
Web Sémantique & XML

Master IPSS 2021/2022



Web Sémantique & XML

Chapitre 2

Introduction au Web sémantique (Semantic Web)

- ▶ Introduction
- ▶ Technologies du Web sémantique
- ▶ Approche en couches



Introduction

Le Web actuel

Le Web actuel

- ▶ Le **Web** dans une **premier temps** est un web de texte et d'images qui sont très utiles pour les utilisateurs.
- ▶ Les ordinateurs jouent un rôle très limité sur ce type du Web. Ils se contentent:
 - ▶ D'indexer les mots-clés.
 - ▶ De transmettent les informations des serveurs aux clients.
 - ▶ ...
- ▶ **Exemple :**
 - ▶ Horaires de trains et horaires d'avion Sont des documents HTML !!!
 - ▶ Comment croiser les deux documents pour un trajet **train** puis **avion** ?



Introduction

Le Web actuel

► **Exemple :**

- Horaires de trains et horaires d'avion Sont des documents HTML !!!
- Comment croiser les deux documents pour un trajet **train** puis **avion** ?
- Les documents HTML ne peuvent être utilisés car les documents HTML sont une *présentation* des données.

**Tout le travail intelligent (sélection, combinaison, agrégation, etc.)
doit être effectué par le lecteur humain.**



Introduction

Vision du Web sémantique

Vision du Web sémantique

- ▶ La vision générale du «Web sémantique» peut être résumée par :
Rendre le Web plus accessible aux ordinateurs.
- ▶ Rendre le Web **plus riche pour les machines**:
 - ▶ il doit être rempli de données "lisibles" et "compréhensibles" par la machine.
- ▶ Un tel Web permettra des choses impossibles sur le "Web actuel":
 - ▶ La recherche ne se limiterait plus à la recherche par mots-clés. Elle devient plus sémantique:
 - ▶ Inclure la recherche des synonymes,
 - ▶ Eviter les homonymes
 - ▶ Tenir compte le contexte et le but de la requête de recherche.



Introduction

Vision du Web sémantique

- ▶ Les sites Web pourraient devenir plus personnalisés :
 - ▶ si les navigateurs étaient en mesure de comprendre le contenu d'une page Web, et l'adapter à des profils d'intérêts personnels.
- ▶ La liaison (linking) pourrait devenir plus sémantique :
 - ▶ en décidant de manière dynamique quelles pages seraient des destinations utiles en fonction des activités de l'utilisateur actuel, au lieu de fournir toujours les mêmes liens pour tous les utilisateurs.
- ▶ Il serait possible « d'intégrer » des informations en provenance de plusieurs sites Web:
 - ▶ au lieu de faire un «copier-coller» chaque fois qu'on trouve des informations utiles sur un site et qu'on souhaite les combiner avec d'autres informations provenant d'un autre sites.



Introduction

Architecture Web du Web sémantique

Architecture Web du Web sémantique

- ▶ Le contenu Web traditionnel est distribué, à la fois en termes de localisation et de propriété :
 - ▶ les pages Web qui se lient entre elles résident sur des serveurs Web différents :
 - ▶ situés à des emplacements physiques différents
 - ▶ appartiennent à différentes parties.
- ▶ Le fait que "tout le monde peut dire n'importe quoi sur n'importe quoi" est un facteur essentiel de la croissance du Web.
 - ▶ plus précisément, quiconque peut se référer à la page Web de n'importe qui sans avoir à négocier au préalable les autorisations.

Un mécanisme similaire est mis en œuvre dans le Web sémantique.

