

Introduction au web sémantique

Michel Gagnon

École polytechnique de montréal

Plan de la présentation

- Définition et description générale
- Les différentes couches du web sémantique
- Exemples d'application
- Conclusion

Plan

- Définition et description générale
- Les différentes couches du web sémantique
- Exemples d'application
- Conclusion

Définition

"The Semantic Web is an **extension** of the current web in which information is given **well-defined meaning**, better enabling computers and people to work in cooperation."

-- Tim Berners-Lee, James Hendler, Ora Lassila, The Semantic Web, *Scientific American*, Mai 2001.

Le web actuel

- Ensemble de documents

Le web sémantique

Le web actuel

- Ensemble de documents

Le web sémantique

- Ensemble de connaissances

Le web actuel

- Ensemble de documents
- Basé essentiellement sur HTML

Le web sémantique

- Ensemble de connaissances

Le web actuel

- Ensemble de documents
- Basé essentiellement sur HTML

Le web sémantique

- Ensemble de connaissances
- Basé sur XML et RDF(S)

Le web actuel

- Ensemble de documents
- Basé essentiellement sur HTML
- Recherche par mots clé

Le web sémantique

- Ensemble de connaissances
- Basé sur XML et RDF(S)

Le web actuel

- Ensemble de documents
- Basé essentiellement sur HTML
- Recherche par mots clé

Le web sémantique

- Ensemble de connaissances
- Basé sur XML et RDF(S)
- Recherche par concepts

Le web actuel

- Ensemble de documents
- Basé essentiellement sur HTML
- Recherche par mots clé
- Utilisable par l'humain

Le web sémantique

- Ensemble de connaissances
- Basé sur XML et RDF(S)
- Recherche par concepts

Le web actuel

- Ensemble de documents
- Basé essentiellement sur HTML
- Recherche par mots clé
- Utilisable par l'humain

Le web sémantique

- Ensemble de connaissances
- Basé sur XML et RDF(S)
- Recherche par concepts
- Utilisable par la machine

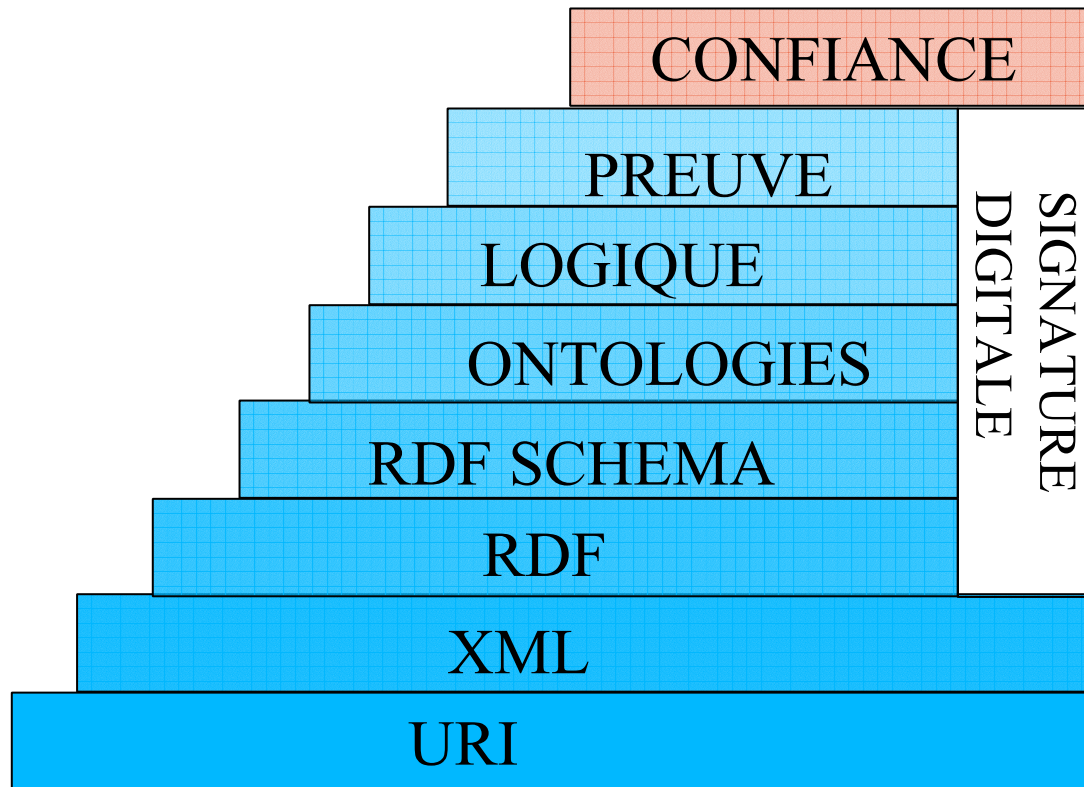
Technologies du web sémantique

- Métadonnées
- Ontologies
- Moteurs d'inférence
- Agents

Plan

- Définition et description générale
- **Les différentes couches du web sémantique**
- Exemples d'application
- Conclusion

Les couches du web sémantique



URI

(universal resource identifier)

Identifie une entité à laquelle on réfère sur le web.

Cette entité n'est pas nécessairement une ressource disponible sur le web.

Exemple:

<http://www.polymtl.ca/professeurs#michelgagnon>

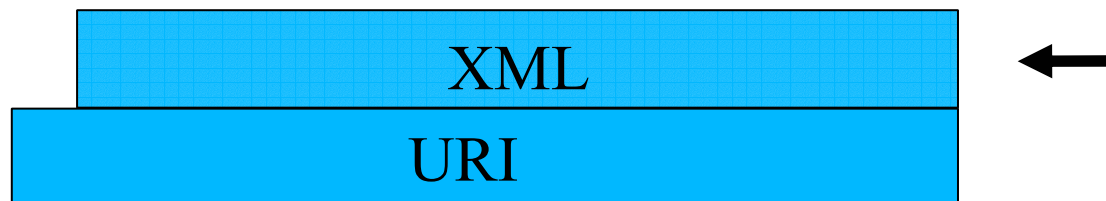
URI (Uniform Resource Identifier)

XML

Langage de balisage extensible

En fait, il s'agit plutôt d'un métalangage qui nous permet de définir nos propres balises pour nos documents.

Norme de facto pour la représentation d'information structurée.



XML - Exemple

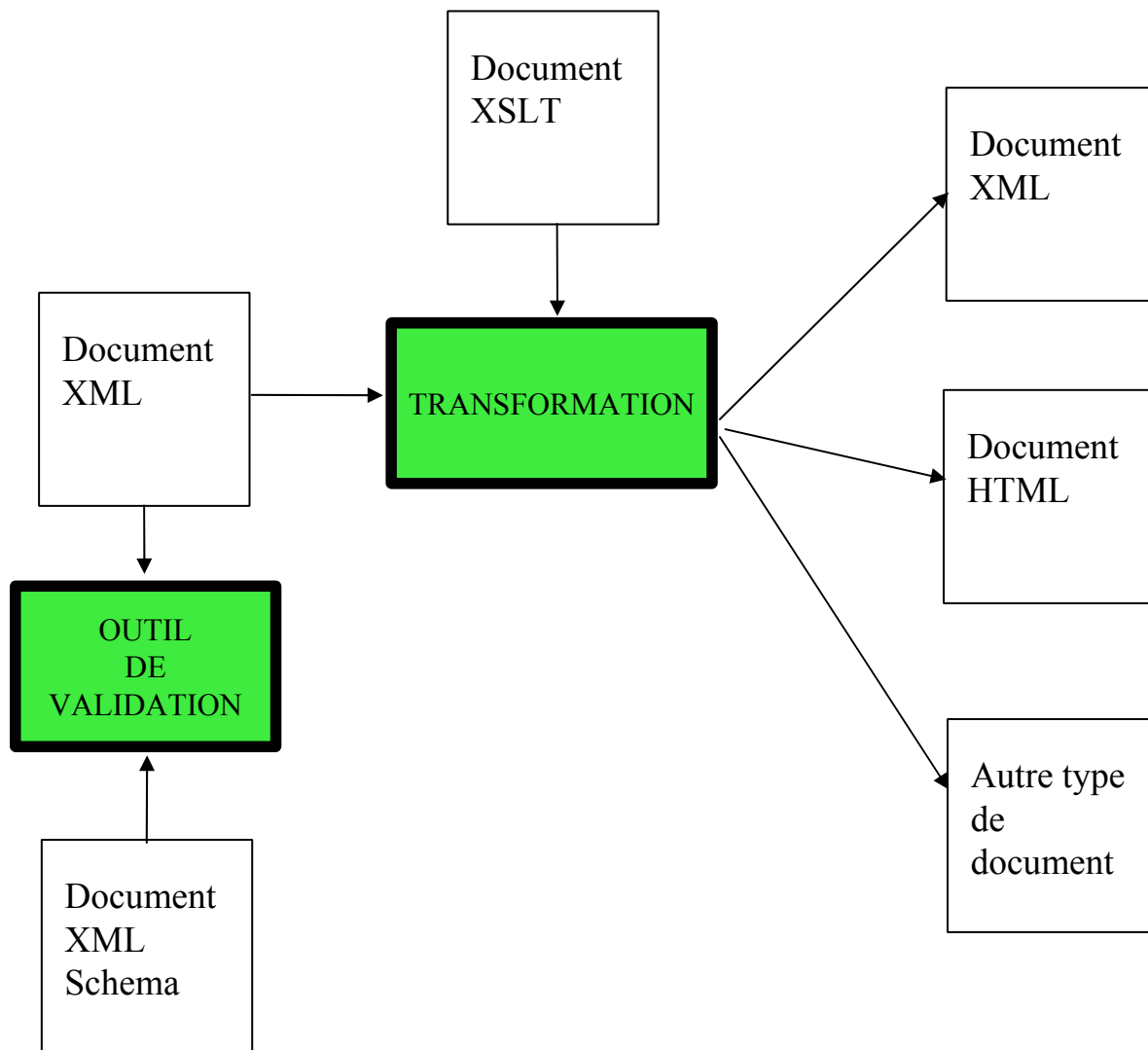
```
<livre>  
  <language> es </language>  
  <title> Cronicas de Bustos Domecq </title>  
  <auteur> Jorge Luis Borges </auteur>  
  <auteur> Adolfo Bioy Casares </auteur>  
  <publisher> Editorial Losada </publisher>  
  <date> 1967 </date>  
  <ISBN> 0525475486 </ISBN>  
</livre>
```

