

# XML, langage de description

La vision du Web sémantique

# En guise d'introduction

À la naissance du « **web** »,

Les réseaux étaient d'une lenteur aujourd'hui oubliée,

Ses « **sites** » étaient peu nombreux, détenus seulement par quelques centres de recherche,

Si leurs adresses pouvaient encore être diffusées via quelques hyperliens dans quelques pages d'annuaire statiques, leur multiplication a rendu cette solution impossible.

Dès la conception du Web ses promoteurs se sont préoccupés de créer et de recommander des ***moyens sémantiquement justifiés*** d'accès aux sites recherchés.

Avec le développement des réseaux les sites ont proliféré,

et de façon pragmatique des solutions de moteurs de recherche lexicaux ont traité la question, avec grands bruits et silences.

Faute d'algorithme efficace de construction, les initiatives pour l'indexation sémantique ont tardé à émerger : *elles ont cependant trouvé bien d'autres usages.*

Aujourd'hui avec les progrès des technologies de traitement automatique des langues, des moteurs de recherche lexicaux intègrent progressivement, à leur façon, des catégorisations sémantiques. (tels que Google Knowledge Graphs)

Les trois catégories d'expression des humains  
et de leurs machines :  
*les Messages et les Descriptions et les Autres*

XML pour toutes modalités d'expression  
verbalisées



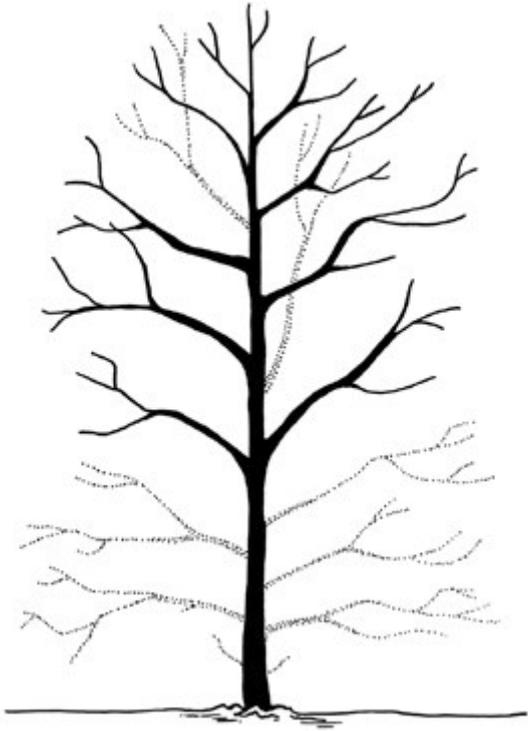
# Ce qu'il est plus rare d'analyser...

- Les Humains échangent trois sortes d'expressions verbalisées:
  - Des **Messages**, **Ordres**, et **Documents**,
    - Structurés, séquentiellement organisés, d'une façon préalablement supposée connue des deux parties:
      - Orateur et auditeur.
      - Écrivain et lecteur.
  - Des **Descriptions**,
    - Qui sont des ensembles d'assertions de **relations** entre des entités.
    - Qui sont non ordonnées, sinon non organisées ;  
par exemple :
      - » Une grammaire est une **description**.
      - » Un schéma est une **description**.
    - Qui sont le moyen par lequel les connaissances sont partagées.
  - les **Autres sortes d'expression verbalisées**
    - Qui restent à découvrir...
- À l'évidence la distinction est complexe, dans la mesure où :
  - Les **Messages**, **ordres**, et **documents** peuvent contenir des **descriptions** .
  - Des **Descriptions** peuvent référer des **messages**, **ordres** et **documents** .
- Les **Messages**, **Ordres**, et **Documents** nécessitent toujours des **Descriptions** initiales ...  
...de la façon dont ils sont structurés, séquentiellement organisés.

# Des arbres et des graphes

- Typiquement...
  - Les **Messages**, **ordres**, et **documents** sont organisés en **arbres** :
    - Les Standards de **documents** tels que GML, SGML et maintenant XML donnent des structures de base pour exprimer des organisations en **arbres**.
  - Les **Descriptions** s'organisent en **graphes** :
    - Les Linguistes, terminologues, ingénieurs des connaissances,
      - Exigent des expressions de **graphes**.
      - Ne sauraient se contenter des structures XML conçues pour des documents .
- Évidemment...
  - Les modèles DTD et XMLschema expriment des **Descriptions** :
    - Unbounded Choices, ID, IDrefs mechanisms permettent des sortes d'expressions de **graphes**.
    - XML Schema, les spécifications XSL sont des descriptions de **graphes**...
  - Mais les structures qu'ils créent sont inefficaces pour l'expression universelle de connaissances.
- Heureusement...
  - Tout **graphe** peut être porté par un **arbre**...

