

# Web ngữ nghĩa

Soạn bởi: Nguyễn Bá Ngọc

## Chương 1

Hà Nội-2021

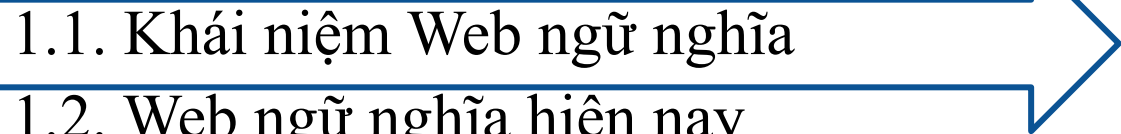
# Chương 1.

## Tổng quan về Web ngữ nghĩa

# Nội dung

- 1.1. Khái niệm Web ngữ nghĩa
- 1.2. Web ngữ nghĩa hiện nay
- 1.3. Đồ thị tri thức
- 1.4. Tổng quan công nghệ Web ngữ nghĩa
- 1.5. Ontology và một số khái niệm quan trọng

# Nội dung

- 
- 1.1. Khái niệm Web ngữ nghĩa
  - 1.2. Web ngữ nghĩa hiện nay
  - 1.3. Đồ thị tri thức
  - 1.4. Tổng quan công nghệ Web ngữ nghĩa
  - 1.5. Ontology và một số khái niệm quan trọng

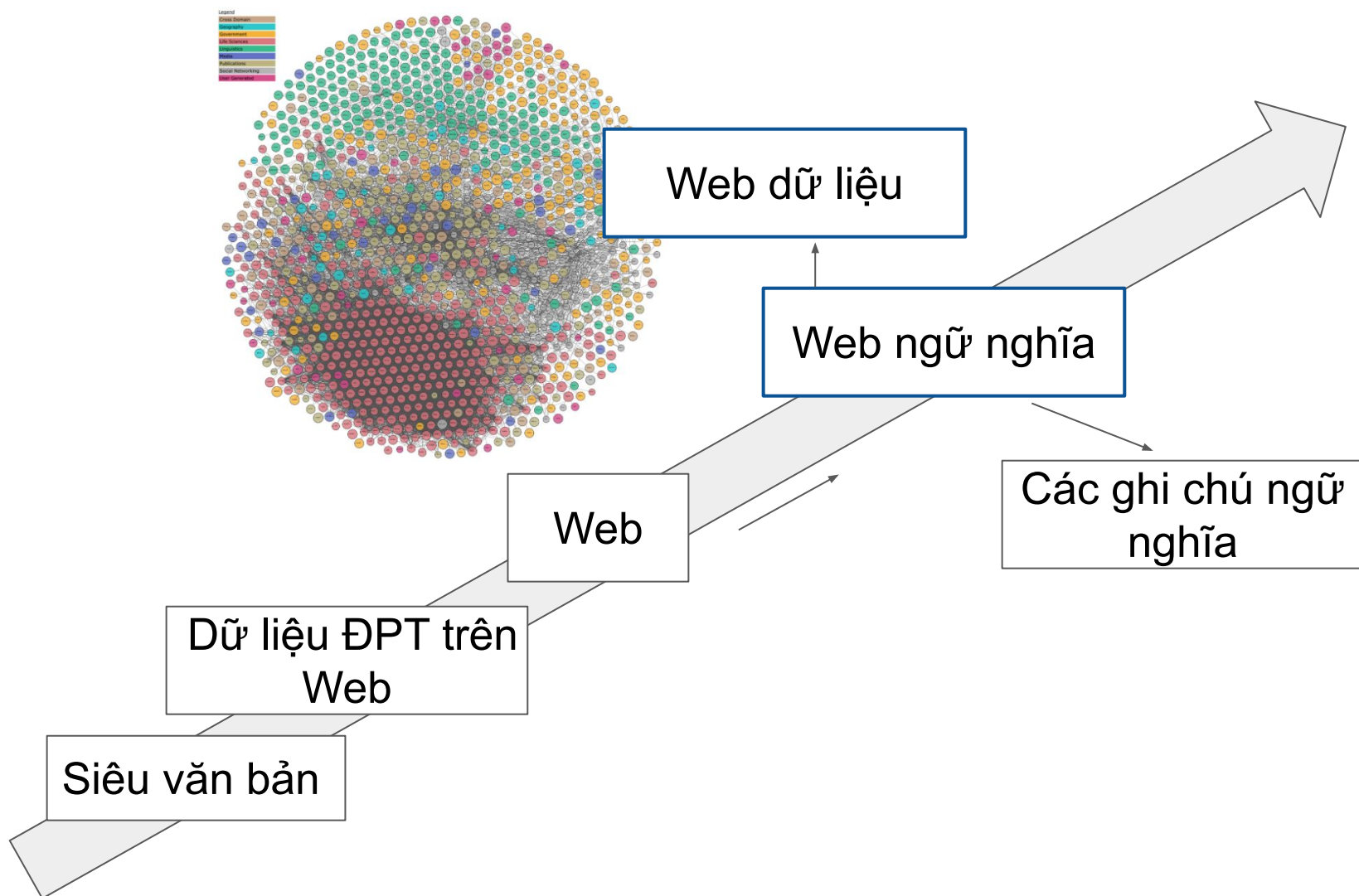
# Tóm lược về Web

- Web là nguồn tri thức lớn của nhân loại
- Tuy nhiên Web có một số hạn chế khiến cho việc xử lý gặp nhiều khó khăn
  - Được thiết kế cho người, không phải cho máy
    - Các nội dung chủ yếu là văn bản, âm thanh, hình ảnh, video
    - Người có thể dễ dàng tiếp nhận và hiểu so với máy tính...
    - ... Tuy nhiên khả năng xử lý ngữ nghĩa bằng máy tính là tiềm năng lớn để phát triển nhiều ứng dụng hữu ích cho con người.

# Tóm lược về Web<sub>(2)</sub>

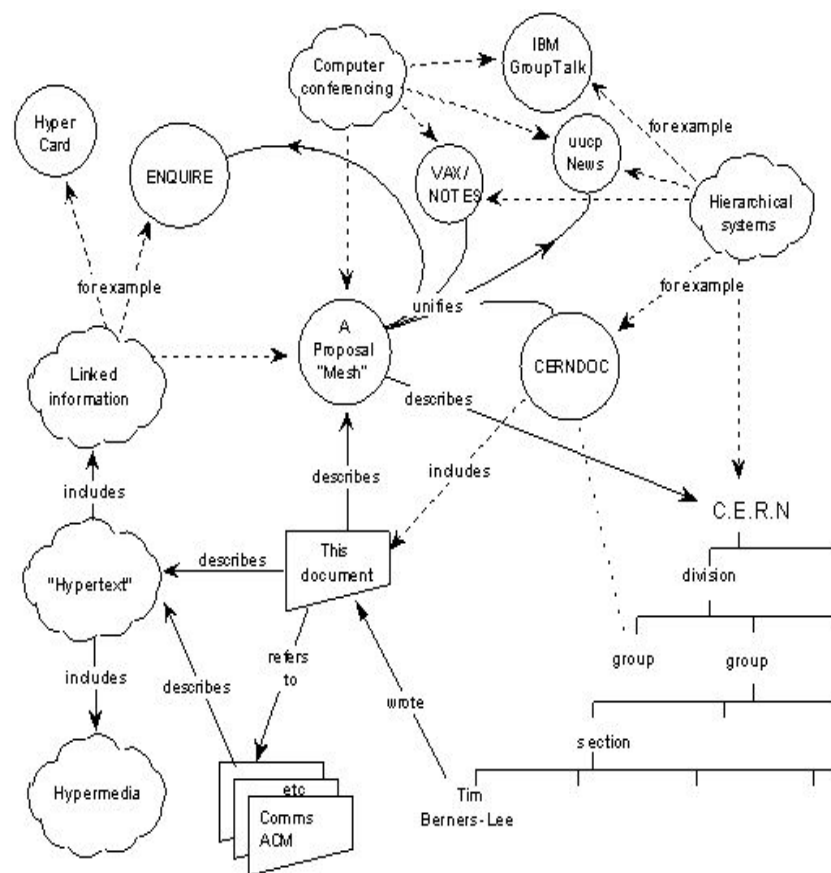
- Truy cập các nội dung Web được thực hiện chủ yếu thông qua cơ chế tìm kiếm thông tin.
  - Với các công nghệ hiện tại:
    - Các truy vấn dựa trên từ khóa (ngôn ngữ tự nhiên) => Danh sách các tài liệu được xếp hạng
    - Người dùng tự rà soát danh sách tài liệu để lọc ra các thông tin phù hợp
  - Trong nhiều trường hợp, người dùng muốn có những kết quả trực tiếp:
    - Ví dụ, Số PI => 3.14159265359
    - => Cần khả năng xử lý ngữ nghĩa bằng máy tính

# Dữ liệu trong môi trường Web



# Sự xuất hiện của Web ngữ nghĩa

- Trong một đề xuất năm 1989 Tim Berners-Lee đã mô tả Web của các mối quan hệ giữa các đối tượng có tên như sự hợp nhất của nhiều hoạt động quản lý thông tin
- Vắn tắt lịch sử
  - Guha MCF (~94)
  - XML+MCF=>RDF(~96)
  - RDF+OO=>RDFS(~99)
  - RDFS+KR=>DAML+OIL(00)
  - Các hoạt động của W3C (01)
  - Quy chuẩn OWL của W3C (03)
  - ...



<http://www.w3.org/History/1989/proposal.html>



# Hai khái niệm Web ngữ nghĩa

- Web ngữ nghĩa 1

- Khái niệm đầu tiên được đưa ra bởi Tim Berners-Lee
- Ý tưởng về Web của thông tin mà máy tính có thể hiểu
- Không giới hạn công nghệ hỗ trợ
- Có thể yêu cầu nhiều TTNT hơn (ví dụ NLP, học máy)
- Người dùng cuối là con người

- Web ngữ nghĩa 2

- Được định hình bởi W3C: Web của dữ liệu
- Sử dụng các quy chuẩn W3C, Ví dụ, RDF, OWL, SPARQL, XML, RIF, v.v.
- Hỗ trợ tương tác máy-máy làm nền tảng cho các ứng dụng phục vụ con người

# Các mục tiêu của Web ngữ nghĩa theo W3C

- Tập trung vào các xử lý bằng máy tính:

*"Web ngữ nghĩa là một mở rộng của Web hiện tại trong đó ý nghĩa của thông tin được thiết lập tường minh, mở ra khả năng cộng tác tốt hơn giữa người và máy tính."*

-- Berners-Lee, Hendler and Lassila, The Semantic Web, Scientific American, 2001

*Máy tính cũng có thể hiểu thông tin trong môi trường Web ngữ nghĩa*



# Vì sao khó xử lý ngữ nghĩa trong HTML

## HTML

`<h1>`

`<span>`

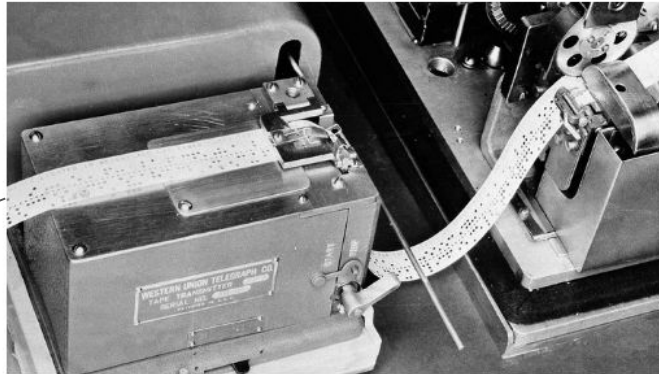
`<picture>`

`<p><i>`

## As We May Think

"Consider a future device ... in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility. It is an enlarged intimate supplement to his memory."

VANNEVAR BUSH JULY 1945 ISSUE



The telegram was a breakthrough in communication technology, which Vannevar Bush imagined could evolve in unprecedented ways. (AP)

*As Director of the Office of Scientific Research and Development, Dr. Vannevar Bush has coordinated the activities of some six thousand leading American scientists in the application of science to warfare. In this significant article he holds up an incentive for scientists when the fighting has ceased. He urges that men of science should then turn to the massive task of making more accessible our bewildering store of knowledge. For years inventions have extended man's physical powers rather than the powers of his mind. Trip hammers that multiply the fists, microscopes that sharpen the eye, and engines of destruction and detection are new results, but not the end results, of modern science. Now, says Dr. Bush, instruments are at hand which, if properly developed, will give man access to and command over the inherited knowledge of the ages. The perfection of these pacific instruments should be the first objective of our scientists as they emerge from their war work. Like Emerson's famous address of 1837 on "The American Scholar," this paper by Dr. Bush calls for a new relationship between thinking man and the sum of our knowledge. — THE EDITOR*

## Ngữ nghĩa

Tiêu đề bài viết

Tác giả & năm xuất bản

Hình minh họa máy điện tín

Lời giới thiệu của người biên soạn

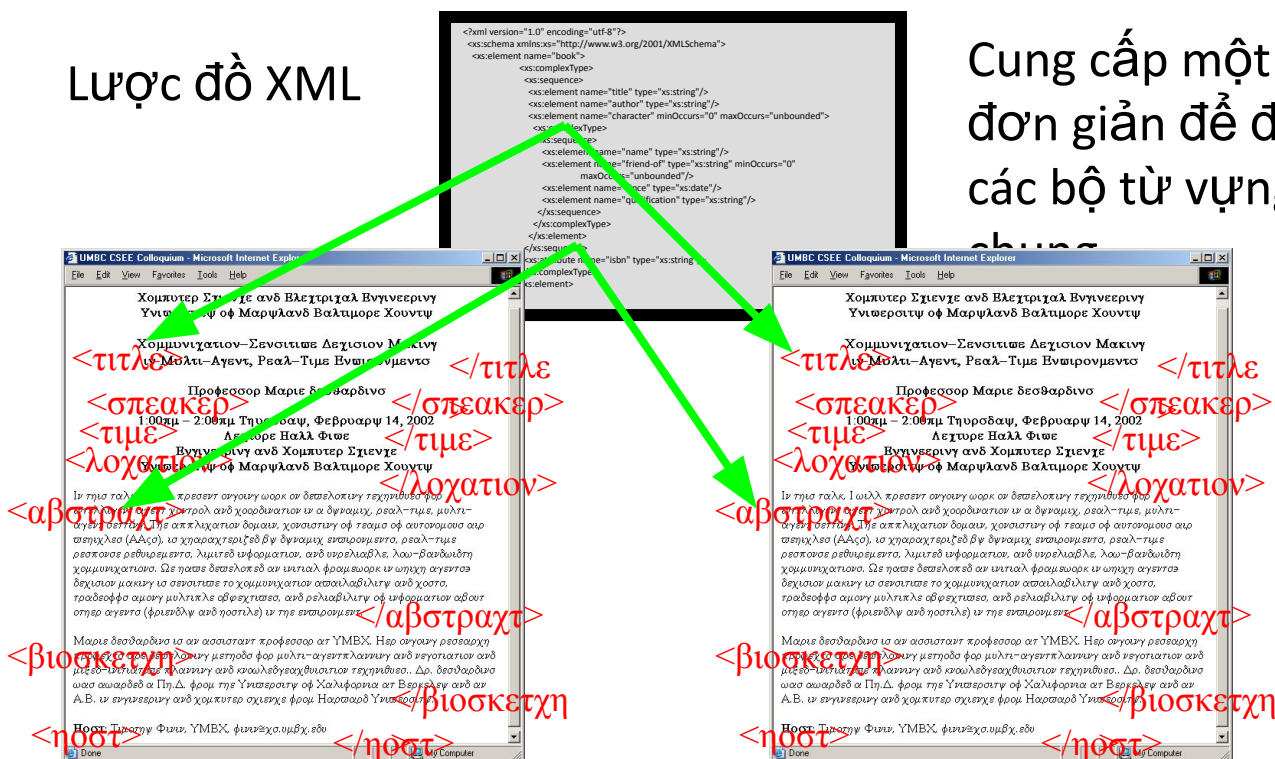
# Giải quyết các vấn đề bằng XML?

