

The background of the slide is a faded, light blue version of the famous Japanese woodblock print 'The Great Wave off Kanagawa' by Katsushika Hokusai. The image shows a massive, curling wave about to crash over a small boat with several figures. In the distance, the snow-capped Mount Fuji is visible under a pale sky. The overall style is traditional Japanese ink and wash, rendered in a monochromatic blue-grey palette.

**MIF24**  
**Annotation et Intelligence Artificielle**

**Yannick Prié et Frank Nack**  
**LIRIS - SILEX**

## Le plan

---

- Notion de représentation de connaissance
- Ontologies
- Annotation textuelle pour le web sémantique + Outils

URL: <http://liris.cnrs.fr/yannick.prie/ens/06-07/TIA/index.html>

## La représentation de connaissance

---

La représentation des connaissances désigne un ensemble d'outils et de technologies destinés d'une part à représenter et d'autre part à organiser le savoir humain pour l'utiliser et le partager.



Modèle conceptuel  
Langage de représentation de connaissances  
Interprétation

## Différents types de connaissances du modèle conceptuel

- Connaissances stratégiques
  - Comment une tâche va être effectuée
- Connaissance du domaine
  - Concepts manipulés et leur relations. Description du domaine, par exemple l'anatomie du corps humain

## Modèle conceptuel

- Connaissances stratégiques + du domaine
- Abstraction permettant d'opérationnaliser une expertise de résolution de problème, une partie d'un monde...

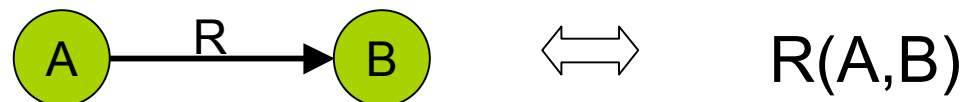
# La représentation de connaissance - Formalisation

## Les outils classiques

- taxonomies ou classifications, qui permettent d'organiser les connaissances sur les objets du monde
- les thésaurii, utilisés en indexation documentaire

## Les outils plus formel et permettant de représenter des connaissances complexes

- les graphes conceptuels  
[Person: John]←-(Agnt)←-[Go]->(Dest)->[City: Aalborg]  
"John is going to Aalborg »
- les réseaux sémantiques



## La représentation de connaissance - Formalisation

Dans une représentation formelle, les connaissances sont représentées par des objets logiques (logique modale, calcul des prédicats, logique temporelle) reliés par

- des propriétés (frames et scripts)
- des axiomes
- des règles.

Pour le développement du Web, en particulier du Web sémantique, on utilise les ontologies.

# La représentation de connaissance - Langage

---

## Notation

- DATR : langage pour décrire la représentation de connaissances lexicales
- **RDF** : modèle de graphe destiné à décrire de façon formelle les ressources Web et leurs métadonnées

## Langages

- **Cycl** langage descriptif utilisé dans Cyc
- **IKL** IKRIS Knowledge Language
- **KIF** Knowledge Interchange Format : langage d'échange de connaissances utilisé dans le cadre des systèmes multi-agents.
- **Loom** production rule facility and a pattern-directed method that supports the definition of object-oriented methods).
- **SKOS** Simple Knowledge Organisation System : famille de langages formels permettant une représentation standard des thésaurus, classifications
- **KM** Knowledge Machine, langage à base de frames pour la représentation des connaissances

# La représentation de connaissance - Langage

---

## Notation

- DATR : langage pour décrire la représentation de connaissances lexicales
- **RDF** : modèle de graphe destiné à décrire de façon formelle les ressources Web et leurs métadonnées

## Langages

- Dublin Core (DC) comprend officiellement 15 éléments de description formels (titre, auteur, éditeur), intellectuels (sujet, description, langue...) et relatifs à la propriété intellectuel).

- **RDFS**
- **OWL**

langage extensible de représentation des connaissances

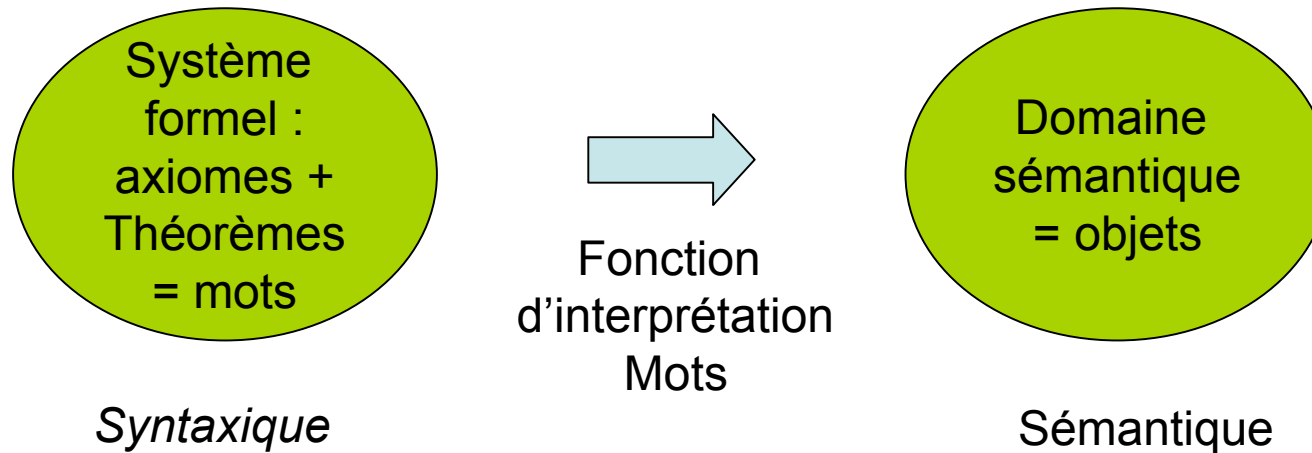
Web Ontology Language : dialecte XML basé sur une syntaxe RDF. Il fournit les moyens pour définir des ontologies Web structurées