# Diversité naturelle des arbres, expressions de graphes



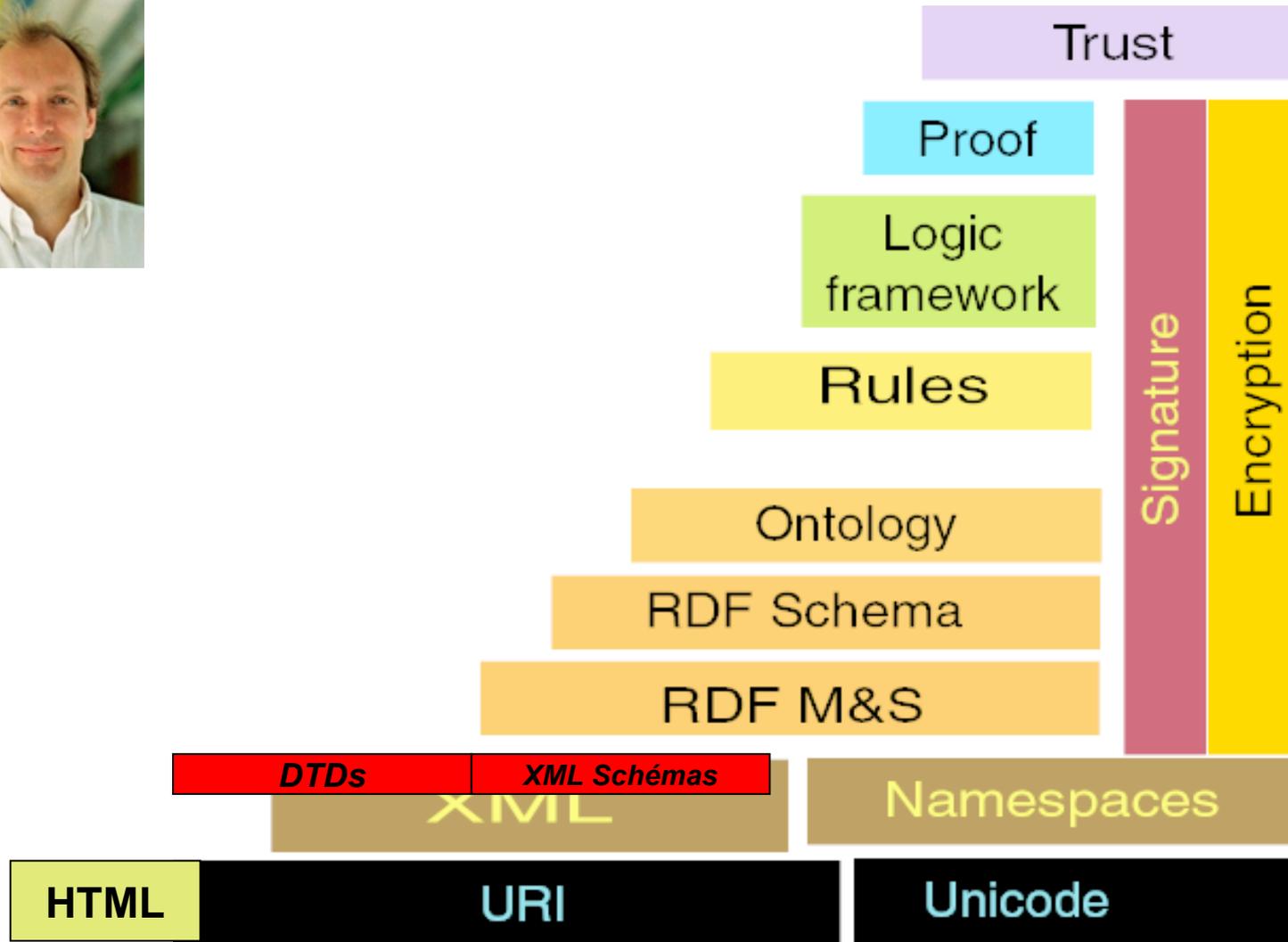
# Du Web Syntaxique au Web Sémantique

Une grande vision d'architecte

# Pyramide à degrés de Djoser, selon Imhotep



# Pyramide à degrés du Web selon Sir Tim Berners Lee



# "Au commencement..."

- ... le monde des ordinateurs était un espace hostile, et l'obscurité couvrait les abysses des systèmes d'exploitation, alors que de méchants vents couraient sur les réseaux propriétaires" ...
- " Alors Tim Berners Lee dit, "*Let there be the Web,*"  
et alors fut le Web... "
- " et le Web n'était qu'un moyen de partager quelques *documents*
  - *publiés en HTML,*
  - *identifiés par des URL* sur un réseau hétérogène d'ordinateurs reliés par des réseaux de réseaux obéissant au *protocole Internet,* via un logiciel *navigateur* "