- Người dùng XML tự tạo tập thẻ riêng cho từng ứng dụng
  - Có thể dùng các thẻ <title>, <author>, v.v..
- XML ≠ biểu diễn ngữ nghĩa dành cho máy
  - Các tên thẻ không chứa đựng ngữ nghĩa
  - DTD hàm chứa rất ít hoặc không hàm chứa ngữ nghĩa
  - Lược đồ XML cung cấp một cơ chế đơn giản để định nghĩa các bộ từ vựng dùng chung: **Hỗ trợ** giải quyết vấn đề xử lý ngữ nghĩa

# Các lược đồ XML

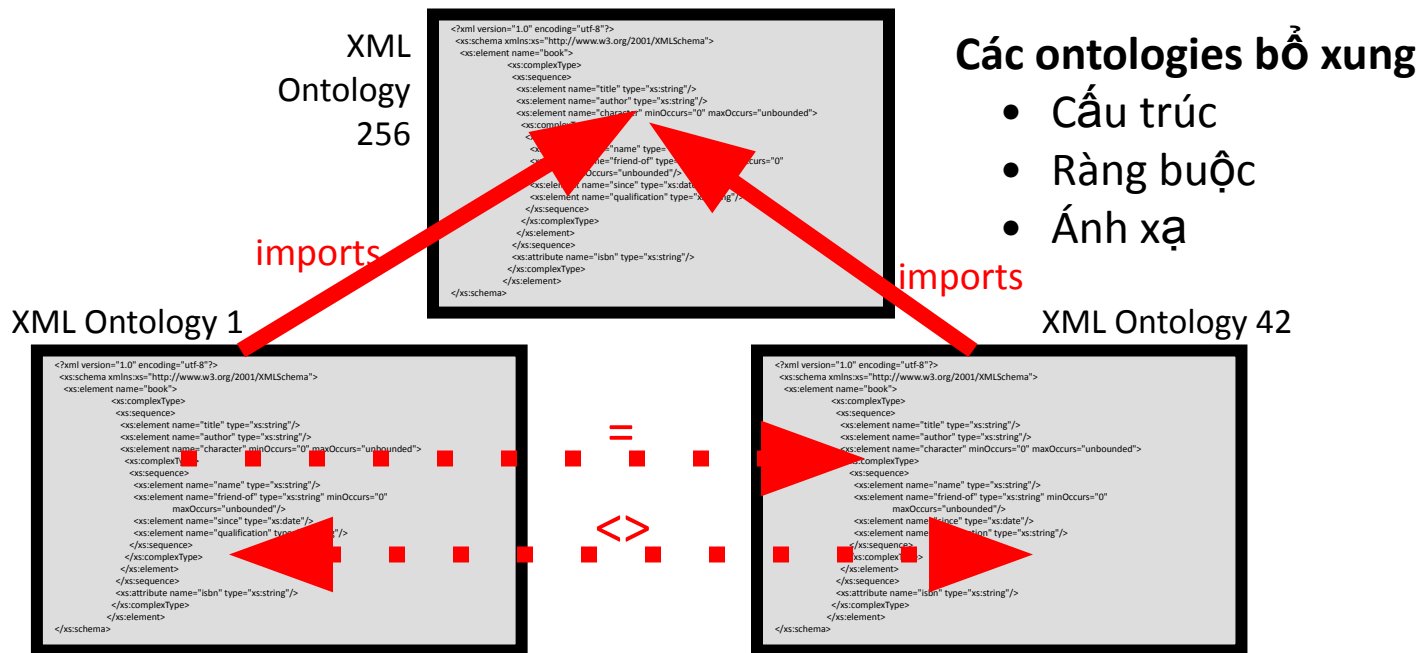
Lược đồ XML

Cung cấp một cơ chế đơn giản để định nghĩa các bộ từ vựng dùng chung



Có thể có quá nhiều lược đồ XML... Nhưng không có cơ chế liên kết các lược đồ XML

# Cần biểu diễn ngữ nghĩa mức cao



## Được biểu diễn như những tài liệu XML

## Có thể liên kết các lược đồ

## Máy tính có thể hiểu ngữ nghĩa (ở một mức độ nhất định)

# Web ngữ nghĩa

- Sử dụng công nghệ Web ngữ nghĩa để chia sẻ dữ liệu và tri thức trong môi trường Web
  - Các bộ dữ liệu được liên kết để hỗ trợ việc tích hợp và kết nối tri thức
- Các hệ thống máy tính chia sẻ dữ liệu và tri thức bằng các quy chuẩn ngôn ngữ và giao thức Web

# Web ngữ nghĩa: Các ngôn ngữ

Các ngôn ngữ thường được chia thành ba thành phần:

1. Cú pháp: Syntax - các quy ước ngữ pháp để tạo các câu trong ngôn ngữ
2. Ngữ nghĩa: Semantic - ánh xạ các câu tới ý nghĩa (có thể là giá trị chân lý theo lý thuyết)
3. Ngữ dụng: Pragmatics - phần còn lại, các khía cạnh ứng dụng (cách vận dụng ngôn ngữ, hiểu biết về thế giới, v.v..)

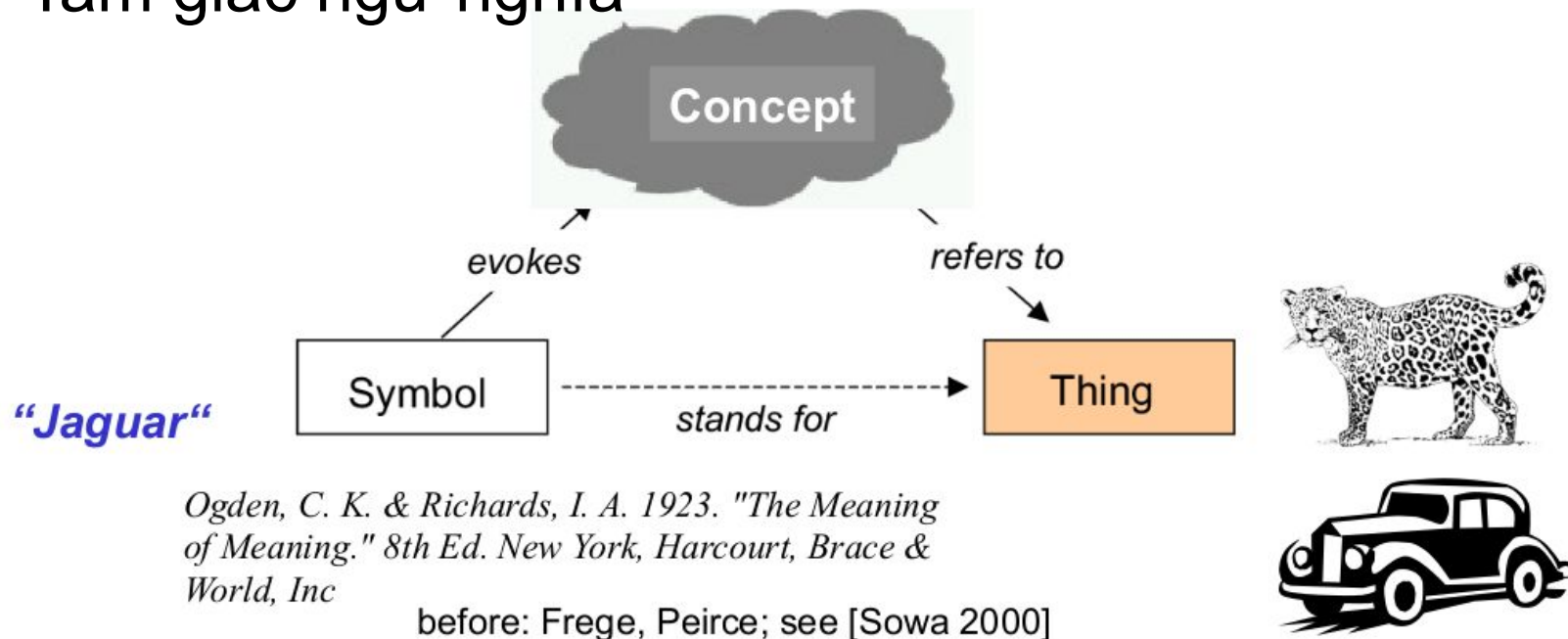


# Ngôn ngữ hình thức: Cú pháp

- Các URIs (Uniform Resource Identifier) chỉ các lớp, các thuộc tính, các đối tượng, các mối liên hệ  
[http://live.dbpedia.org/resource/Alan\\_Turing](http://live.dbpedia.org/resource/Alan_Turing)  
<http://schema.org/Scientist>  
<http://www.w3.org/1999/02/22-rdf-syntax-ns#type>
  - Sử dụng chuỗi ký tự cho nội dung văn bản
  - Sử dụng bộ ba để tạo các câu
- dbpedia:Alan\_Turing    rdfs:type    schema:Scientist .
- “Alan Turing is a scientist” - "Alan Turing là một nhà khoa học"

# Ngôn ngữ hình thức: Ngữ nghĩa

## Tam giác ngữ nghĩa

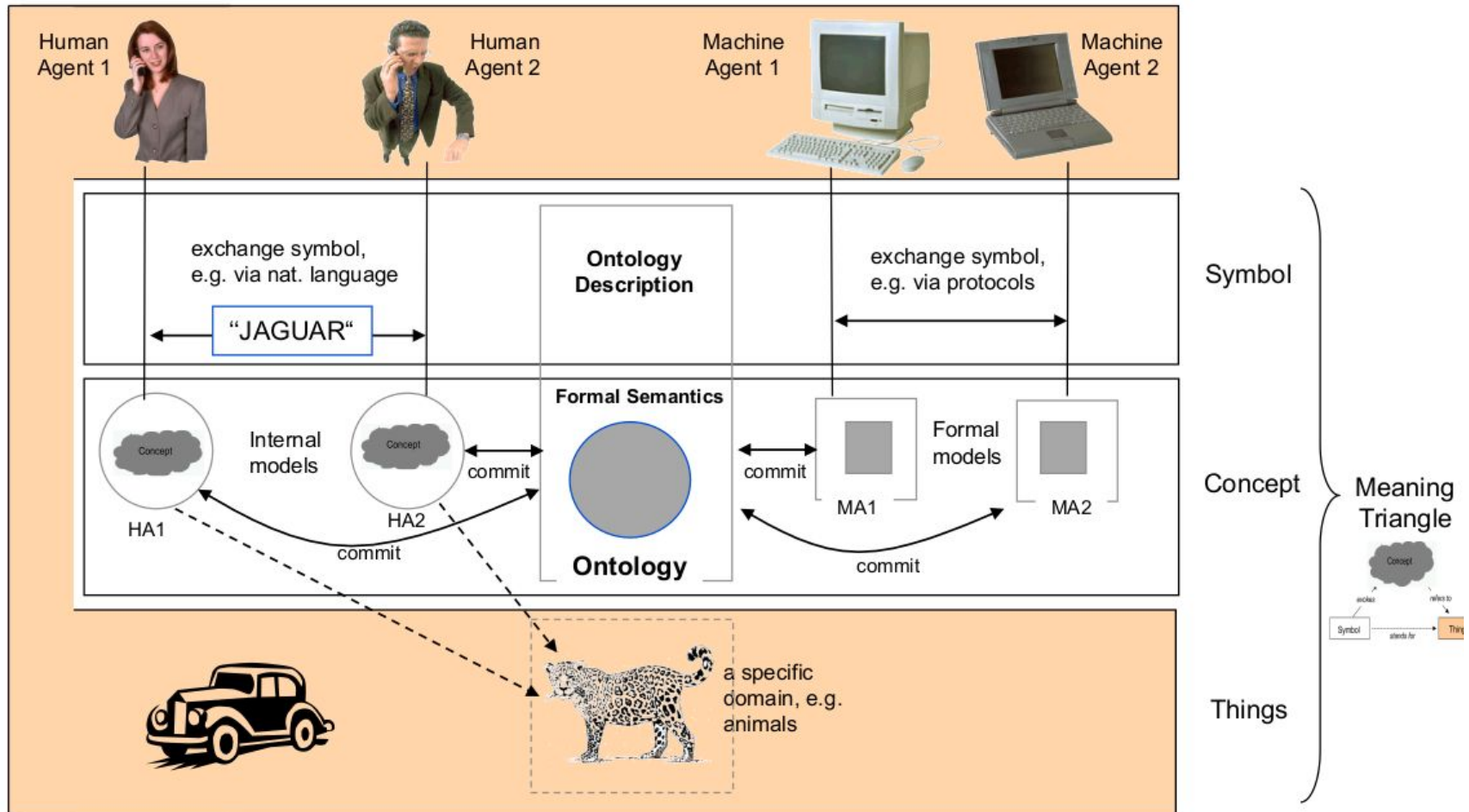


[Carole Goble, Nigel Shadbolt, Ontologies and the Grid Tutorial]

Trong giao tiếp của người, các khái niệm là trung gian kết nối từ (hoặc tối thiểu là ký hiệu) với các thực thể.