XML – Exemple avec espace de nommage

```
<mg:livre
  xmlns:bib="http://www.exemple.org/bib"
  xmlns:mg="http://www.polymtl.ca/michelgagnon/"
  xmlns:dc="http://purl.org/dc/elements/1.1">
  <dc:language> es </dc:language>
  <dc:title> Cronicas de Bustos Domecq </dc:title>
  <mg:auteur> Jorge Luis Borges </mg:auteur>
  <mg:auteur> Adolfo Bioy Casares </mg:auteur>
  <dc:publisher> Editorial Losada </dc:publisher>
  <dc:date> 1967 </dc:date>
  <bib:ISBN> 0525475486 </bib:ISBN>
</mg:livre>
```

XML - Technologies

- Grande quantité de parsers disponibles
- DTD, XML Schema
- Langages de transformation XSL (XSLT et XSL-FO)
- Xlink et XPointer, pour référer à une section précise d'un document

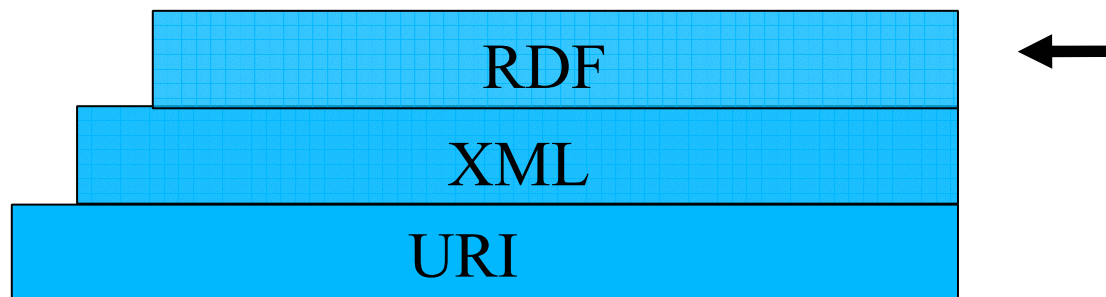


RDF

Modèle de données pour décrire des ressources du web

Graphe:

- les noeuds représentent des ressources
- les arcs représentent des relations entre ces ressources

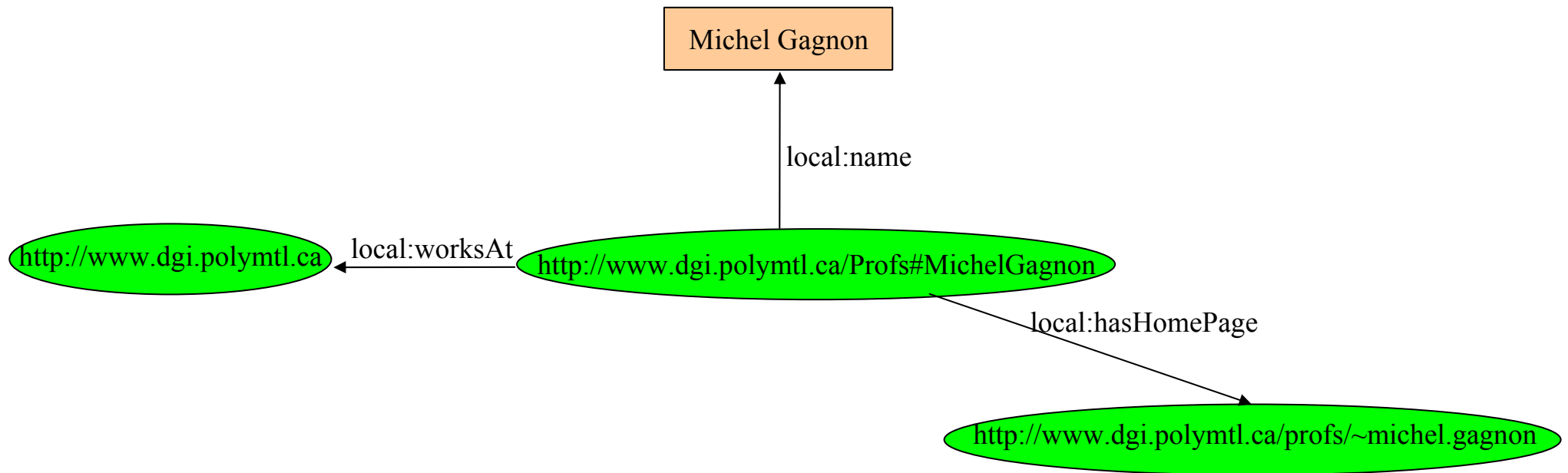


RDF

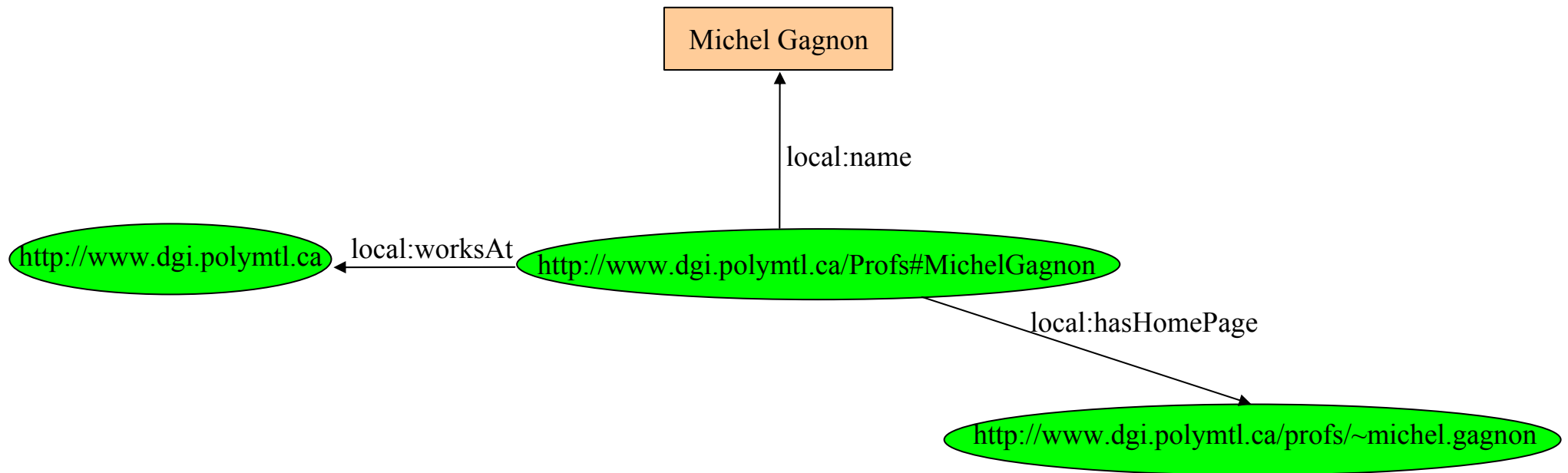
(Resource Description Framework)

- Le graphe est représenté par un ensemble d'énoncés (statements)
- Un énoncé est un triplet $\langle S, P, O \rangle$, où
 - S est le sujet
 - P est le prédicat (une propriété)
 - O est l'objet (la valeur de la propriété pour le sujet en question)