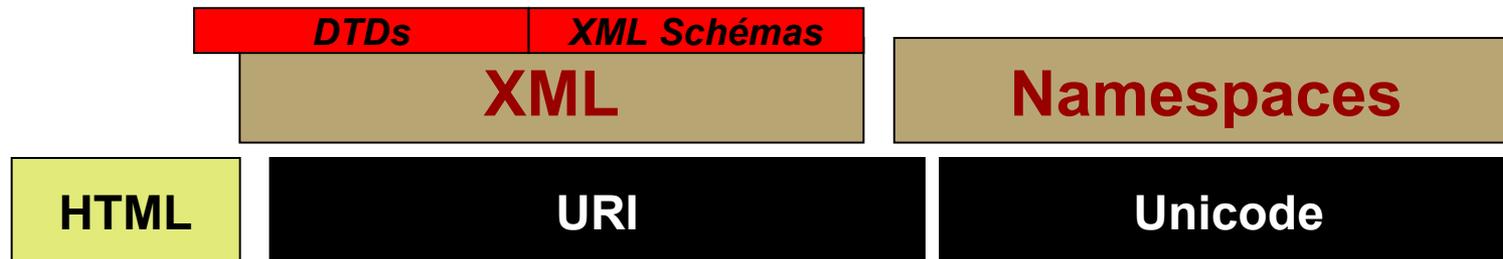
HTML

URI

Unicode

# S'en est suivi la recommandation...

- De principes de structuration adaptés pour des documents de toutes sortes :
  - Organisés par une structure emboîtée de segments, délimités par des balises,
  - Exprimés en méta-langage XML, est adaptation de SGML,
    - dont HTML est un modèle (DTD).



# Un document est un « arbre d'objets typés »

## NOTE D'ETUDE 2EME STADE

E M E LM 99 0102

### Pièce N° 1 Note de synthèse

#### 1.1 origine de la modification

##### 1.1.1 Historique

Depuis sa création, les objectifs en matière de sûreté et de disponibilité du système ABC ont évolué. A ce titre la CCE 3.14159<sup>2</sup> a confié au SEPTEN une étude précisant pour les différents états des tranches les conditions d'utilisation des BZZ. A la suite de l'étude SEPTEN le CTE du 30/02/1995 donne son accord pour la réalisation du xxx de yyy sur zzz

##### 1.1.2 Énoncé du problème - But recherché

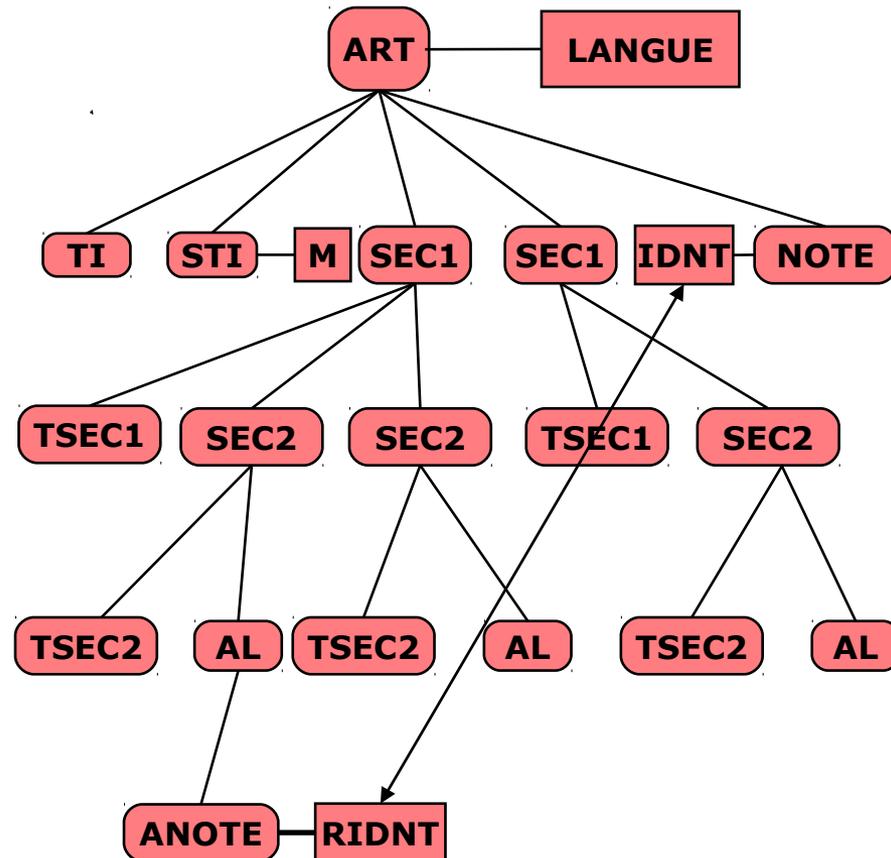
La note d'étude E-XX/FC 2.17 ind. A attire l'attention sur le fait que le temps de blahblah important dû à la complexité du système de verrouillage est paradoxal pour un système de secours. Faisant suite à cette conclusion le SEPTEN (note d'étude E-XX/FC 2.18) ind. A que blahblah...