# Ngôn ngữ hình thức: Ngữ nghĩa (2)

## Giao tiếp người và máy



Maedche et al., 2002

# Ngôn ngữ hình thức: Ngữ nghĩa<sub>(3)</sub>

## Ngữ nghĩa trong Web ngữ nghĩa

- Ánh xạ URI tới các thực thể mà nó chỉ định trong "thế giới": Thực và ảo
  - Các URI là duy nhất
- Cung cấp ý nghĩa xác định mà máy tính có thể xử lý
- Ngữ nghĩa của một số URIs cho phép thực hiện suy diễn
  - Quan hệ parent (là phụ mẫu) là nghịch đảo của quan hệ child (là con)
  - `schema:parent owl:inverse schema:child`

# Ngôn ngữ hình thức: Ngữ dụng

- Ngoài giá trị chân lý (các câu kiểm chứng thông tin), ngữ nghĩa còn bao hàm nhiều khía cạnh khác
- Cần tính đến cả các giao thức, ngữ cảnh, v.v..
  - Một số được xử lý bởi Giao thức Web (GET, POST)
  - Một số được xử lý bằng các giao thức chuyên dụng (ví dụ, truy vấn SPARQL)
  - Một số có các cơ sở tri thức chứa các liên kết với thế giới (ví dụ DBpedia) để hỗ trợ xác định các thực thể phổ biến.

# Nội dung

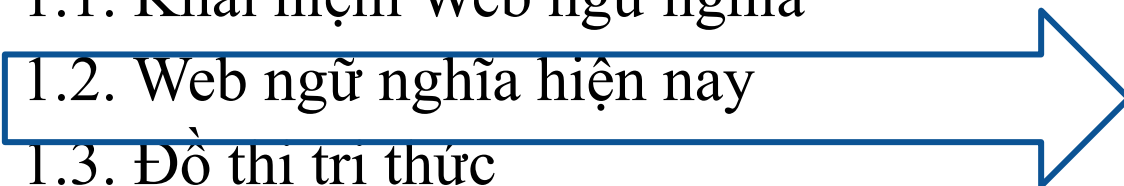
1.1. Khái niệm Web ngữ nghĩa

1.2. Web ngữ nghĩa hiện nay

1.3. Đồ thị tri thức

1.4. Tổng quan công nghệ Web ngữ nghĩa

1.5. Ontology và một số khái niệm quan trọng



# Trạng thái hiện tại

- Phiên bản Web ngữ nghĩa của W3C đã được phát triển liên tục
- Các ngôn ngữ và quy chuẩn Web được sử dụng trong nhiều hệ thống và dịch vụ Web:
  - Google và Facebook phát hiện và sử dụng (một số) dữ liệu RDF ở dạng nhúng trong các trang HTML
  - Google, Yahoo, Microsoft và Yandex tạo lập Schema.org và phát triển những bộ từ vựng dùng chung
  - lod-cloud.net có nhiều bộ dữ liệu mở ở định dạng RDF
  - V.V..

# Các dịch vụ Web hiện có

- Google có thể là một ví dụ tiêu biểu, tuy nhiên các công ty khác cũng có những giải pháp tương đương:
  - 2010 Google thu tóm MediaWeb và cơ sở tri thức Freebase của nó
  - 2014: Freebase có 1.2 tỷ dữ kiện về khoảng 43 triệu thực thể
  - 2015+: Đồ thị tri thức của Google, được cập nhật tự động bằng các kỹ thuật trích rút thông tin





florence price



[Tất cả](#) [Hình ảnh](#) [Video](#) [Tin tức](#) [Mua sắm](#) [Thêm](#) [Cài đặt](#) [Công cụ](#)

Khoảng 137.000.000 kết quả (0,46 giây)

## Florence Price

Born	<b>Florence</b> Beatrice Smith April 9, 1887 Little Rock, Arkansas, United States
Died	June 3, 1953 (aged 66) Chicago, Illinois, United States
Occupation	Musical composer pianist organist music teacher
Years active	1899–1952

[5 hàng khác](#)

[en.wikipedia.org](https://en.wikipedia.org/wiki/Florence_Price) > [wiki](#) > [Florence\\_Price](#)

[Florence Price - Wikipedia](#)

[Giới thiệu về đoạn trích nổi bật](#) • [Phản hồi](#)

## Mọi người cũng hỏi

What is Florence price known for?

**Florence Price** was the first African-American woman to have her music performed by a major symphony orchestra. In 1933, the Chicago Symphony Orchestra gave the world premiere of Symphony No. 1 by a then little-known composer named **Florence Price**. 21 thg 1, 2019

[www.npr.org](https://www.npr.org/2019/01/21/revisiting-the-pioneering-co...) > 2019/01/21 > revisiting-the-pioneering-co...

[Revisiting The Pioneering Composer Florence Price : NPR](#)



## Florence Price

Nhà soạn nhạc

## Có trên

Spotify

Deezer

Được dịch từ tiếng Anh - Florence Beatrice Price là một nhà soạn nhạc cổ điển người Mỹ gốc Phi, nghệ sĩ piano, nhà tổ chức và giáo viên âm nhạc. Giá được ghi nhận là người phụ nữ Mỹ gốc Phi đầu tiên được công nhận là nhà soạn nhạc giao hưởng, và là người đầu tiên có một tác phẩm được chơi bởi một dàn nhạc lớn. [Wikipedia \(tiếng Anh\)](#)

Xem mô tả gốc ▾

**Ngày/nơi sinh:** 9 tháng 4, 1887, Little Rock, Arkansas, Hoa Kỳ

# Facebook



## Introduction

The [Open Graph protocol](#) enables any web page to become a rich object in a social graph. For instance, this is used on Facebook to allow any web page to have the same functionality as any other object on Facebook.

While many different technologies and schemas exist and could be combined together, there isn't a single technology which provides enough information to richly represent any web page within the social graph. The Open Graph protocol builds on these existing technologies and gives developers one thing to implement. Developer simplicity is a key goal of the Open Graph protocol which has informed many of the [technical design decisions](#).

---

## Basic Metadata

To turn your web pages into graph objects, you need to add basic metadata to your page. We've based the initial version of the protocol on [RDFa](#) which means that you'll place additional `<meta>` tags in the `<head>` of your web page. The four required properties for every page are:

- `og:title` - The title of your object as it should appear within the graph, e.g., "The Rock".
- `og:type` - The [type](#) of your object, e.g., "video.movie". Depending on the type you specify, other properties may also be required.
- `og:image` - An image URL which should represent your object within the graph.
- `og:url` - The canonical URL of your object that will be used as its permanent ID in the graph, e.g., "https://www.imdb.com/title/tt0117500/".

As an example, the following is the Open Graph protocol markup for [The Rock on IMDB](#):

```
<html prefix="og: https://ogp.me/ns#">
<head>
<title>The Rock (1996)</title>
<meta property="og:title" content="The Rock" />
```

# Apple

*Siri sử dụng đồ  
thị tri thức để  
hiểu yêu cầu của  
người dùng*

*Trợ lý ảo*

## Calls and Texts

**Siri lets you stay connected  
without lifting a finger.**

Siri can make calls or send texts for you whether you are driving, have your hands full, or are simply on the go.<sup>1</sup> It can even announce your messages on your AirPods.<sup>2</sup> It also offers proactive suggestions — like texting someone that you're running late for a meeting — so you can stay in touch effortlessly.<sup>3</sup>

"Hey Siri, call Mom  
on speaker"



Siri reminds you to make  
the calls that matter.



"Text Donna 'I'm on the  
way exclamation point!'"

# Các dữ kiện cho cơ sở tri thức từ đâu tới?

- Các cơ sở tri thức như DBpedia và Freebase đã bắt đầu bằng cách biểu diễn dữ liệu từ Wikipedia bằng các lược đồ riêng
- Các công nghệ Web ngữ nghĩa giống như mã nguồn mở trong biểu diễn tri thức
  - Vẫn đang liên tục được phát triển
- Trích xuất dữ liệu từ các tài liệu văn bản phi cấu trúc, ví dụ, các bài viết, bài báo, v.v.

# Cơ sở tri thức mở: DBpedia

- DBpedia là một cơ sở tri thức mở dựa trên các công nghệ Web ngữ nghĩa
  - 800 triệu dữ kiện về khoảng 4.6 triệu thực thể từ Wikipedia tiếng anh, có dữ liệu trong khoảng 21 ngôn ngữ
  - Hỗ trợ tích hợp 90 tỉ dữ kiện từ hơn 1000 tập dữ liệu RDF trong không gian dữ liệu liên kết mở.

# DBPedia

Dữ liệu từ  
Wikipedia  
trong RDF

SPARQL Explorer for <http://dbpedia.org/sparql>

SPARQL:

```
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX dc: <http://purl.org/dc/elements/1.1/>
PREFIX : <http://dbpedia.org/resource/>
PREFIX dbpedia2: <http://dbpedia.org/property/>
PREFIX dbpedia: <http://dbpedia.org/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>

SELECT distinct ?soccerplayer ?countryOfBirth ?team ?countryOfTeam ?stadiumcapacity
{
  ?soccerplayer a dbo:SoccerPlayer ;
    dbo:position|dbp:position <http://dbpedia.org/resource/Goalkeeper_(association_football)> ;
    dbo:birthPlace|dbo:country* ?countryOfBirth ;
    #dbo:number 13 ;
    dbo:team ?team .
    ?team dbo:capacity ?stadiumcapacity ; dbo:ground ?countryOfTeam .
    ?countryOfBirth a dbo:Country ; dbo:populationTotal ?population .
    ?countryOfTeam a dbo:Country .
  FILTER (?countryOfTeam != ?countryOfBirth)
  FILTER (?stadiumcapacity > 30000)
  FILTER (?population > 10000000)
} order by ?soccerplayer
```

Results:

SPARQL results:

soccerplayer	countryOfBirth	team	countryOfTeam	stadiumcapacity
:Abdesslam_Benabdellah <a href="#">↗</a>	:Algeria <a href="#">↗</a>	:Wydad_Casablanca <a href="#">↗</a>	:Morocco <a href="#">↗</a>	67000
:Ailton_Moraes_Michellon <a href="#">↗</a>	:Brazil <a href="#">↗</a>	:FC_Red_Bull_Salzburg <a href="#">↗</a>	:Austria <a href="#">↗</a>	31000
:Alain_Gouaméné <a href="#">↗</a>	:Ivory_Coast <a href="#">↗</a>	:Raja_Casablanca <a href="#">↗</a>	:Morocco <a href="#">↗</a>	67000
:Allan_McGregor <a href="#">↗</a>	:United_Kingdom <a href="#">↗</a>	:Beşiktaş_J.K. <a href="#">↗</a>	:Turkey <a href="#">↗</a>	41903
:Anthony_Scribe <a href="#">↗</a>	:France <a href="#">↗</a>	:FC_Dinamo_Tbilisi <a href="#">↗</a>	:Georgia_(country) <a href="#">↗</a>	54549
:Brahim_Zaari <a href="#">↗</a>	:Netherlands <a href="#">↗</a>	:Raja_Casablanca <a href="#">↗</a>	:Morocco <a href="#">↗</a>	67000
:Bréiner_Castillo <a href="#">↗</a>	:Colombia <a href="#">↗</a>	:Deportivo_Táchira <a href="#">↗</a>	:Venezuela <a href="#">↗</a>	38755
:Carlos_Luis_Morales <a href="#">↗</a>	:Ecuador <a href="#">↗</a>	:Club_Atlético_Independiente <a href="#">↗</a>	:Argentina <a href="#">↗</a>	48069
:Carlos_Navarro_Montoya <a href="#">↗</a>	:Colombia <a href="#">↗</a>	:Club_Atlético_Independiente <a href="#">↗</a>	:Argentina <a href="#">↗</a>	48069
:Cristián_Muñoz <a href="#">↗</a>	:Argentina <a href="#">↗</a>	:Colo-Colo <a href="#">↗</a>	:Chile <a href="#">↗</a>	47000
:Daniel_Ferreira <a href="#">↗</a>	:Argentina <a href="#">↗</a>	:FBC_Melgar <a href="#">↗</a>	:Peru <a href="#">↗</a>	60000
:David_Bílik <a href="#">↗</a>	:Czech_Republic <a href="#">↗</a>	:Karşıyaka_S.K. <a href="#">↗</a>	:Turkey <a href="#">↗</a>	51295
:David_Loria <a href="#">↗</a>	:Kazakhstan <a href="#">↗</a>	:Karşıyaka_S.K. <a href="#">↗</a>	:Turkey <a href="#">↗</a>	51295
:Denys_Boykó <a href="#">↗</a>	:Ukraine <a href="#">↗</a>	:Beşiktaş_J.K. <a href="#">↗</a>	:Turkey <a href="#">↗</a>	41903
:Eddie_Gustafsson <a href="#">↗</a>	:United_States <a href="#">↗</a>	:FC_Red_Bull_Salzburg <a href="#">↗</a>	:Austria <a href="#">↗</a>	31000
:Emilian_Dolha <a href="#">↗</a>	:Romania <a href="#">↗</a>	:Lech_Poznań <a href="#">↗</a>	:Poland <a href="#">↗</a>	43269
:Eusebio_Acasuzo <a href="#">↗</a>	:Peru <a href="#">↗</a>	:Club_Bolívar <a href="#">↗</a>	:Bolivia <a href="#">↗</a>	42000
:Faryd_Mondragón <a href="#">↗</a>	:Colombia <a href="#">↗</a>	:Real_Zaragoza <a href="#">↗</a>	:Spain <a href="#">↗</a>	34596
:Faryd_Mondragón <a href="#">↗</a>	:Colombia <a href="#">↗</a>	:Club_Atlético_Independiente <a href="#">↗</a>	:Argentina <a href="#">↗</a>	48069
:Federico_Vilar <a href="#">↗</a>	:Argentina <a href="#">↗</a>	:Club_Atlas <a href="#">↗</a>	:Mexico <a href="#">↗</a>	54500
:Fernando_Martinuzzi <a href="#">↗</a>	:Argentina <a href="#">↗</a>	:Real_Garcilaso <a href="#">↗</a>	:Peru <a href="#">↗</a>	45000
:Fábio André da Silva <a href="#">↗</a>	:Portugal <a href="#">↗</a>	:Servette_FC <a href="#">↗</a>	:Switzerland <a href="#">↗</a>	30084



# Wikidata

- Hướng tới tạo một cơ sở tri thức có thể đọc và cập nhật bởi người và máy.
- Giữ vai trò trung tâm lưu trữ dữ liệu có cấu trúc cho các bộ Bách khoa toàn thư khác như Wikipedia, Wikivoyage, Wiktionary, Wikisource, v.v..