- Topic Maps

Les cartes topiques (ou cartes de thèmes) sont un outil très général de représentation des connaissances, dont le but est d'agrèger autour d'un point unique d'indexation (appelé topic) toutes les informations disponibles concernant un sujet donné, et de relier ces points par un réseau sémantique de relations appelées associations.



## La représentation de connaissance - Interprétation



- Possibilité de notion de vérité dans le domaine sémantique = possibilité de « bonne interprétation »
  - Idéalement théorème  $\Rightarrow$  interprétation vraie

- Modèles informatisables répertoriant les concepts et les notions d'un domaine en les formalisant
- Point de vue de l'ingénierie : en vue de leur utilisation dans des réalisations informatiques.
  - De ce point de vue, elles reflètent un certain niveau de connaissance sur un domaine
- Langages de représentation d'ontologies sur le Web
  - OWL, RDFS, RDF

## Ontologie - Composants typiques

---

- Concepts ou classes
  - *Enseignant, Cours, CoursInformatique*
  - Relations de généralisation / spécialisation entre concepts
- Relations entre concepts ou « propriétés objets »
  - *X enseigne Y*
  - ObjectProperties en OWL
- Les attributs des concepts ou « propriétés types de données »
  - *X a comme e-mail « frank.nack@liris.cnrs.fr »*
  - DataTypesProperties en OWL

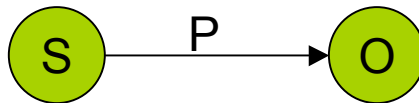
- XML : couche de transport syntaxique
  - transformation, échange, stockage, import/export
- RDF : réseau sémantique de base
  - triplets relationnels...
  - modèle de connaissances : description, intégration
- RDFS : langage de classes, simple
  - représentation de structures conceptuelles et d'ontologies simples avec description hiérarchique des concepts et des propriétés
- OWL : langage de représentation d'ontologies plus complexes
  - RDFS + propriété complexes, types définis,

## Ontologie - RDF

---

Resource Description Framework (RDF) est un modèle de graphe destiné à décrire de façon formelle les ressources Web et leurs métadonnées, de façon à permettre le traitement automatique de telles descriptions.

Un document structuré en RDF est un ensemble de triplets. Un triplet RDF est une association:

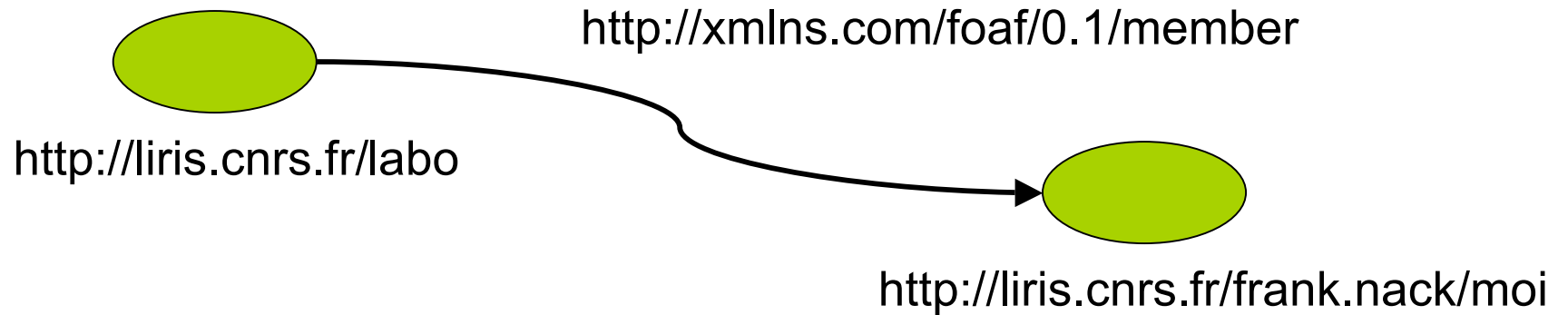


- Le sujet représente la ressource à décrire.
- Le prédicat représente un type de propriété applicable à cette ressource
- L'objet représente une donnée ou une autre ressource

