



Convertir le Trésor de la Langue Française en Ontolex-Lemon

Sina Ahmadi ( @sina_ahm)

National University of Ireland Galway

Analyse et Traitement Informatique de la Langue Française (ATILF)

Le 13 décembre 2021 - Nancy



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015.

Aperçu

1. Une introduction au Web sémantique
2. Le TLFi en Ontolex-Lemon
3. Conclusion et les travaux futurs



Introduction



Introduction

Internet

- Un vaste réseau de réseaux et d'infrastructures de mise en réseau
- Le TCP/IP constitue la base de la communication de données
- Bob Kahn et Vint Cerf sont considérés comme « les pères d'Internet »

World Wide Web (WWW, le Web, la Toile)

- Mécanisme d'accès à l'information sur Internet
- HTTP est le seul langage parlé sur le Web
- Les pages Web sont liées les unes aux autres via des hyperliens
- Tim Berners-Lee est crédité comme le fondateur du web en 1989

L'Internet et le Web

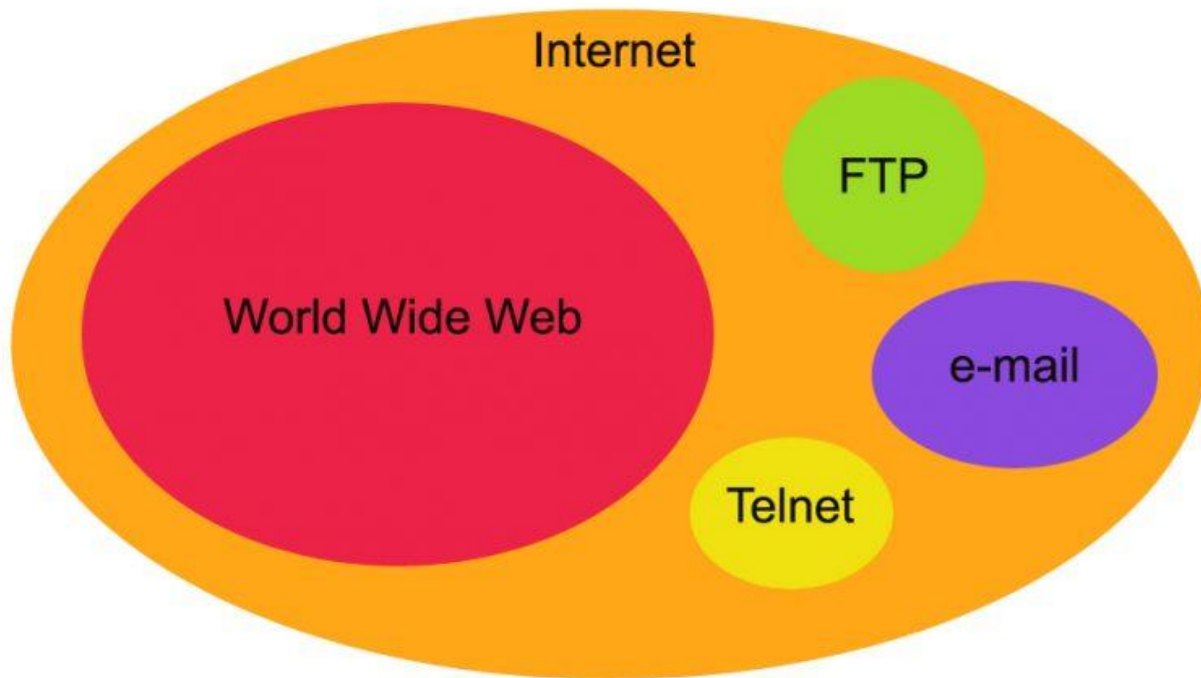


L'Internet et le Web



L'Internet et le Web

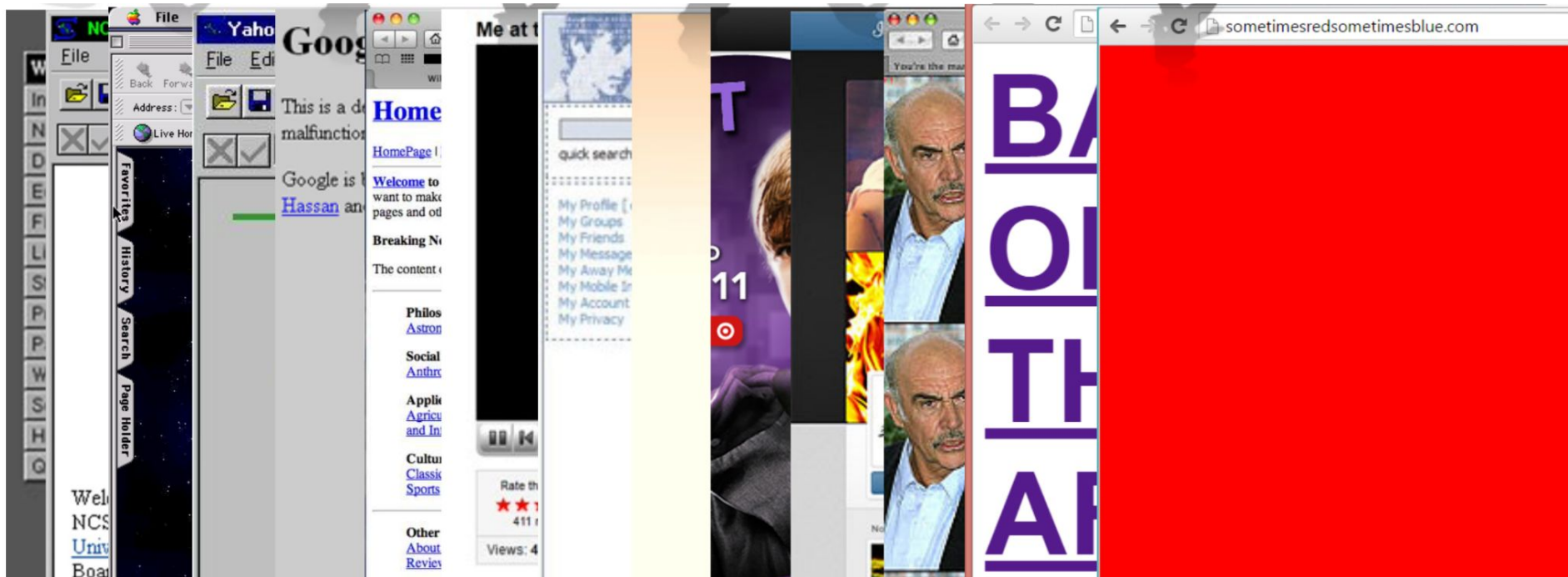
Le Web est un sous-ensemble de l'Internet



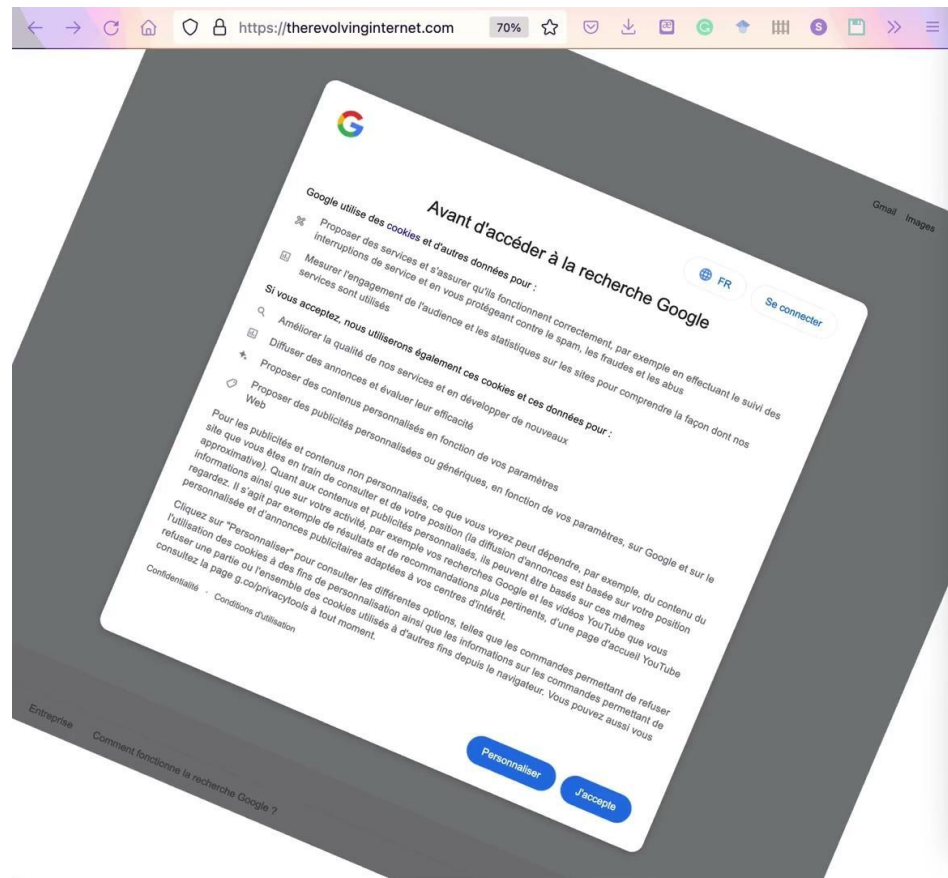
* Source de l'image :

<https://www.techpluto.com/difference-between-internet-and-world-wide-web/>

L'évolution du Web



Le Web actuel est fantastique !



Google est aussi très bien

Google

louis de funès date de naissance


[All](#)
[Images](#)
[Videos](#)
[News](#)
[Shopping](#)
[More](#)
[Tools](#)

About 92,200 results (1.00 seconds)

Tip: Search for **English** results only. You can specify your search language in [Preferences](#)

Louis de Funès / Date of birth

July 31, 1914



πώς να |

- πωσ να χασω κιλα
- πωσ να βγαλω λεφτα
- πωσ να καθυστερησω την περιοδο
- πωσ να ενεργηθω αμεσα
- πωσ να χασω λιπος
- πωσ να ψηλωσω
- πωσ να μεινω εγκυος

while my guitar gently weeps lyrics

[All](#)
[Videos](#)
[Images](#)
[News](#)
[Maps](#)
[More](#)
[Tools](#)

About 2,620,000 results (0.71 seconds)

While My Guitar Gently Weeps

[The Beatles](#)

Ok, one, two, three, four

I look at you all see the love there that's sleeping
While my guitar gently weeps

I look at the floor and I see it needs sweeping
Still my guitar gently weeps.
I don't know why nobody told you
How to unfold your love
I don't know how someone controlled you
They bought and sold you

I look at the world and I notice it's turning
While my guitar gently weeps

With every mistake we must surely be learning
Still my guitar gently weeps

I don't know how you were diverted
You were perverted too
I don't know how you were inverted

Source: [Musixmatch](#)

Songwriters: John Lennon / George Harrison / Paul James McCartney

While My Guitar Gently Weeps lyrics © Harrisongs Ltd

L'avenir du Web ?



A quoi ressemblera le Web dans 3 décennies ?





1- Introduction

Quel est le problème avec le Web actuel ?