RDF - Exemple

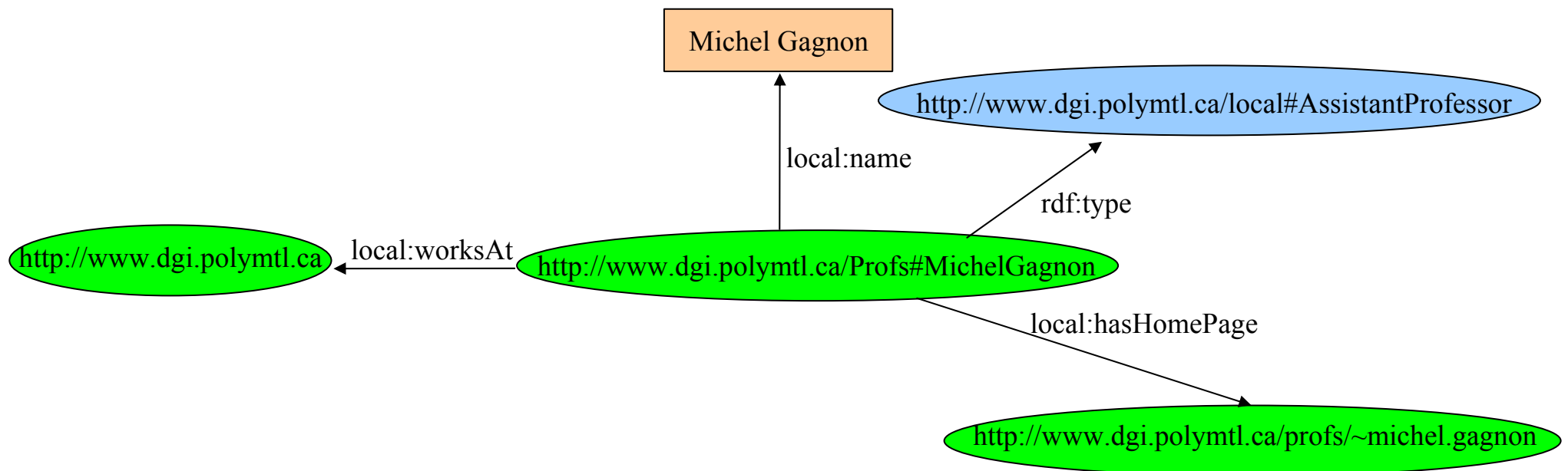


RDF - Exemple



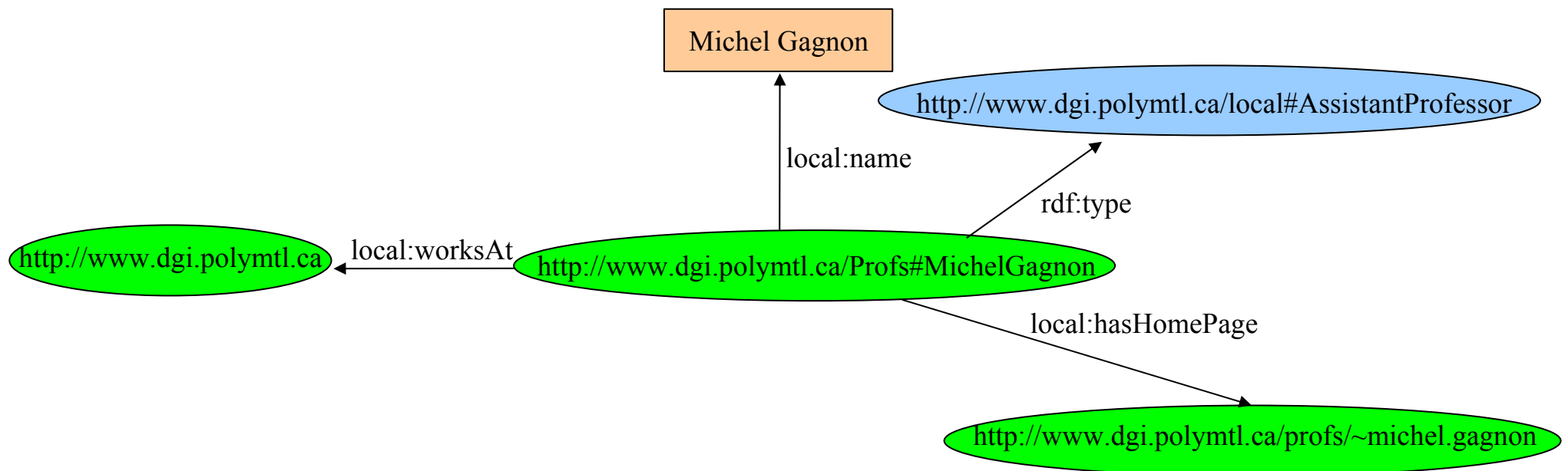
```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:local="http://www.dgi.polymtl.ca/local">
  <rdf:Description rdf:about="http://www.dgi.polymtl.ca/Profs#MichelGagnon">
    <local:hasHomePage resource="http://www.dgi.polymtl.ca/profs/~michel.gagnon"/>
    <local:worksAt resource="http://www.dgi.polymtl.ca"/>
    <local:name>Michel Gagnon</local:name>
  </rdf:Description>
</rdf:RDF>
```

RDF – Exemple (suite)



```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:local="http://www.dgi.polymtl.ca/local">
  <rdf:Description rdf:about="http://www.dgi.polymtl.ca/Profs#MichelGagnon">
    <rdf:type rdf:resource="http://www.dgi.polymtl.ca/local#AssistantProfessor"/>
    <local:hasHomePage resource="http://www.dgi.polymtl.ca/profs/~michel.gagnon"/>
    <local:worksAt resource="http://www.dgi.polymtl.ca"/>
    <local:name>Michel Gagnon</local:name>
  </rdf:Description>
</rdf:RDF>
```

RDF – Exemple (suite)

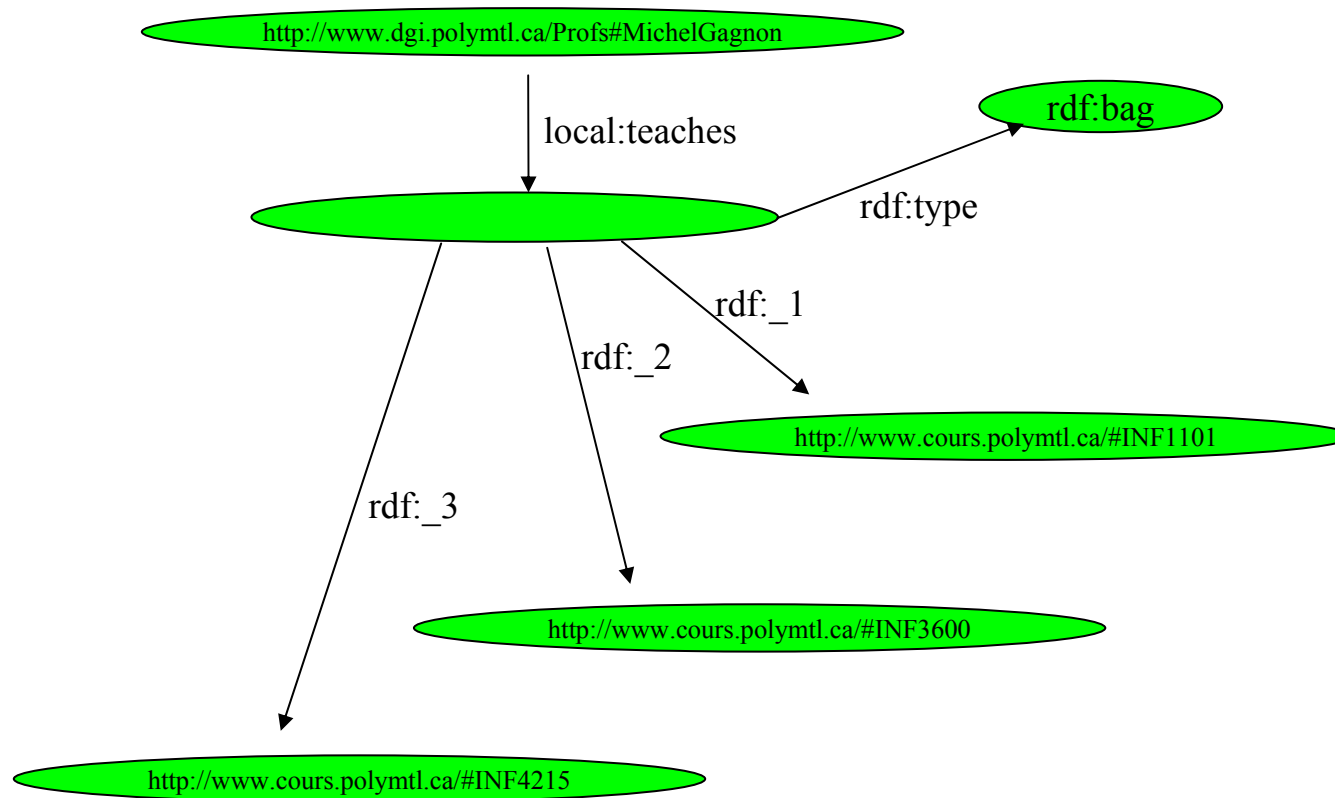


```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:local="http://www.dgi.polymtl.ca/local">
  <local:AssistantProfessor rdf:about="http://www.dgi.polymtl.ca/Profs#MichelGagnon">
    <local:hasHomePage resource="http://www.dgi.polymtl.ca/profs/~michel.gagnon"/>
    <local:worksAt resource="http://www.dgi.polymtl.ca"/>
    <local:name>Michel Gagnon</local:name>
  </local:AssistantProfessor>
</rdf:RDF>
```

RDF – autres caractéristiques

- Conteneurs: bag, seq, alt
- Collections (listes)
- Réification: un énoncé (statement) est lui-même une ressource qui peut faire partie d'une description
- Possibilité de définir des noeuds « anonymes »

RDF – exemple de conteneur



RDF – exemple de conteneur

```
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:local="http://www.dgi.polymtl.ca/">
```

```
<rdf:Description rdf:about="http://www.dgi.polymtl.ca/Profs#MichelGagnon">
```

```
<local:teaches>
```

```
<rdf:Bag>
```

```
<rdf:_1 rdf:resource="http://www.cours.polymtl.ca/INF1101"/>
```

```
<rdf:_2 rdf:resource="http://www.cours.polymtl.ca/INF3600"/>
```

```
<rdf:_3 rdf:resource="http://www.cours.polymtl.ca/INF4215"/>
```