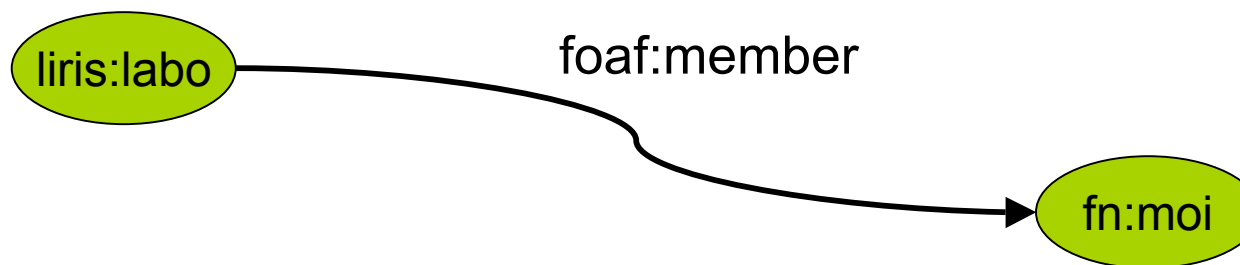
# Ontologie - RDF

---

## RDF triplet



## RDF notation



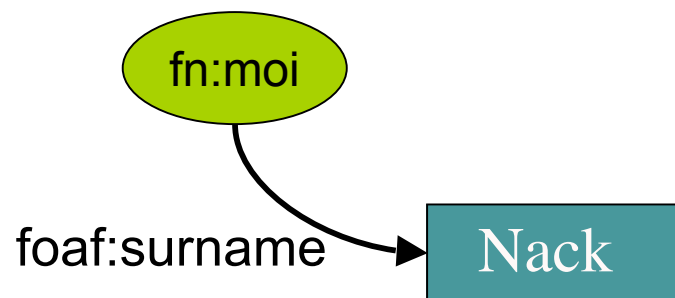
liris: `http://liris.cnrs.fr/`  
foaf: `http://xmlns.com/foaf/0.1/`  
fn: `http://liris.cnrs.fr/frank.nack/`

## Ontologie - RDF

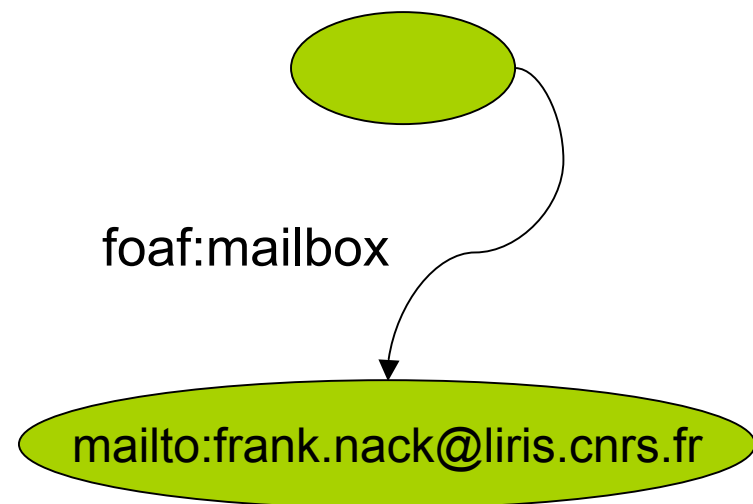
---

Le sujet, et l'objet dans le cas où c'est une ressource, peuvent être identifiés par une URI ou être des nœuds anonymes. Le prédicat est nécessairement identifié par une URI.