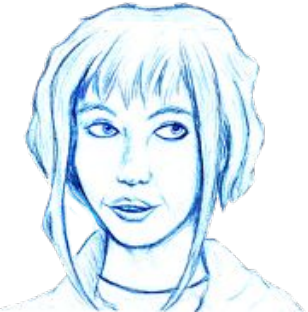
Les problèmes avec le Web actuel



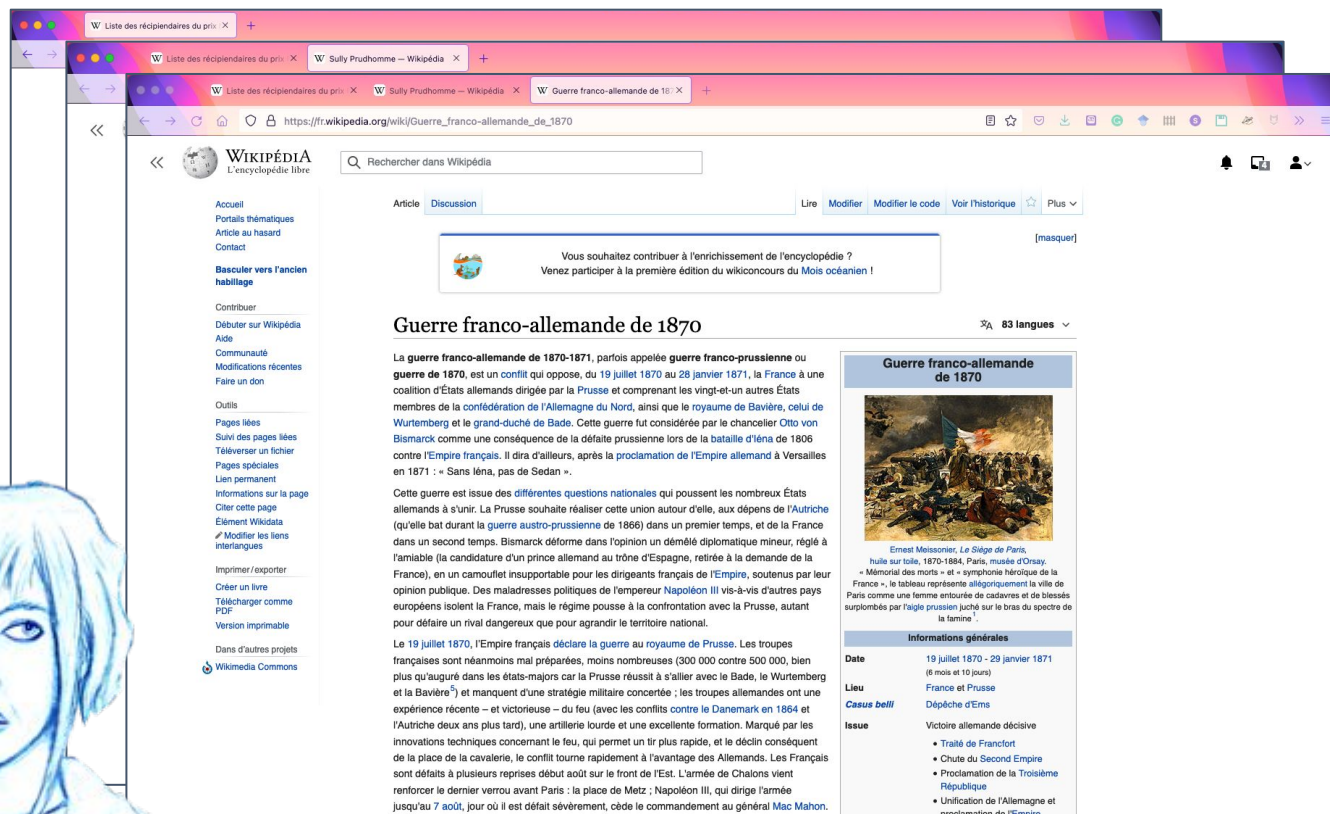
Si on voudrait trouver :

- Les lauréats du prix Nobel de littérature
- qui ont combattu dans au moins une guerre
- L'année où ils ont remporté le prix
- Et l'année où la ou les guerres ont commencé

« Comment répondre à tout ça ? »



Beaucoup d'onglets Wikipédia...



The screenshot shows a web browser with multiple tabs open, including 'Liste des récipiendaires du prix...', 'Sully Prudhomme - Wikipédia', and 'Guerre franco-allemande de 1870'. The active tab displays the Wikipedia article for 'Guerre franco-allemande de 1870'. The article text is in French and discusses the conflict between France and Prussia in 1870-1871. It mentions the coalition of German states, the role of Otto von Bismarck, and the outcome of the war. The article is written in French and includes a section for 'Informations générales' (General information) at the bottom right.

Informations générales

Date	19 juillet 1870 - 29 janvier 1871 (6 mois et 10 jours)
Lieu	France et Prusse
Causa belli	Dépêche d'Éms
Issue	Victoire allemande décisive <ul style="list-style-type: none"> • Traité de Francfort • Chute du Second Empire • Proclamation de la Troisième République • Unification de l'Allemagne et proclamation de l'Empire



L'algorithme du lauréat

results := \emptyset

for all *nobel-lit-winner* **in** *wiki-list*

year-prize := year of *nobel-lit-winner*

wars := **search** “guerre”, “conflit”, “bataille”, “combat”, “lutte armée” **in** *nobel-lit-winner*

for all *war* **in** *wars*

year-war := starting year of *war*

add *nobel-lit-winner*, *year-prize*, *war*, *year-war* **to** *results*

end

end

return *results*



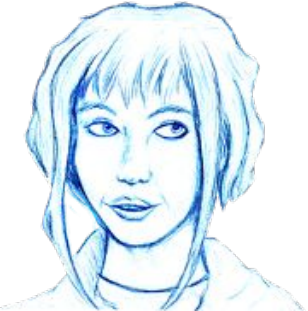
Les problèmes avec le Web actuel



Si on voudrait trouver :

- Les lauréats du prix Nobel de littérature
- qui ont combattu dans au moins une guerre
- L'année où ils ont remporté le prix
- Et l'année où la ou les guerres ont commencé

mais pourquoi cette requête est-elle difficile sur le Web actuel ?



Le Web actuel est centré sur les documents

http://ex.org/France

France


La France est un État souverain transcontinental.

La capitale de la france est [Paris](#).

http://ex.org/Paris

Paris


Paris est la capitale de la [France](#).

La ville de Paris comptait 2,187 millions d'habitants en 2020.

```

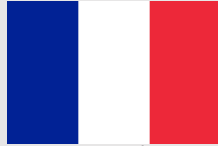
<html>
  <body>
    <h1>France</h1>
    <div class="flag"></div>
    <p>La France est un État souverain transcontinental.</p>
    <p>La capitale de la france est <a href="http://ex.org/Paris">Paris</a>.</p>
  </body>
</html>

```

Cela a du sens pour les humains (plus ou moins)

<http://ex.org/France>

France



La France est un État souverain
transcontinental.

La capitale de la france est [Paris](#).

<http://ex.org/Paris>

Paris




Paris est la capitale de la [France](#).

La ville de Paris comptait 2,187
millions d'habitants.



... si on parle la même langue

<http://ex.org/ckb/فهرانسسا>



فهرانسسا

فهرانسسا ولاتیکی سهر بهخو له ئه‌ورووپایه.

پایتەختی ولاتی فهرانسسا پاریسه.

<http://ex.org/ckb/پاریس>



پاریس

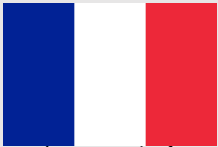
پاریس پایتەختی ولاتی فهرانسایه.

ریژەهی نیشتەجێیانی ئەم ولاتی له سالی ۲۰۲۰دا ۲،۱۸۷ ملیۆن کەس بووه.



C'est encore pire pour les machines !

<http://ex.org/ckb/فهرانسنا>



فهرانسنا

فهرانسنا ولاتیکی سهر بهخۆ له نهو رووپایه.

پایتهختی ولاتی فهرانسنا پاریسه.

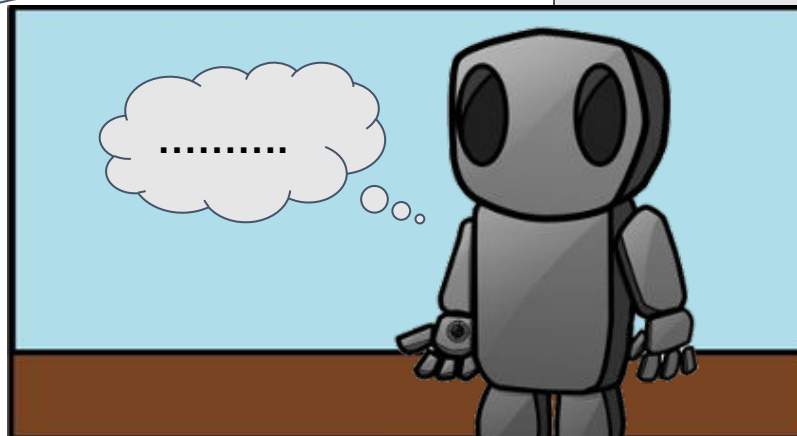
<http://ex.org/ckb/پاریس>



پاریس


پاریس پایتهختی ولاتی فهرانسنایه.

رێژهی نیشتهجێانی ئهم ولاتی له سالی ۲۰۲۰ تا ۲۰۱۷، ۲ ملیۆن کەس بووه.



Est-ce que les machines peuvent «comprendre» le Web?

http://ex.org/France

 **فهرانسا**

فهرانسا ولاتیکی سهر بهخۆ له ئهورووپایه.

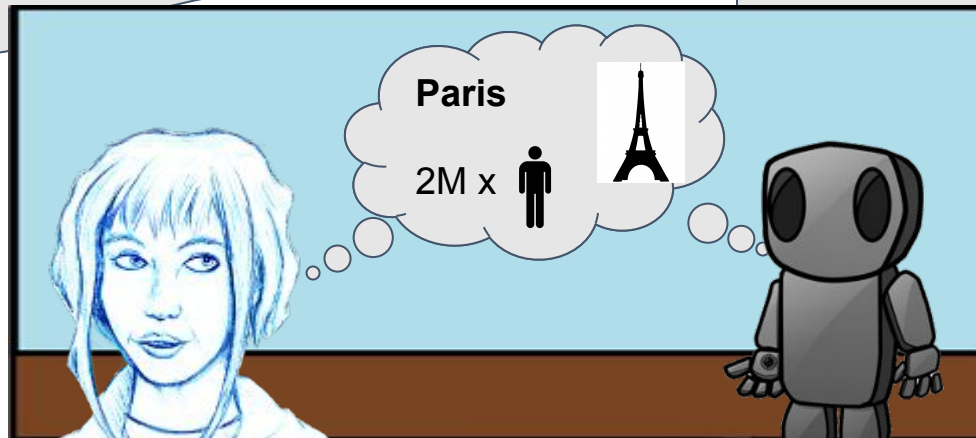
پایتهختی ولاتی فهرانسا پاریسه

http://ex.org/Paris

 **پاریس**

پاریس پایتهختی ولاتی فهرانسایه.

ریژهی نیشهجینیانی ئهم ولاتی له سالی ۲۰۲۰ تا ۲۰۱۸ ۲ ملیۆن کەس بووه.