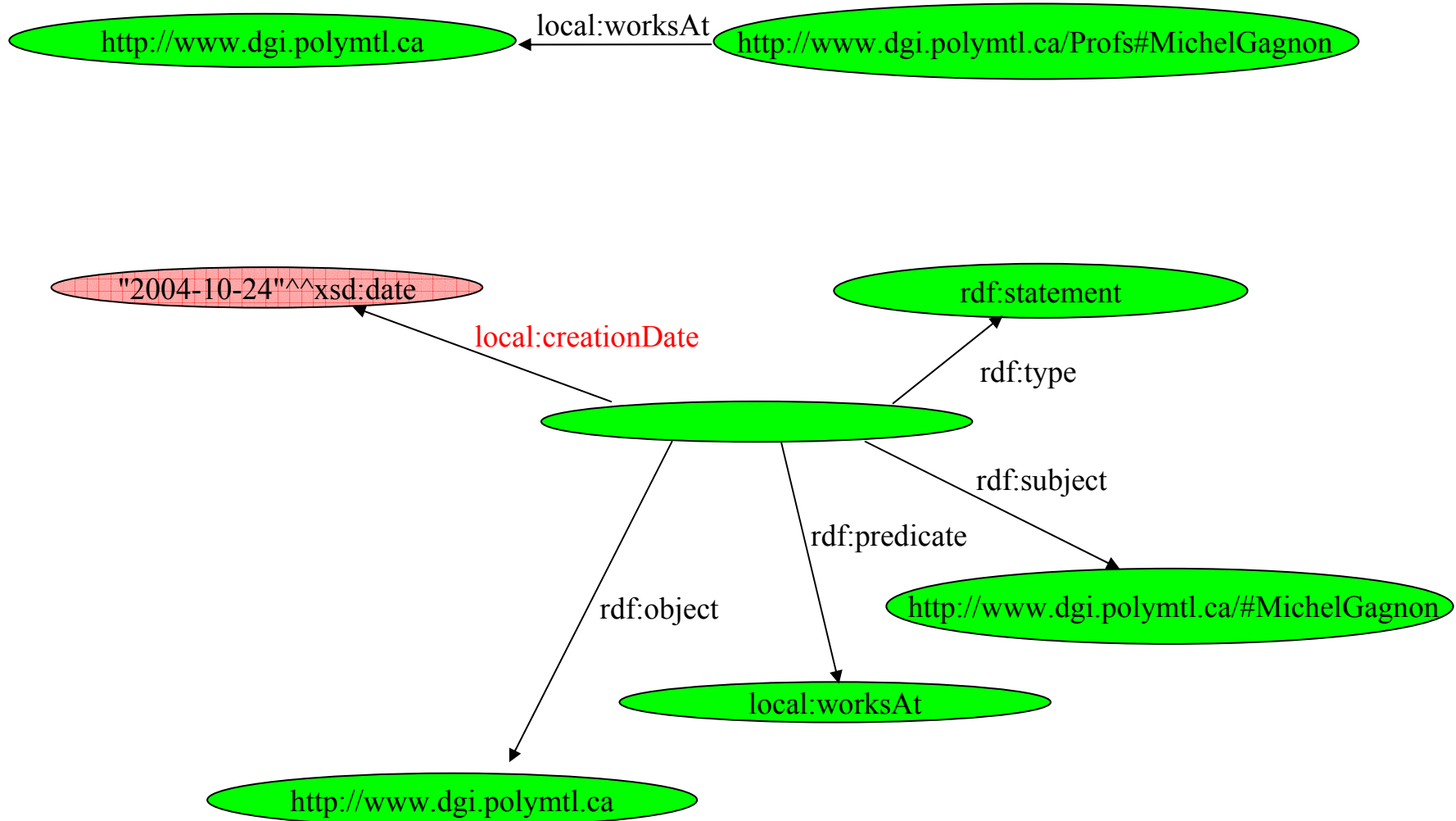
```
</rdf:Bag>
```

```
</local:teaches>
```

```
</rdf:Description>
```

```
</rdf:RDF>
```

RDF – exemple de réification



RDF - Remarques sur la réification

- RDF ne fournit aucun moyen d'identifier un triplet original avec l'énoncé réifié
- RDF ne fournit pas de mécanisme pour attribuer des valeurs de vérités aux énoncés
- Il s'agit d'une caractéristique qui donne beaucoup (trop?) de pouvoir expressif au langage

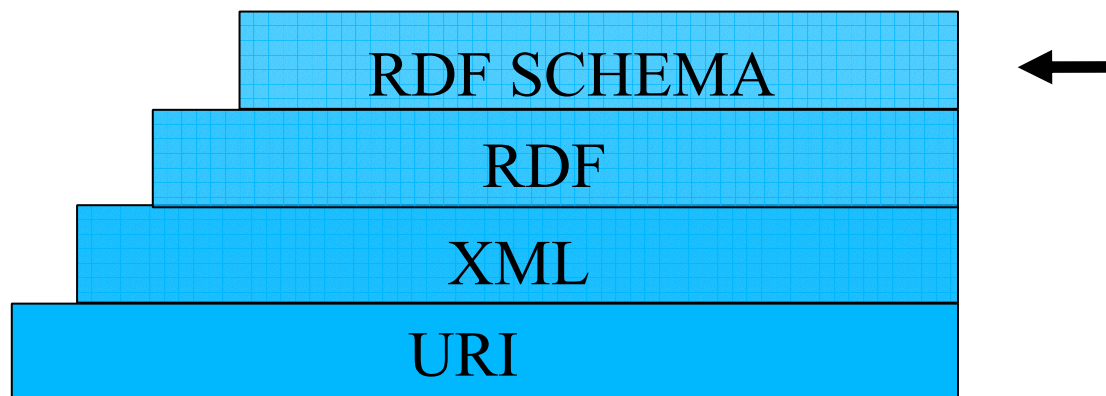
RDF - Remarques sur les URI

- Une ressource est toujours identifiée par une URI
- Si une URI correspond à une URL, on ne sait pas laquelle des deux situations suivantes est la bonne:
 - la ressource est le document associé à l'URL
 - la ressource est une entité abstraite, éventuellement décrite par le document associé à l'URL
- D'autres modèles, comme Topic Map, désambigüisent cette situation