#### 1.2 Description de la modification

##### 1.2.1 Solution retenue

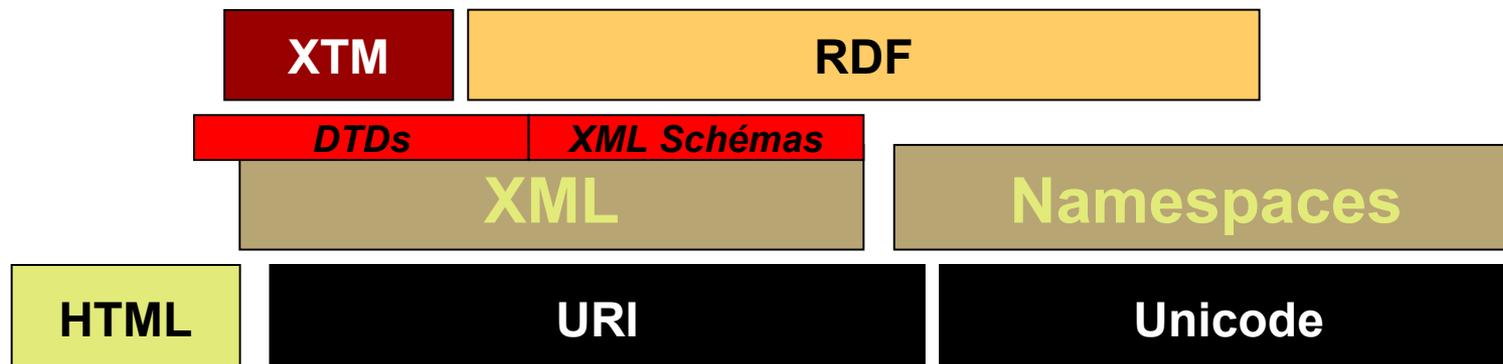
Le XXX des YYY se traduit par la mise en place de ZZZ dans les armoires d'aiguillage permettant de ligner le ABC vers la totalité des tableaux RRR du site

La logique de verrouillage associée au système ABC a pour but de garantir la sécurité du personnel, lors des manœuvres d'exploitation.  
blahblah



# Soyez fertiles et multipliez vous, emplissez la terre et soumettez la... ...ainsi firent les sites web, ils crûrent, se multiplièrent et envahirent le globe

- Chercher, découvrir les informations publiées, faire un index des documents devint un enjeu,
  - auquel ont répondu les moteurs de recherche,
    - Avec beaucoup de **bruit** et aussi de **silence**.
- un index structuré raisonné, formant un réseau de concepts ne se décrit pas par une simple structure arborescente.
  - Mais les **Topic Maps** (norme ISO) et **RDF** (standard W3C) donnent deux approches de représentation de réseaux de concepts exprimables en XML



# la vieille question de topologie de clôtures ?

- Des piquets ou des intervalles, là est la question !  
Une clôture peut être considérée comme,
  - Un réseau de piquets reliés ensemble, ou comme,
  - Des barrières intervalles reliant des piquets.
- RDF et les Topic Maps diffèrent par leur approche :
  - les linguistes préfèrent ordinairement celle des “piquets” (**XTM**).
  - l’Intelligence Artificielle exploite plutôt celle des “intervalles” (**RDF**).