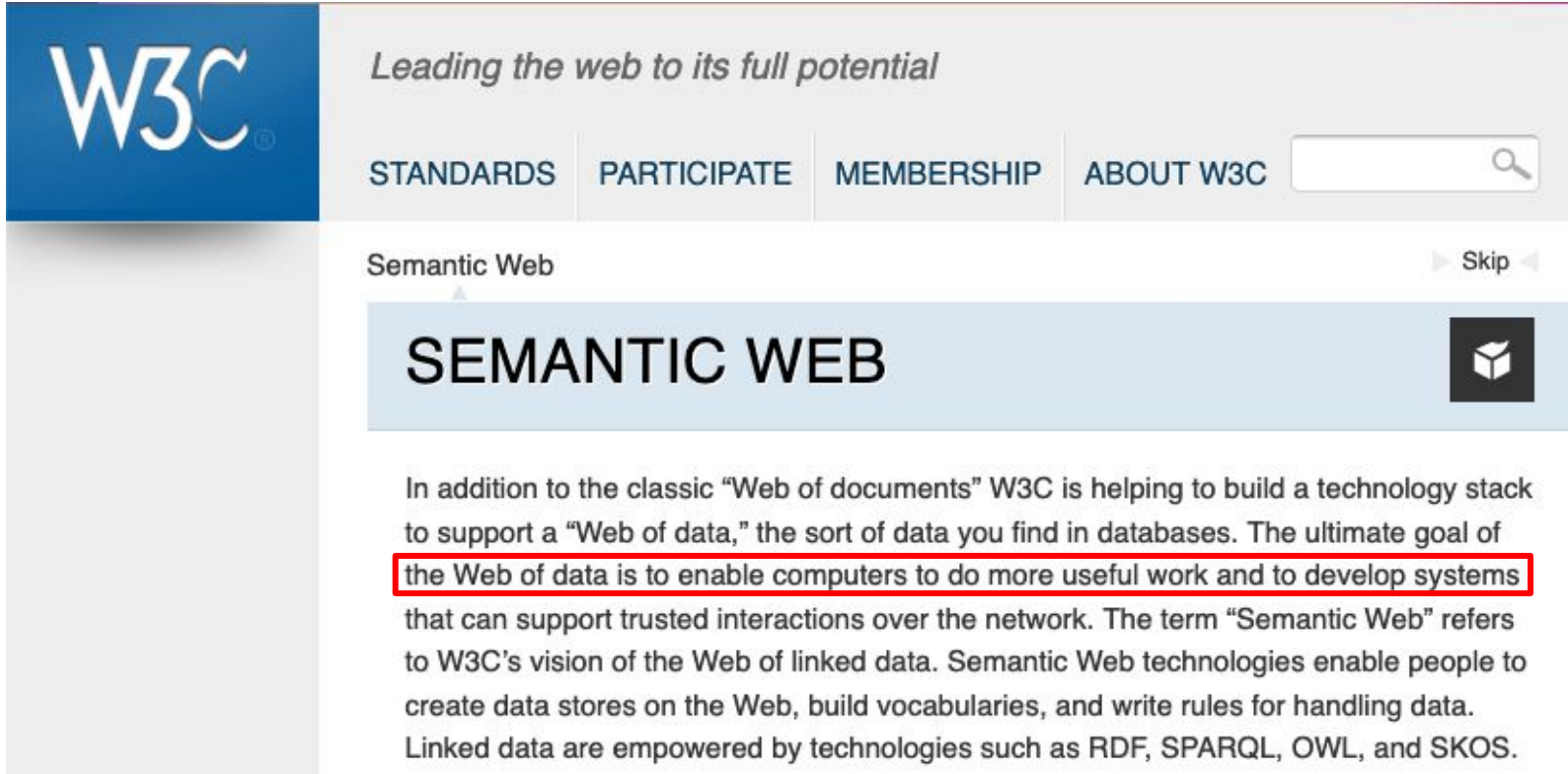


1- Introduction

Le Web sémantique



Le « Web sémantique »



The screenshot shows the W3C website header with the logo and navigation links: STANDARDS, PARTICIPATE, MEMBERSHIP, and ABOUT W3C. Below the header, the 'Semantic Web' section is highlighted with a blue background and the title 'SEMANTIC WEB' in large, bold letters. A red box highlights the sentence: 'The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network.' The text continues to describe the Semantic Web as a 'Web of linked data' and lists technologies like RDF, SPARQL, OWL, and SKOS.

W3C[®]

Leading the web to its full potential

STANDARDS PARTICIPATE MEMBERSHIP ABOUT W3C

Semantic Web Skip

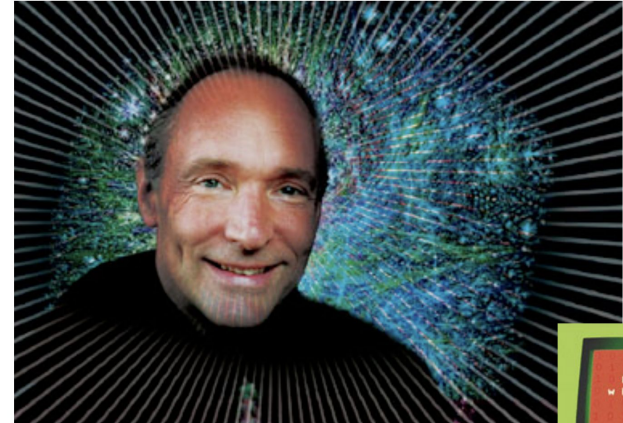
SEMANTIC WEB

In addition to the classic “Web of documents” W3C is helping to build a technology stack to support a “Web of data,” the sort of data you find in databases. The ultimate goal of the Web of data is to enable computers to do more useful work and to develop systems that can support trusted interactions over the network. The term “Semantic Web” refers to W3C’s vision of the Web of linked data. Semantic Web technologies enable people to create data stores on the Web, build vocabularies, and write rules for handling data. Linked data are empowered by technologies such as RDF, SPARQL, OWL, and SKOS.

Le Web sémantique, c'est...

La vision de Tim Berners-Lee ?

« Je rêve d'un Web [dans lequel les ordinateurs] deviennent capables d'analyser toutes les données du Web : le contenu, les liens, et les transactions entre personnes et ordinateurs. »



-- Tim Berners-Lee, Weaving the Web (2000)

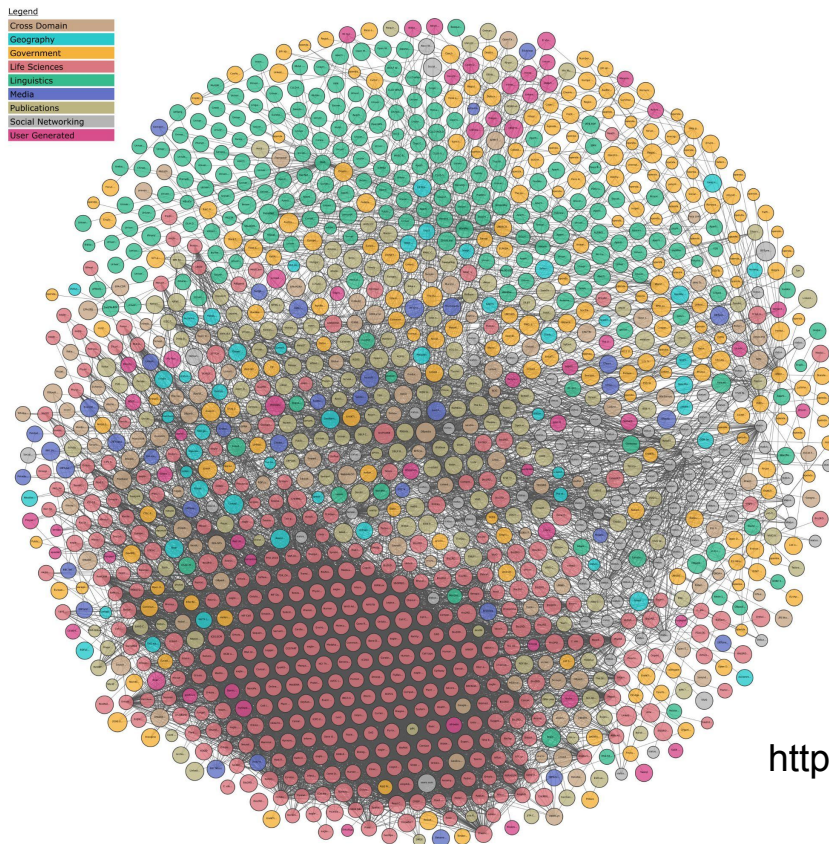
Le Web sémantique, c'est...

Des machines intelligentes qui donnent du sens au Web ?



Le Web sémantique, c'est...

Relier ensemble des données ouvertes ?



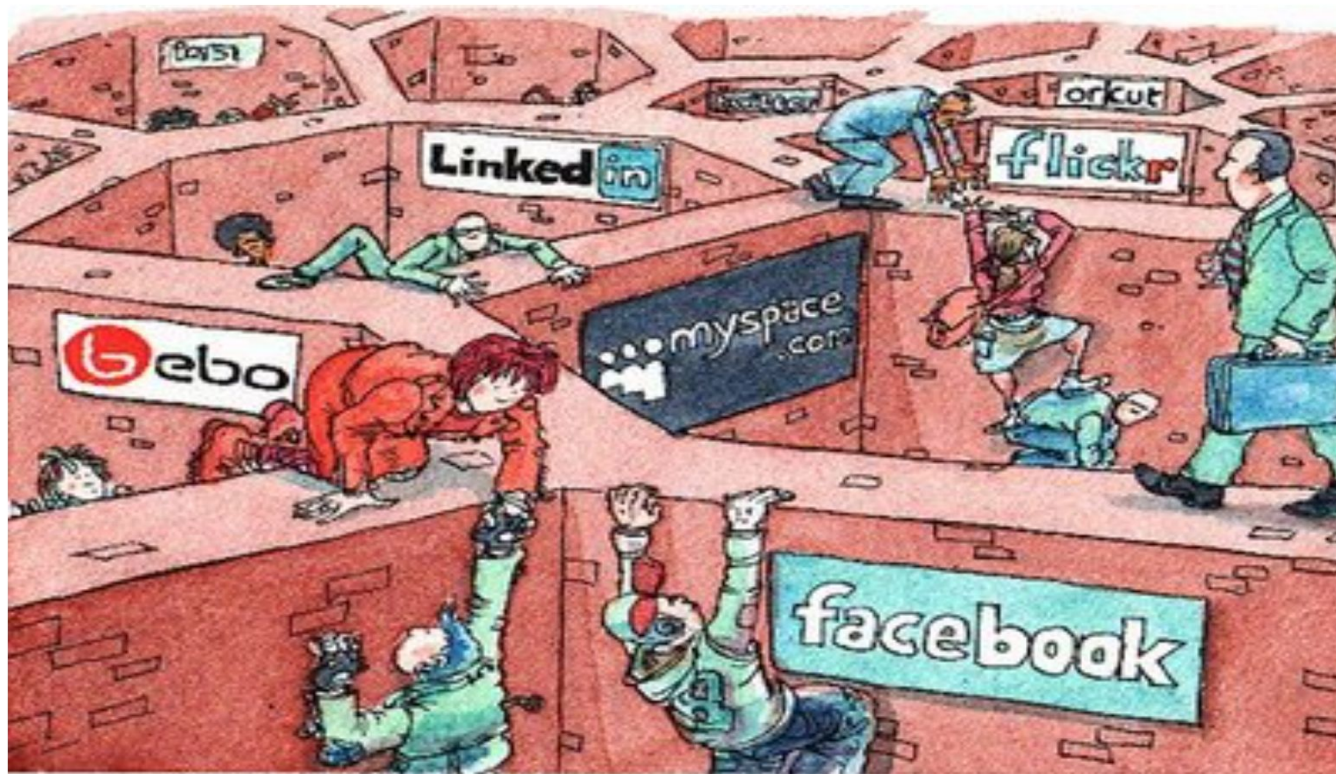
<https://lod-cloud.net/>

Le Web sémantique, c'est... Nommer tout ?