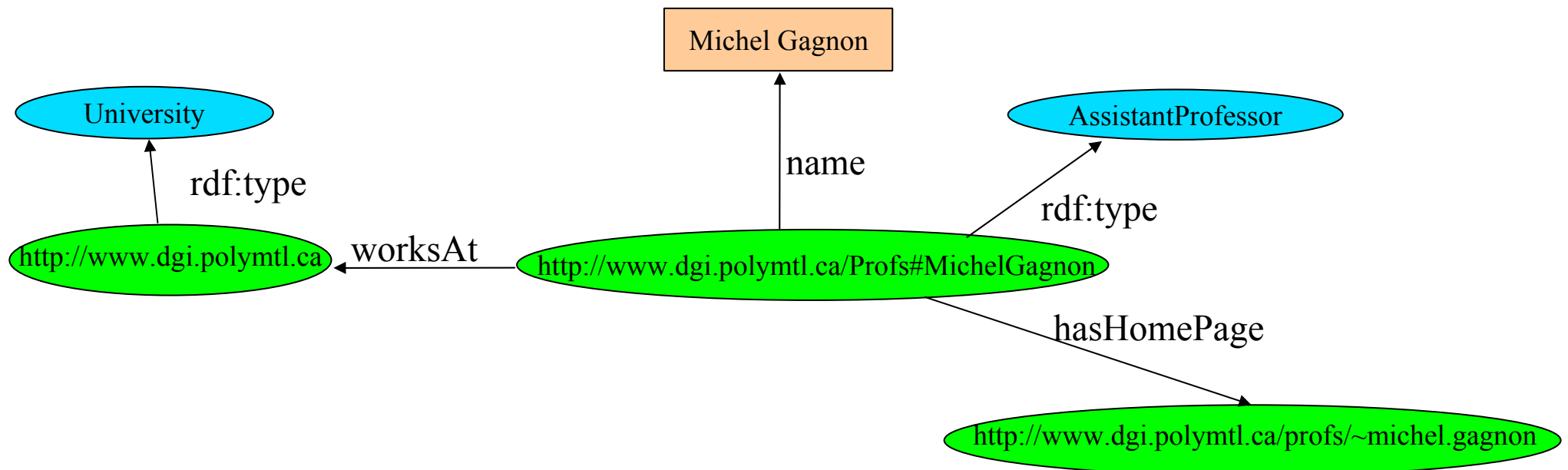
RDF Schema

Définition des hiérarchies de classes et de propriétés

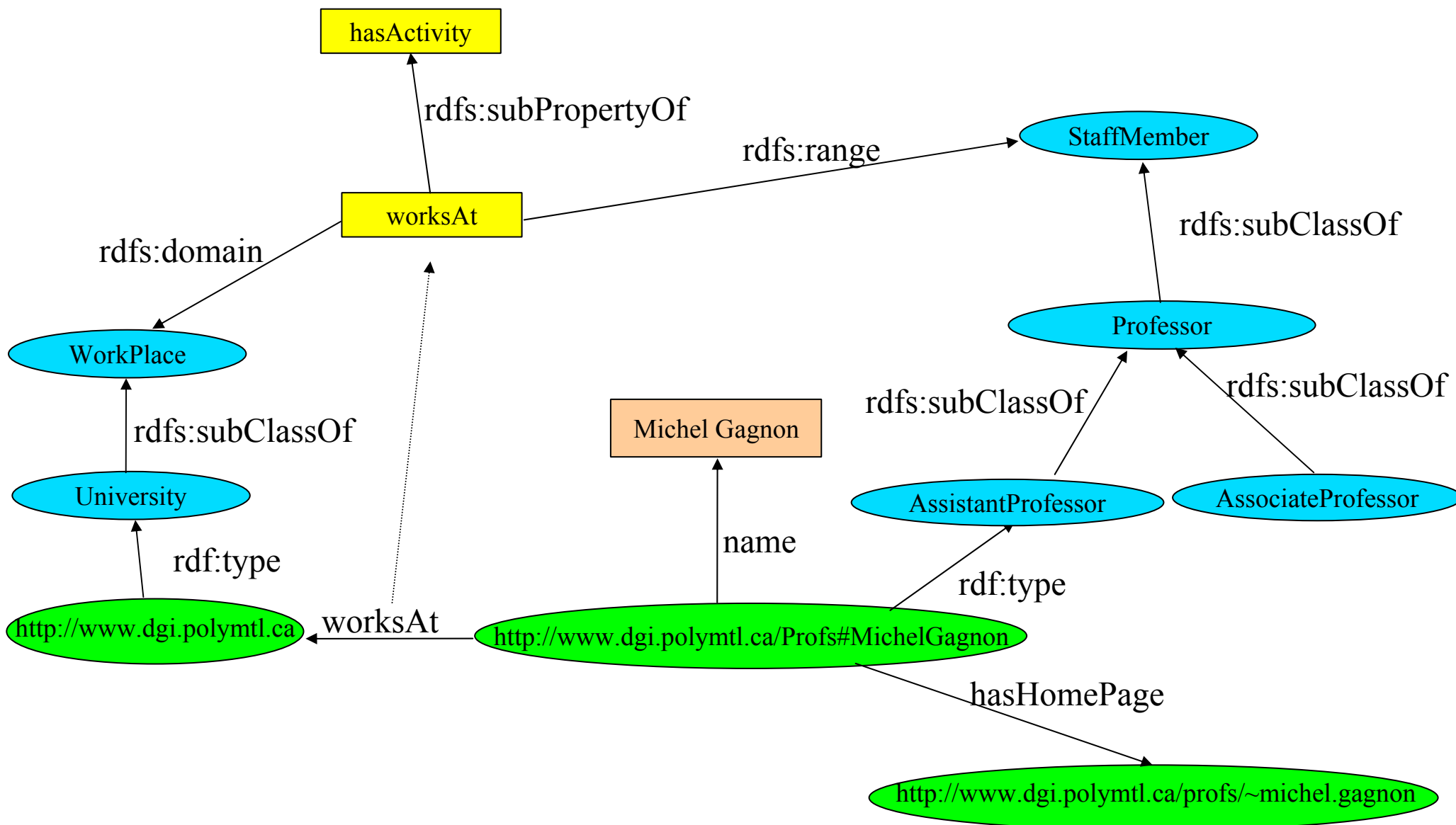
Spécification des domaines et images des propriétés



RDF Schema - Exemple



RDF Schema - Exemple



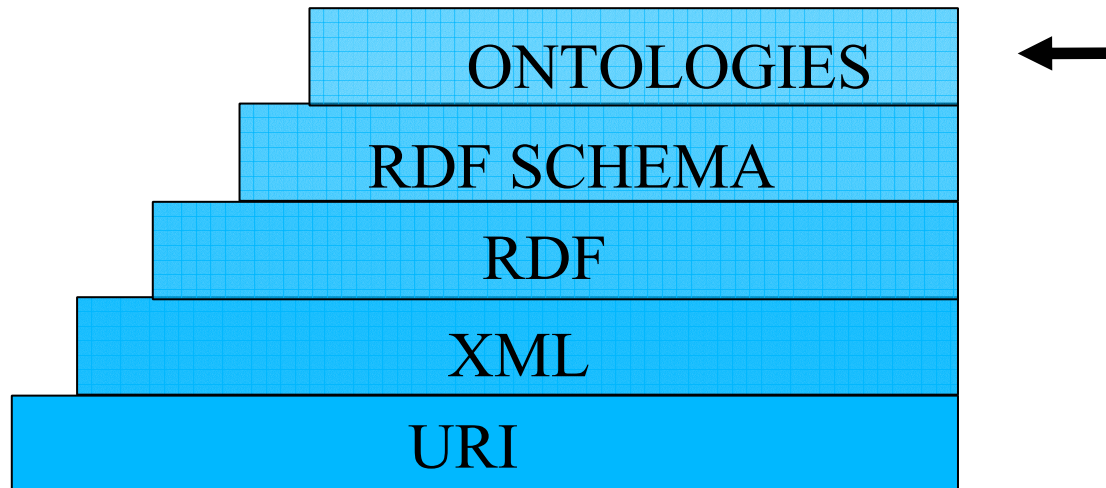
RDF(S) - Caractéristiques

- Les propriétés sont globales
- Ne permet pas de définir les caractéristiques spécifiques à une classe
- Il existe des langages de requête pour RDF(S)

Ontologies

An ontology is an explicit and formal specification of a conceptualization.

- T.R. Gruber



Ontologies – caractéristiques

- Restriction de cardinalité
- Définition de classe par spécification de restrictions sur des propriétés
- Identification de classes disjointes
- Définition de classe par combinaison booléenne
- Caractérisation de certaines propriétés (transitivité, fonctions, inverses)

OWL: le langage du web sémantique

- Vient en trois saveurs:

– OWL Lite

– OWL DL

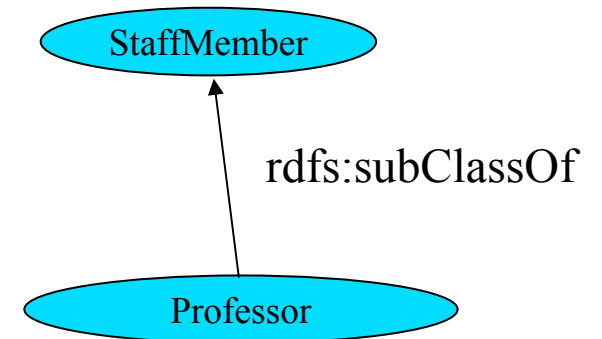
– OWL Full

↑
Pouvoir expressif
↓

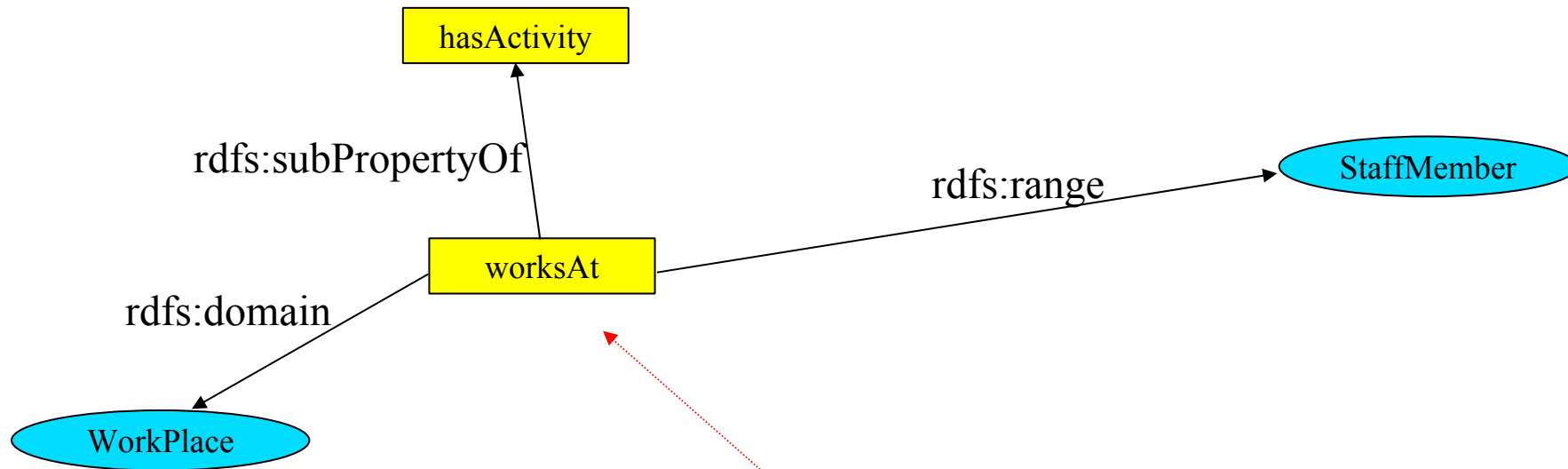
↓
Décidabilité
↑

OWL – definition de classe simple

```
<owl:Class rdf:ID="Professor">  
  <rdfs:subClassOf>  
    <owl:Class rdf:about="#StaffMember"/>  
  </rdfs:subClassOf>  
</owl:Class>
```