### RDF littéral

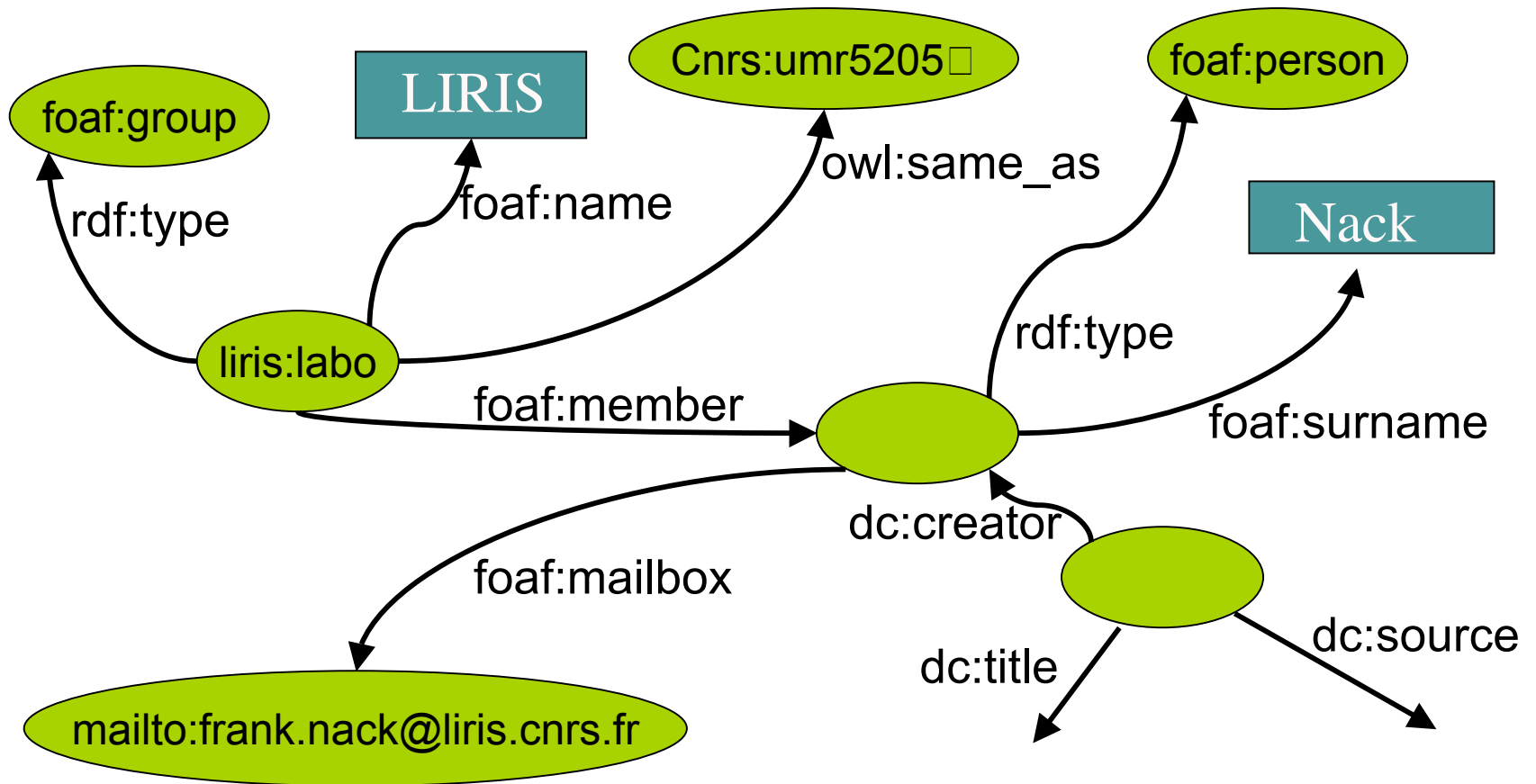


### RDF nœud vierge



## Ontologie - RDF

Un document RDF ainsi formé correspond à un multi-graphe orienté étiqueté. Chaque triplet correspond alors à un arc orienté dont le label est le prédicat, le nœud source est le sujet et le nœud cible est l'objet.





## Ontologie - RDF

---

La structure de RDF est extrêmement générique et sert de base à un certain nombre de schémas ou vocabulaires dédiés à des applications spécifiques.

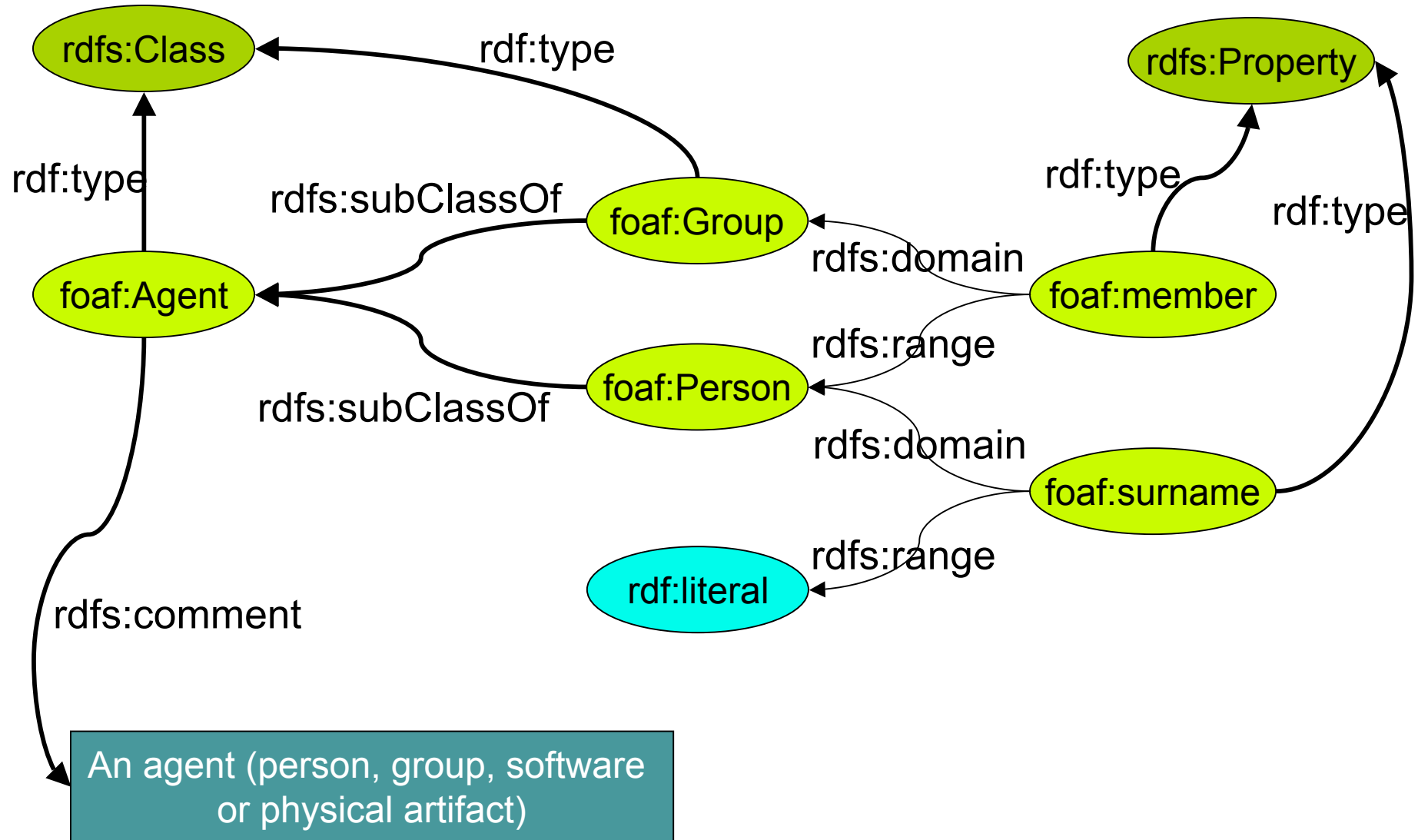
Une partie de ces vocabulaires est spécifiée par le W3C, comme les langages d'ontologie RDFS et OWL, ou le langage SKOS pour la représentation des thésaurus et autres vocabulaires structurés.