"Now! ... That should clear up
a few things around here!"

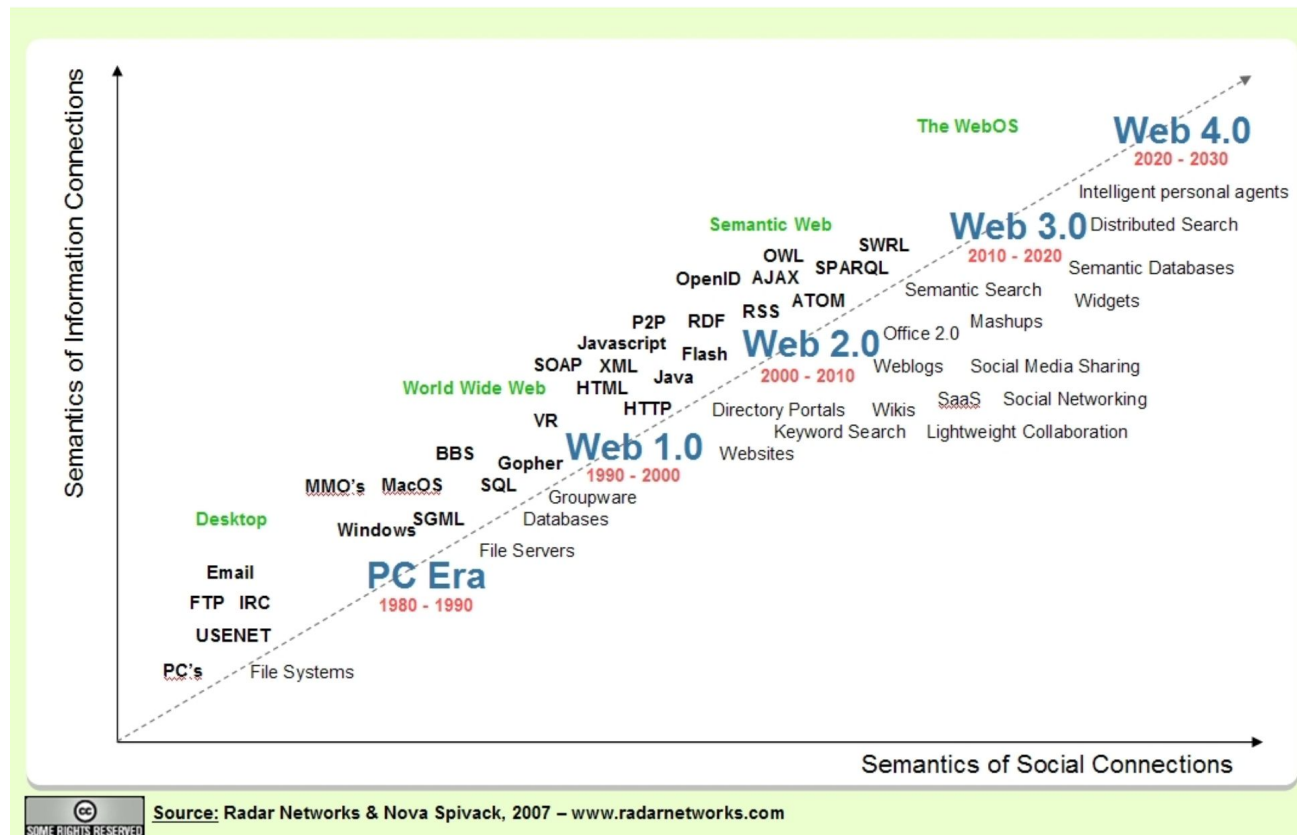
Le Web sémantique, c'est... Abattre les murs entre les sites ?



L'image de
<http://www.digitaltimes.ie>

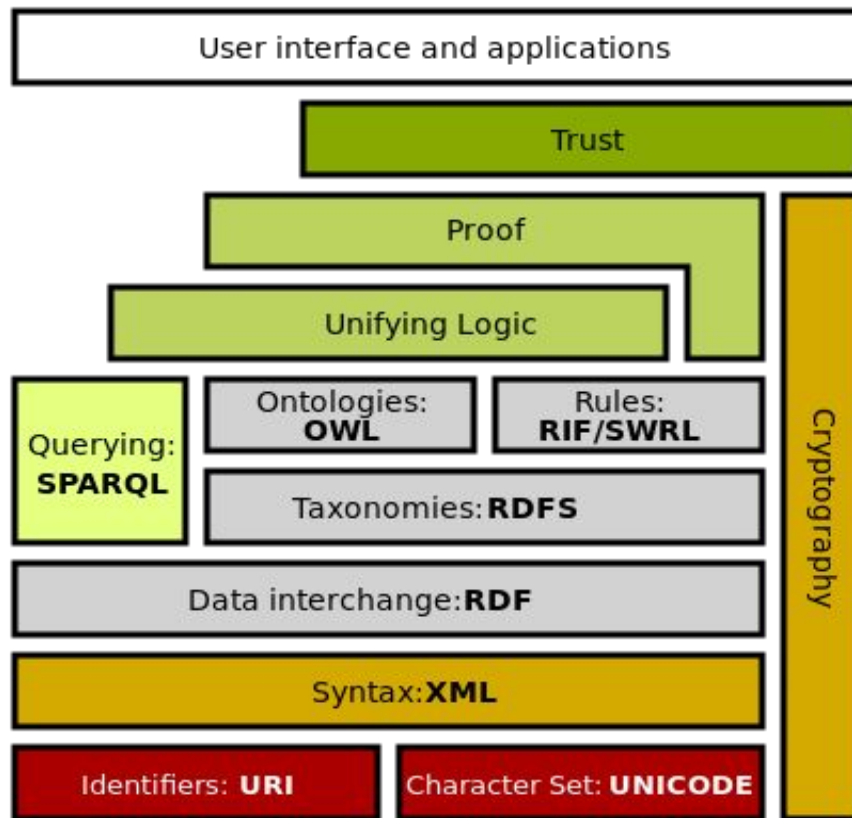
Le Web sémantique, c'est...

Le Web 4.0 ?



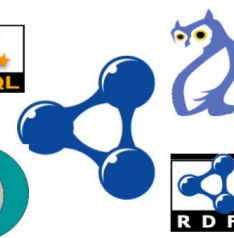
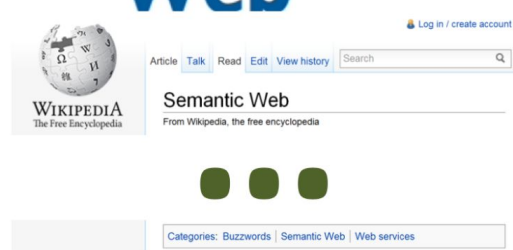
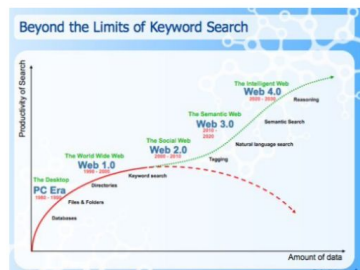
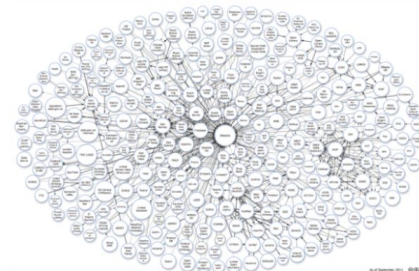
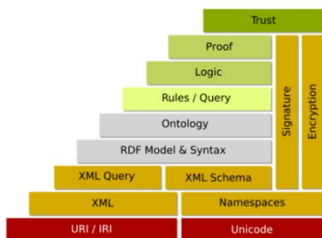
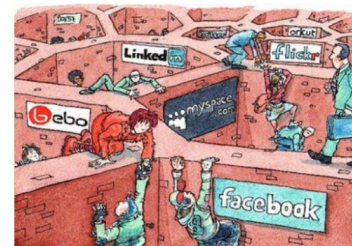
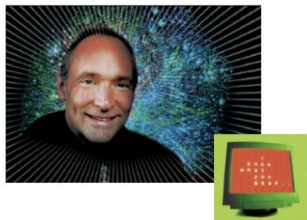
Le Web sémantique, c'est...

Un ensemble de technologies et de standards ?



Source : W3C,
T Berners-Lee,
Ivan Herman

Le Web sémantique, c'est...



Le Web sémantique : les principes

Les trois principaux standards du Web sémantique

1. RDF : un modèle de **triplets** pour décrire et connecter des ressources anonymes ou identifiées par un URI
2. SPARQL : un langage de requête sur les graphes RDF
3. OWL : Le langage de schéma, ou langage de représentation des connaissances du Web sémantique



1- Introduction

RDF



RDF (*Resource Description Framework*)



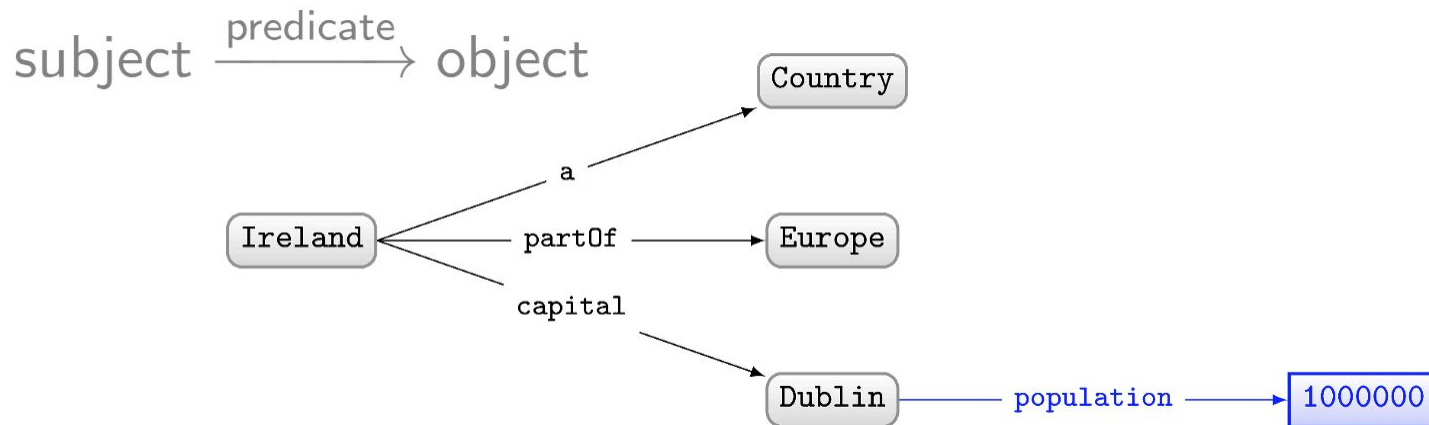
- Modèle de données pour décrire des ressources du web
- C'est un graphe :
 - les noeuds représentent des ressources
 - les arcs représentent des relations entre les ressources
- Triplet RDF <sujet, prédicat, objet>
 - Le sujet décrit une entité (identifié par une URI)
 - Le prédicat associe au sujet une propriété (identifiée par une URI)
 - L'objet donne une valeur à la propriété. Une valeur est soit une ressource identifiée par une URI, soit une valeur primitive
- Plusieurs formats pour écrire du RDF tels que N-Triples, RDF/XML, RDFa, JSON-LD et Turtle