OWL – Restriction de cardinalité

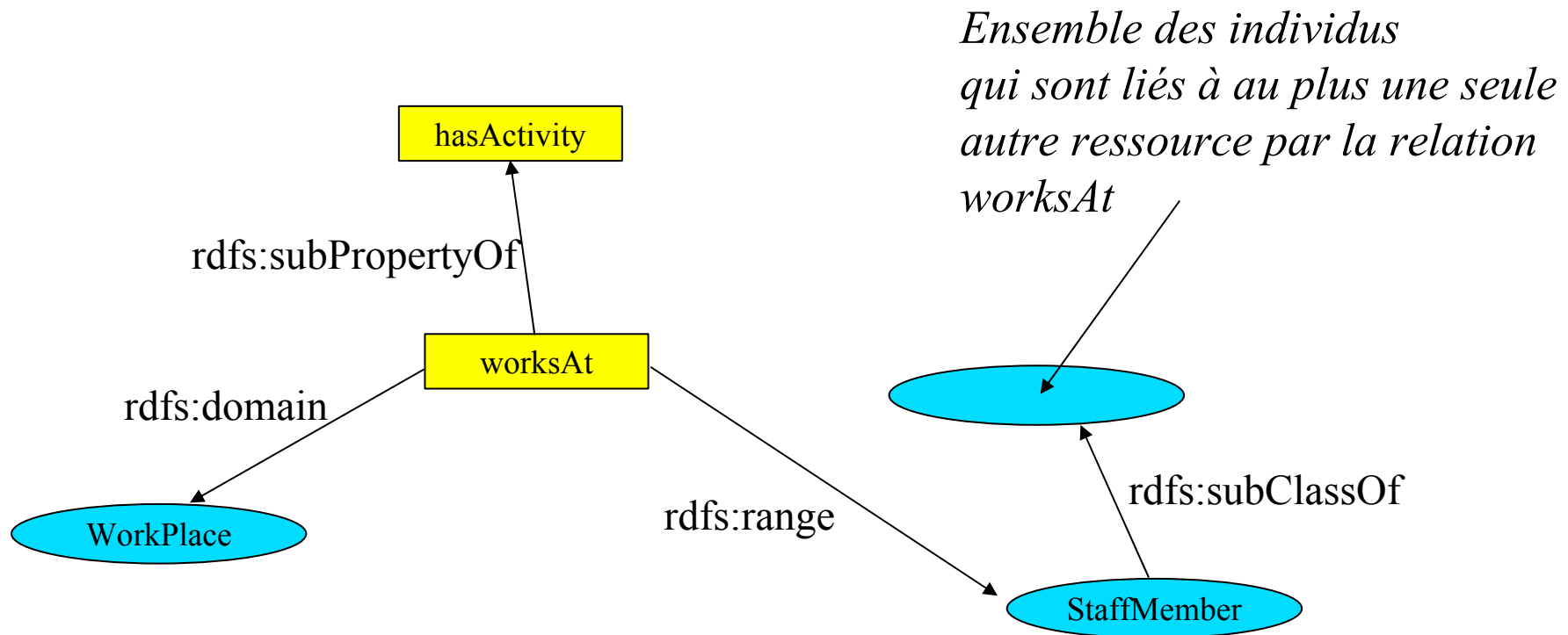


On voudrait pouvoir spécifier que personne ne peut travailler à plus d'un endroit.

Sans restriction de cardinalité

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#">
  <owl:Ontology rdf:about="www.dgi.polymtl.ca/Staff"/>
  <owl:Class rdf:ID="StaffMember"/>
  <owl:Class rdf:ID="WorkPlace"/>
  <owl:ObjectProperty rdf:ID="worksAt">
    <rdfs:range rdf:resource="#WorkPlace"/>
    <rdfs:domain rdf:resource="#StaffMember"/>
    <rdfs:subPropertyOf>
      <owl:ObjectProperty rdf:ID="hasActivity"/>
    </rdfs:subPropertyOf>
  </owl:ObjectProperty>
</rdf:RDF>
```

OWL – Restriction de cardinalité



Avec restriction de cardinalité

```
<?xml version="1.0"?>
<rdf:RDF
  xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#"
  xmlns:owl="http://www.w3.org/2002/07/owl#">
  <owl:Ontology rdf:about="www.dgi.polymtl.ca/Staff"/>
  <owl:Class rdf:ID="StaffMember"/>
  <rdfs:subClassOf>
    <owl:Restriction>
      <owl:maxCardinality rdf:datatype="http://www.w3.org/2001/XMLSchema#int">1</owl:maxCardinality>
      <owl:onProperty>
        <owl:ObjectProperty rdf:about="#worksAt"/>
      </owl:onProperty>
    </owl:Restriction>
  </rdfs:subClassOf>
</owl:Class>
  <owl:Class rdf:ID="WorkPlace"/>
  <owl:ObjectProperty rdf:ID="worksAt">
    <rdfs:range rdf:resource="#WorkPlace"/>
    <rdfs:domain rdf:resource="#StaffMember"/>
    <rdfs:subPropertyOf>
      <owl:ObjectProperty rdf:ID="hasActivity"/>
    </rdfs:subPropertyOf>
  </owl:ObjectProperty>
</rdf:RDF>
```

OWL – Définition de classe par restrictions sur des propriétés

On veut représenter le concept de professeur occupé, défini de la manière suivante:

un professeur qui enseigne plus de trois cours.

On le représente par l'intersection de deux classes:

$\text{BusyProfessor} \equiv \text{Professor} \cap (\text{cardinalité}(\text{teaches}) \leq 3)$

Classe BusyProfessor