**RDFS** ou RDF Schema est un langage extensible de représentation des connaissances. Il fournit des éléments de base pour la définition d'ontologies ou vocabulaires destinés à structurer des ressources RDF.

### Principales caractéristiques de RDFS

- Classes (concepts)
- Propriétés (rôles, domain, range)
- Possibilité de les structurer en hiérarchie de spécialisation (sous-classes, récursivité)
- Contraintes sur le domaine et la portée des propriétés (limité).

# Ontologie - RDFS



## Classes et sous classes

rdfs:Class permet de déclarer une ressource RDF comme une classe pour d'autres ressources.

La définition de rdfs:Class est récursive. rdfs:Class est la classe de toutes les classes.

- rdfs:subClassOf permet de définir des hiérarchies de classes.

## Propriétés

RDFS précise la notion de propriété définie par RDF en lui ajoutant les notions de "domain" et "range"

- \* rdfs:domain définit la classe des sujets supportant une propriété.
- \* rdfs:range définit la classe ou le type de données des valeurs de la propriété.

Les hiérarchies de classes supportent l'héritage des "domain" et "range" des propriétés.

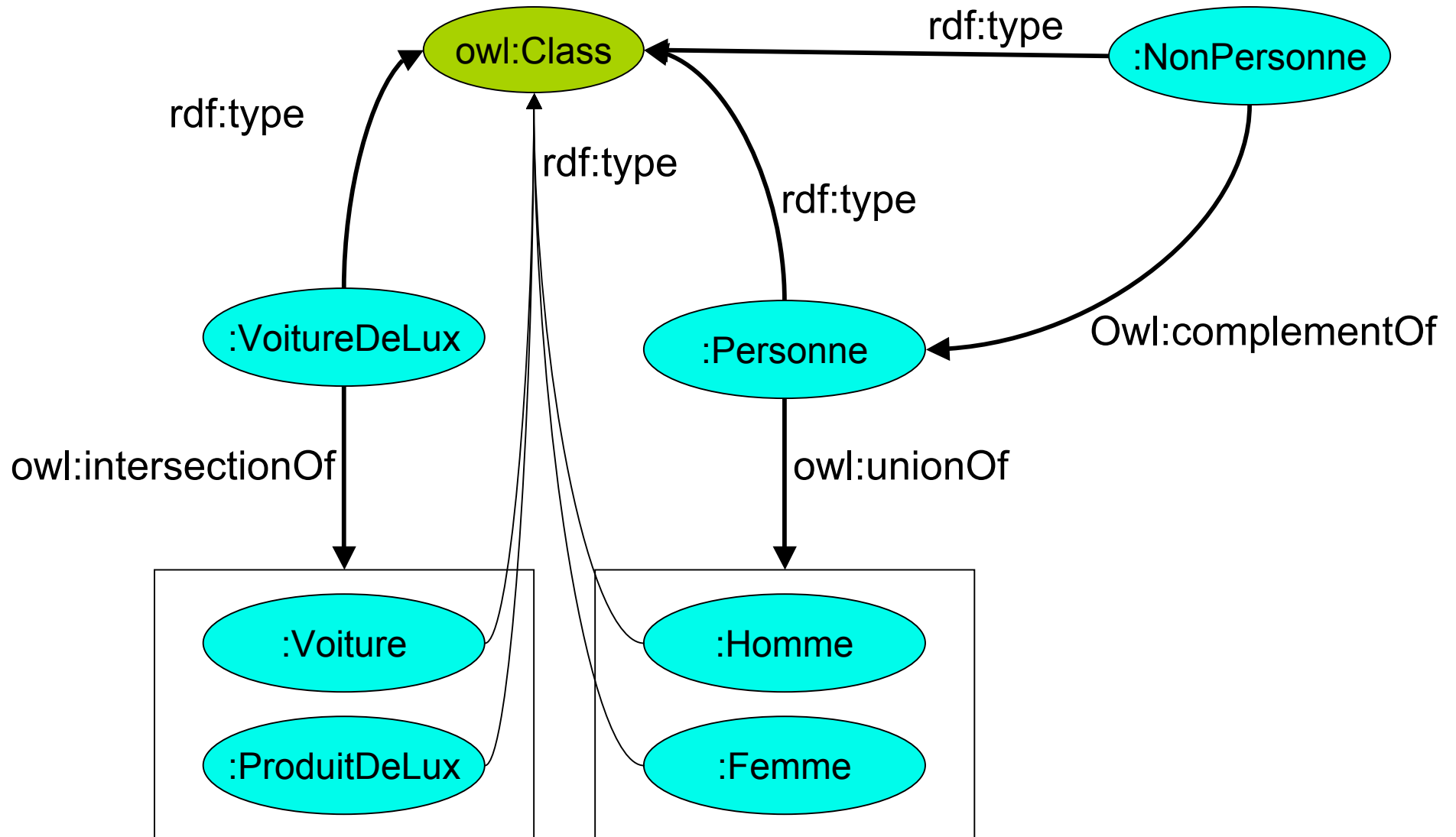
### Problème

- les familles de termes (classes, propriétés, individus) ne sont pas distinctes