Modéliser tout en triplet

<i>subject</i>	<i>predicate</i>	<i>object</i>
Ireland	partOf	Europe
Ireland	a	Country
Ireland	capital	Dublin

Concaténer pour « intégrer » de nouvelles données

<i>subject</i>	<i>predicate</i>	<i>object</i>
Ireland	partOf	Europe
Ireland	a	Country
Ireland	capital	Dublin
Dublin	population	1,000,000

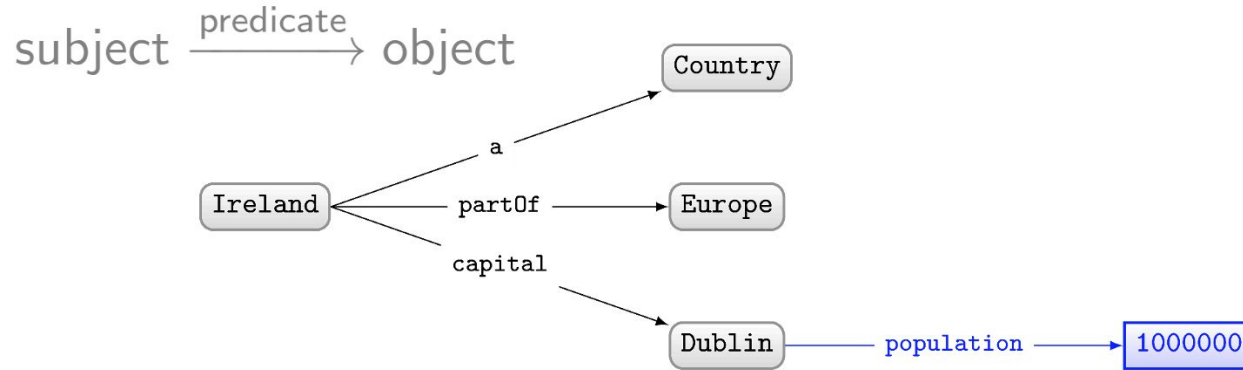


Besoin de symboles/identifiants non ambigus

- **URL: Uniform Resource Location**
 - L'emplacement d'une ressource sur le Web
 - `http://ex.org/Dubl%C3%ADn.html`
- **URI: Uniform Resource Identifier (RDF 1.0)**
 - N'a pas besoin d'être un emplacement, peut aussi être un nom
 - `http://ex.org/Dubl%C3%ADn`
- **IRI: Internationalised Resource Identifier (RDF 1.1)**
 - Un URI qui autorise les caractères Unicode
 - `http://ex.org/Dublín`

<i>subject</i>	<i>predicate</i>	<i>object</i>
[IRI, Blank Node]	[IRI]	[IRI, Blank Node, Literal]

Les triplets



<<http://ex.org/Ireland>> <<http://ex.org/a>> <<http://ex.org/Country>> .

<<http://ex.org/Ireland>> <<http://ex.org/partOf>> <<http://ex.org/Europe>> .

<<http://ex.org/Ireland>> <<http://ex.org/capital>> <<http://ex.org/Dublin>>

.

<<http://ex.org/Dublin>> <<http://ex.org/population>> "1000000" .

Beaucoup trop d'URI ? Utilisons des préfixes

<http://ex.org/Dublin> ↔ **ex:Dublin**

“**ex:**” désigne un préfixe pour <http://ex.org/>

Donc, `ex:Ireland` `ex:Capital` `ex:Dublin` .

Les préfixes fréquemment utilisés :

Prefix	Value
<code>rdf:</code>	<code>http://www.w3.org/1999/02/22-rdf-syntax-ns#</code>
<code>xsd:</code>	<code>http://www.w3.org/2001/XMLSchema#</code>
<code>rdfs:</code>	<code>http://www.w3.org/2000/01/rdf-schema#</code>
<code>owl:</code>	<code>http://www.w3.org/2002/07/owl#</code>

C'est quoi Wikidata ?

The image shows the Wikidata main page. On the left is a sidebar with the Wikidata logo and a list of links: Main page, Community portal, Project chat, Create a new Item, Create a new Lexeme, Recent changes, Random Item, Query Service, Nearby, Help, Donate, Tools, What links here, Related changes, and Special pages. The main content area has a navigation bar with 'Main Page', 'Discussion', 'Read', 'View source', and 'View history', followed by a search bar labeled 'Search Wikidata'. The central text reads 'Welcome to Wikidata' and 'the free knowledge base with 89,078,643 data items that anyone can edit.' Below this are links for 'Introduction • Project Chat • Community Portal • Help'. A green button says 'Welcome!' and a blue button says 'Learn about data'. A network diagram is overlaid on the page, with nodes and lines connecting various terms: 'open', 'multilingual', 'free', 'collaborative', and 'Want to help translate? Translate the missing messages.'

WIKIDATA

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

Tools
What links here
Related changes
Special pages

Main Page Discussion Read View source View history Search Wikidata

open

multilingual

free

collaborative

Welcome to Wikidata

the free knowledge base with 89,078,643 data items that anyone can edit.

Introduction • Project Chat • Community Portal • Help

Want to help translate? Translate the missing messages.

Welcome!

Learn about data



2- Données liées



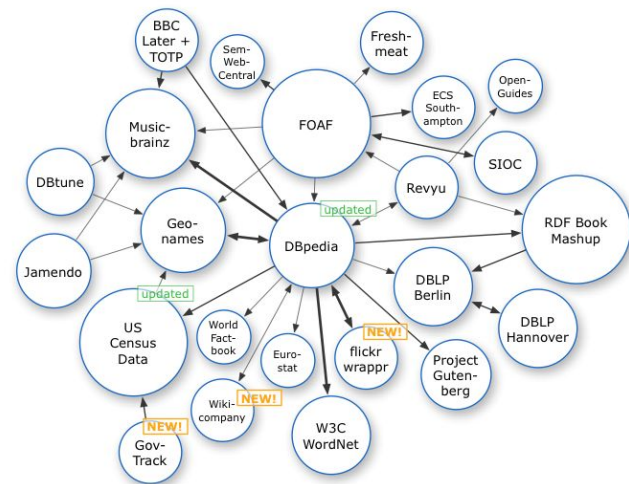
Le Web sémantique (avant 2006)



Et si on pouvait connecter et lier toutes ces ressources ?

Le Web des données

- Les sources des données font appel à des vocabulaires partagés:
 - FOAF, GoodRelations
 - SKOS, DC, MADS, OAI ORE, FRBR
 - WGS84 GEO, Geonames, Event, Time
- Les promesses des données ouvertes liées
 - représentation et modélisation
 - interopérabilité, à la fois structurelle et co
 - fédération
 - écosystème
 - expressivité
 - importation dynamique

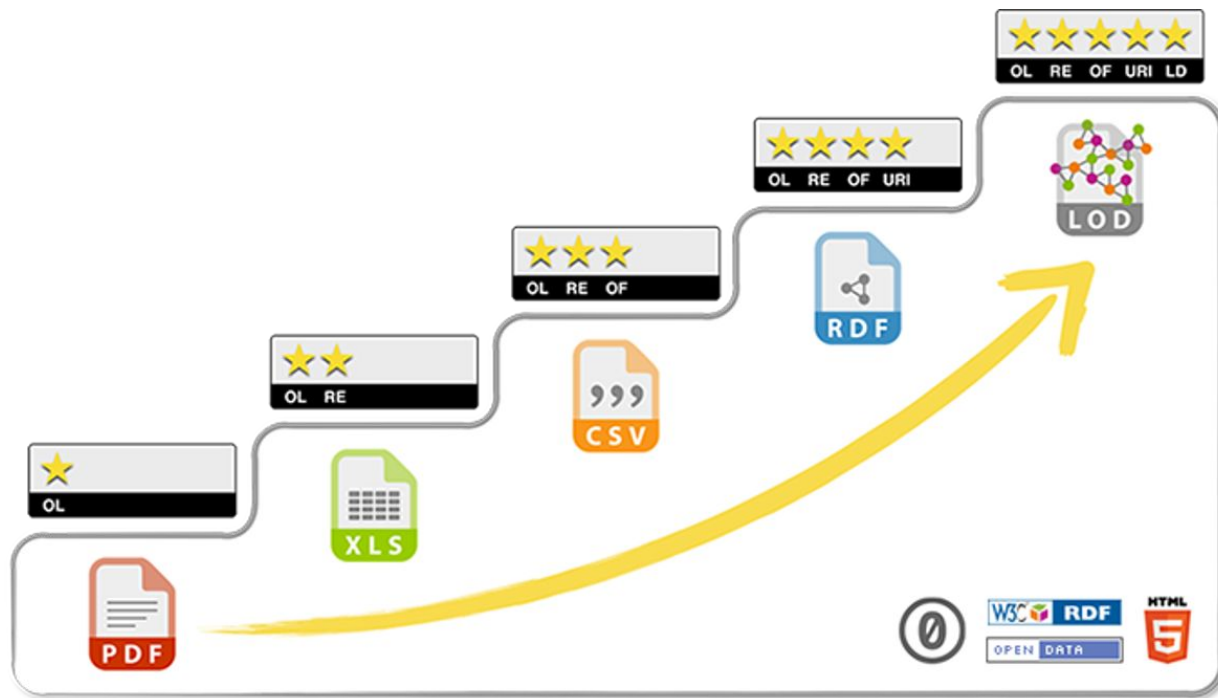


Les quatre principes de données liées

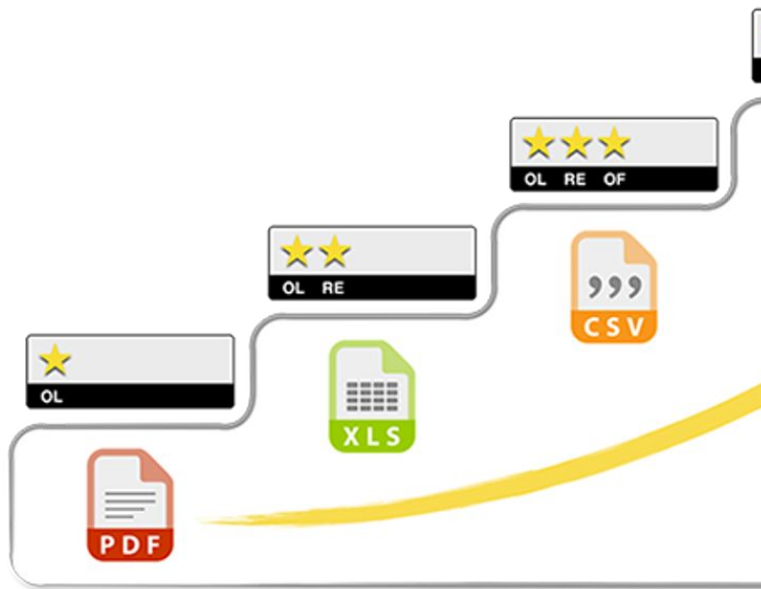
1. utiliser des adresses URI uniques pour identifier les choses
2. utiliser des adresses URI HTTP qui existent sur le Web (des URL, donc)
3. fournir à travers l'adresse URI des renseignements exploitables, lisibles par les humains et par les machines, en s'appuyant sur des formats ouverts comme RDF ou SparQL.
4. mailler l'adresse URI initiale en lui associant des adresses URI externes et ce pour améliorer la découverte d'autres informations sur le Web.

