
WHERE {
    ?person    rdf:type  ex:Person ;
               ex:name   ?name ;
               ex:age    ?age .
    FILTER (xsd:integer(?age) >= 18)
}
```



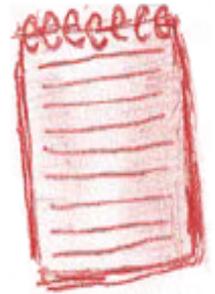
Teste de valores

- Testa e compara constantes, variáveis e expressões
- Comparadores: $<$, $>$, $=$, $<=$, $>=$, $!=$
- Expressões regulares: `regex(?x, "A.*")`
- Testa valores de variáveis: `isURI(?x)`, `isBlank(?x)`, `isLiteral(?x)`, `bound(?x)`
- Podem ser usadas funções sobre literais, numéricos, datas.



Cadeias de caracteres e literais

<code>CONTAINS(lit₁, lit₂)</code>	string inclusion
<code>STRSTARTS(lit₁, lit₂)</code>	string inclusion
<code>STRENDIS(lit₁, lit₂)</code>	string inclusion
<code>STRDT(value, type)</code>	create literal with datatype
<code>STRLANG(value, lang)</code>	create literal with language
<code>CONCAT(lit₁, ..., lit_n)</code>	concatenate strings
<code>SUBSTR(lit, start [, length])</code>	extract substring
<code>ENCODE_FOR_URI(str)</code>	encode string for URI
<code>UCASE(str)</code> , <code>LCASE(str)</code>	change case
<code>STRLEN(str)</code>	string length



Outras funções

YEAR(Date), MONTH(Date), DAY(Date)
HOURS(Date), MINUTES(Date), SECONDS(Date)
NOW()



ABS(Val), CEIL(Val), FLOOR(Val), ROUND(Val)
isNumeric(Val)
RAND()



COALESCE(val₁, ..., val_n)
IRI(str), URI(str)
BNODE(ID)



Conectores Lógicos

- And: &&
- Or: ||
- Not: !
- ()



Mais informações

- Apache Jena SPARQL Tutorial
<https://jena.apache.org/tutorials/sparql.html>
- [SPARQL By Example - World Wide Web Consortium](#)
- [A Linguagem de Consulta SPARQL](#)
- ...



Linked Data vs. Semantic Web (original)

- ▶ *“Em contraste com a visão completa da Web Semântica, os dados ligados referem-se principalmente à publicação de dados estruturados no RDF usando URIs, em vez de se concentrar no nível ontológico ou de inferência. Essa simplificação - assim como a Web simplificou as abordagens acadêmicas estabelecidas dos sistemas de hipertexto - reduz a barreira de entrada para os provedores de dados, portanto, promove uma adoção generalizada.”* (Hausenblas, 2009)
- ▶ Dados ligados trata do uso de tecnologias da web semântica na publicação de dados estruturados na Web e na definição de enlaces entre fontes de dados.



- ▶ A mensagem principal da comunidade de dados ligados → uma abordagem bottom-up para alavancar a Publicação na Web Semântica, onde a ênfase não está nas linguagens, mas na vinculação de dados estruturados na web, facilitando a interoperabilidade. (Hogan, 2014).
- ▶ Essa filosofia bottom-up é melhor condensada pelo esquema de 5 estrelas dos dados abertos ligados (LOD) (Berners-Lee, 2006).

★ PUBLISH DATA ON THE WEB UNDER AN OPEN LICENSE
★★ PUBLISH STRUCTURED DATA
★★★ USE NON-PROPRIETARY FORMATS
★★★★ USE URIS TO IDENTIFY THINGS
★★★★★ LINK YOUR DATA TO OTHER DATA





Linked Open Data



- ▶ Para promover os princípios de dados ligados → W3C criou um projeto comunitário chamado de “Dados abertos interligados” (LOD). Inspirado no movimento de dados abertos.
- ▶ O objetivo do projeto de dados ligados é duplo:
 1. Introduzir os benefícios do RDF e as tecnologias de web semântica ao movimento de dados abertos;
 2. Alavancar a Web de dados criando, publicando e interligando exportações RDF desses conjuntos de dados abertos.
- ▶ Um conjunto completo de diretrizes de publicação sobre os princípios fundamentais é descrito em (Heath, Bizer; 2011).