XTM  
Topic Maps

RDF  
Resource Description  
Framework

# Exemple d'index en Topic Maps: un prototype d'indexation de contenu documentaire

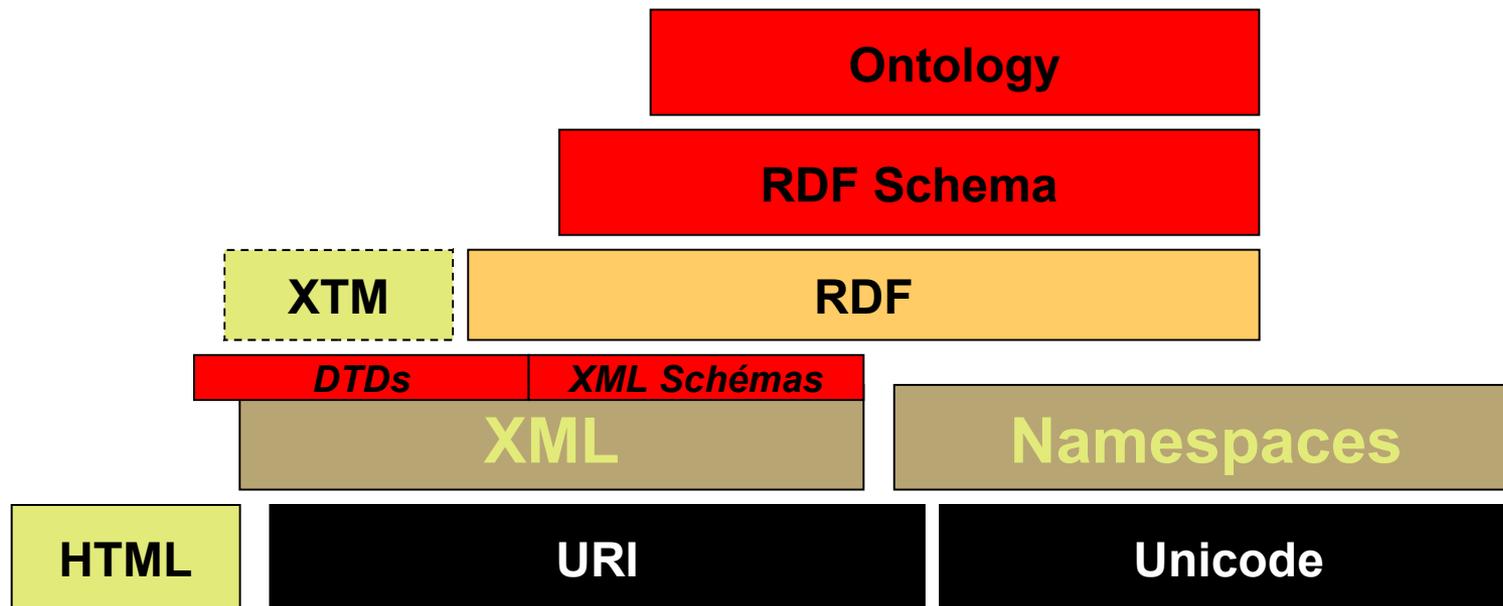
The screenshot displays a software interface with several windows:

- Test XMT:** A hierarchical tree view of documents. The root is 'Dossiers des Systèmes Élémentaires', followed by '1300p4' and 'ASG'. Under 'ASG', there are folders for 'Document N°1' through 'Document N°11' and a folder 'RRI'. The 'ASG 121 VD' folder is expanded, showing a list of chapters from 'CHAPITRE 4' to 'CHAPITRE 14'.
- NaviDoc HTML Viewer:** A window showing a table of contents with 'ASG 121 VD' selected. The main content area displays text:
  - les clapets antiretour (121, 122, 123, 124VD) restent fonctionnels sur les quatre lignes pour éviter d'aggraver une brèche secondaire (diamètre de la tuyauterie ASG : DN100).
  - la ligne alimentant un G.V ne soit pas détériorée par fouetterment ou jet provenant d'une tuyauterie secondaire associée à l'un des trois autres G.V (elle doit rester fonctionnelle). Le maintien de la ligne
- NaviDoc 2D Viewer:** A window showing a detailed technical diagram of a piping system. The diagram includes various components like valves (013LP, 014LP), pumps (031PO), and tanks (041TC). It also shows flow directions and labels like 'RPE' and 'ATMOSP'. A control panel at the bottom right includes buttons for 'Zoom avant', 'Impression', and 'Zoom arrière', along with a timestamp: 'Auteur: ba', 'Date: 08/06/01', 'Heure: 17:58'.



# Deux descriptions structurelles complémentaires

- **XML Schema et DTDs**  
pour la description de structures de documents.
- **RDF Schema et OWL**  
pour la description sémantique de réseaux de connaissances.



