



**Semantic  
Web**

terminologie médicale  
et web sémantique

*une application concrète*

# VERS UN WEB « SÉMANTIQUE »

- ⊙ Le web est une source d'informations incontournable
- ⊙ Pourtant, ces informations (mal structurées) sont difficilement exploitables par les ordinateurs
- ⊙ En effet, elles n'ont pour eux aucun **sens** !

*“To a user, this has become an exciting world, but there is very little machine-readable information there.”*

Tim Berners-Lee, 1994.

- ⊙ La solution serait donc un web *sémantique*

# LES TECHNOLOGIES DU WEB SÉMANTIQUE

- ⊙ Le web sémantique ne s'oppose pas au web classique, mais l'enrichit grâce à de **nouvelles façons de structurer et décrire l'information**
- ⊙ Ses principaux « outils » sont **RDF** (triplets), **SPARQL** (requêtes), **OWL** (ontologies)
- ⊙ Ces outils se complètent et constituent la base des **Linked (Open) Data** ou données liées (libres)
- ⊙ Les données peuvent ainsi être comprises à la fois par des êtres humains et par des machines

# EXEMPLE D'APPLICATION : UNE TERMINOLOGIE MULTILINGUE

N25.9	Disorder resulting from impaired renal tubular function, unspecified	U28	Limited function/disability(U)
D64.9	Anaemia, unspecified	B	Blood,
T78.3	Angioneurotic oedema	A92	Allergy/allergic reaction NC
R45.0	Nervousness	P01	Feeling anxious/nervous/te
R07.2	Precordial pain	K02	Pressure/tightness of hear

- ⊙ Les informations, collectées dans le cadre du projet Meriterm, sont présentées de façon **tabulaire** – idéale notamment pour donner une vue d'ensemble
- ⊙ Mais elles sont difficiles à interpréter, à intégrer dans d'autres modèles, à réutiliser ailleurs, etc.
- ⊙ C'est encore pire dans le cas de pages web, de documents textes ou PDF, etc. !

- ③ **RDF** (Resource Description Framework) propose une structuration de l'information radicalement différente : le **triplet**
- ③ Le *triplet* est une relation atomique (une affirmation) qui unit un **sujet** à un **objet** par un **prédicat** (de façon directionnelle)



# LE TRIPLET

- ◎ *Sujet* : ce à propos de quoi on affirme quelque chose  
*Prédicat* : la « propriété » qui unit sujet et objet  
*Objet* : la valeur attribuée à cette propriété
- ◎ Le triplet se conçoit très bien si l'on se réfère au langage naturel :  
Paris [sujet] est la capitale de [prédicat] la France [objet]
- ◎ L'objet d'un triplet peut évidemment être le sujet d'un autre triplet et les prédicats sont souvent réutilisés :  
France [sujet] estMembreDe [prédicat] UE [objet]  
Espagne [sujet] estMembreDe [prédicat] UE [objet]  
Japon [sujet] estMembreDe [prédicat] ONU [objet]

# TRIPLETS : AVANTAGES, INCONVÉNIENTS

- ⊙ Le triplet est à la fois **simple** et **puissant** ; il facilite considérablement le partage de l'information (sélection de l'information pertinente), y compris entre des domaines de spécialité différents
- ⊙ UK aLangueOfficielle **anglais** (administration)  
**anglais** estDansFamilleLangue germanique (linguistique)  
hamlet estEcritEn **anglais** (littérature)
- ⊙ En revanche, le nombre de triplets nécessaires est généralement élevé (+ de 13 000 dans notre cas)

# CRÉER DES TRIPLETS

- ⊙ Il n'est pas concevable d'écrire tous ces triplets à la main!
- ⊙ On peut toutefois les générer automatiquement à l'aide de programmes informatiques adéquats
- ⊙ L'information est illisible pour un être humain, mais idéale pour un ordinateur (lignes de texte)
- ⊙ (Note : en réalité, les triplets sont « abstraits », mais on peut les représenter selon plusieurs formalismes)

# EXEMPLE D'APPLICATION : NOS TRIPLETS

```

- <owl:Class rdf:about="http://meritem.org/heartfailure/heartfailure.rdf#48"> ← sujet #1
- <Vocabulary:hasLS> ← prédicat #1
- <TMF:LS rdf:about="http://meritem.org/heartfailure/heartfailure.rdf#48_LS_FR"> ← objet #1  sujet #2
- <Vocabulary:hasTS> ← prédicat #2
- <TMF:TS rdf:about="http://meritem.org/heartfailure/heartfailure.rdf#48_TS_FR_PREF"> ← objet #2  sujet #3
  <Vocabulary:isTerm xml:lang="fr">coronarographie</Vocabulary:isTerm>
  <Vocabulary:hasOriginatingInstitution rdf:resource="http://meritem.org/heartfailure/heartfailure.rdf#SSMG"/> ...
  <Vocabulary:hasOriginatingDatabaseName rdf:resource="http://meritem.org/heartfailure/heartfailure.rdf#RBP"/>
  <Vocabulary:hasNormativeAuthorization rdf:resource="http://meritem.org/heartfailure/datacategories.owl#preferredTerm"/>
  <Vocabulary:hasIdentifier>764762014.fr.pref</Vocabulary:hasIdentifier>
  <Vocabulary:hasElementWorkingStatus rdf:resource="http://meritem.org/heartfailure/datacategories.owl#pending"/>
  <Vocabulary:hasDomainExpert rdf:resource="http://meritem.org/heartfailure/heartfailure.rdf#Marc_Jamouille"/>
  </TMF:TS>
  <Vocabulary:hasTS>
  <Vocabulary:hasLanguageIdentifier rdf:resource="http://meritem.org/heartfailure/datacategories.owl#fr"/>
  <Vocabulary:hasElementWorkingStatus rdf:resource="http://meritem.org/heartfailure/datacategories.owl#pending"/>
  </TMF:LS>
</Vocabulary:hasLS>