The screenshot shows the Wikidata Main Page. At the top is the Wikidata logo and a navigation bar with links: Main Page, Discussion, Read, View source, View history, and a search box labeled 'Search Wikidata'. Below the navigation bar is a large central banner with the text 'Welcome to Wikidata' and 'the free knowledge base with 91,786,616 data items that anyone can edit.' The banner also includes links for 'Introduction', 'Project Chat', 'Community Portal', and 'Help'. To the left of the banner 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, Random Lexeme, Tools, What links here, Related changes, Special pages, Permanent link, Page information, Wikidata item, In other projects, Wikimedia Commons, MediaWiki, Meta-Wiki, Wikispecies, Wikibooks, Wikimania, Wikinews, Wikipedia, Wikiquote, Wikisource, Wikiversity, Wikivoyage, Wiktionary, In Wikipedia, and العربية. To the right of the banner is a section titled 'Welcome!' with text about Wikidata being a free and open knowledge base, acting as central storage for structured data, and providing support to many other sites. Below this is a section titled 'Learn about data' with text about developing and improving data literacy, and a section titled 'Get involved' with text about visiting the community portal and learning about Wikidata. The page also features a network diagram in the background with nodes and connecting lines.

Wikidata

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  
Wikidata item

In other projects  
Wikimedia Commons  
MediaWiki  
Meta-Wiki  
Wikispecies  
Wikibooks  
Wikimania  
Wikinews  
Wikipedia  
Wikiquote  
Wikisource  
Wikiversity  
Wikivoyage  
Wiktionary

In Wikipedia  
العربية

Main Page Discussion Read View source View history Search Wikidata

Welcome to Wikidata

the free knowledge base with 91,786,616 data items that anyone can edit.

Introduction • Project Chat • Community Portal • Help

Want to help translate? Translate the missing messages.

Welcome!

Wikidata is a free and open knowledge base that can be read and edited by both humans and machines.

Wikidata acts as central storage for the **structured data** of its Wikimedia sister projects including Wikipedia, Wikivoyage, Wiktionary, Wikisource, and others.

Wikidata also provides support to many other sites and services beyond just Wikimedia projects! The content of Wikidata is available under a [free license](#), exported using [standard formats](#), and can be [interlinked to other open data sets](#) on the linked data web.

Learn about data

New to the wonderful world of data? [Develop and improve your data literacy](#) through content designed to get you up to speed and feeling comfortable with the fundamentals in no time.

Item: *Earth* (Q2)

Property: *highest point* (P610)

custom value:

Get involved

For a complete starters' guide, visit the [community portal](#).

Learn about Wikidata

# Nội dung

1.1. Khái niệm Web ngữ nghĩa

1.2. Web ngữ nghĩa hiện nay

1.3. Đồ thị tri thức

1.4. Tổng quan công nghệ Web ngữ nghĩa

1.5. Ontology và một số khái niệm quan trọng



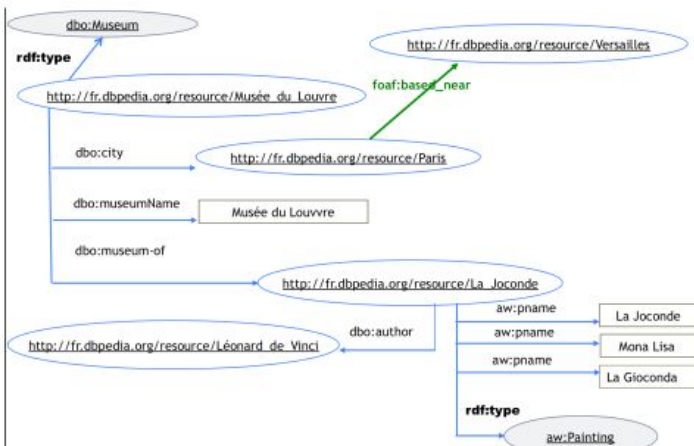
# Khái niệm đồ thị tri thức

- Đồ thị tri thức là một mạng lưới của các thực thể, thuộc tính, kiểu ngữ nghĩa và mối quan hệ giữa các thực thể. -> Đồ thị RDF
  - "Knowledge graphs are large networks of entities, their semantic types, properties, and relationships between entities" [Journal of Web Semantics]
- Đồ thị tri thức có thể được biểu diễn như đồ thị RDF, bao gồm một tập các bộ ba RDF.
  - [Farber et al.]

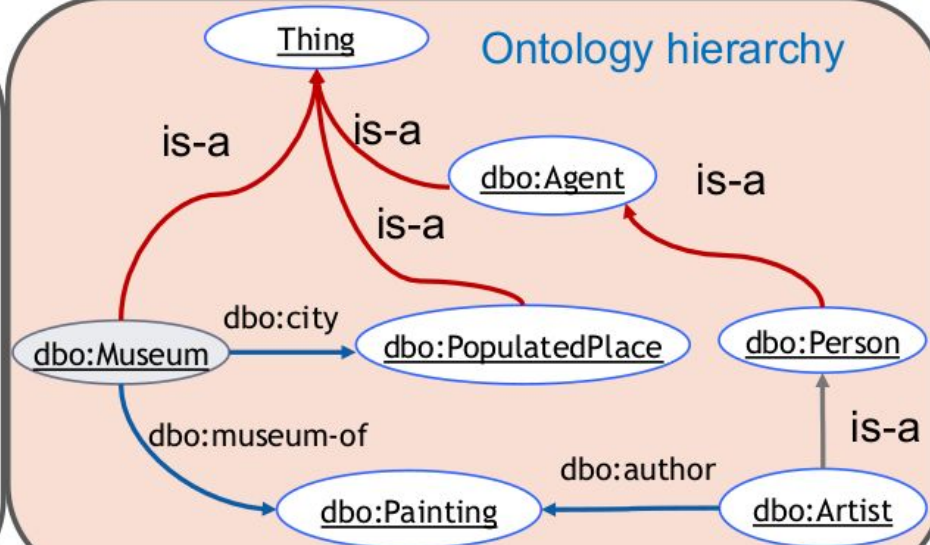
*Đồ thị tri thức ~ Một cơ sở tri thức dạng đồ thị.*

# Đồ thị tri thức (KG)

## RDF Graphs



## Ontology hierarchy



## Querying (SPARQL)

```
PREFIX dbo: <http://dbpedia.org/ontology#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
SELECT ?m ?p
WHERE { ?m rdf:type dbo:Museum . ?m dbo:museum-of ?p }
```

## Reasoners: (Pellet, Fact++, Hermit, etc.)

- KG saturation: infer whatever can be inferred from the KG.
- KG consistency checking: no contradictions
- KG repairing
- ...

## Ontology axioms and rules

```
owl:equivalentClass(dbo:Municipality, dbo:Place)
owl:equivalentClass(dbo:Place, dbo:Wikidata:Q532)
owl:equivalentClass(dbo:Village, dbo:PopulatedPlace)
owl:equivalentClass(dbo:PopulatedPlace, dbo:Municipality)
owl:disjointClass(dbo:PopulatedPlace, dbo:Artist)
owl:disjointClass(dbo:PopulatedPlace, dbo:Painting)
owl:FunctionalProperty(dbo:city)
owl:InverseFunctionalProperty(dbo:museum-of)
```

```
dbo:birthPlace(X, Y) => dbo:citizenOf(X, Y)
dbo:parentOf(X, Y) => dbo:child(Y, X)
```

# Mở rộng & Khai phá

- Mở rộng: các đồ thị tri thức luôn cần được cập nhật
  - Liên kết dữ liệu (phát hiện thực thể, các trùng lặp, ...)
  - Dự đoán liên kết: Thêm mới liên kết
  - Khớp nối ontology: Kết nối các đồ thị
  - Dự đoán/suy diễn các dữ kiện bị bỏ qua
- Kiểm tra: Đồ thị tri thức có thể chứa lỗi
  - Kiểm tra liên kết
  - Xử lý lỗi và các vấn đề đa nghĩa
- Khai phá: Liệu có thể phát hiện tri thức mới?
  - Phát hiện tri thức
  - Suy diễn và lập kế hoạch tự động

# Ví dụ truy vấn trên Wikidata

## *Country populations together with total city populations*

Map

218 results in 4388 ms

</> Code

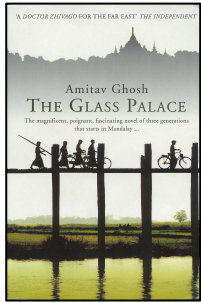
Download

Link

Search

country	countryLabel	population	totalCityPopulation	ratio
<a href="#">Q148</a>	People's Republic of China	1409517397	115388039	12.21545499182978575448
<a href="#">Q668</a>	India	1326093247	147242392	9.00619195998934871963
<a href="#">Q30</a>	United States of America	328239523	10936541	30.0131022230886346972
<a href="#">Q15180</a>	Soviet Union	293047571	28500	10282.37091228070175438596
<a href="#">Q252</a>	Indonesia	270625568	10961704	24.6882754724995310948
<a href="#">Q843</a>	Pakistan	216565318	43101219	5.02457524461199113649
<a href="#">Q155</a>	Brazil	210147125	90381055	2.3251236113585972193
<a href="#">Q1033</a>	Nigeria	190886311	44761481	4.26452178827595092307
<a href="#">Q34266</a>	Russian Empire	181537800	69825	2599.89688506981740064447
<a href="#">Q34266</a>	Russian Empire	178378800	69825	2554.65520945220193340494
<a href="#">Q902</a>	Bangladesh	164669751	22627219	7.27750728005947173623
<a href="#">Q159</a>	Russia	146804372	1135276	129.31161409207981142912
<a href="#">Q96</a>	Mexico	130526945	34239370.789	3.81218877544134299716
<a href="#">Q17</a>	Japan	126434565	23318712	5.42202180806555696558
<a href="#">Q34266</a>	Russian Empire	125640021	69825	1799.35583243823845327605
<a href="#">Q7318</a>	Nazi Germany	109518183	7741	14147.80816431985531585067
<a href="#">Q115</a>	Ethiopia	104957438	2612647	40.17283544236936715905
<a href="#">Q79</a>	Egypt	94798827	28436468	3.33370610583564738068
<a href="#">Q881</a>	Vietnam	94660000	13370442	7.07979586613516591299
<a href="#">Q974</a>	Democratic Republic of the Congo	86790567	17871025	4.85649631176723215372

# Tích hợp dữ liệu: Ví dụ dữ liệu minh họa



## Book

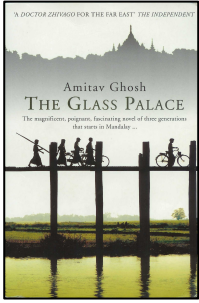
ID	Author	Title	Publisher	Year
ISBN 0-00-6511409-X	id_xyz	The Glass Palace	id_qpr	2000

## Author

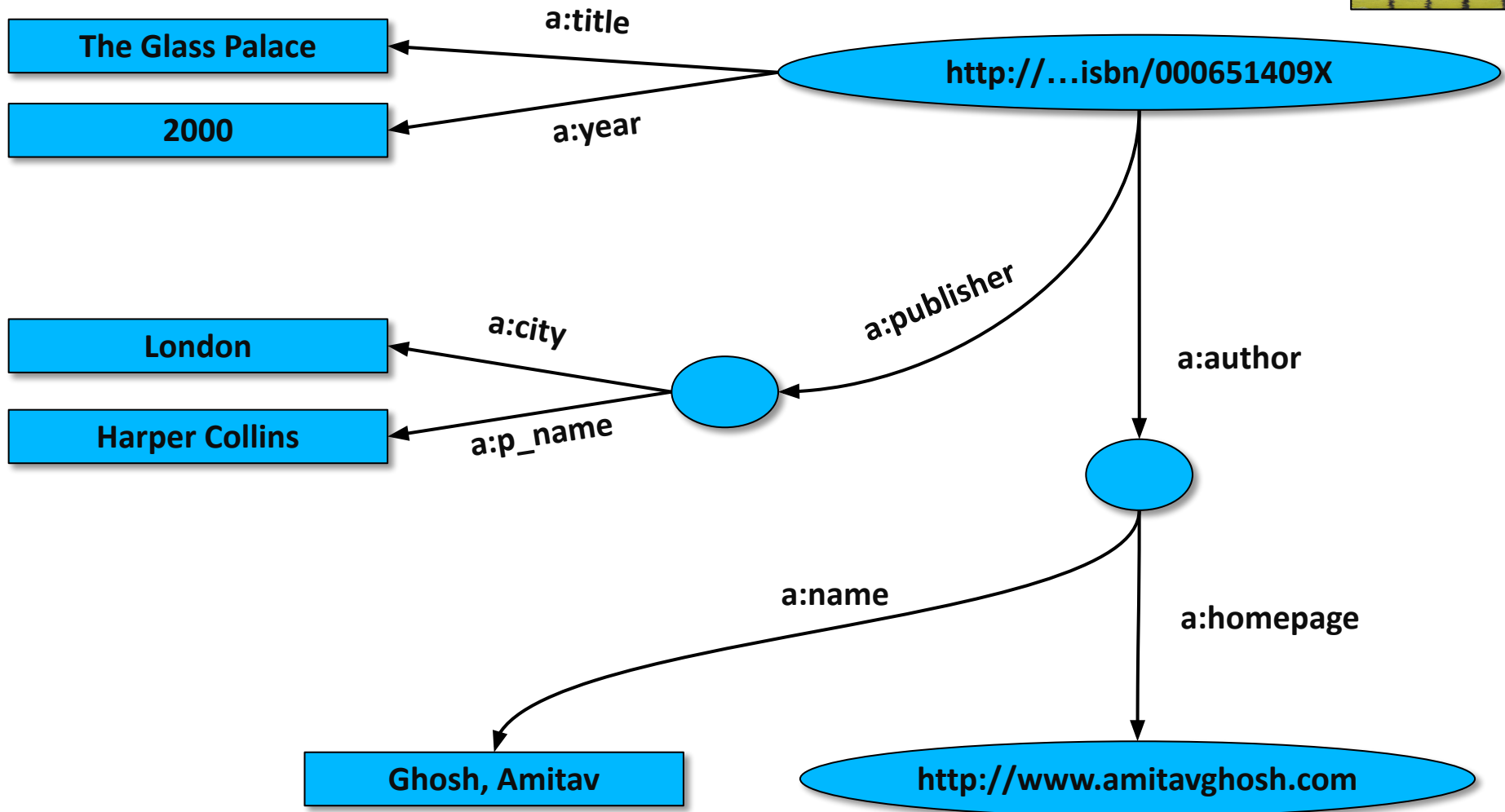
ID	Name	Homepage
id_xyz	Ghosh, Amitav	<a href="http://www.amitavghosh.com">http://www.amitavghosh.com</a>