Diretrizes principais

- ▶ Práticas de desreferenciação: Qualquer URI HTTP deve ser desreferenciável, o que significa que os clientes HTTP podem consultar o URI usando o protocolo HTTP e recuperar uma descrição do recurso identificado pelo URI (Seja HTML-humanos ou RDF-máquinas)
- ▶ Ligando Aliases
- ▶ Descrevendo os Termos de Vocabulário
- ▶ Fornecimento de SPARQL Endpoints



-
- ▶ A comunidade de dados abertos ligados está crescendo. Editores de dados ligados inclui atualmente a BBC, the New York Times, Sears, Freebase (pertence a Google), o governo do Reino Unido, etc.
 - ▶ Para acompanhar o crescimento dos conjuntos de dados publicados, Cyganiak e Jentzch começaram em 2007 o “Diagrama de Nuvem de dados abertos ligados”. Veja Figura.
 - ▶ Cada nó representa um conjunto de dados com um URI desreferenciável que recuperam documentos RDF. O dataset deve conter no mínimo containig at least 1000 triplas e 50 enlaces a conjuntos de dados externos.
 - ▶ In 2014, continha 570 datasets com alguns bilhões de triplas e vários milhões de entidades.
 - ▶ A nuvem atualmente contem **1301** conjuntos de dados com **16,283** enlaces (Maio 2020)
-



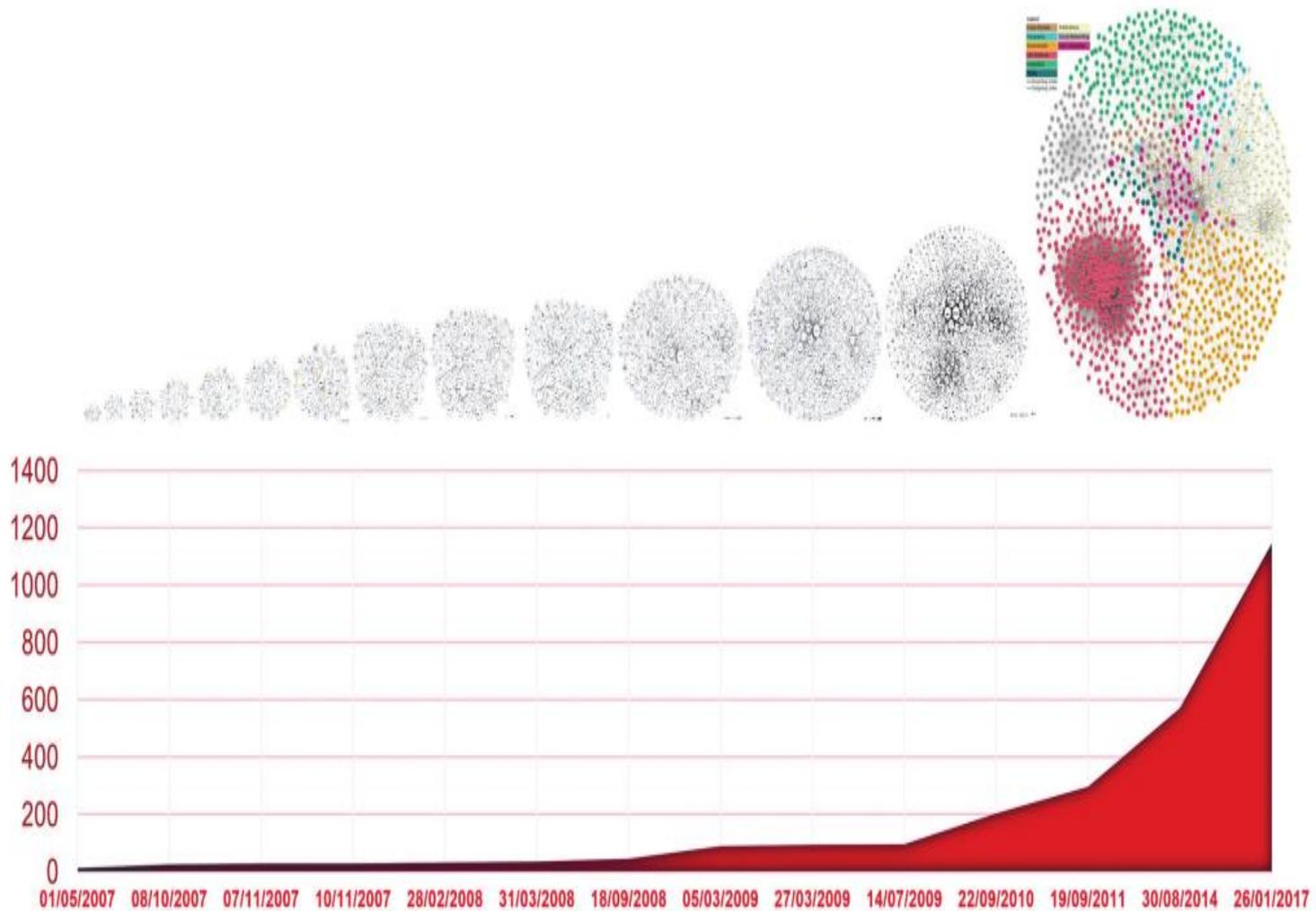
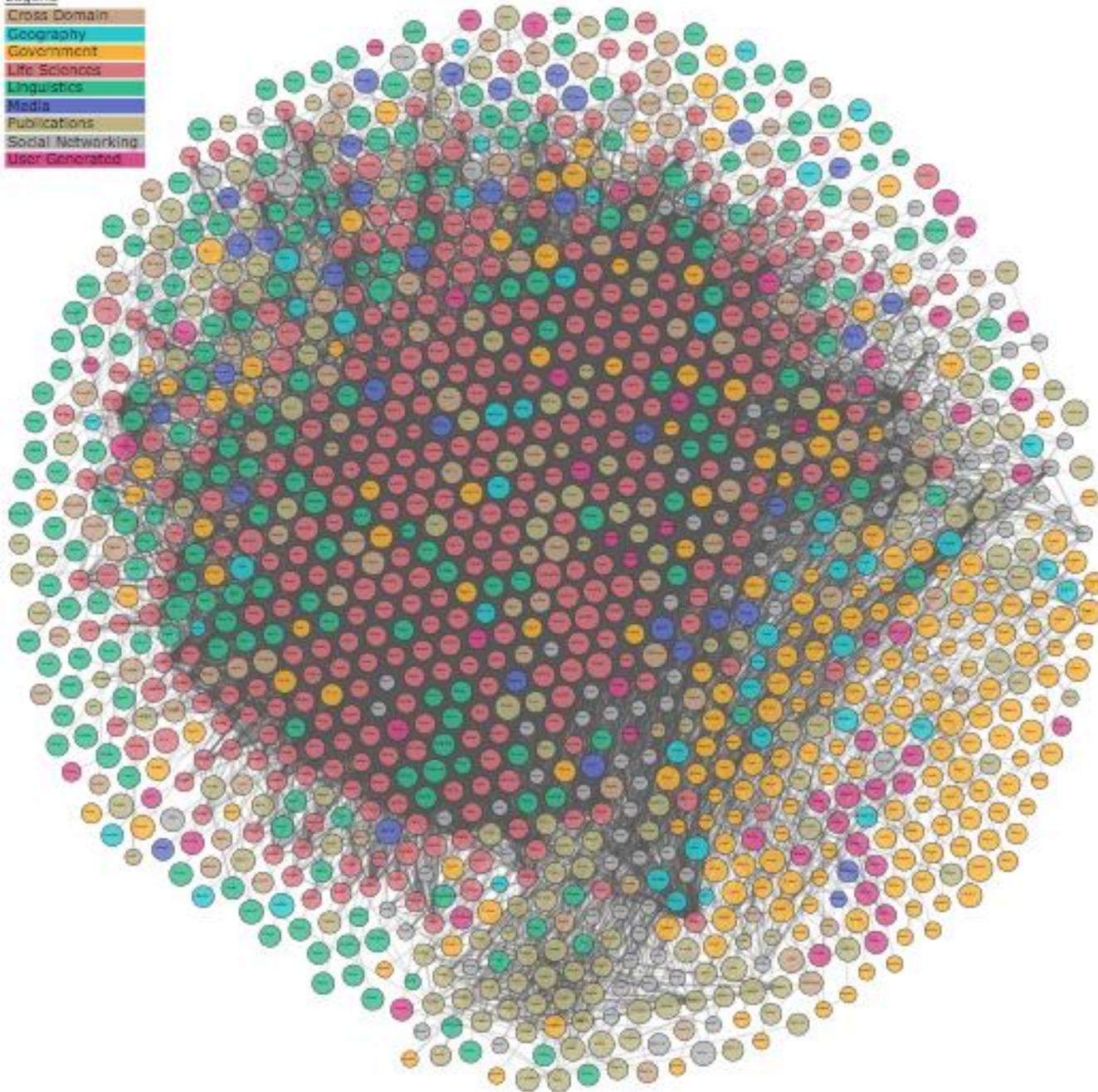


Figure 1.1 Number of linked open datasets on the Web in the Linked Open Data Cloud. From the Linked Open Data Cloud at lod-cloud.net.

Fonte: Allemang, D. et al., 2020

Legend

- Cross Domain
- Geography
- Government
- Life Sciences
- Linguistics
- Media
- Publications
- Social Networking
- User Generated



Grafos de Conhecimento

- Em 2012, um novo termo apareceu, quando Google lançou seu Grafo de Conhecimento (GC). Peças do GC, podem ser vistas quando se faz buscas por entidades importantes no Google.