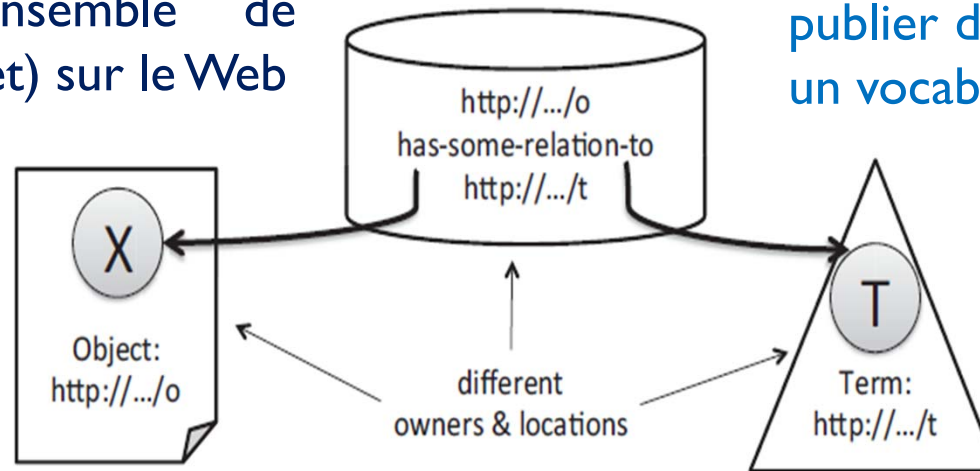


Introduction

Architecture Web du Web sémantique

une première partie peut publier un ensemble de données (DataSet) sur le Web

une seconde partie peut publier de façon indépendante un vocabulaire de termes



un tiers peut décider d'annoter les objets de la première partie avec les termes publiés par la deuxième partie, sans demander l'autorisation de l'une d'entre elles et sans même que l'un ou l'autre en soit informé.

C'est ce découplage qui constitue l'essentiel de la nature Web du Web Sémantique.

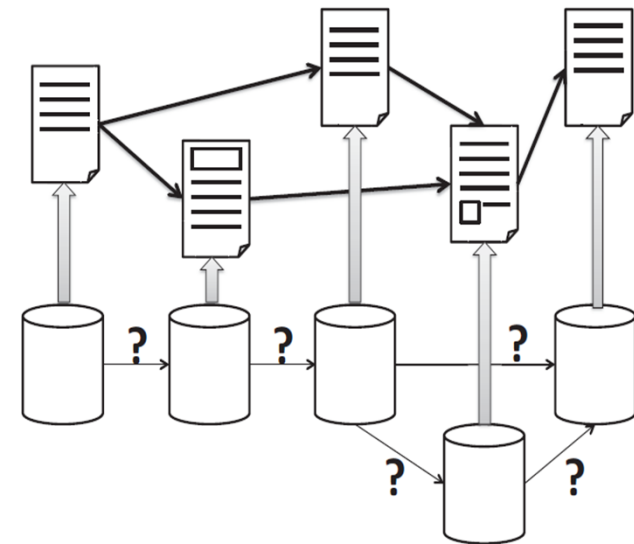


Introduction

Décisions de conception pour le WS

Décisions de conception pour le WS

- ▶ le «web de textes et d'images», non structuré, repose en réalité sur une très grande quantité de données structurées et semi-structurées.
- ▶ La grande majorité du contenu Web est générée à partir de bases de données contenant des ensembles de données soigneusement structurés.
- ▶ Cependant, cette structure est presque complètement perdue lors de la publication de ces données structurées sous forme de pages HTML lisibles par un humain.





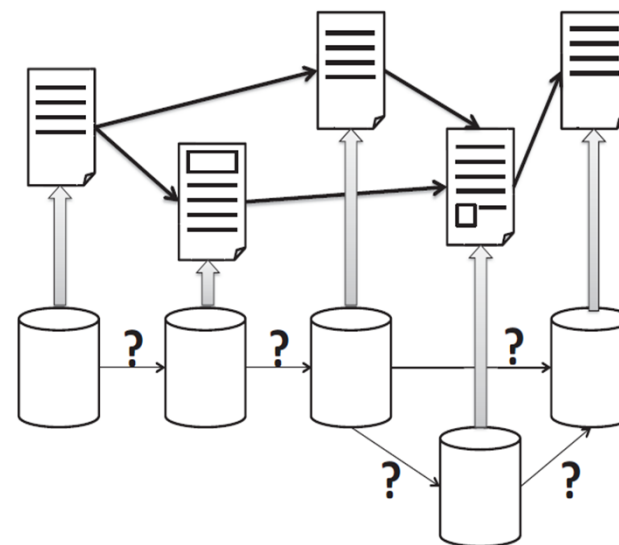
Introduction

Décisions de conception pour le WS

- **Une idée clé** pour faire de grands progrès vers la vision d'un Web plus sémantique:



pouvoir publier et interconnecter les ensembles de données structurés au lieu de simplement publier et interconnecter les pages HTML après la perte d'une grande partie de la structure.