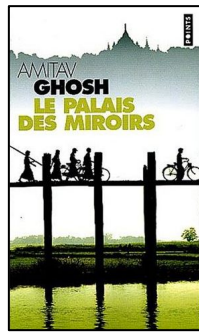
## Publisher

ID	Publisher's name	City
id_qpr	Harper Collins	London

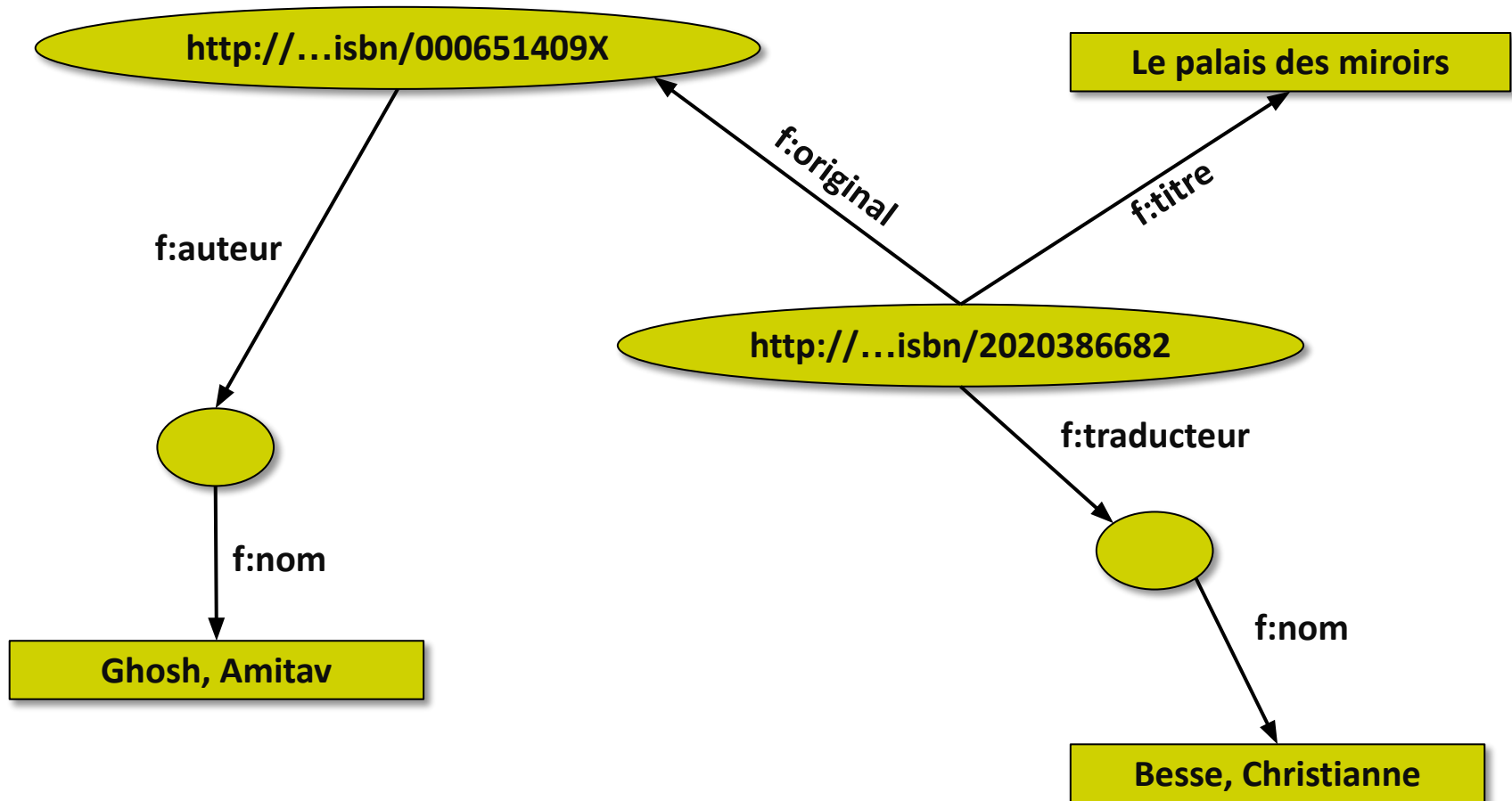


# Tích hợp dữ liệu: Biểu diễn đồ thị

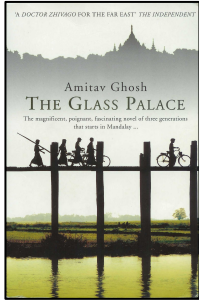
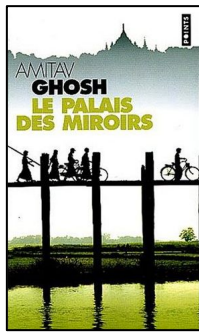
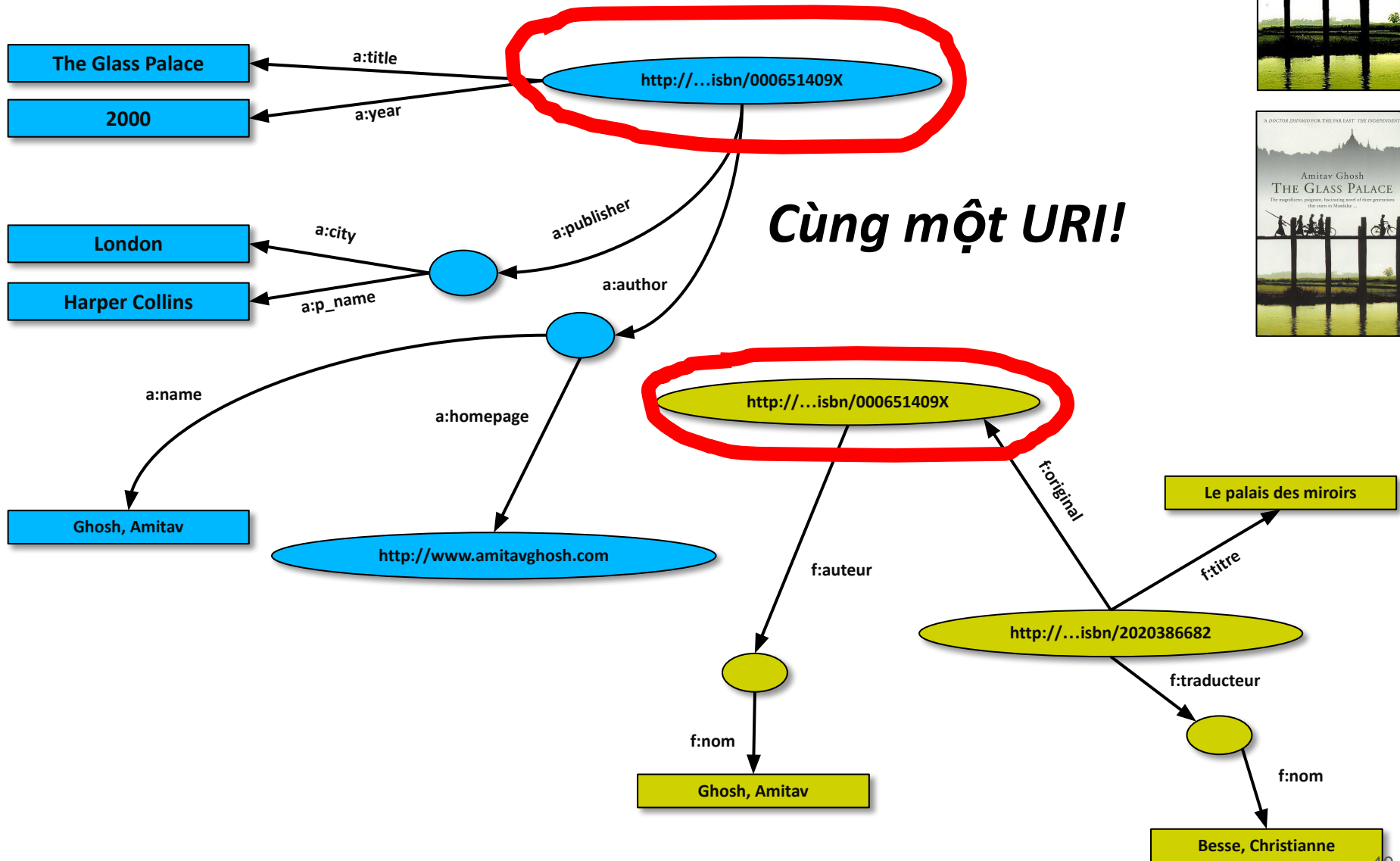