# Des Ontologies... ...formelles

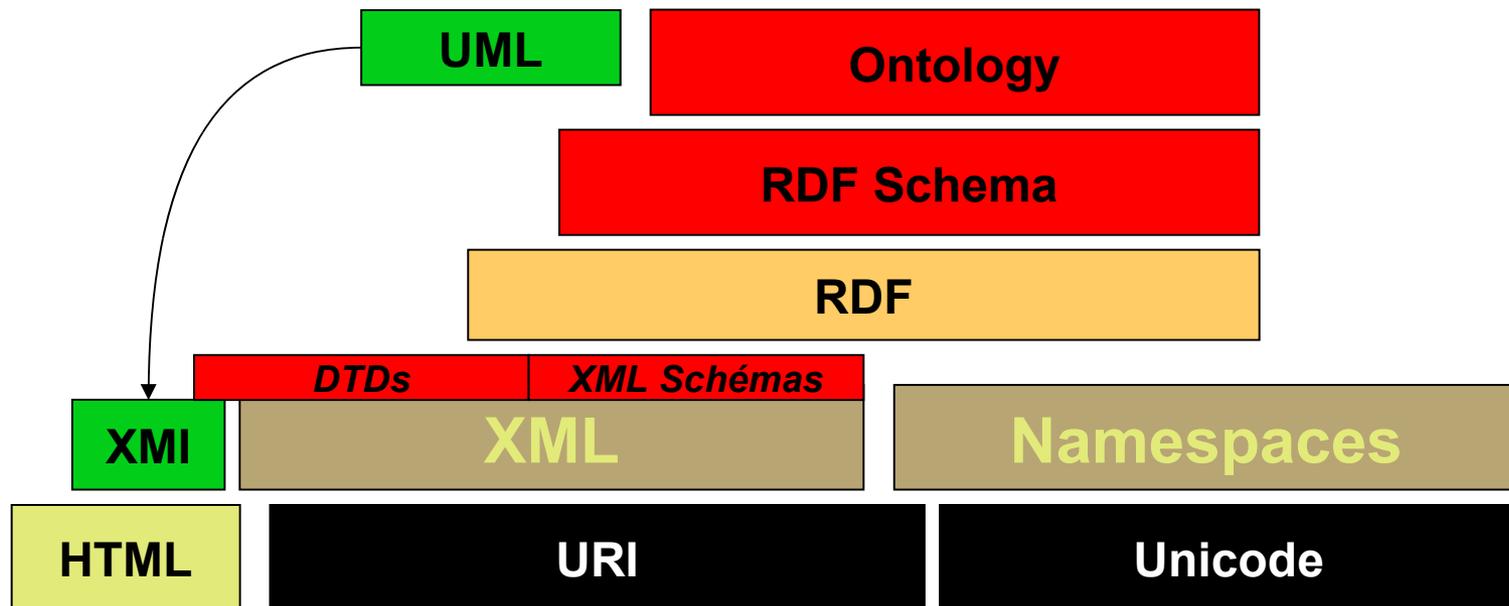
- En philosophie, l'**ontologie** est une branche de la métaphysique.
  - « *Partie de la philosophie qui a pour objet l'être en tant qu'être, qui étudie les propriétés générales de l'être* » (Dictionnaire de l'Académie Française)

## Ontology

- En informatique, une **ontologie formelle** est une expression de la compréhension de réalités, appelée connaissances (...ce qui est la même chose)
  - elle décrit un ensemble structuré de concepts.
  - les concepts sont organisés dans un graphe dont les relations peuvent être :
    - des relations sémantiques ;
    - des relations de composition et d'héritage (au sens objet)
  - les connaissances référant une (ou des) ontologie(s) utilisent sans ambiguïté son (ou leurs) vocabulaire(s).
- En **médecine**, l'**ontologie** s'intéresse à la **genèse des maladies**.