Introduction

Du Web vers le Web sémantique

Du Web vers le Web sémantique

- ▶ Le Web sémantique suit différents principes de conception, :
 1. Rendre les données structurées et semi-structurées disponibles dans des formats normalisés sur le Web;
 2. Rendre non seulement l'ensemble de données, mais également les éléments de données individuels et leurs relations accessibles sur le Web;
 3. Décrire la sémantique voulue de ces données dans un formalisme, afin que cette sémantique puisse être traitée par des machines.



Introduction

Du Web vers le Web sémantique

- ▶ Certaines étapes importantes sont nécessaires pour que la vision du web sémantique devient une réalité:
 1. Se mettre d'accord sur une syntaxe standard pour représenter les données et les métadonnées.
 2. Avoir un accord sur les vocabulaires de métadonnées pour pouvoir partager la sémantique souhaitée des données.
 3. Publier les données dans les formats de l'étape 1, en utilisant les vocabulaires de l'étape 2.



Introduction

Technologie de base pour le WS

Technologie de base pour le WS

- ▶ La mise en œuvre des trois principes de conception suscite leurs traductions vers les technologies actuelles:
 1. Utilisez des **graphes étiquetés** comme modèle de données pour représenter les objets et leurs relations. Le modèle «**Resource Description Framework**» **RDF** est utilisé comme formalisme.
 2. Utilisez des **identificateurs Web** (Internationalized Resource Identifier – **IRI**) pour identifier les éléments de données individuels et leurs relations.
 3. Utiliser des **ontologies** comme modèle de données pour représenter formellement la sémantique voulue des données. **RDF Schema** et **OWL** (Web Ontology Language) sont utilisés à cette fin.



Technologies du Web sémantique

Métadonnées

Métadonnées

- Soit le code HTML suivant :

```
<h1> Centre de physiothérapie Agilitas </h1>
```

```
Bienvenue à la page d'accueil du centre de physiothérapie Agilitas.  
Avez-vous des douleurs? Laissez notre personnel Lisa Davenport,  
Kelly Townsend (notre charmante secrétaire) et Steve Matthews  
prendre soin de vous.
```

```
<h2>horaire des consultations</h2>
```

```
Lun 11am - 7pm<br> Mar 11am - 7pm<br>
```

```
Mer 3pm - 7pm<br> Jeu 11am - 7pm<br>
```

```
Ven 11am - 3pm<p>
```

```
Veillez noter que nous n'avons pas des consultations pendant les  
semaines du <a href="...">State of Origin</a> games.
```



Technologies du Web sémantique

Métadonnées

- ▶ Pour les personnes, les informations sont présentées de manière satisfaisante.
- ▶ Mais les machines auront des problèmes :
 - ▶ Une machine pourrait être capable d'identifier le personnel du centre. mais il aura du mal à distinguer les thérapeutes du secrétaire.
 - ▶ Une machine aura du mal à trouver les heures de consultation exactes (pour lesquelles il faudrait suivre le lien).



Technologies du Web sémantique

Métadonnées

- ▶ Pour résoudre ces problèmes, l'approche du WS:
 - ▶ Propose d'attaquer le problème du côté de la page Web. Si HTML est remplacé par des langues plus appropriées, les pages Web peuvent porter leur contenu sur leur manche.
 - ▶ En plus de contenir des informations de formatage, ceux-ci pourraient également contenir des informations sur leur contenu.



Technologies du Web sémantique

Métadonnées

- ▶ La première étape dans cette direction est le langage **XML (eXtensible Markup Language)**.

```
<société>  
  <treatmentProposé>Physiotherapy</treatmentProposé>  
  <nonSociété> Centre de physiothérapie Agilitas </nonSociété>  
  <personnel>  
    <thérapeute>Lisa Davenport</thérapeute>  
    <thérapeute>Steve Matthews</thérapeute>  
    <secrétaire>Kelly Townsend</secrétaire>  
  </personnel >  
</société >
```

- ▶ Cette représentation est beaucoup plus facile à traiter par des machines.