# Tích hợp dữ liệu: Phiên bản tiếng Pháp

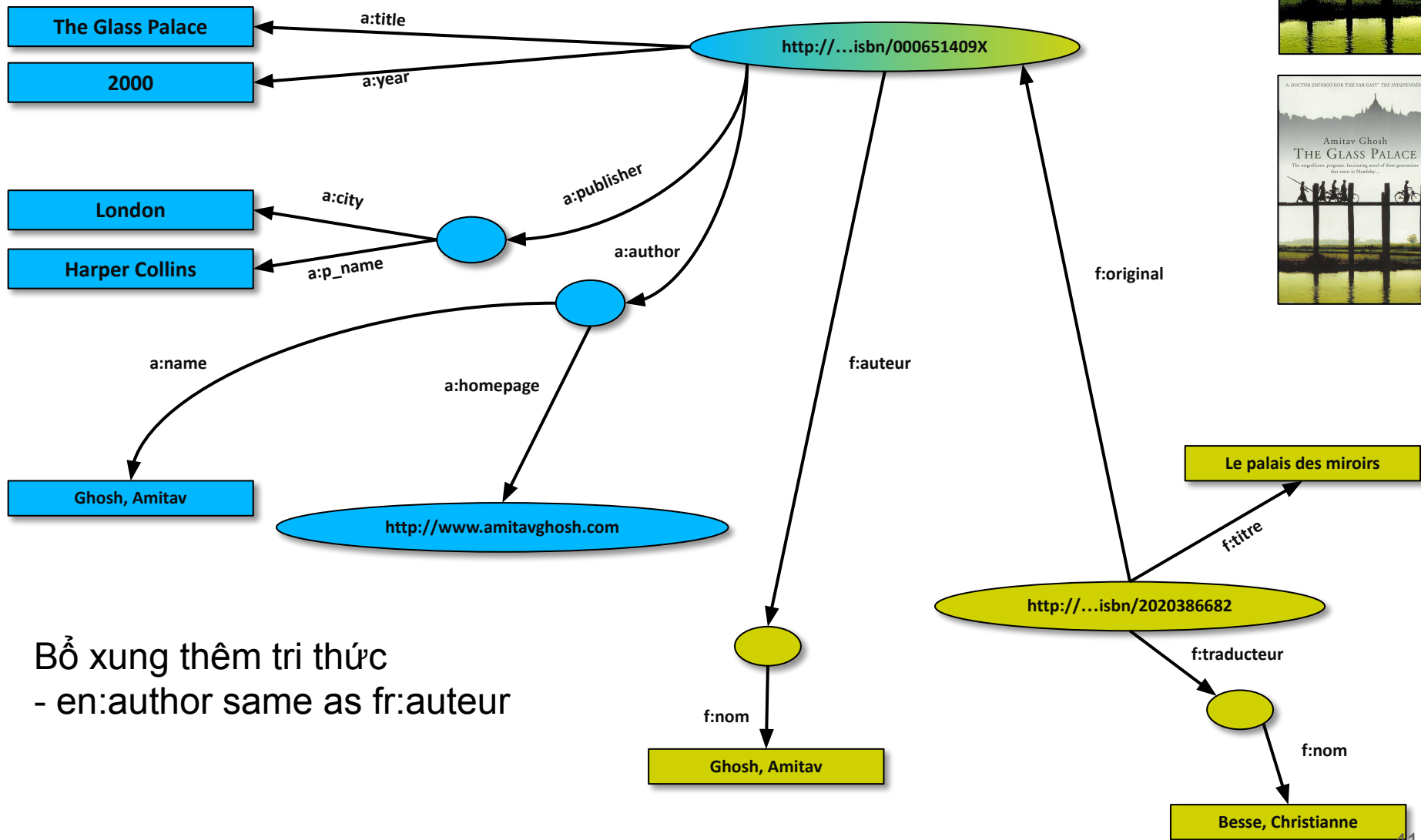


# Tích hợp hai đồ thị



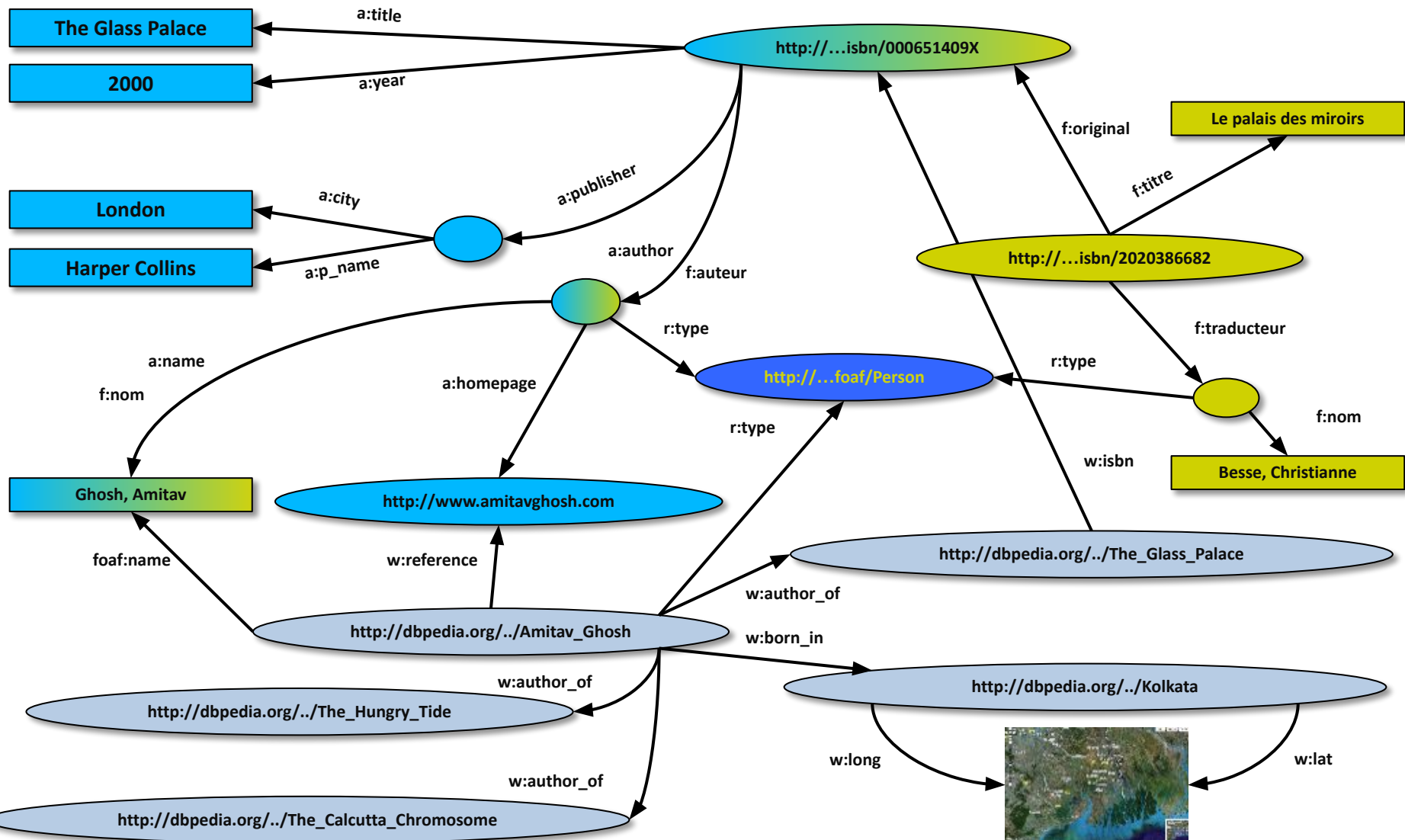


# Kết quả tích hợp



Bổ xung thêm tri thức  
- en:author same as fr:auteur

# Tích hợp với Wikidata



*Làm sao để xác định các lớp, thuộc tính, thực thể tương đương?*

# Web ngữ nghĩa & đồ thị tri thức

- Web ngữ nghĩa cung cấp nền tảng công nghệ để hiện thực hóa biểu diễn và tích hợp cơ sở tri thức dạng đồ thị
  - Các đồ thị tri thức quy mô rất lớn, ở quy mô Web
  - Các CSTT giữ vai trò thiết yếu để liên kết và tích hợp, ví dụ, Wikidata, DBpedia, v.v..
- Các ontologies, luật, v.v.. bổ xung thêm tri thức vào các tập dữ liệu
- Cho phép thực hiện nhiều truy vấn phức tạp và hữu ích.

# Nội dung

1.1. Khái niệm Web ngữ nghĩa

1.2. Web ngữ nghĩa hiện nay

1.3. Đồ thị tri thức

1.4. Tổng quan công nghệ Web ngữ nghĩa

1.5. Ontology và một số khái niệm quan trọng

# Các công nghệ Web ngữ nghĩa

- Các quy chuẩn W3C
  - RDF, RDFS, RDFa, OWL, SPARQL, RIF, R2R, v.v..
- Các công cụ và hệ thống thông dụng - Thương mại, miễn phí và mã nguồn mở
  - Trình soạn thảo ontology, lưu trữ bộ ba, mô tơ suy diễn, v.v.
- Các ontologies và các bộ dữ liệu thông dụng
  - Foaf, DBpedia, SKOS, PROV, v.v.
- Các hệ thống hạ tầng
  - Tìm kiếm, ontology siêu dữ liệu, dịch vụ liên kết
- Các quy chuẩn bên ngoài W3C: Schema.org, Freebase, ...

# Các ngôn ngữ biểu diễn tri thức

- Biểu diễn tri thức và suy diễn (KR&R) luôn là một phần quan trọng của TTNT và nhiều lĩnh vực khác
- Nhiều cách tiếp cận đã được xây dựng, triển khai và phát triển từ những năm 1960
- Hầu hết các giải pháp là các giải pháp riêng, chỉ được sử dụng bởi người phát triển
- Bắt đầu từ những năm 1990, ngôn ngữ biểu diễn tri thức dùng chung để hỗ trợ tái sử dụng tri thức và các hệ cơ sở tri thức phân tán được quan tâm phát triển
- Các ngôn ngữ Web ngữ nghĩa (ví dụ, OWL) hiện là sự khái quát hóa của những ý tưởng này
  - Không có ngôn ngữ biểu diễn tri thức nào khác được sử dụng phổ biến

# Các câu hỏi

- Cơ sở dữ liệu (CSDL) vs. cơ sở tri thức (CSTT)
  - CSDL thường có lược đồ dữ liệu (tri thức đơn giản) và rất nhiều dữ liệu
  - CSTT có thể có lược đồ phức tạp (các ontologies) và có thể có nhiều dữ liệu hoặc không có nhiều dữ liệu
- CSTT hỗ trợ suy diễn, ví dụ,

$\text{parent}(?x, ?y) \Rightarrow \text{person}(?x), \text{person}(?y), \text{child}(?y, ?x), \text{older}(?x, ?y), ?x \neq ?y$   
 $\text{Parent}(\text{john}, \text{mary}) \Rightarrow \text{person}(\text{john}), \text{child}(\text{mary}, \text{john}), \dots$

# Các câu hỏi<sub>(2)</sub>

Hệ quả của việc sử dụng các cấu trúc khác nhau để biểu diễn dữ liệu hoặc tri thức là gì?

- Ngôn ngữ tự nhiên
- Mã nguồn lập trình
- Quan hệ vs. Đồ thị vs. Đối tượng
- Lô-gic vs. Luật vs. Tiến trình
- Mạng nơ-ron
- Tenxơ



# Các câu hỏi<sub>(3)</sub>

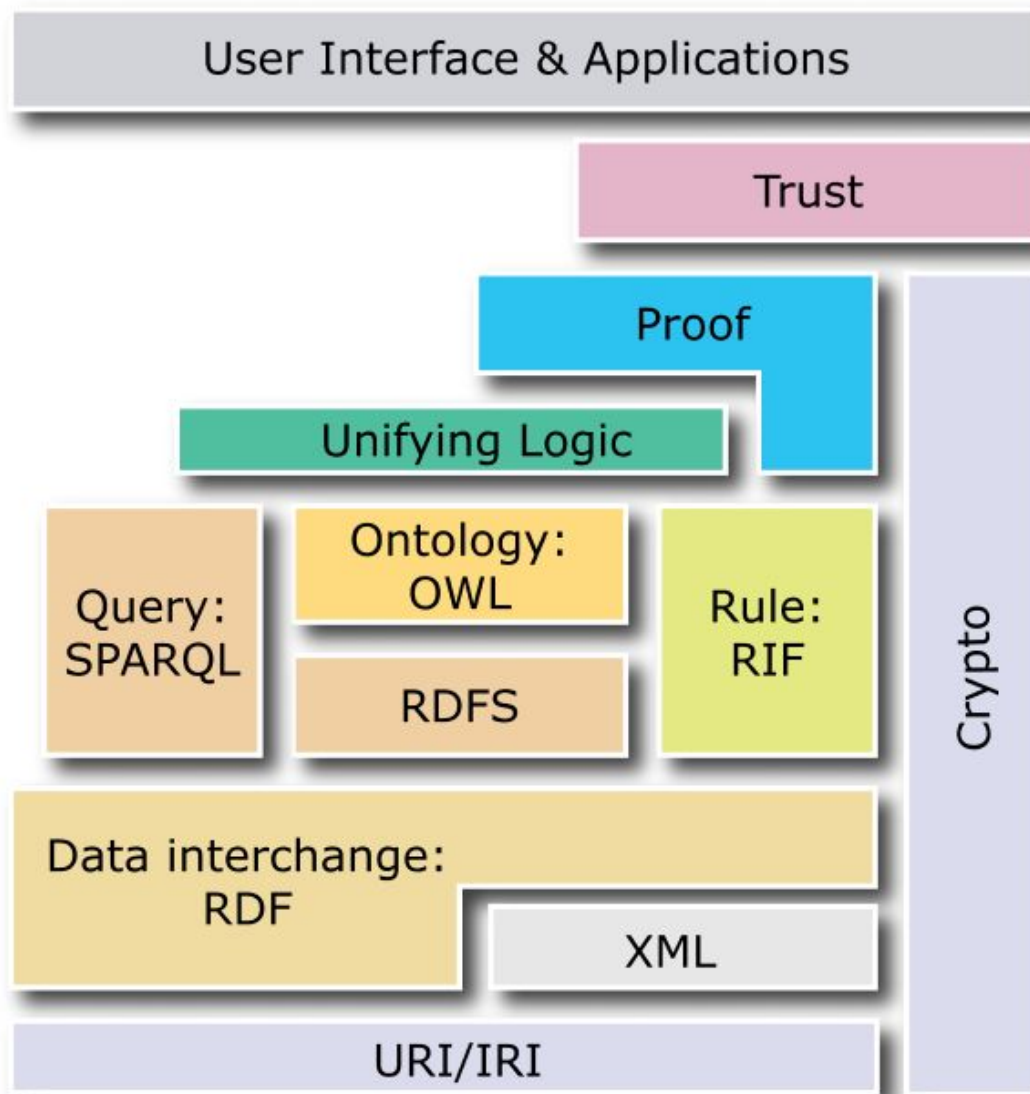
Các mô hình ngữ nghĩa:

- Lô-gic tính toán là một lựa chọn phổ biến
  - $\text{man(NVA)}, \forall x \text{ man}(x) \Rightarrow \text{mortal}(x)$
  - Lô-gic có một số giới hạn: Các dữ kiện, các quan hệ, và "luật" chỉ có thể là đúng (true) hoặc sai (false)
- Có thể cần biểu diễn và suy diễn với xác suất hoặc các dữ kiện mờ hoặc tri thức
- Có thể cần xử lý các dữ kiện hoặc tri thức biến động

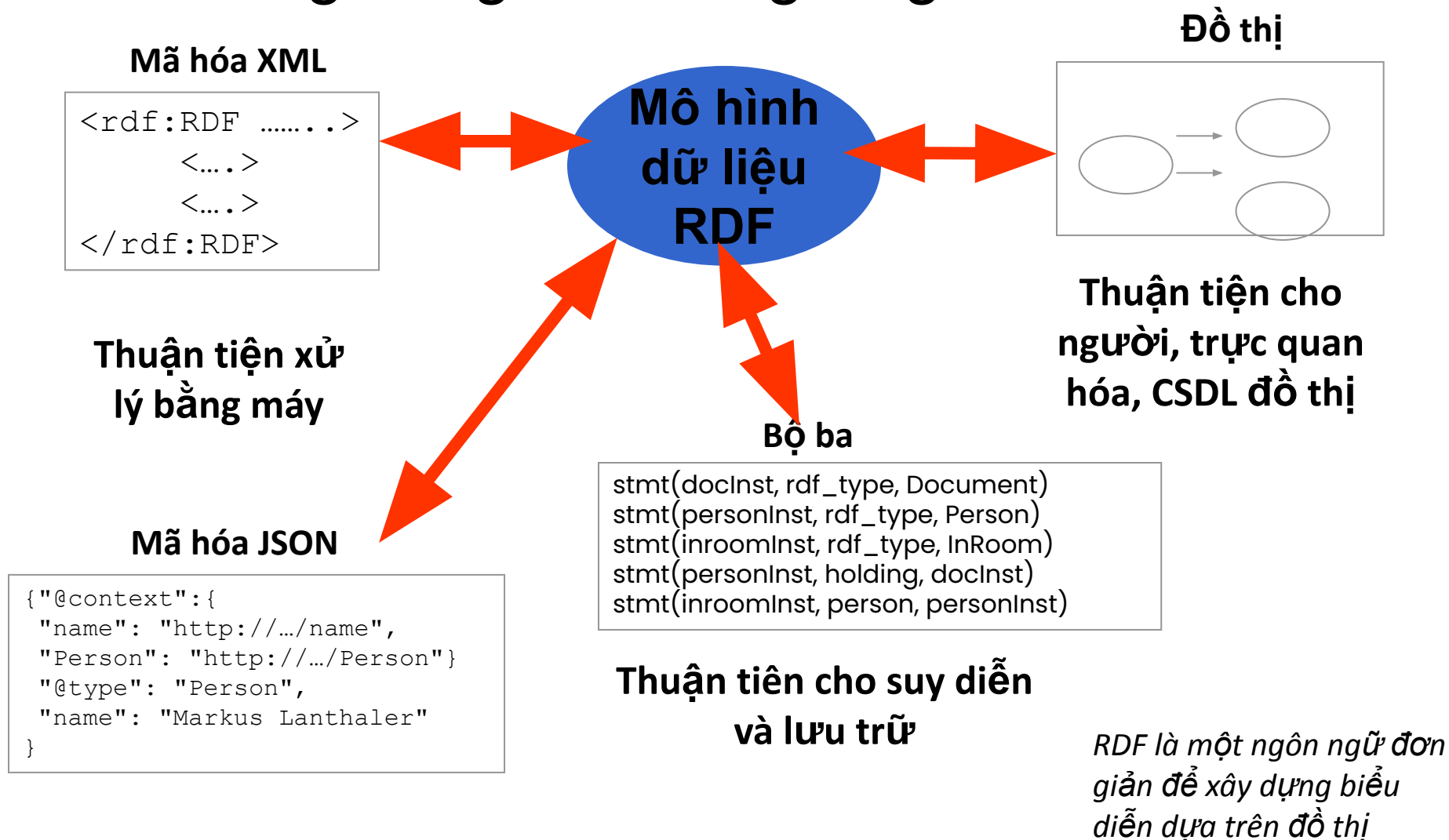
# Các công nghệ Web ngữ nghĩa

- Các công nghệ cơ bản sử dụng Lô-gic tính toán làm nền tảng ngữ nghĩa
  - Đơn giản, hành vi xác định, giải thuật suy diễn hiệu quả
  - - Không có cơ chế xác suất
- Tri thức được biểu diễn như đồ thị
  - Đơn giản, công cụ hỗ trợ tốt
  - - Có thể quá đơn giản

# Ngăn xếp công nghệ Web ngữ nghĩa W3C



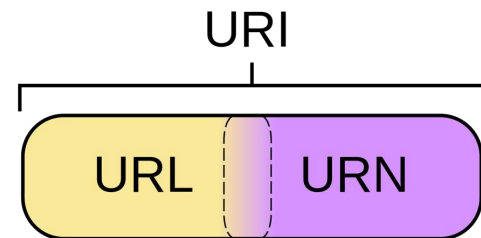
# RDF là ngôn ngữ Web ngữ nghĩa đầu tiên