Source : <http://www.w3.org/DesignIssues/LinkedData.html> (traduction en français : https://fr.wikipedia.org/wiki/Web_des_donn%C3%A9es)

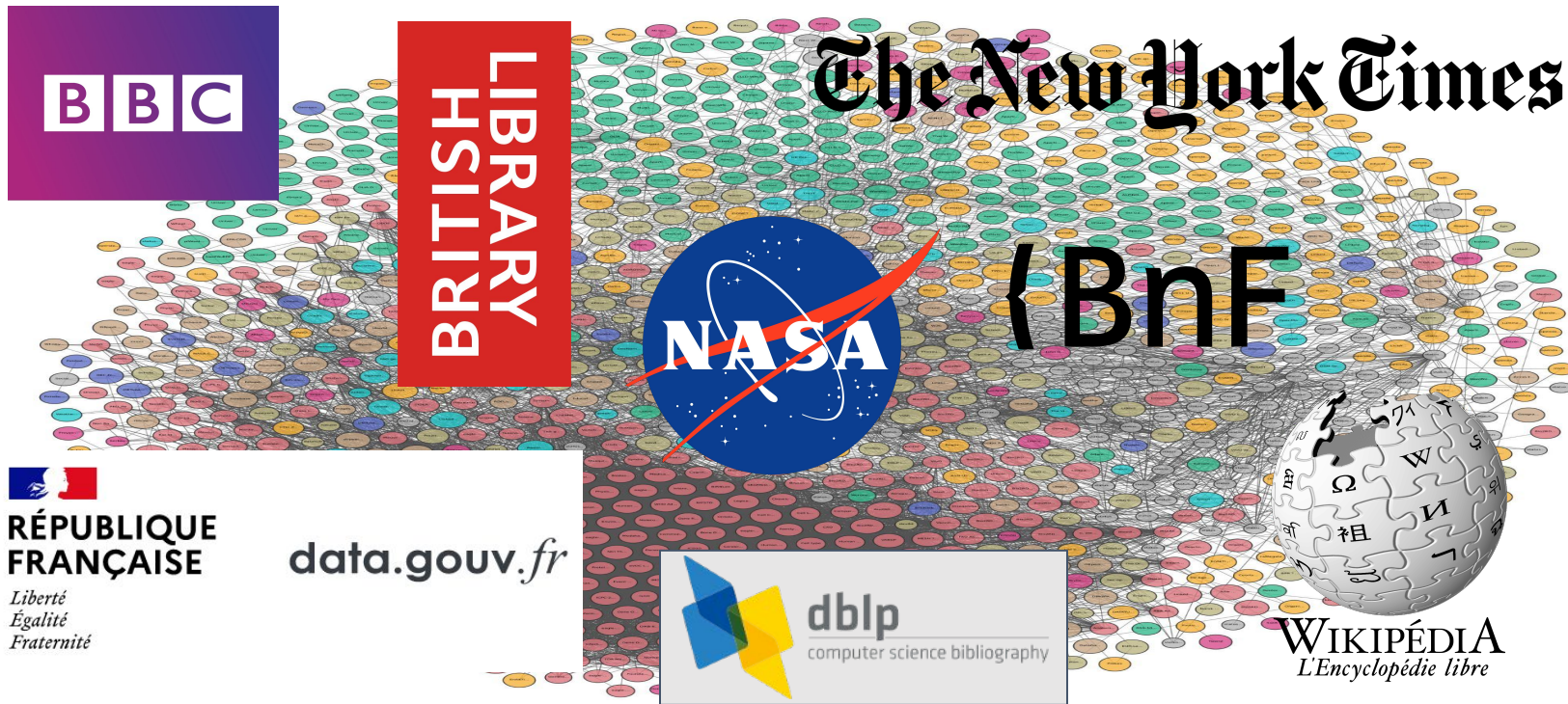
Vos données ouvertes liées sont-elles 5 étoiles ?



Vos données ouvertes liées sont-elles 5 étoiles ?



Le Linked Data Cloud



The Linked Open Data Cloud from lod-cloud.net



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 731015.



Un exemple de la BNF

@prefix foaf: <<http://xmlns.com/foaf/0.1/>> .

@prefix rdagroup2elements: <<http://RDVocab.info/ElementsGr2/>> .

@prefix xfoaf: <<http://www.foafrealm.org/xfoaf/0.1/>> .

@prefix owl: <<http://www.w3.org/2002/07/owl#>> .

<<http://data.bnf.fr/ark:/12148/cb11928016k#foaf:Person>> a foaf:Person;

 rdagroup2elements:dateOfBirth "08-02-1828" ;

 rdagroup2elements:dateOfDeath "24-03-1905" ;

 rdagroup2elements:placeOfBirth "Nantes" ;

 rdagroup2elements:placeOfDeath "Amiens" ;

 xfoaf:nationality <<http://id.loc.gov/vocabulary/countries/fr>> ;

owl:sameAs <http://dbpedia.org/resource/Jules_Verne>;

 foaf:name "Jules Verne";

 foaf:page <http://data.bnf.fr/11928016/jules_verne/> .



2- Données liées

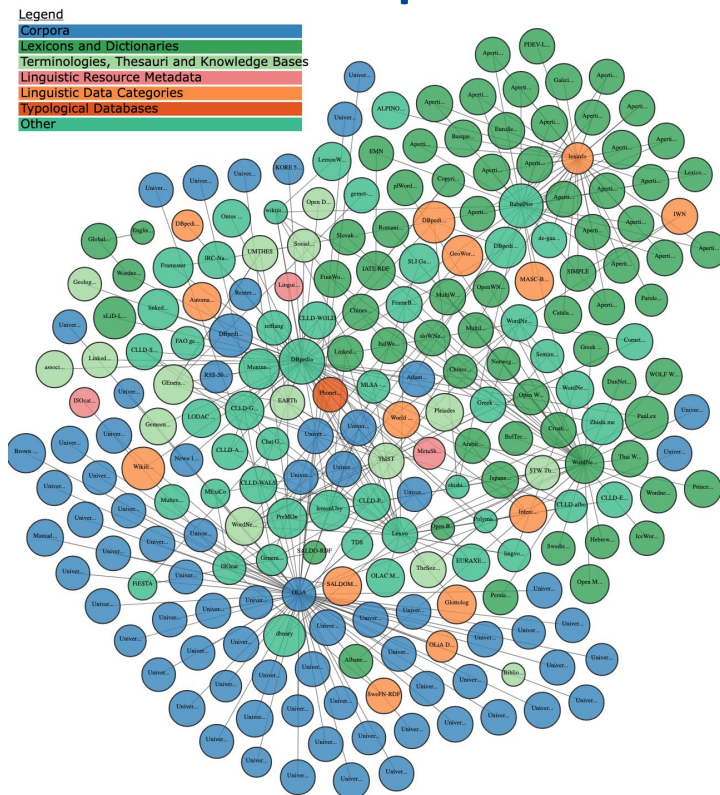
Ontolex-Lemon



Données linguistiques liées (ouvertes)

- Données linguistiques
 - Dictionnaires, corpus, typologies, etc.
- Données liées
 - Ce qui fait référence à d'autres jeux de données
 - Utilisation des normes (W3C), par ex. RDF
- Données ouvertes
 - Licences ouvertes, par ex. Creative Commons

The Linguistic Linked Open Data Cloud



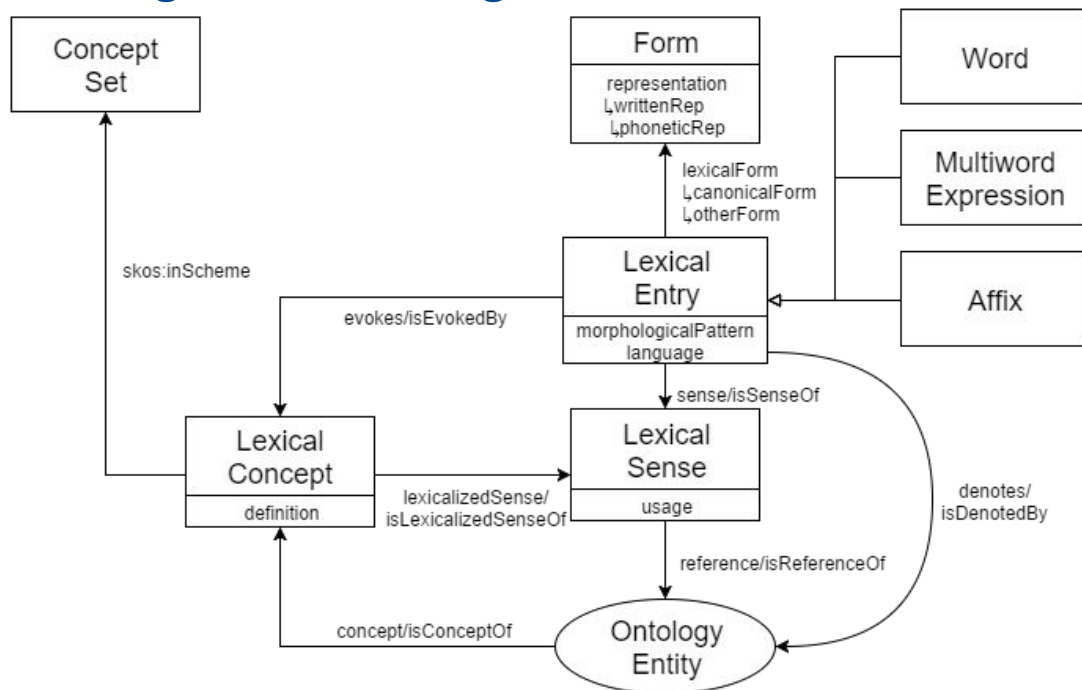
The Linguistic Linked Open Data Cloud from lod-cloud.net



<https://linguistic-lod.org/>

Ontolex-Lemon

- Modèle de représentation des dictionnaires en RDF
- S'intègre aux ontologies en OWL



Rejoignez la communauté
Ontolex-Lemon :
<http://www.w3.org/community/ontolex>

Ontolex-Lemon : les caractéristiques

1. OWL & RDF
2. Multilingualisme
 - a. Prend en charge n'importe quelle langue
 - b. Ne pas faire d'hypothèses spécifiques à une langue
 - c. Traduction et variation
3. Sémantique par référence : les références capture des informations sémantiques
4. Ouverture
5. Réutiliser les autres standards, tels que OWL, RDF, SKOS, Dublin Core, LMF et TMF

Espaces de
nommage

Ontolex-Lemon : un exemple

```
@prefix ontollex: <http://www.w3.org/ns/lemon/ontollex#> .
```

```
@prefix skos: <http://www.w3.org/2004/02/skos/core#> .
```

```
<#cat> a ontollex:Word ;
```

```
ontollex:canonicalForm [  
  ontollex:writtenRep "chat"@fr ] ;
```

Lemme

```
ontollex:denotes [  
  skos:definition "Petit mammifère familial à poil doux, aux yeux  
oblongs et brillants, à oreilles triangulaires, aux griffes  
rétractiles."@fr ] .
```

Sens



2- Données liées

Le TLFi en Ontolex-Lemon



Le TLFi

Le TLFi (<https://www.atilf.fr/ressources/tlfi/>) est une des plus importantes ressources lexicographiques du français.