Technologies du Web sémantique

Métadonnées

- ▶ Cependant, XML reste au niveau syntaxique, car il décrit la structure de l'information, mais pas son sens ou sa signification.
- ▶ Le "langage" de base du Web sémantique est RDF. C'est un "langage" qui permet des déclarations sur des informations.
- ▶ Dans l'exemple précédent, ces déclarations seront sous forme:

```
Société A  offre physiothérapie.  
Le nom de A  est "Centre de physiothérapie Agilitas".  
Lisa Davenport  est une  thérapeute.  
Lisa Davenport  travaille pour  A.  
...
```
- ▶ Les **métadonnées** fait référence à des données sur des données.
- ▶ Les métadonnées saisissent une partie de la signification des données, d'où le terme sémantique dans Web sémantique.



Technologies du Web sémantique

Ontologies

Ontologies

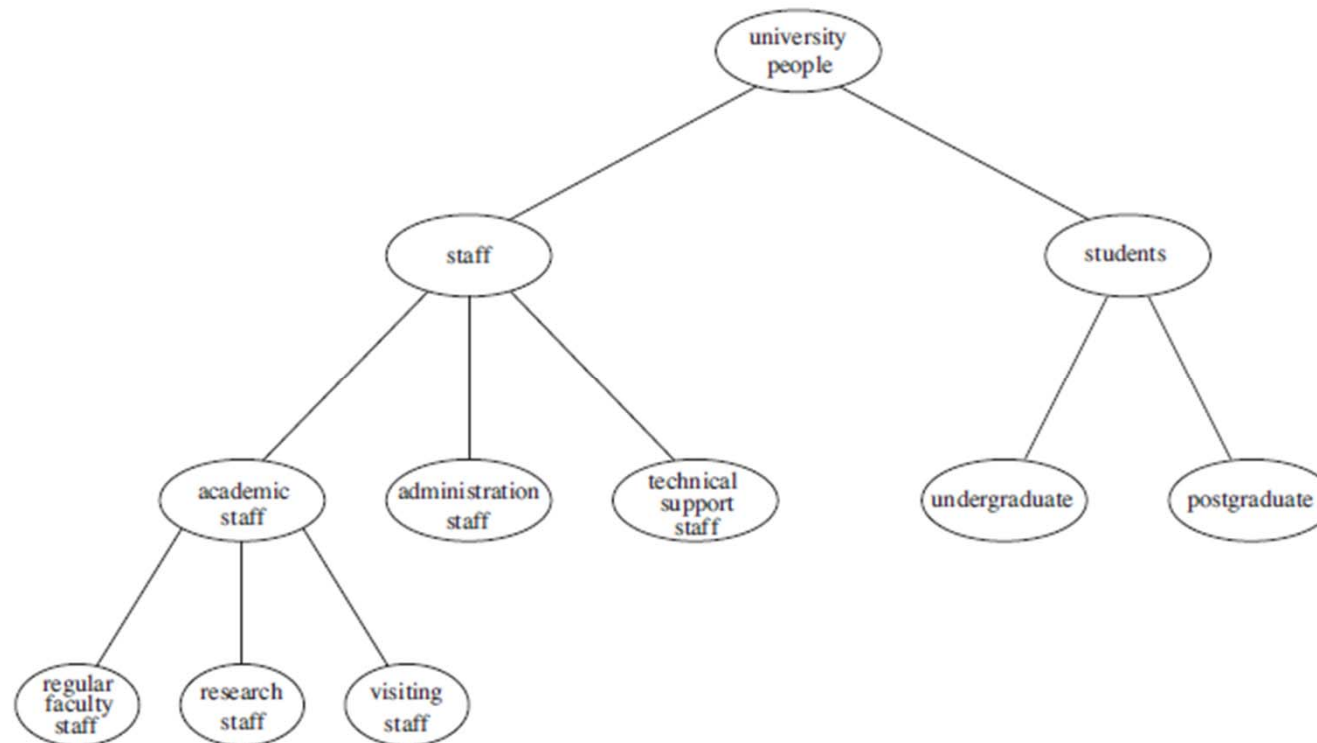
- ▶ Le terme **ontologie** provient à l'origine de la philosophie.
C'est parmi les mots détournés par l'informatique et se voit attribuer un sens technique assez différent de celui d'origine.
- ▶ En informatique, une première définition :
« **Une ontologie est une spécification explicite et formelle d'une conceptualisation** ».
- ▶ En général, une ontologie **décrit formellement** un domaine de discours.
Elle consiste en une **liste finie de termes** et des **relations entre ces termes**.
- ▶ Les relations incluent généralement des hiérarchies de classes.