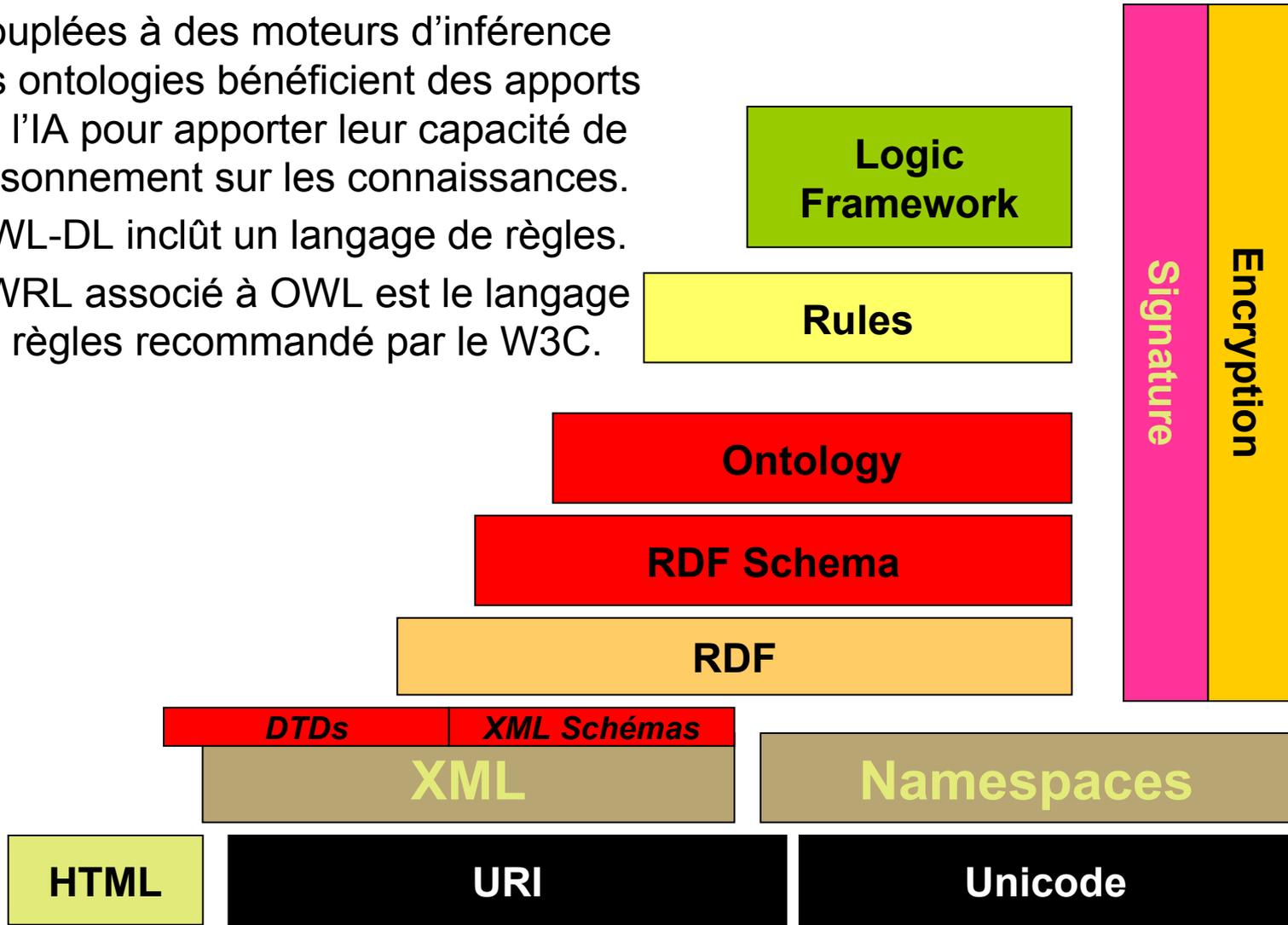
# Situer UML...

- Un diagramme de classes UML est descriptible comme une ontologie non peuplée..
- La spécification du langage XMI de sérialisation en XML de UML le restreint au niveau d'un dialecte XML.