```
<owl:Class rdf:ID="BusyProfessor">
  <owl:equivalentClass>
    <owl:Class>
      <owl:intersectionOf rdf:parseType="Collection">
        <owl:Class rdf:about="#Professor"/>
        <owl:Restriction>
          <owl:onProperty>
            <owl:ObjectProperty rdf:about="#teaches"/>
          </owl:onProperty>
          <owl:minCardinality
            rdf:datatype="http://www.w3.org/2001/XMLSchema#int"> 3
          </owl:minCardinality>
        </owl:Restriction>
      </owl:intersectionOf>
    </owl:Class>
  </owl:equivalentClass>
</owl:Class>
```

OWL DL

- Exemples de restrictions de OWL-DL:
 - Une classe ne peut pas être instance d'une autre classe
 - La plupart des éléments du vocabulaire de RDF(S) ne sont pas valides en OWL-DL
 - Toutes les classes doivent être déclarées, ce qui rend l'exemple suivant invalide si la classe Humain n'est pas définie quelque part:

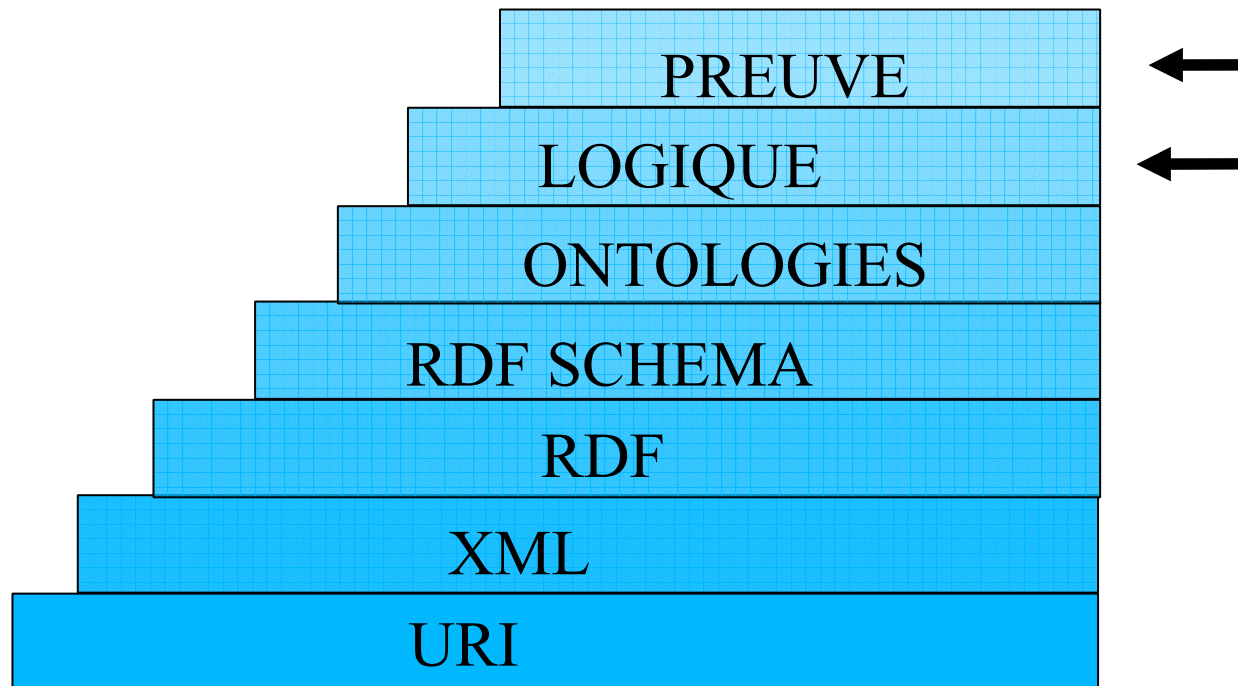
```
<owl:Class rdf:ID= "Femme">  
  <rdfs:subClassOf rdf:resource="Humain"/>  
</owl:Class>
```

OWL Lite

- Exemples de restrictions de OWL Lite:
 - Cardinalité peut être seulement 0 ou 1
 - On ne peut pas définir une classe par énumération
 - On ne peut pas déclarer des classes disjointes
 - Limites sur l'utilisation de `rdfs:subClassOf`

Logique et preuves

Il faut pouvoir faire des inférences.
Il faut aussi pouvoir les expliquer



Logique et preuves

- La logique est un langage permettant d'exprimer des « règles » de raisonnement
- Ces règles permettent de déduire de nouveaux faits à partir des faits existants
- Une preuve est en quelque sorte une suite d'applications de règles qui permettent de déduire un nouveau fait

Exemple de règles

SI

instance(X, Professor) ET
possede_diplome(X, PhD)

ALORS

peut_etre_membre(X, JuryDoctorat)

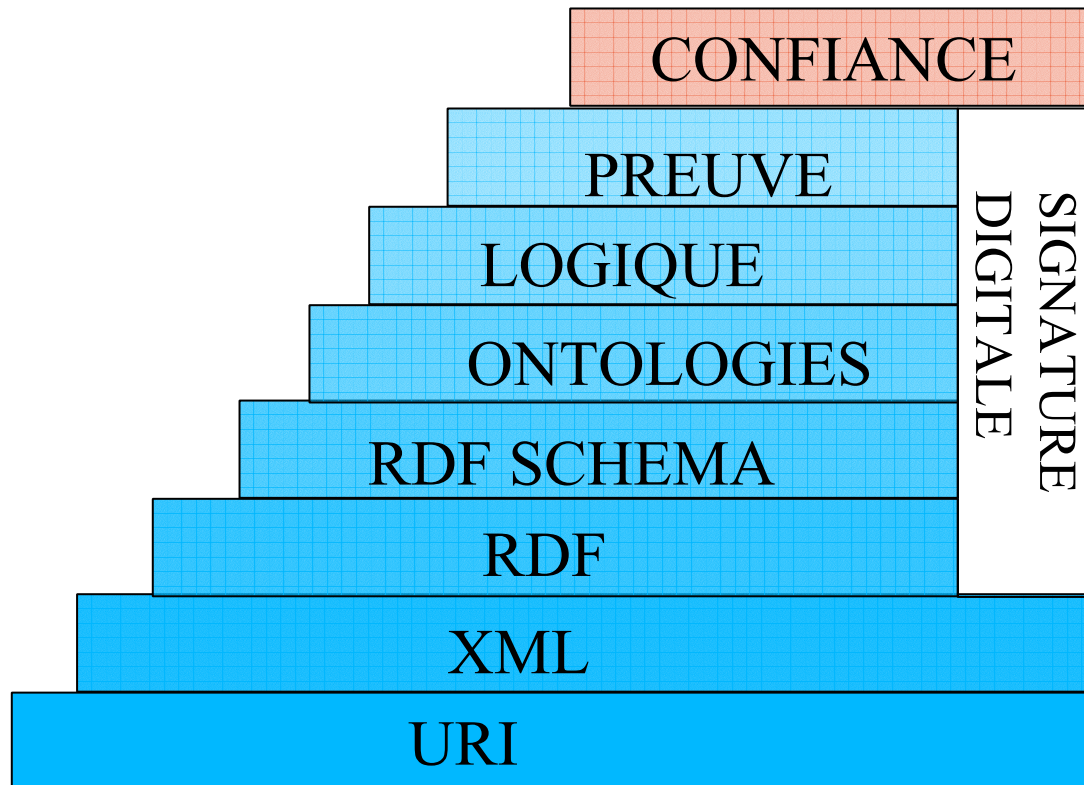
SI

peut_etre_membre(X, JuryDoctorat) ET
disponible(X)

ALORS

membre_potentiel(X, JuryDoctorat)

Niveau de confiance



Niveau de confiance

- Si on utilise des agents pour prendre des décisions à notre place, il faudrait qu'on puisse avoir « confiance » aux résultats
- Pour ce faire, il faut que l'agent utilisé puisse:
 - expliquer clairement comment il arrive à ses conclusions (preuve)
 - garantir la fiabilité et l'origine des informations utilisées (signature digitale)

Plan de la présentation

- Définition et description générale
- Les différentes couches du web sémantique
- Exemples d'application
- Conclusion

Annotations partagées: Annotea

- Proposé par le W3C:

<http://www.w3.org/2001/Annotea/>

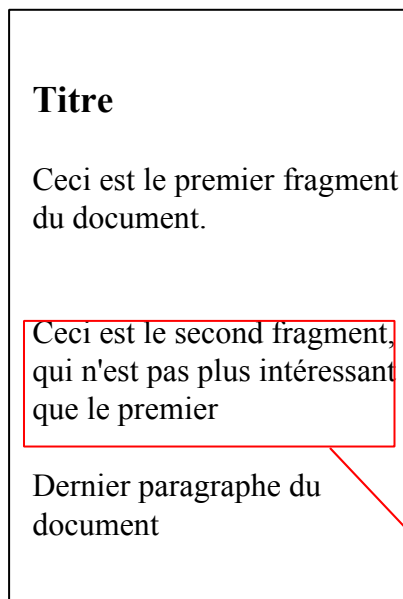
- N'importe qui peut annoter un document électronique disponible sur le web
- Les annotations sont stockées en format RDF sur un ou plusieurs serveurs
- Lorsqu'on récupère un document du web, on peut demander de récupérer aussi les annotations qui y sont associées

Annotea

- Une annotation peut référer au document entier ou à un fragment de celui-ci
- Pour identifier le fragment annoté, on utilise le langage XPointer, qui permet de spécifier le « chemin » jusqu'au fragment
- Une annotation est essentiellement un texte, mais plusieurs informations, comme son type et son créateur, y sont associées
- Plusieurs types d'annotation ont été définis

Annotea

Document



Stocker annotations

Récupérer annotations

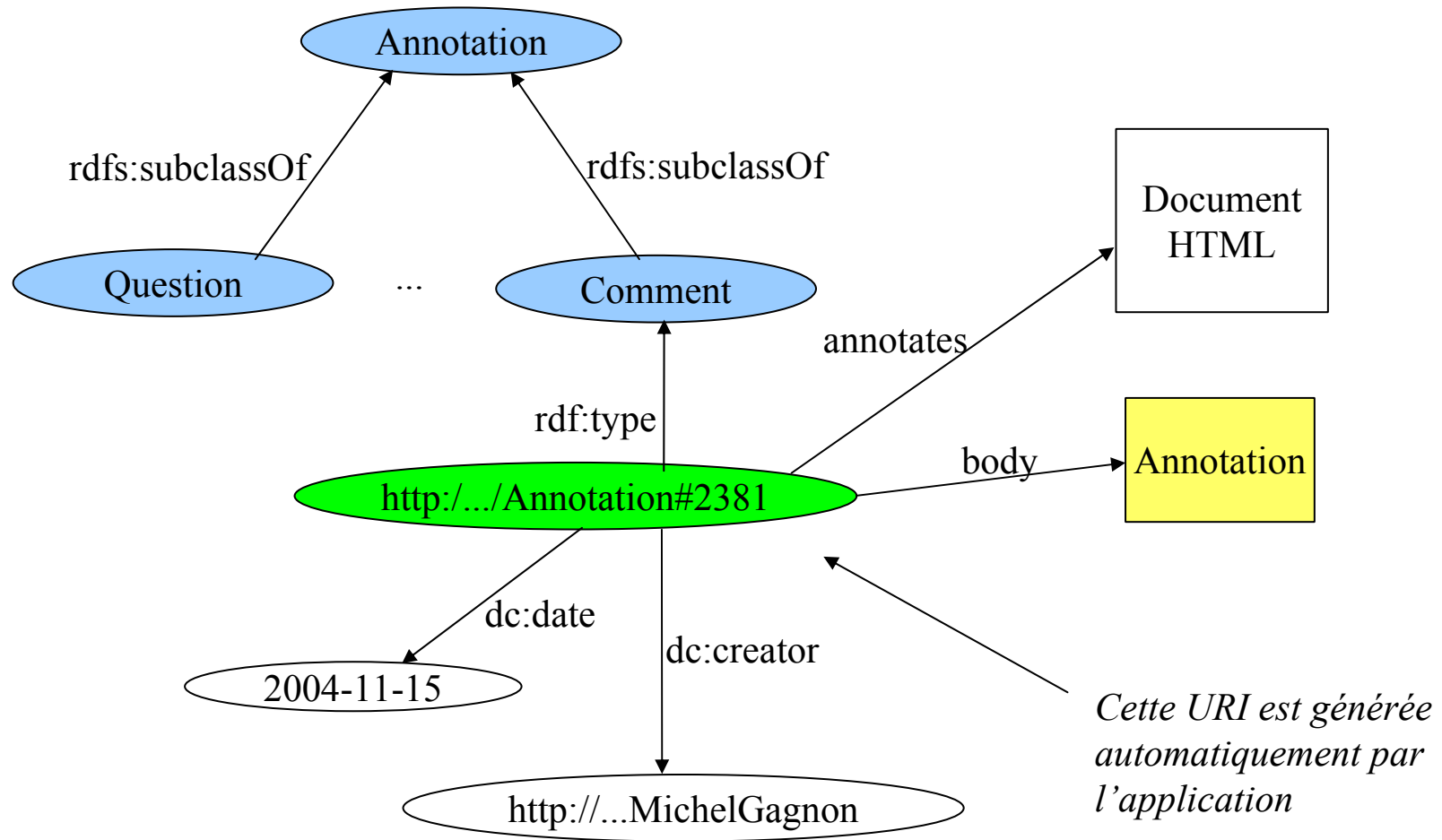
BD RDF

BD RDF

Cette section est redondante

Annotation

Annotea – exemple d'annotation



Recherche sémantique: TAP

- Projet mis de l'avant par
 - Knowledge Systems Laboratory (Stanford)
 - Knowledge Management Group (IBM Almaden)
 - W3C
- Site du projet:
 - <http://tap.stanford.edu/>

TAP

- Propose un protocole GetData, pour récupérer sur le web des informations en format RDF
- Ce protocole permet d'identifier une ressource par ses propriétés plutôt que son URI: référence par description
- Deux serveurs peuvent utiliser les descriptions pour identifier une même ressource qu'ils nomment de manière différente
- Cela suppose le partage d'un vocabulaire pour les descriptions

TAP

- Une requête selon le protocole GetData contient les deux éléments suivants:
 - Une description d'une ressource
 - Une URI désignant une propriété
- On s'attend à ce qu'un serveur recevant une telle requête retourne la valeur associée à cette propriété pour la ressource en question
- La requête est formulée dans le langage SOAP (langage défini pour l'accès aux services web)

TAP – exemple de requête GetData