Il contient:

- **100 000** entrées
- **270 000** définitions
- **430 000** exemples
- du XIVème au XXème siècle

La micro-structure de ce dictionnaire est enrichie par plusieurs types d'informations, notamment des **sens** et des **définitions**, des **exemples d'usage**, des **étymologies**, des **indications d'emplois** et de **domaine général**, ainsi que des **locutions**.

Le TLFi

Le TLFi (<https://www.atilf.fr/ressources>)
importantes ressources lexicographiques

Il contient:

- 100 000 entrées
- 270 000 définitions
- 430 000 exemples
- du XIV^{ème} au XX^{ème} siècle

La micro-structure de ce dictionnaire
d'informations, notamment des sens
d'usage, des étymologies, des indices
général, ainsi que des locutions.

■ **BAROQUE**, adj.

A. [En parlant d'un style, d'une époque, d'une œuvre monumentale]

1. Domaine des B.-A.

– **ARCHIT.** Qui est caractéristique de la période qui a suivi la Renaissance classique. *Style baroque, architecture baroque* :

- 1. Soleure. La belle cathédrale **baroque** (...) toute blanche et frisée, heureuse, et même un peu triomphante avec ses statues dorées et ses inscriptions latines largement déployées au-dessus des chapelles. GREEN, *Journal*, 1946-50, p. 118.

– **MUS.** Qui est caractéristique de la période musicale propre à l'Allemagne, à l'Angleterre, à l'Italie, et qui s'étend de 1580 à 1760.

♦ [En parlant des œuvres et des artistes] Qui illustre les principes esthétiques de ces périodes. *Musique baroque*.

– **PEINTURE** :

- 2. [Maurice Denis] On a discerné en lui l'influence conjugée de Fra Angelico et de Poussin. Il faudrait ajouter celle de la Renaissance italienne, de Rome plus que de Florence, du clair-obscur bolonais et des peintres **baroques**. *Arts et litt. dans la société contemp.*, 1936, p. 1807.

– P. anal., **LITT.** Qui appartient à l'époque littéraire qui, en France, correspond aux règnes de Henri IV et Louis XIII.

2. Plus gén. [En parlant d'un artiste, ou d'une création artistique] Qui par son style rappelle celui de la période baroque :

- 3. Ses allégories [de J. B. Rousseau], sont jugées tout d'une voix : **baroques**, métaphysiques, sophistiquées, sèches, inextricables, nul défaut n'y manque. SAINT-BEUVE, *Portraits littér.*, t. 1, 1844-64, p. 136.
- 4. Ariadne auf Naxos, (...), accentue plus encore l'évolution du musicien vers le style **baroque** avec son *maniérisme*, ses *bizarries*, où la délicatesse, la grâce de Mozart se mêlent à la bouffonnerie débridée de la Commedia Dell'Arte. R. DUMESNIL, *Hist. illustrée du théâtre lyrique*, 1953, p. 171.

3. *Emploi subst.*

a) [L'art, le style qui appartiennent à la période baroque] :

- 5. Hier, discussion inutile sur le **Baroque** (...) Toujours cette éternelle confusion entre le **baroque** et le rococo (...) On refuse au vrai **baroque** la gravité religieuse alors qu'elle est très évidente à la chapelle de la Sorbonne ou au Val-de-Grâce; ... GREEN, *Journal*, 1946-50, p. 142.
- 6. Dès que l'art d'assouvissement point, notre répulsion apparaît. D'où notre admiration du grand **baroque** créateur, de Michel-Ange au Greco, et notre dédain du **baroque** établi; ... MALRAUX, *Les Voix du silence*, 1951, p. 526.

– [L'art, le style, la manière qui rappellent ceux de l'époque baroque] :

- 7. Le **baroque** de son dessin forcé [de Van Gogh] est plus proche des proues scandinaves ou des plaques scythes que de Rubens... MALRAUX, *Les Voix du silence*, 1951, p. 576.

b) Artiste dont le style rappelle cette période :

- 8. Il fallut la mort de Péguy pour familiariser Barrès avec l'idée qu'il n'était autre chose qu'un **baroque**... J. et J. THARAUD, *Pour les fidèles de Péguy*, 1928, p. 86.

B.- **JOAILL.** [Se dit d'une perle] Qui est de forme irrégulière, d'une rondeur imparfaite :

- 9. En dehors de son orient et de son lustre attrayants, une perle doit avoir une couleur homogène et une rondeur aussi parfaites que possible; les beaux spécimens se vendent à la pièce (en grains); les perles **baroques** se vendent en sachets au poids. A. et N. METTA, *Les Pierres précieuses*, 1960, p. 120.

– *Emploi subst.* :

- 10. ... contre les verrières étaient suspendus quelques colliers de grenat, d'ambre, de **baroques**, de corail, etc., objets de négoce de Judas le Lapidaire, ... P. BOREL, *Champavert*, Dina la belle juive, 1833, p. 134.

Le TLFi

- Il est disponible sous format XML avec une DTD associée.
- Accessible sur plusieurs sites tels que :
 - <https://www.cnrtl.fr/definition/>
 - <http://atilf.atilf.fr/>
 - <https://www.le-tresor-de-la-langue.fr/> (!)
- Malgré l'utilisation généralisée, il est difficilement intégrable et interrogeable en TAL

```

1  <?xml version="1.0" encoding="ISO-8859-1"?>
2  <dico>
3    <art id="5485">
4      <ved>
5        <mot>
6          <da>
7            <R>AMBARVALE, AMBARVALES, AMBARVALIES,</R>
8          </da>
9        </mot>
10       <cod>
11         <da>
12           <R>adj. et subst. fém. plur.</R>
13         </da>
14       </cod>
15     </ved>
16     <sync>
17       <H>
18         <parmarge/>
19         <B>
20           <dom>
21             <da>
22               <I>ANTIQ. ROMAINE</I>
23             </da>
24           </dom>
25         </B>
26       <H>
27         <parah>
28           <da>
29             <G>A._</G>
30           </da>
31         </parah>
32       <B>
33         <ind>
34           <da>
35             <I>Emploi adj.</I>
36           </da>
37         </ind>

```

Conversion en Ontolex-Lemon

1. La structure XML du dictionnaire est parcourue et les éléments essentiels en sont extraits
2. L'extraction des données se concentre actuellement uniquement sur
 - les lemmes
 - les catégories grammaticales
 - les sens
 - les définitions
3. Ces informations sont ensuite converties en Ontolex-Lemon (toujours en RDF)



Actuellement, 32 073 entrées du TLFi sont converties en Ontolex-Lemon.

Le TLFi en Ontolex-Lemon

```

1 <https://www.cnrtl.fr/definition/11597> a ontolex:LexicalEntry ;
2   rdfs:label "baroque"@fr ;
3   lexinfo:partOfSpeech lexinfo:adjective ;
4   ontolex:sense <https://www.cnrtl.fr/definition/11597#A._1.UND-1>,
      <https://www.cnrtl.fr/definition/11597#A._1.UND-2>,
      <https://www.cnrtl.fr/definition/11597#A._1.UND-3>,
      <https://www.cnrtl.fr/definition/11597#A._3.UND-6>,
      <https://www.cnrtl.fr/definition/11597#UND-9>.
5   <https://www.cnrtl.fr/definition/11597#A._1.UND-1> skos:definition
      "Qui est caractéristique de la période qui a suivi la
      Renaissance classique."@fr .
6   <https://www.cnrtl.fr/definition/11597#A._1.UND-2> skos:definition
      "Qui est caractéristique de la période musicale propre à
      l'Allemagne, à l'Angleterre, à l'Italie, et qui s'étend de 1580 à
      1760."@fr .
7   <https://www.cnrtl.fr/definition/11597#A._1.UND-3> skos:definition
      "Qui appartient à l'époque littéraire qui, en France, correspond
      aux règnes de Henri IV et Louis XIII."@fr .
8   <https://www.cnrtl.fr/definition/11597#A._3.UND-6> skos:definition
      "Artiste dont le style rappelle cette période"@fr .
9   <https://www.cnrtl.fr/definition/11597#UND-9> skos:definition "Qui
      est de forme irrégulière, d'une rondeur imparfaite"@fr .
  