- Dans RDFS il n'y a pas de termes complexes, et notamment pas de négation
- on ne peut pas exprimer de contradiction en RDFS
- les mécanismes d'inférence restent décidables

**Ontology Web Language** : fournit les moyens pour définir des ontologies Web structurées.

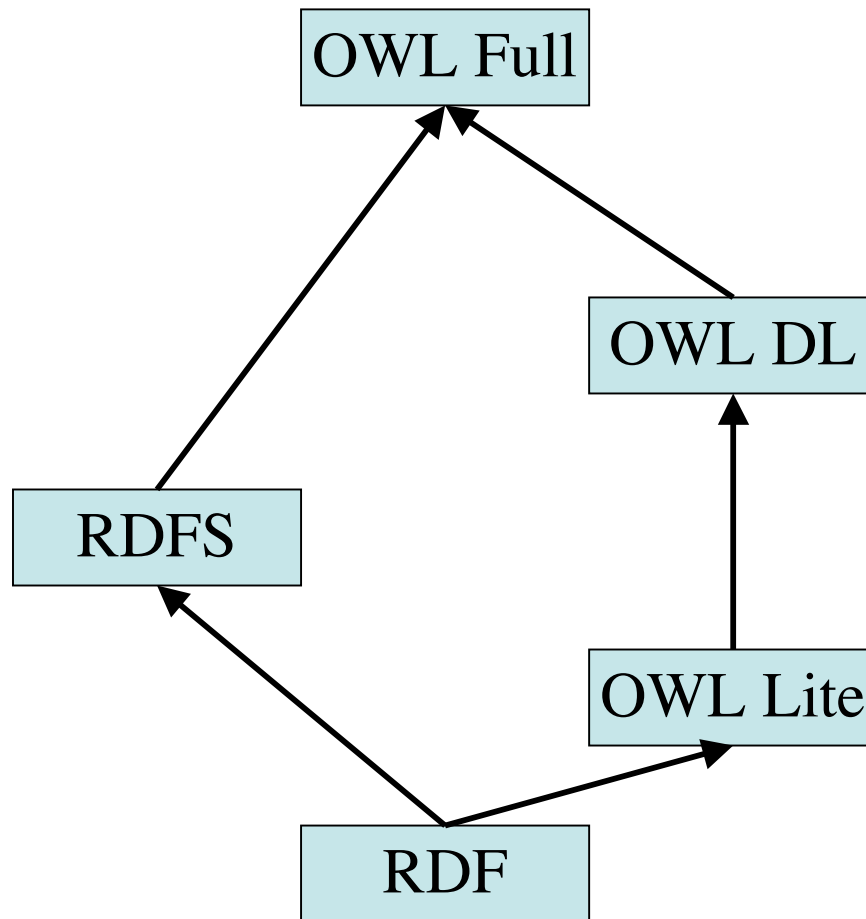
- Contraintes diverses ou axiomes restreignant l'interprétation des éléments
  - Ajoute les concepts de classes équivalentes, de propriété équivalente, d'égalité de deux ressources, de leurs différences, du contraire, de symétrie et de cardinalité
- Construction de classes complexes
  - Utilisation de connecteurs logiques
  - Définitions
  - *Ces constructions sont utilisées par les mécanismes inférentiels*

# Ontologie - OWL



# Ontologie - OWL

OWL définit trois sous-langages, du moins expressif au plus expressif :



OWL Full est destiné aux utilisateurs qui veulent un maximum d'expressivité et la liberté syntaxique de RDF mais sans garantie de calculabilité.

OWL DL assure une expressivité maximum et la complétude des calculs (toutes les conclusions peuvent être calculées) et de décidabilité (tous les calculs s'effectuent en un temps fini).

OWL Lite supporte les hiérarchies de classification et les contraintes simples.