```
<?xml version="1.0" encoding="UTF-8"?>
<SOAP-ENV:Envelope
  xmlns:SOAP-ENV="http://schemas.xmlsoap.org/soap/envelope/"
  SOAP-ENV:encodingStyle="http://schemas.xmlsoap.org/soap/encoding/"
  xmlns:xsi="http://www.w3.org/1999/XMLSchema-instance"
  xmlns:xsd="http://www.w3.org/1999/XMLSchema"
  xmlns:a="http://tap.stanford.edu/">
  <SOAP-ENV:Body>
    <GetData>
      <a:Musician>
        <a:title>Yo Yo Ma</a:title>
        <a:title>Ma, Yo-Yo</a:title>
        <a:playsInstrument
          ressource="http://tap.stanford.edu/kb/CelloInstrument"/>
        <a:oid namespace="a">MusicianMa,_Yo-Yo</a:oid>
      </a:Musician>
      <a:concertSchedule/>
    </GetData>
  </SOAP-ENV:Body>
</SOAP-ENV:Envelope>
```

TAP pour recherche sémantique

- TAP a créé une base de connaissances contenant diverses données (musiciens, films, compagnies, etc.)
- Quand un usager soumet un terme à l'outil de recherche, ce terme est recherché dans la base de connaissances
- Si on le trouve, on oriente la recherche selon le type du concept et ses propriétés
- Par exemple, si l'usager soumet « Yo Yo MA », la base de connaissances permettra de savoir qu'il s'agit d'un musicien, ce qui amènera à rechercher des informations comme le calendrier de ses concerts

Communauté de personnes: FOAF

- Friend-of-a-friend, une idée originale de Libby Miller et Dan Brickley

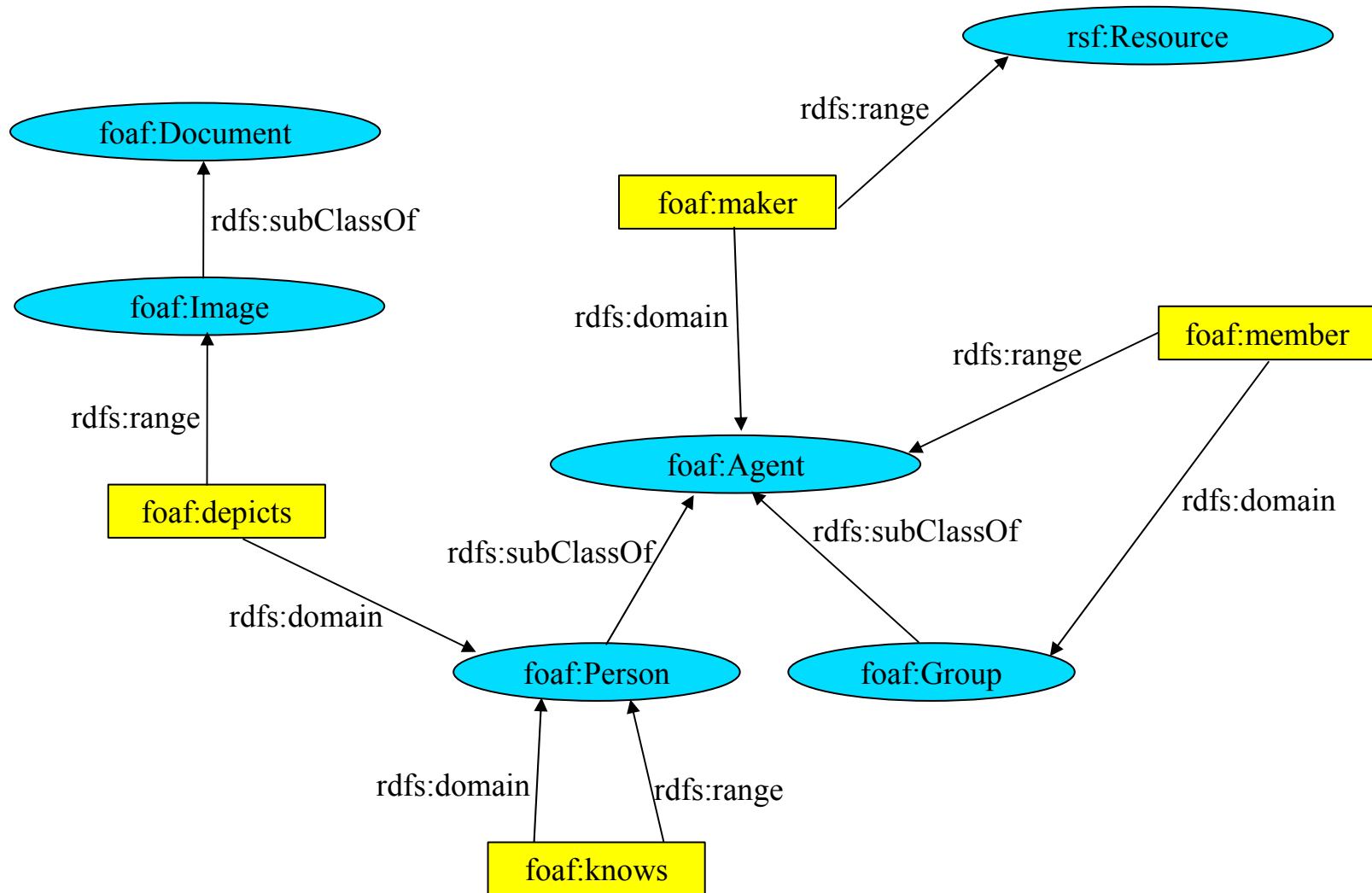
<http://www.foaf-project.org>

- L'idée consiste à publiciser sur le web une description de soi-même, en quelque sorte une page personnelle lisible par la machine
- Cette description est faite en RDF, en utilisant un vocabulaire pré-défini par les créateurs de FOAF
- FOAF fournit un utilitaire pour créer cette description: FOAF-A-Matic

FOAF

- Tout comme avec TAP, l'identification se fait à partir des descriptions, et non à l'aide d'une URI
- Certaines propriétés, comme l'adresse électronique, permettent d'identifier clairement la personne
- Dans la description, il est possible de spécifier d'autres personnes que l'on connaît
- Ainsi, un outil peut être utilisé pour parcourir ces descriptions, et passer d'une personne à une autre, par exemple

FOAF – vocabulaire (extrait)



FOAF - exemple

```
<rdf:RDF xmlns:foaf="http://xmlns.com/foaf/0.1/">
  <foaf:Person rdf:about="http...professeurs#MichelGagnon">
    <foaf:name>Michel Gagnon</foaf:name>
    <foaf:title>Mr</foaf:title>
    <foaf:firstName>Michel</foaf:firstName>
    <foaf:family_name>Gagnon</foaf:family_name>
    <foaf:mbox_sha1sum> 569c16042374161d8 </foaf:mbox_sha1sum>
    <foaf:homepage
      rdf:resource="http://www.professeurs.polymtl.ca/michel.gagnon"/>
    <foaf:knows>
      <foaf:Person rdf:ID="">
        <foaf:name>Eliana de Mattos Pinto Coelho</foaf:name>
        <foaf:title>Mrs.</foaf:title>
        <foaf:firstName>Eliana</foaf:firstName>
        <foaf:family_name>de Mattos Pinto Coelho</foaf:family_nname>
        <foaf:mbox_sha1sum> 1c2d418f5b6689b9337a</foaf:mbox_sha1sum>
      </foaf:Person>
    </foaf:knows>
  </foaf:Person>
</rdf:RDF>
```

Site web sémantique: Mindswap

- Se présente comme le premier site web sémantique:
<http://www.mindswap.org>
- Toutes les informations sont contenues dans une base de données RDF
- Les mêmes informations sont disponibles en RDF/XML, dans des fichiers publics
- Des vues HTML sont créées automatiquement à partir de la base de données RDF
- Plusieurs utilitaires pour créer les informations sémantiques, notamment à partir de documents électroniques existants

Autres applications

- E-learning
- Intranets
- Services web
- Swoogle (outil de recherche)
- Outils semi-automatiques d'annotation (Amilcare, Melita, MnM, etc.)

Plan

- Définition et description générale
- Les différentes couches du web sémantique
- Exemples d'application
- **Conclusion**

Problèmes spécifiques à résoudre

- Indexation et évaluation des ontologies
- Création, documentation et maintenance des ontologies
- Automatisation du processus d'annotation
- Sécurité
- Développement des agents intelligents

Résumé

Web actuel

Résumé

Web actuel

+

Ontologies

Résumé

Web actuel

+

Ontologies

+

Agents

Résumé

Web actuel

+

Ontologies

+

Agents

=

Web sémantique

Conclusion

Le web sémantique

Conclusion

Le web sémantique

est un projet ambitieux,

Conclusion

Le web sémantique

est un projet ambitieux,

mais souhaitable

Conclusion

Le web sémantique

est un projet ambitieux,

mais souhaitable

et réalisable!