```



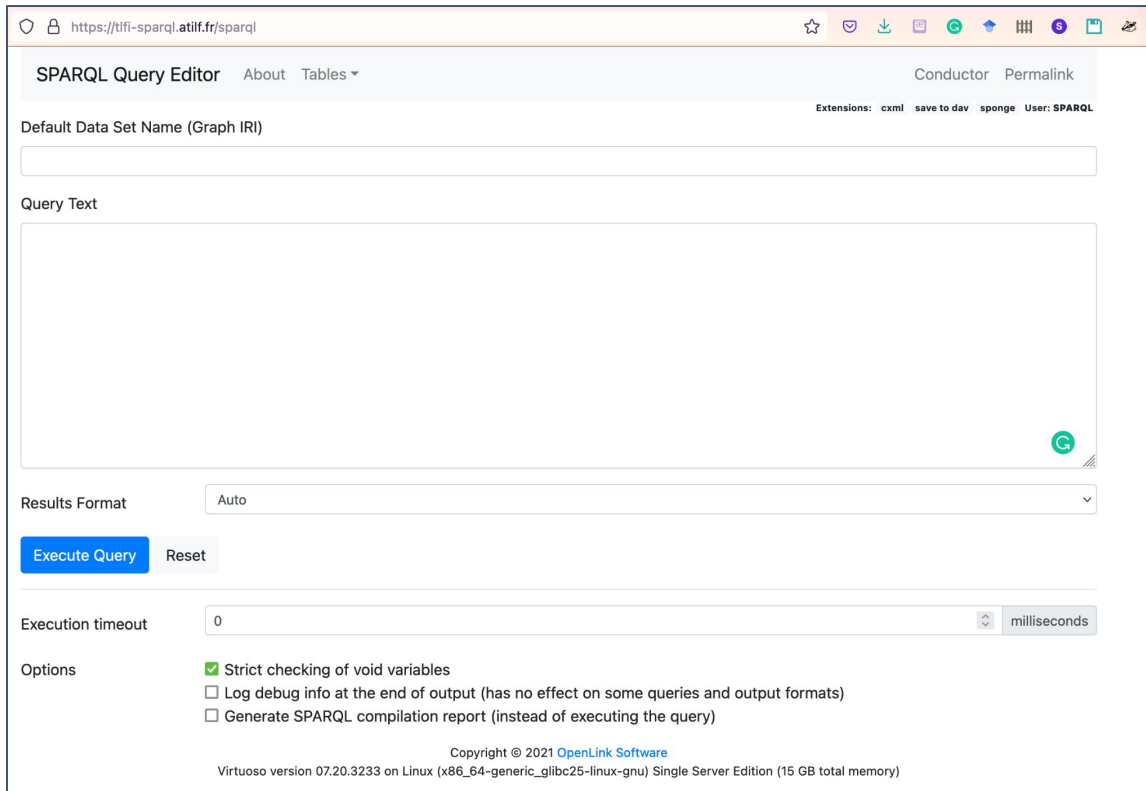
2- Données liées

Quelques requêtes



Le point d'accès SPARQL du TLFi

<https://tlfi-sparql.atilf.fr/sparql>



The screenshot shows a web browser window with the URL <https://tlfi-sparql.atilf.fr/sparql>. The page is titled "SPARQL Query Editor" and includes a navigation bar with "About" and "Tables" links. The main interface features a "Default Data Set Name (Graph IRI)" input field, a large "Query Text" area with a green circular icon in the bottom right corner, and a "Results Format" dropdown menu set to "Auto". Below these are two buttons: "Execute Query" (in blue) and "Reset". Further down is an "Execution timeout" input field set to "0" milliseconds. The "Options" section includes three checkboxes: "Strict checking of void variables" (checked), "Log debug info at the end of output (has no effect on some queries and output formats)", and "Generate SPARQL compilation report (instead of executing the query)". At the bottom, the footer text reads: "Copyright © 2021 OpenLink Software" and "Virtuoso version 07.20.3233 on Linux (x86_64-generic_glibc25-linux-gnu) Single Server Edition (15 GB total memory)".

1- Récupérer toutes les entrées lexicales

prefix ontolex: <http://www.w3.org/ns/lemon/ontolex#>

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

SELECT ?entree

WHERE {

 ?entree a ontolex:LexicalEntry ;

 rdfs:label ?mot .

} **LIMIT 100**

[Lancer la requête](#)

2- Récupérer la partie du discours d'un mot

prefix ontolox: <http://www.w3.org/ns/lemon/ontolox#>

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

prefix lexinfo: <http://www.lexinfo.net/ontology/2.0/lexinfo#>

SELECT ?mot ?etiquette

WHERE {

VALUES ?mot {'ballon'@fr}

 ?entree a ontolox:LexicalEntry , ontolox:Word ;

 rdfs:label ?mot .

 ?entree lexinfo:partOfSpeech ?etiquette .

} **LIMIT 100**

Lancer la requête

3- Compter le nombre de sens par entrée (et trier)

prefix ontalex: <http://www.w3.org/ns/lemon/ontalex#>

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

SELECT ?entree ?mot (**COUNT**(**DISTINCT** ?sense) **AS** ?count)

WHERE {

 ?entree a ontalex:LexicalEntry , ontalex:Word ;

 rdfs:label ?mot ;

 ontalex:sense ?sense .

}

ORDER BY DESC(?count)

[Lancer la requête](#)

4- Récupérer les mots avec POS, sens & définitions

```
prefix ontlex: <http://www.w3.org/ns/lemon/ontlex#>  
prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#>  
prefix skos: <http://www.w3.org/2004/02/skos/core#>  
prefix lexinfo: <http://www.lexinfo.net/ontology/2.0/lexinfo#>
```

```
SELECT * WHERE {  
  ?entree a ontlex:LexicalEntry ;  
    rdfs:label ?mot ;  
    lexinfo:partOfSpeech ?etiquette ;  
    ontlex:sense ?sense .  
  ?sense skos:definition ?def .  
} LIMIT 100
```

[Lancer la requête](#)

5- Les définitions qui contiennent le mot « grec » ou « grecque »

prefix ontollex: <http://www.w3.org/ns/lemon/ontollex#>

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

prefix skos: <http://www.w3.org/2004/02/skos/core#>

prefix lexinfo: <http://www.lexinfo.net/ontology/2.0/lexinfo#>

SELECT * WHERE {

 ?entree a ontollex:LexicalEntry ;

 rdfs:label ?mot ;

 lexinfo:partOfSpeech ?etiquette ;

 ontollex:sense ?sense .

 ?sense skos:definition ?def .

FILTER(REGEX(?def, ".*grec.*") || regex(?def, ".*grecque.*"))

} LIMIT 100

[Lancer la requête](#)

6- Les substantifs masculins finissant par « *tion* », « *ie* », « *ique* », « *aison* » ou « *sion* »

prefix ontolex: <http://www.w3.org/ns/lemon/ontolex#>

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

prefix skos: <http://www.w3.org/2004/02/skos/core#>

prefix lexinfo: <http://www.lexinfo.net/ontology/2.0/lexinfo#>

SELECT * WHERE {

 ?entree a ontolex:LexicalEntry ;

 rdfs:label ?mot ;

 lexinfo:partOfSpeech lexinfo:noun ;

 lexinfo:gender lexinfo:masculine .

FILTER (REGEX(?mot, '^.*(tion|ie|ique|aison|sion)\$'))

} LIMIT 100

[Lancer la requête](#)



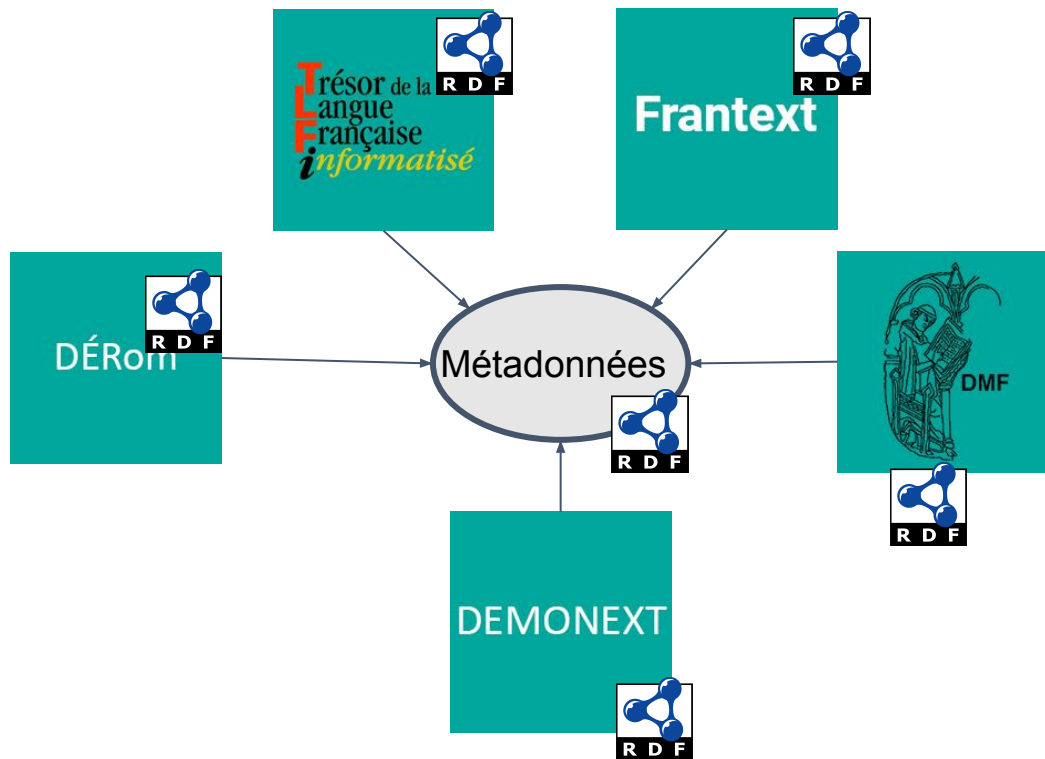
Conclusion et travaux futurs



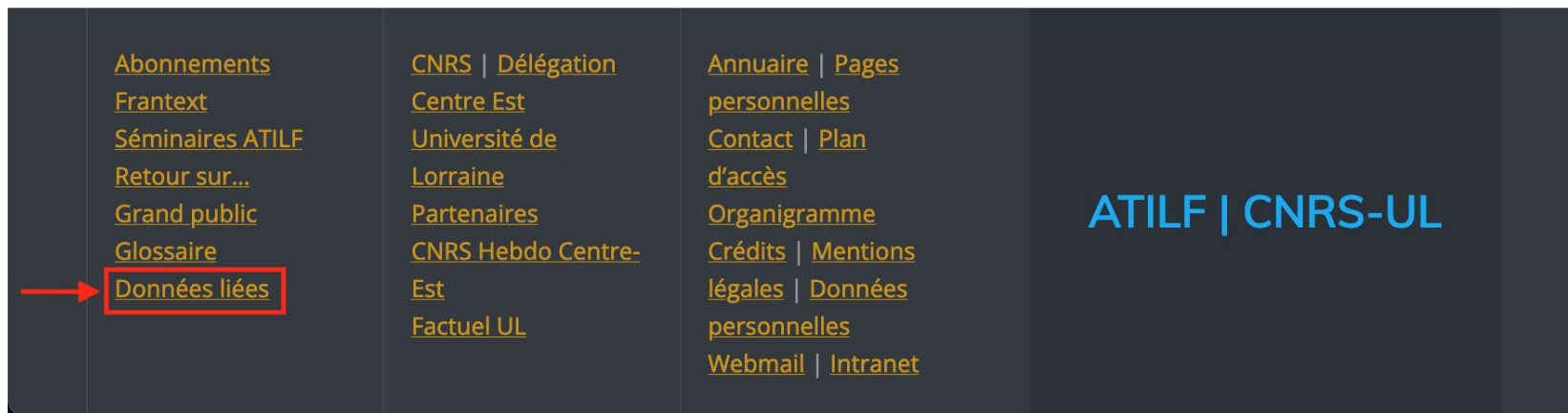
4. Conclusion

- Le Web sémantique est important parce qu'il
 - supprime les documents fastidieux en ajoutant une couche d'application pour avoir accès direct à la connaissance
 - RDF pour décrire les données, SparQL pour les retrouver
 - Les données liés permettent d'augmenter l'interopérabilité entre les ressources aussi
- Un format sémantique des ressources linguistiques à l'ATILF, y compris le TLFi, permet l'utilisation et l'intégration plus généralisée et facile
- Encore au début du chemin !

Travaux futurs : sémantisation des ressources actuelles



Travaux futurs : ATILF et données liées ?



Questions ? 😊

Remerciements

- Merci à Aidan Hogan pour ses diapos sur le Web sémantique (<https://aidanhogan.com/teaching/cc7220-1-2020/>)
- Merci à John McCrae et Thierry Declerck pour leurs contenus sur le Ontolex-Lemon et le LLOD
- Merci à Cyril PESTEL et Etienne PETITJEAN de m'avoir aidé mettre en place le point d'accès SPARQL
- Merci surtout à Mathieu CONSTANT, Bruno GUILLAUME et Karën FORT de m'avoir accueilli
- Et mercis à vous tous et toutes à l'ATILF 