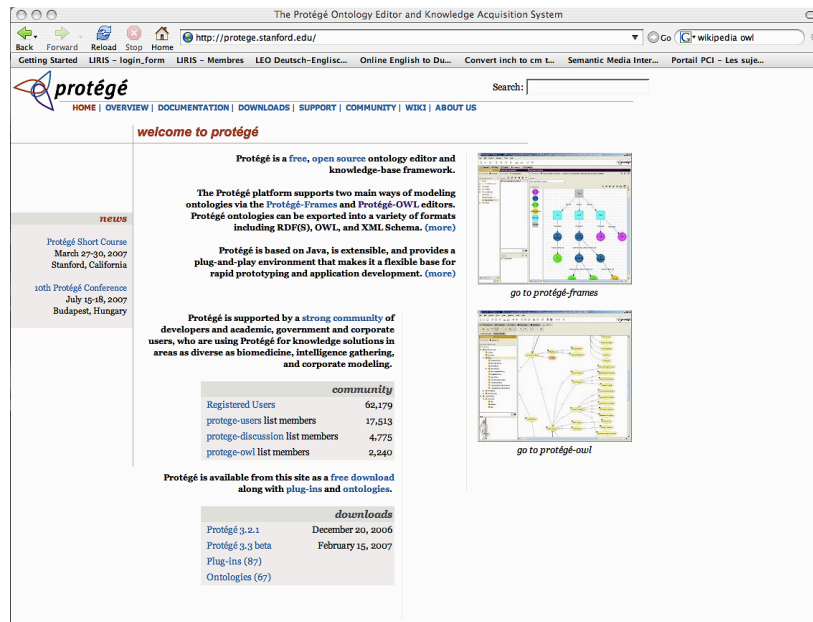


## Ontologie - application

---

- Recherche de ressources basée sur les structures relationnelles entre concepts
  - versus recherche basée sur des mots-clés
- Navigation sémantique
  - versus navigation basée sur des liens “anonymes” pour les logiciels
- Génération sémantique de ressources, personnalisation
  - versus une seule “taille” pour tout le monde
- Question - réponse en combinant des résultats
  - versus simple recherche de document
- Services
  - versus langages de descriptions de service, négociation, composition de services ...

# Annotation textuelle pour le web sémantique - Protégé



**Protégé** est un éditeur d'ontologies et un système d'acquisition de connaissances gratuit et open source.

Comme Eclipse, Protégé est un framework dans lequel on peut ajouter de nombreux plugins.


Lien: <http://protege.stanford.edu/>

# Annotation textuelle pour le web sémantique - Sesame

**Sesame** est un framework RDF open source qui supporte les inférences et les requêtes sur les schémas RDF.

Sesame a été conçu de sorte à être flexible. Il peut être déployé sur une grande variété de systèmes de stockage (bases de données relationnelles, en mémoire, systèmes de fichiers, indexation de mots-clés, etc.)

Il offre un grand nombre d'outils aux développeurs pour exploiter la puissance de RDF et des Schémas RDF, tels que l'API d'accès flexible, qui supporte à la fois les accès locaux et distants (au travers de HTTP ou RMI), et plusieurs langages de requêtes, dont SeRQL qui est le plus puissant.

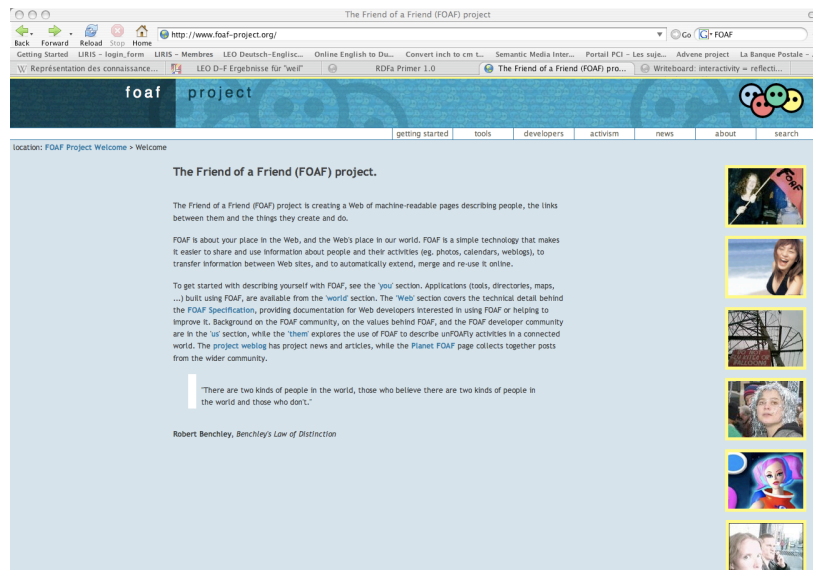


The screenshot shows the openRDF.org website with the following content:

- Navigation:** Back, Forward, Reload, Stop, Home. URL: http://www.openrdf.org/index.jsp
- Header:** openRDF.org, ...home of Sesame, open source from ADUNA
- Left Sidebar:** Site contents, News, About, License, Download, Plugins and Extensions, Documentation, Forums & mailing lists, Issue tracker, Support, Related sites, Acknowledgements, Live demo.
- Main Content:**
  - News:** Sesame 2.0-beta1 released (February 29, 2007). Text: "We are ecstatic to be able to announce the first beta release of Sesame 2.0! Sesame 2.0-beta1 marks the end of architectural changes to Sesame 2 and allows us to focus on adding features and fixing bugs..."
  - Repository, Sail and Query APIs stable:** "We have moved from alpha-stage to beta-stage, meaning that the core APIs, the interfaces and method signatures, are now frozen and stable..."
  - Improved Context Support:** "We have improved the way Sesame handles contexts, allowing developers to freely access any combination of zero, one or more contexts in a single repository..."
  - Sesame 2.0 Web Client:** "beta1 features the first release of a web client for Sesame servers..."
- Right Sidebar:** Ads by Google, RDF data access with SPARQL query language, RacerPro for Semantic Web, Edit RDF Documents, Want to move to Java?, Apache Web Services Stack.