Technologies du Web sémantique

Ontologies

- ▶ Par exemple:
 - ▶ dans une université, les membres du personnel, les étudiants, les cours, les amphithéâtres et les disciplines sont des concepts importants.
 - ▶ tous les professeurs sont des membres du personnel.





Technologies du Web sémantique

Ontologies

- ▶ Outre les relations de hiérarchie, les ontologies peuvent inclure des informations telles que :
 - ▶ Propriétés
(X enseigne Y),
 - ▶ Restrictions de valeur
(seuls les membres du corps enseignant peuvent enseigner des cours)
 - ▶ Déclarations de disjonction
(le corps professoral et le personnel d'administration sont en général disjoints)
 - ▶ Spécifications des relations logiques entre les objets
(chaque département doit inclure au moins dix membres du corps professoral).



Technologies du Web sémantique

Ontologies

- ▶ Parmi les langages d'ontologie les plus importants pour le WS on trouve:
 - ▶ **RDF Schema :**
 - ▶ Décrire les propriétés et les classes par des ressources RDF,
 - ▶ Définir les hiérarchies existantes de ces propriétés et de ces classes.
 - ▶ Exiger le domaine (Domaine) et la plage(Range) des propriétés.
 - ▶ **OWL :**
 - ▶ Plus riche que RDF Schema.
 - ▶ De plus, OWL permet de définir d'autres relations entre les classes (par exemple, la disjonction), les cardinalités (par exemple, «exactement un»), l'égalité, le typage plus riche de propriétés, les caractéristiques de propriétés (par exemple, la symétrie).



Technologies du Web sémantique

Logique

Logique

- ▶ En général, nous utilisons la logique pour exprimer les règles d'inférences.
- ▶ L'inférence est le fait pour une application informatique, de prendre des décisions ou de déduire de nouvelles informations à partir d'informations et de ressources déjà existantes sur le web.
- ▶ Les règles désignent l'ensemble des assertions déduites de l'aspect sémantique des ontologies.



Technologies du Web sémantique

Logique

- ▶ Par exemple :
 - ▶ supposons que nous sachions que tous les professeurs sont des membres du corps professoral,
 - ▶ que tous les membres du corps professoral sont des membres du personnel et que Michael est un professeur.

Dans la logique des prédicats :

```
prof(X)      --->  faculty(X)
faculty(X)   --->  staff(X)
prof(michael)
```

On peut alors en déduire :

```
faculty(michael)
staff(michael)
prof(X) ---> staff(X)
```




Technologies du Web sémantique

Logique

- ▶ Ainsi, la logique peut être utilisée pour découvrir des connaissances ontologiques.
- ▶ Elle peut également aider à découvrir des relations et des incohérences inattendues.
- ▶ Plusieurs langages ont été définis à ce niveau : **Langage d'inférence** et **Langages de règles**.
- ▶ De plus, la logique est plus générale que les ontologies.
- ▶ Elle peut également être utilisée pour prendre des décisions et sélectionner des actions.
 - ▶ Par exemple, un agent de magasin peut décider d'accorder une réduction à un client fidèle en fonction de la règle
`loyalCustomer(X) ---> discount(X, 5%)`
- ▶ Elle peut fournir des explications aux conclusions et prouver la pertinence de l'information retournée .