# Mô hình dữ liệu RDF

- Một tài liệu RDF là một tập không thứ tự các câu, mỗi câu có cú pháp đơn giản bao gồm chủ ngữ (subject), khẳng định (predicate) và object (đối tượng)
  - Câu RDF = Bộ ba
- Mỗi bộ ba có thể được biểu diễn như một cạnh nối hai đỉnh có nhãn trong đồ thị
- Các câu mô tả các thuộc tính của các tài nguyên Web
- Tài nguyên là đối tượng có thể được xác định bằng một URI
  - - Văn bản, hình ảnh, đoạn văn trên Web
  - ví dụ, [https://vi.wikipedia.org/wiki/Alan\\_Turing](https://vi.wikipedia.org/wiki/Alan_Turing)
  - hoặc một cuốn sách: isbn://5031-4444-3333
- Các thuộc tính đồng thời cũng là các tài nguyên (URIs)

# URIs giữ vai trò nền tảng



- URI = (Uniform Resource Identifier)
  - Tập hợp các tên/địa chỉ là những chuỗi ký tự ngắn tham chiếu các tài nguyên
  - URLs (Uniform Resource Locators) là tập con của URIs, được sử dụng cho các tài nguyên có thể truy cập được trên Web
- URIs nhìn giống URLs, thường có chỉ số đoạn để chỉ tới một phần của tài liệu
  - `http://foo.com/bar/mumble.html#pitch`

# URIs giữ vai trò nền tảng<sub>(2)</sub>

- URIs có ý nghĩa xác định, không nhập nhằng về nghĩa như các từ ngôn ngữ tự nhiên -- Web cung cấp một không gian tên toàn cầu
- Chúng ta có thể sử dụng URI để tham chiếu bất cứ thứ gì, ví dụ, một khái niệm, thực thể, sự kiện hoặc quan hệ
- Chúng ta thường giả định hai thứ có URIs giống nhau (có thể xuất hiện ở hai ngữ cảnh khác nhau) là một

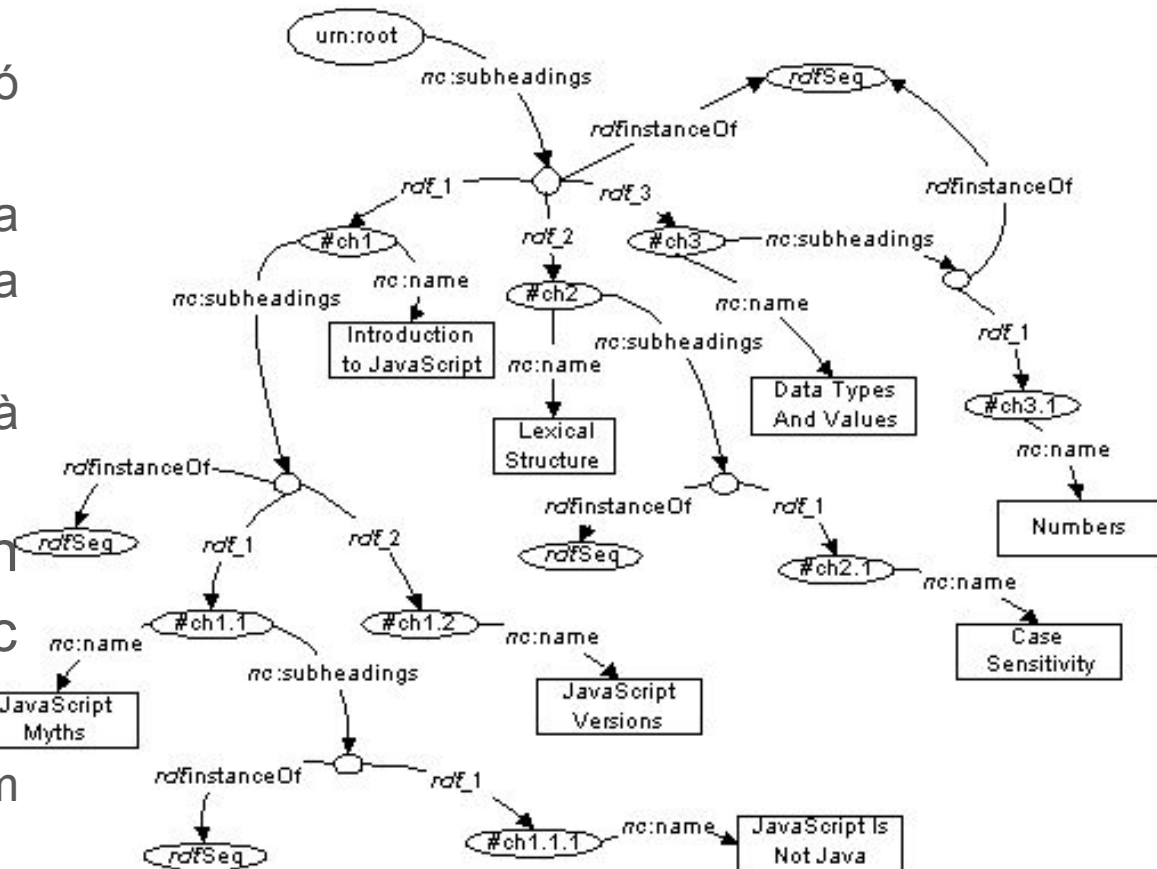
# Nghĩa của URI là gì?

- Trong một số trường hợp URIs chỉ định một tài nguyên Web
- Đôi khi là những thực thể trong thế giới thực
  - Ví dụ, <https://www.hust.edu.vn/> là 1 Đại học hàng đầu của Việt Nam
- Các URIs tốt thường không thay đổi
  - <http://www.w3.org/Provider/Style/URI>

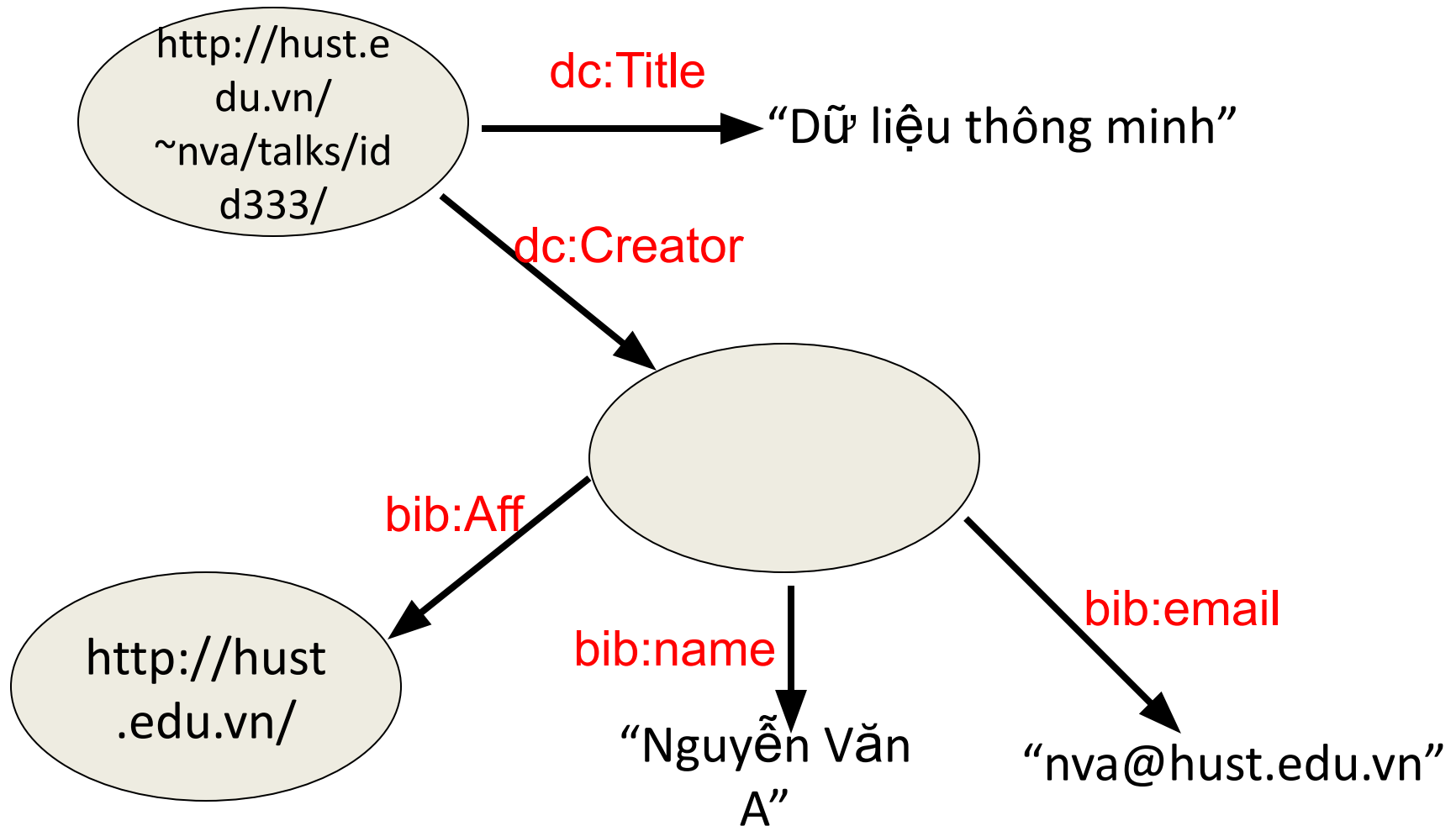


# Đồ thị RDF

- Biểu diễn chủ ngữ và đối tượng như các nút, khẳng định như cạnh có hướng
  - Thu được đồ thị có hướng có gán nhãn
  - Chủ ngữ của một bộ ba có thể là đối tượng của một bộ ba khác
  - Đối tượng có thể là chuỗi ký tự
- Đồ thị đơn giản hơn CSDL quan hệ và các đối tượng
  - Hàm chứa cả ưu điểm và nhược điểm



# Ví dụ Đồ thị RDF đơn giản



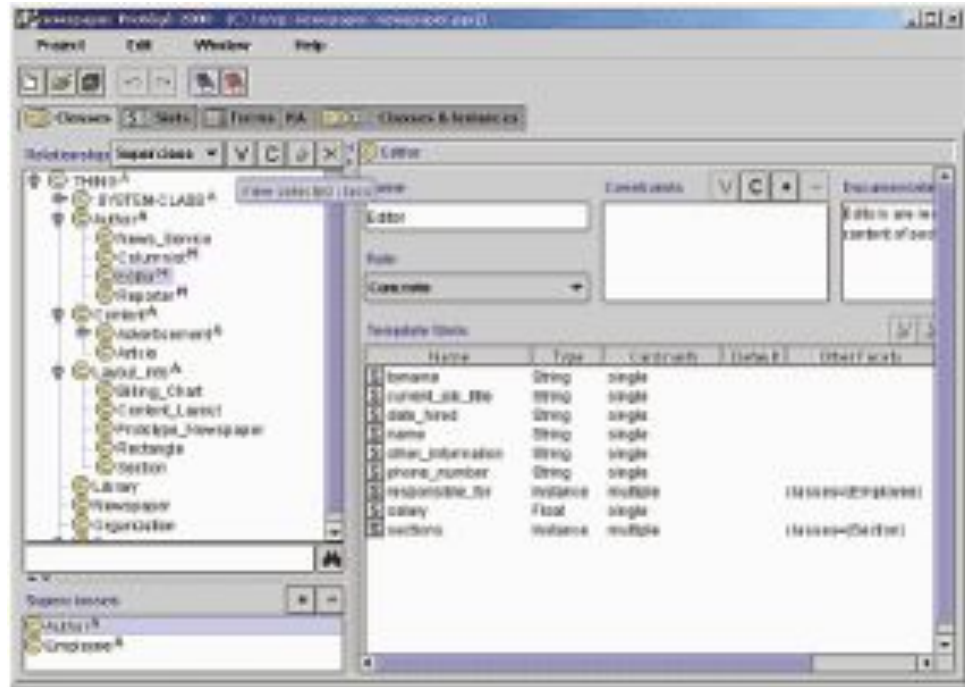
*Các dữ liệu chỉ có tính minh họa*

# Lưu trữ

- Đồ thị là một mô hình trừu tượng, chúng ta cần lưu trữ nó như văn bản vì nhiều lý do, ví dụ, hiển thị, soạn thảo, trao đổi, v.v...
- Có nhiều quy chuẩn lưu trữ cho dữ liệu RDF
- Các quy chuẩn quan trọng nhất: XML, Turtle, N-Triples, JSON-LD
- Hầu hết các công cụ Web ngữ nghĩa đều có thể đọc hoặc viết bất kỳ định dạng nào trong số các định dạng được liệt kê

# Lược đồ RDF (RDFS)

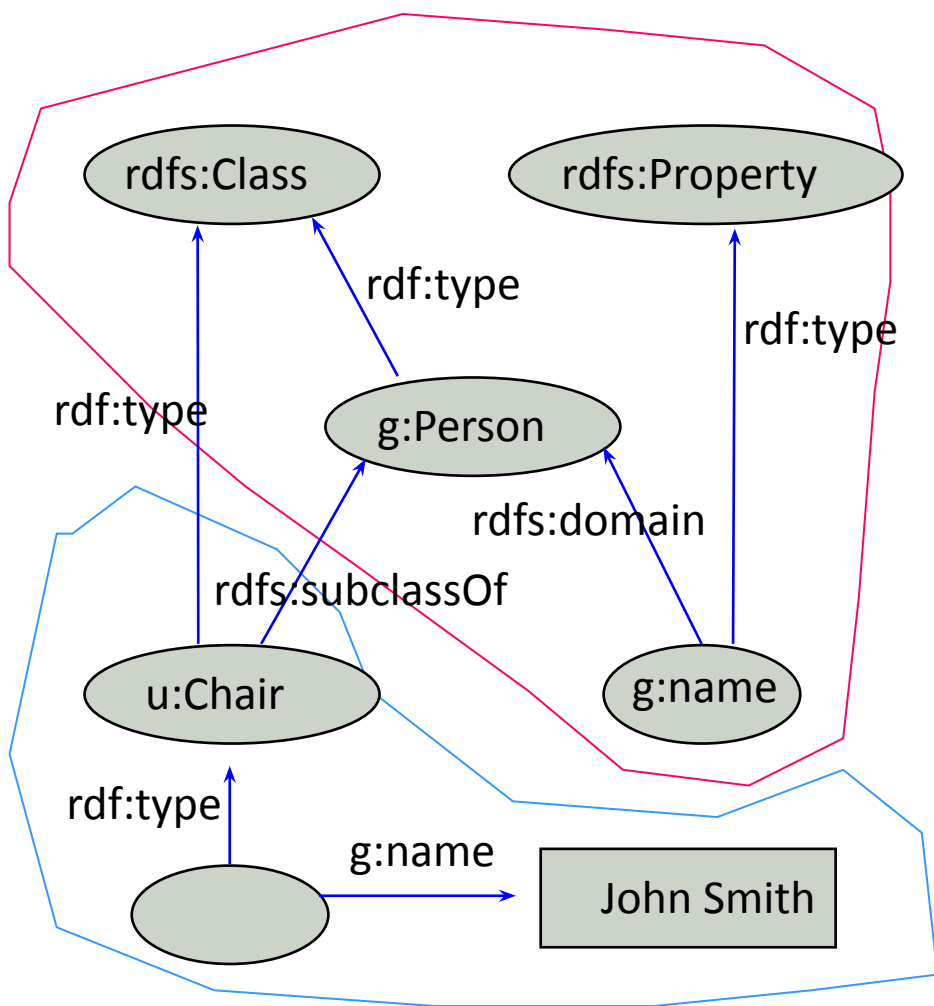
- Lược đồ RDF bổ xung các cây phân cấp cho các lớp & các thuộc tính
  - Lớp con (subClass) và thuộc tính con (subProperty)
- Thêm một số siêu dữ liệu
  - Ràng buộc miền và khoảng trên các thuộc tính
- Nhiều công cụ KG phổ biến có thể nhập và xuất tài liệu RDFS



## Trình soạn thảo cơ sở tri thức Protégé của Stanford

- Java, mã nguồn mở
- Khả mở, nhiều plug-ins
- Cung cấp các tính năng suy diễn

# RDF và lược đồ RDF



## Thông tin mức Lược đồ

@prefix rdf:

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

@prefix rdfs:

<http://www.w3.org/2000/01/rdf-schema#> .