- Veja Infobox ligado à busca.
- Google oferece uma API para acessar o conteúdo do GC.

Figure 2. Google Knowledge Graph node as shown after searching on google.com for the term "Kofi Annan."



The image shows a screenshot of a Google Knowledge Graph node for Kofi Annan. At the top, there is a red banner with white text that reads: "Figure 2. Google Knowledge Graph node as shown after searching on google.com for the term 'Kofi Annan.'" Below the banner, the node is displayed. On the left, the name "Kofi Annan" is written in a large, bold font, with "Ghanaian diplomat" underneath it. To the right of the name is a small portrait of Kofi Annan. Below the name and portrait, there is a paragraph of text: "Kofi Atta Annan was a Ghanaian diplomat who served as the seventh Secretary-General of the United Nations from January 1997 to December 2006. Annan and the UN were the co-recipients of the 2001 Nobel Peace Prize. Wikipedia". Below this paragraph, there are several lines of text providing biographical information: "Born: April 8, 1938, Kumasi, Ghana", "Died: August 18, 2018, Bern, Switzerland", "Full name: Kofi Atta Annan", "Nationality: Ghanaian", "Education: University of Geneva, MDRE", and "Spouse: Nane Maria Annan (m. 1964–2018), Tsi Akakye (m. 1965–1983)".

Grafos de Conhecimento

- Tecnologias de GC têm achado um lugar importante na indústria. Além do Google, Microsoft, IBM, Facebook e eBay (Noy, N., et al, 2019).
 - Visão de grafo de conhecimento (um novo enquadramento das ideia da web semântica?). Diferenças:
 - visão de abertura dos dados ligados muda nos GC que são liderados pela indústria e não são verdadeiramente abertos. A nuvem de dados ligados é dificilmente uma entidade concisa, ela consiste de subgrafos individuais fracamente interligados, cada um governado pela sua própria estrutura, esquema de representação, etc. GCs são entendidos como artefatos muito mais consistentes internamente e mais fortemente controlados.
-
- ▶ Transição de uma pesquisa acadêmica (web de dados) para o seu uso na indústria (GC)

Grafos de Conhecimento

- O emprego de abstrações do conhecimento baseadas em grafos tem permitido uma ligação com a área de Aprendizado de Máquina. **Veja fig.** Aprendizado de Máquina e especificamente "Deep Learning", estão sendo usados com GC para completar grafos (adição de relacionamentos), tratar com dados "noisy", Graph Embedding, etc. GCs estão sendo usados como instrumentos na área de EAI (IA Explicável). (Hitzler, P. et al, 2020)