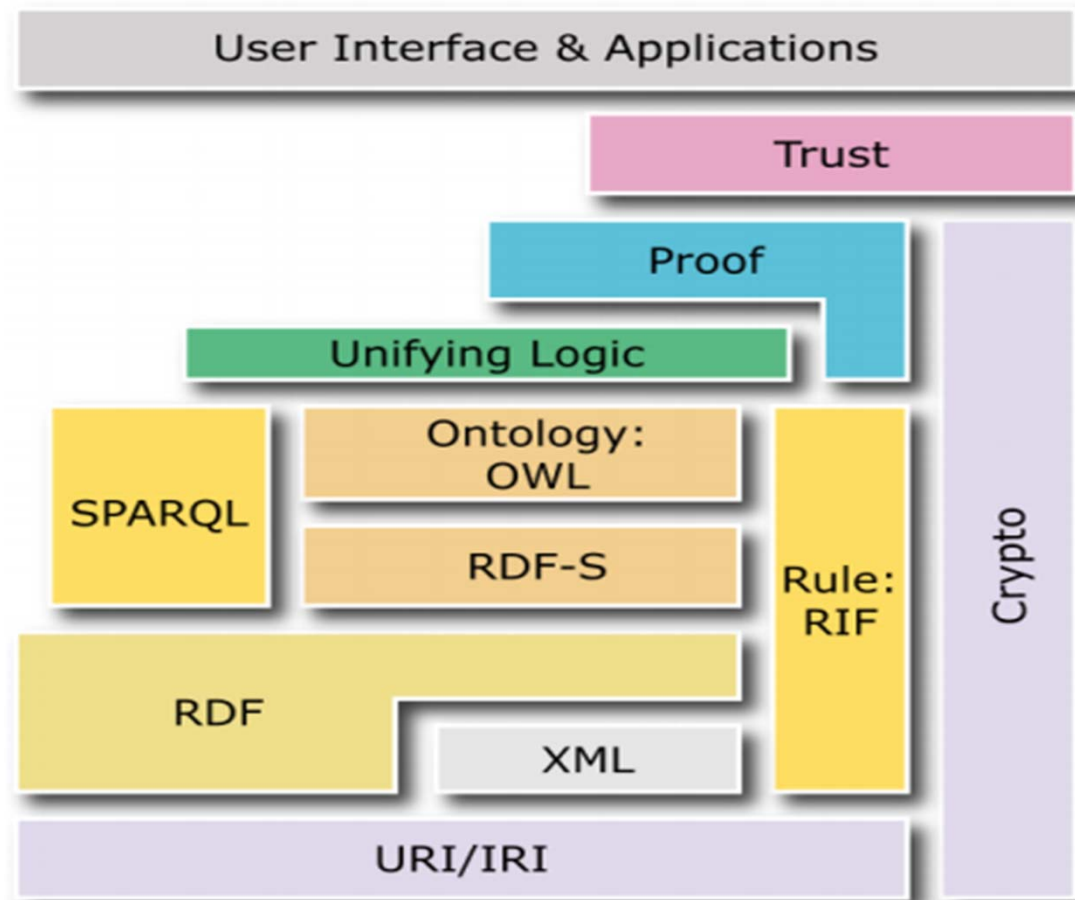
Approche en couches

- ▶ Le développement du Web sémantique se fait par étapes, chaque étape créant une couche superposée à une autre couche.
- ▶ Pour construire une couche au-dessus d'une autre, il convient de suivre deux principes:
 - ▶ **Compatibilité descendante.**
Les agents connaissant parfaitement une couche doivent également pouvoir interpréter et utiliser les informations écrites à des niveaux inférieurs.
Par exemple, les agents conscients de la sémantique de OWL peuvent tirer pleinement parti des informations écrites dans RDF et RDF Schema.
 - ▶ **Compréhension partielle ascendante.**
Les agents pleinement conscients d'une couche soient en mesure de tirer parti au moins partiellement des informations situées à des niveaux plus élevés.
Par exemple, un agent connaissant uniquement les sémantiques de RDF Schema et RDF peut interpréter en partie les connaissances écrites en OWL, en ignorant les éléments qui vont au-delà de schéma RDF et RDF.