@prefix g: <http://schema.org/gen> .

@prefix u: <http://schema.org/univ> .

```
g:name rdf:type rdfs:Property;  
rdfs:domain g:Person .
```

```
u:Chair rdfs:subClassOf g:Person .
```

## Thông tin mức đối tượng

```
_john rdf:type u:Chair;  
g:name "John Smith" .
```

*!Có thể mô tả bất cứ điều gì, kể cả nội dung sai.*

# RDFS hỗ trợ các suy diễn đơn giản

- Một ontology RDFs và một vài câu RDF có thể ngầm định thêm một số câu RDF khác
  - Được coi như một phần của mô hình dữ liệu
- Không đúng với dữ liệu XML

```
@prefix rdfs: <http://www...>.
@prefix : <...genesis.n3>.
:parent rdfs:domain :Person;
        rdfs:range :Person.
:mother
  rdfs:subProperty parent;
  rdfs:domain :Woman.
:eve :mother :cain.
```



```
:parent a rdf:Property.
:Person a rdf:Class.
:Woman rdfs:subClassOf Person.
:mother a rdf:Property.
:eve a :Person;
      a :Woman;
      :parent :cain.
:cain a :Person.
```

# Liệu RDF(S) có tốt hơn XML?

- Phụ thuộc vào từng tình huống ứng dụng cụ thể
- Mô hình XML
  - cây phân cấp chặt chẽ
  - các ứng dụng có thể dựa trên vị trí trong cây
  - cú pháp và cấu trúc tương đối đơn giản
  - không dễ kết hợp thành cây
- Mô hình RDF
  - Một tập hợp linh động của các liên kết
  - Các ứng dụng có thể thực hiện các tìm kiếm tương tự CSDL
  - không dễ để khôi phục cấu trúc
  - dễ dàng hợp nhất các liên hệ thành một tập lớn
  - tốt cho việc tích hợp thông tin từ nhiều nguồn

# Một số hạn chế của RDFS

- Không có khả năng mô tả chi tiết các tài nguyên
  - Không thể giới hạn ràng buộc domain và range trong phạm vi hẹp hơn: Không thể mô tả range của hasChild là cat khi áp dụng cho cat và là dog khi áp dụng cho dog
  - Không có ràng buộc tồn tại/cơ sở: Không thể mô tả một cái xe đạp có hai bánh
  - Không có các thuộc tính bắc cầu, nghịch đảo hoặc đối xứng: Không thể nói isPartOf là một thuộc tính bắc cầu, hasPart là nghịch đảo của isPartOf hoặc touches là đối xứng
- Sẽ cần đến các từ khóa RDF mới để thực hiện những mô tả này



# Ngôn ngữ OWL của W3C

- Dự án DARPA, DAML+OIL, tiền thân của OWL
- OWL được công bố như một quy chuẩn W3C, 2/10/04
- Ba mức OWL đã được xác định theo thứ tự giảm dần độ phức tạp và đồng thời là khả năng diễn đạt
  - OWL Full - Đầy đủ nhất
  - OWL DL - (Lô-gic mô tả) giới thiệu các giới hạn
  - OWL Lite - Ngôn ngữ mức bắt đầu, hướng tới tính đơn giản và dễ triển khai
- Owl 2 trở thành quy chuẩn W3C trong năm 2009, được cập nhật năm 2012

# OWL $\Leftrightarrow$ RDF

- Một tài liệu OWL là một tập các câu RDF
  - OWL xác định ngữ nghĩa cho một số câu cụ thể
- Thêm các tính năng thông dụng cho Lô-gic mô tả, ví dụ, ràng buộc cơ sở, định nghĩa lớp, tương đương, các lớp không giao, v.v..
- Hỗ trợ các ontologies như những đối tượng (ví dụ, import, phiên bản, ...)
- Một mô-đun suy diễn OWL đầy đủ phức tạp hơn nhiều so với suy diễn RDFS

# OWL $\Leftrightarrow$ RDF<sub>(2)</sub>

- RDF cho phép chúng ta mô tả dữ liệu
- RDFS thêm vào khả năng bổ sung một số dữ liệu ở mức lược đồ
- OWL mở rộng cho phép nhiều thông tin mức lược đồ hơn
- Chúng ta thường sử dụng RDFS và OWL để định nghĩa ontologies lĩnh vực (các lược đồ)
- Và sau đó sử dụng những ontologies đó để đưa ra thông tin về các đối tượng

# Dữ liệu ngữ nghĩa nhúng trong HTML

- Nhúng dữ liệu ngữ nghĩa trong HTML cho phép cả người và máy tính cùng hiểu văn bản
  - RDFa là một quy chuẩn nhúng RDF trong HTML như các thuộc tính của thẻ
  - JSON-LD là quy chuẩn để nhúng RDF trong một định dạng tương thích với JSON
- Facebook tìm kiếm các câu RDFa được nhúng sử dụng bộ từ vựng opengraph (og)

# Các công cụ phần mềm

- Có nhiều trình soạn thảo Ontology
  - Giao diện trực quan, hỗ trợ suy diễn
  - ví dụ, Protégé
- Lưu trữ bộ ba: Cơ sở dữ liệu bộ ba
  - Thường hỗ trợ các API riêng và thường có cổng tiếp nhận truy vấn SPARQL
  - Có thể thực hiện suy diễn: Tức thì hoặc theo yêu cầu
  - Ví dụ, Apache Jena, GraphDB
- Nền tảng và thư viện: Cho các ngôn ngữ lập trình
  - Java: Jena, hỗ trợ lưu trữ, SPARQL, suy diễn, v.v.

# Nội dung

1.1. Khái niệm Web ngữ nghĩa

1.2. Web ngữ nghĩa hiện nay

1.3. Đồ thị tri thức

1.4. Tổng quan công nghệ Web ngữ nghĩa

1.5. Ontology và một số khái niệm quan trọng



# Khái niệm Ontology trong triết học

- Chỉ tồn tại ở dạng số ít (không có ontologies)
- Được định nghĩa bởi các nhà triết học như bản chất tự nhiên của sự sống và sự tồn tại
  - Từ thế kỷ 5 TCN, Empedocles đã chia thế giới thành bốn nguyên tố - Đất, lửa, nước và không khí
    - Có thể được hiểu như công cụ để phân loại và hệ thống hóa tri thức
- Từ ontology trong triết học được sử dụng lần đầu tiên trong thế kỷ 17.

*Trong khoa học máy tính, có nhiều cách định nghĩa khái niệm ontology*

# Một số định nghĩa ontology trong KHMT

- Ontology là 1 tập từ khóa có cấu trúc phân cấp để mô tả một lĩnh vực và có thể được sử dụng như cấu trúc khung cho một cơ sở tri thức
  - *An ontology is a hierarchically structured set of terms for describing a domain that can be used as a skeletal foundation for a knowledge base. (Swartout, Patil, Knight, Russ)*
- Trong môi trường Web ngữ nghĩa, các bộ từ vựng xác định các khái niệm và các mối quan hệ được sử dụng để mô tả một lĩnh vực. (W3.org).
  - Ontology = bộ từ vựng (vocabularies)
  - Từ ontology thường được sử dụng cho các bộ từ vựng phức tạp và có tính hình thức, trong khi khái niệm bộ từ vựng được sử dụng thường xuyên hơn khi không yêu cầu các ràng buộc hình thức. (w3.org).



# Một số định nghĩa ontology trong KHMT<sub>(2)</sub>

Gruber (1993)

"Ontology là một

**đặc tả hình thức**  $\Rightarrow$  biểu diễn cho máy tính

của một **hệ khái niệm**  $\Rightarrow$  các khái niệm và các quan hệ

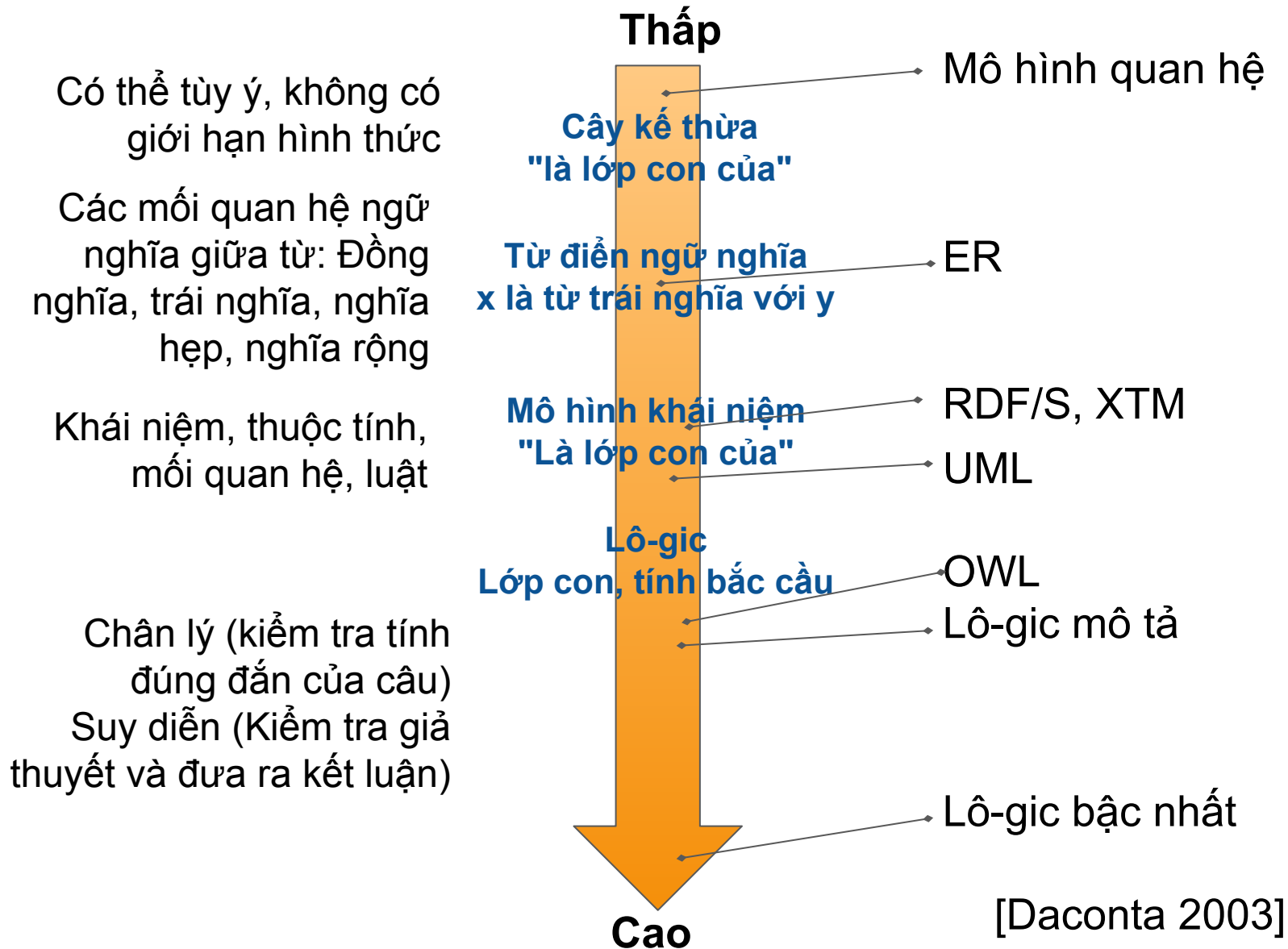
**dùng chung**  $\Rightarrow$  dựa trên các thỏa thuận

của một **lĩnh vực** ứng dụng  $\Rightarrow$  ngữ cảnh của khái niệm

# Các mức ontology

- Cây kế thừa - Taxonomy (tri thức với cấu trúc cây tối thiểu), quan hệ là lớp con của
- Từ điển ngữ nghĩa - Thesaurus, sử dụng tập hợp cố định các quan hệ (đồng nghĩa, trái nghĩa, mở rộng, thu hẹp, v.v..), ví dụ WordNet.
- Mô hình khái niệm - Conceptual Model, biểu diễn các tri thức đơn giản
- Lý thuyết Lô-gic - Các tri thức phức tạp.

# Hàm lượng ngữ nghĩa



# Phân biệt sơ đồ quan hệ và ontology

- Mục đích chính của sơ đồ quan hệ là tổ chức dữ liệu trong cơ sở dữ liệu
  - Các mối quan hệ giữa các thực thể được ngầm định bởi đối tượng sử dụng: người/chương trình máy tính.
    - Trong CSDL quan hệ chỉ có liên kết khóa chính-khóa ngoại
  - Nếu người/chương trình máy tính không hiểu ngữ nghĩa của dữ liệu thì không biểu diễn được các quan hệ.
- Ontology
  - Các mối quan hệ được định nghĩa tường minh bằng các ngôn ngữ hình thức để cả người và chương trình máy tính đều có thể biểu diễn được