Lien: <http://www.openrdf.org/index.jsp>

# Annotation textuelle pour le web sémantique - FOAF



**FOAF** (Friend of a friend, que l'on peut traduire par « l'ami d'un ami ») est un vocabulaire RDF permettant de décrire des personnes et les relations qu'elles entretiennent entre elles. Elle est une application du web sémantique.

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

# Annotation textuelle pour le web sémantique - FOAF

---

## Le fichier foaf.rdf de Frank

```
<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:foaf="http://xmlns.com/foaf/0.1/"
  xmlns:admin="http://webns.net/mvcb/">
<foaf:PersonalProfileDocument rdf:about="">
  <foaf:maker rdf:resource="#me"/>
  <foaf:primaryTopic rdf:resource="#me"/>
  <admin:generatorAgent rdf:resource="http://www.ldodds.com/foaf/foaf-a-matic"/>
  <admin:errorReportsTo rdf:resource="mailto:leigh@ldodds.com"/>
</foaf:PersonalProfileDocument>
<foaf:Person rdf:ID="me">
<foaf:name>Frank Nack</foaf:name>
<foaf:title>Dr</foaf:title>
<foaf:givenname>Frank</foaf:givenname>
<foaf:family_name>Nack</foaf:family_name>
<foaf:mbox_sha1sum>8f0313821db50cdafc5dbb35db477292f90b58f8</foaf:mbox_sha1sum>
<foaf:homepage rdf:resource="https://liris.cnrs.fr/membres?idn=fnack"/>
<foaf:workplaceHomepage rdf:resource="https://liris.cnrs.fr/membres?idn=fnack"/>
<foaf:knows>
<foaf:Person>
<foaf:name>Yannick Prie</foaf:name>
<foaf:mbox_sha1sum>8c4529d6f8246933a2f1738e596cc83418944a23</foaf:mbox_sha1sum></
foaf:Person></foaf:knows></foaf:Person>
</rdf:RDF>
```

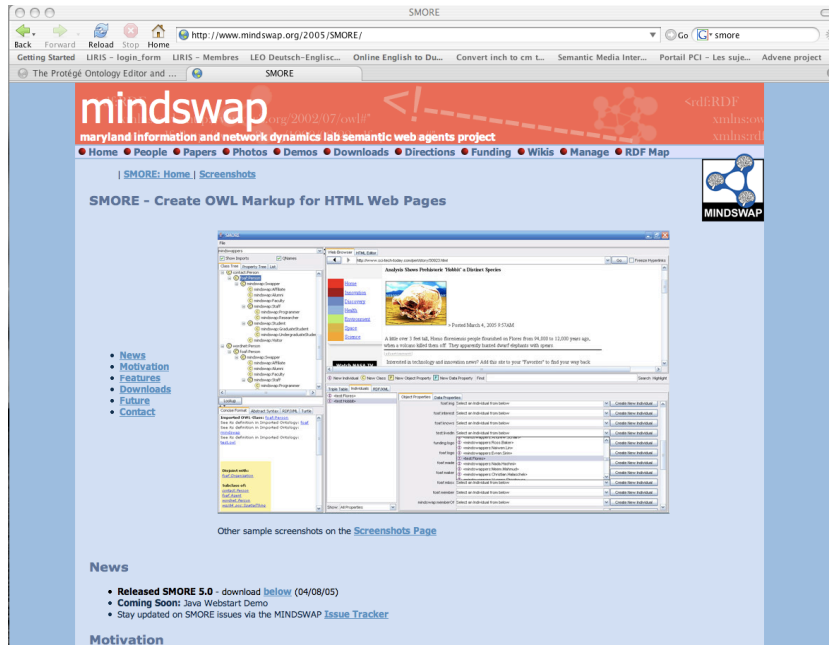
# Annotation textuelle pour le web sémantique - ESWC



Lien: <http://www.eswc2006.org/technologies/index.php>

Des application et services offrent à l'occasion de ESWC (European Semantic Web Conference)

# Annotation textuelle pour le web sémantique - Smore



Lien: <http://www.mindswap.org/2005/SMORE/>

**Smore** est un environnement développé par mindswap pour

- Permettre à l'utilisateur de marquer des documents web avec des connaissances limitées aux termes OWL et à sa syntaxe.

- Fournir un moyen d'utiliser les Classes, les Properties, et les Individuals d'une ontologie existante, de faire de l'édition d'ontologies limités et même de créer une nouvelle ontologie en utilisant les termes de documents web.