Approche en couches

- Les principales couches de la conception de la vision du Web sémantique sont matérialisées sous forme d'un ensemble de standards:





Approche en couches

Interaction:

intégrer les standards du Web sémantique avec les autres standards du Web

Confiance:

permet de suivre les données, leur provenance,...

Raisonnement:

RDF Schema & OWL

Interrogation:

SPARQL: Requêtes sur les Graphe RDF,

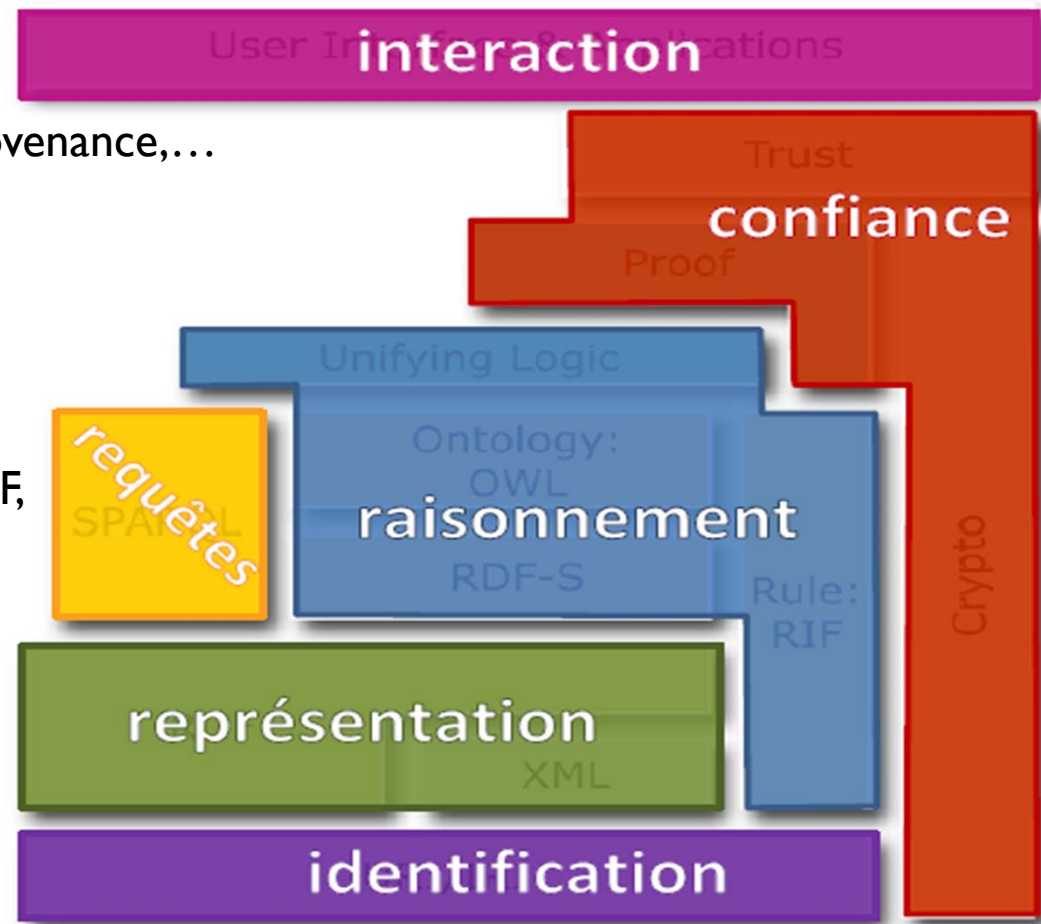
Représentation:

modèle de données : RDF.

Syntaxes : XML, Turtle, N-Triples, ...

Identification :

URI (ou IRI).





Chapitres

- ▶ **Chapitre 1**
XML
- ▶ **Chapitre 2**
Introduction au Web sémantique (Semantic Web)
- ▶ **Chapitre 3**
RDF & Syntaxes RDF
- ▶ **Chapitre 4**
SPARQL
- ▶ **Chapitre 5**
RDF Schema (RDFS) & OWL