```

- ③ **SPARQL** (SPARQL Protocol and RDF Query Language) permet d'effectuer des **requêtes** complexes et raffinées sur les triplets et retourne tous les résultats répondant à une ou plusieurs affirmation(s)
  - ③ *?etat* estMembreDe UE +  
*?etat* aLangueOfficielle francais
- => *?etat* est un ensemble correspondant aux réponses "France", "Belgique", "Suisse", etc.

# EXEMPLE D'APPLICATION : REQUÊTES POSSIBLES

- ⊙ Quels sont tous les concepts présents dans UMLS, mais pas dans ICPC2 ?
- ⊙ Quels sont tous les concepts dont le terme standardisé diffère du terme préféré (en anglais) ?

<a href="http://meritem.org/heartfailure/heartfailure.rdf#30">http://meritem.org/heartfailure/heartfailure.rdf#30</a>	Malignant neoplastic disease	Cancer
<a href="http://meritem.org/heartfailure/heartfailure.rdf#169">http://meritem.org/heartfailure/heartfailure.rdf#169</a>	Conduction disorder of the heart	Heart rhythm disorders
<a href="http://meritem.org/heartfailure/heartfailure.rdf#58">http://meritem.org/heartfailure/heartfailure.rdf#58</a>	Precordial pain	chest pain that may correspond to angina

111 résultats

- ⊙ Quel est le concept qui correspond au terme français *anémie* ?
- ⊙ Quel est le code SNOMED-CT du concept dont le code ICD10 est *D64.9* ?

# UNE REQUÊTE SPARQL

```
SELECT ?sujet ?code_uml  
WHERE {  
  ?sujet hf:hasClassificationCode ?classification_code_uml .  
  ?classification_code_uml hf:hasOriginatingDatabaseName hf:UMLS .  
  ?classification_code_uml hf:hasValue ?code_uml .  
  
  OPTIONAL {  
    ?sujet hf:hasClassificationCode ?classification_code_icpc .  
    ?classification_code_icpc hf:hasOriginatingDatabaseName hf:ICPC2  
  }  
  FILTER(! bound(?classification_code_icpc))  
}
```

- ⊙ La permissivité des triplets peut être problématique
- ⊙ On peut établir des règles grâce aux **ontologies**

*“An ontology is an explicit specification of conceptualization.”*

T. Gruber

- ⊙ Si un triplet enfreint une de ces règles, alors l'ontologie n'est plus **cohérente** : il faut la corriger
- ⊙ Les ontologies peuvent être décrites au moyen du langage **OWL** (Web Ontology Language)
- ⊙ Les ontologies bien décrites permettent aussi de **raisonner** automatiquement sur les données, c'est-à-dire d'**inférer** des faits nouveaux

# ORGANISATION D'UNE ONTOLOGIE

- ⊙ Les ontologies sont constituées de propriétés, de **classes** hiérarchisées et d'*instances* qui les composent

## Symptômes

*fièvre*

*toux*

## Maladies

### Maladies neurologiques

*épilepsie*

### Maladies pulmonaires

*asthme*

*pneumonie*

# INTÉRÊT DES ONTOLOGIES

- ◎ OWL nous a permis de représenter efficacement notre **modèle terminologique** : chaque *entrée terminologique* peut comporter plusieurs *sections langues* ; chaque *code de classification* peut avoir au plus une *source*, etc.
- ◎ Ce modèle s'appuie sur des **normes internationales** : TMF (ISO16642) et isoCat (ISO12620)
- ◎ Pour nous,  
chaque entrée terminologique = une classe  
chaque terme = une instance

- ③ Les triplets RDF permettent aussi d'établir des **liens** avec d'autres ressources déjà existantes, ce qui favorise la diffusion des données et augmente leur intérêt
- ③ Ainsi, nous avons lié nos entrées terminologiques à des concepts équivalents dans SNOMED-CT et ICD10

```
<owl:equivalentClass rdf:resource="http://purl.bioontology.org/ontology/SNOMEDCT/33367005"/>
```

- ③ Les ressources produites (triplets, ontologie) sont finalement placées en **accès libre sur Internet** ; chacun peut les réutiliser, effectuer ses propres requêtes, etc.

**MERCI !**