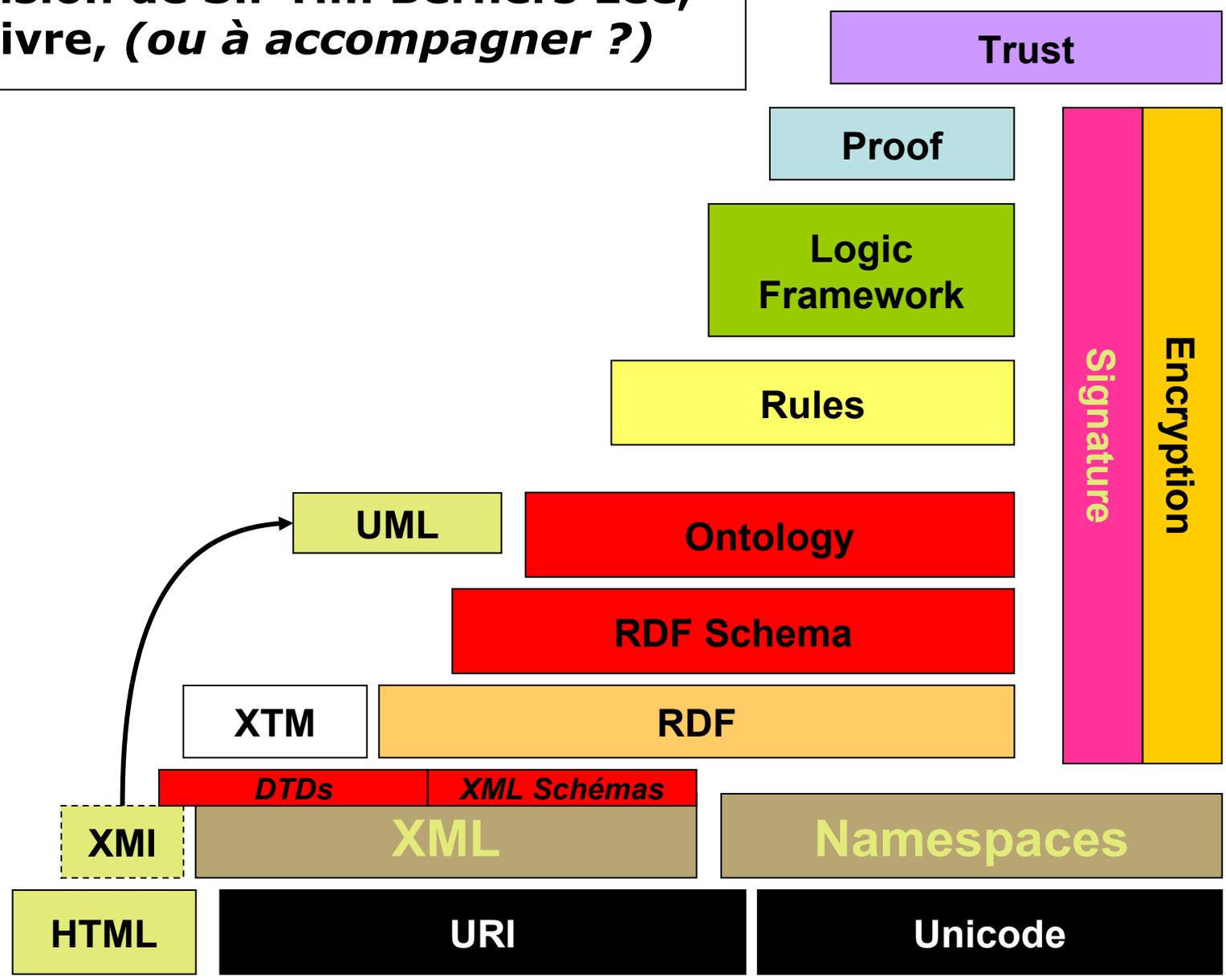


# Le peuplement d'une ontologie bénéficie de la logique de sa description

- Couplées à des moteurs d'inférence les ontologies bénéficient des apports de l'IA pour apporter leur capacité de raisonnement sur les connaissances.
- OWL-DL inclût un langage de règles.
- SWRL associé à OWL est le langage de règles recommandé par le W3C.



# La vision de Sir Tim Berners Lee, à suivre, (ou à accompagner ?)



## Perspective, en guise de conclusion

*” People keep asking what Web 3.0 is.*

*I think maybe when you've got an overlay of scalable vector graphics - everything rippling and folding and looking misty - on Web 2.0 and access to a semantic Web integrated across a huge space of data, you'll have access to an unbelievable data resource. ”*

Tim Berners-Lee,

Fin du module

# En guise de postface

- L'organisation des systèmes d'information est devenue un **sujet de préoccupation**.
- En se popularisant, l'informatique a échappé au contrôle des Directions Informatiques.
  - Informatisé après tous les autres **métiers**, le commerce, en devenant électronique a changé le monde, accéléré son fonctionnement, et accéléré le changement de son fonctionnement.
  - La réduction des constantes de temps élimine toute possibilité pratique de planification périodique de Schémas Directeurs Informatiques
  - Pour des raisons légales comme d'efficacité, toute organisation voit émerger la nécessité nouvelle de formaliser son Système d'Information et d'y asservir ses Systèmes Informatiques,
- L'urbanisation de SI devient une discipline qui progressivement se formalise
    - Avec des méthodes d'analyse cartésienne du SI par niveau de complexité,
    - Avec des **méta modèles**, formalismes graphiques d'expression,
      - **Objectifs, Processus, Procédures, Acteurs, Applications, Infrastructures.**
    - De façon récursive, en intégrant son propre **processus** continu de formalisation.
    - En explicitant les qualités attendues d'un système d'information.
  - Apportant des méthodes formelles de description, le Web Sémantique est éligible pour outiller l'urbanisation des systèmes d'information.