- O processamento de LN, está ligado à área de de GC em aplicações de atendimento a consultas (query answering), criação de GC a partir de textos, entre outros.

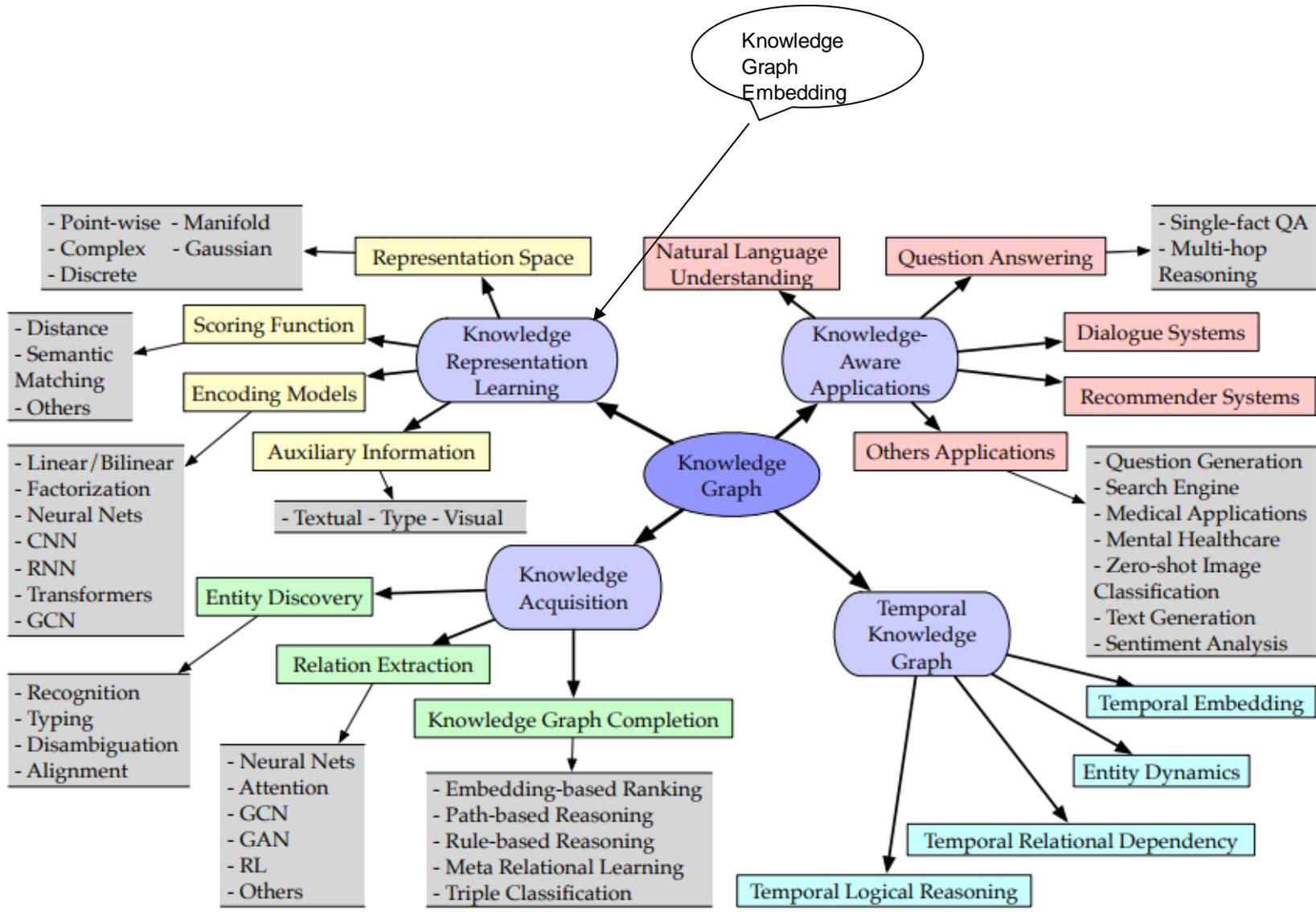


Fig. 2: Categorization of research on knowledge graphs

- O campo de BDs, está ligado aos GC, especificamente os bancos de dados de grafos (tipo de sistemas NoSQL) como uma alternativa de gerenciamento e armazenamento de grandes GCs. Sistemas como Neo4J, GraphDB, etc., estão sendo usados para o armazenameto e gerenciamento de GCs.
- Mais informações em Hogan, A., et al.,2022 e Costabello, L. et al., 2020.



Referências

- Allemang, D. et al. Semantic Web and the working Ontologist. 3ed. Edition. ACM Press. 2020.
- T. Berners-Lee. Linked Data—Design Issues, 2006. <http://www.w3.org/DesignIssues/LinkedData.html>.
- Bush, V. with his Differential Analyzer, c. 1935. MIT Museum by Revol Web - CC BY-SA 2.0
<https://www.flickr.com/photos/revolweb/3985110824/>
- Costabello, L. et al. Knowledge Graph Embeddings Tutorial: From Theory to Practice. <https://kge-tutorial-ecai2020.github.io/>
- Complex information processing: a file structure for the complex, the changing and the indeterminate, T. H. Nelson, ACM, 1965
- Gandon, F. Web of Data Course. Coursera.
<https://www.coursera.org/learn/web-data/home/welcome>



Referências

- Heath, T.; Bizer, C. Linked Data: Evolving the Web into a Global Data Space. Morgan & Claypool Pubs. Series, 2011.
- M. Hausenblas. Exploiting Linked Data to Build Web Applications. IEEE Internet Computing, 13(4):68–73, 2009.
- Hitzler, P. et al. Neural-Symbolic integration and the Semantic Web. Semantic Web Journal. Vol. 11. No. 1. 2020.
- Hitzler, P. A Review of the Semantic Web Field. Communications of the ACM. Vol. 64, No. 2, 2021.
- Hogan, A. Linked Data & the Semantic Web Standards. In: Harth, A. et. al (eds.) Linked Data Management. CRC Press, 2014.



- Hogan, A. et al. Knowledge Graphs. Morgan and Claypool, Pubs. 2022.
- Ji, S. et al. A Survey on Knowledge Graphs: Representation, Acquisition and Applications. IEEE Transactions on Neural Networks and Learning Systems. Vol. 33, No. 2. 2022.
- Noy, N., et al. Industry-scale knowledge graphs lessons and challenges. Communications of the ACM. Vol. 62, No. 8. 2019.

