Université de Montréal

Web Sémantique

un survol

Guy Lapalme

IFT3225

rali

Production de l'information n'est plus *vraiment* un problème...

- près de 100% des documents d'affaires ou scientifiques sont produits électroniquement
- journaux
- radio et télé (streaming, podcast)
- divertissement (DVD, CD, MP3...)
- personnel
  - caméras numériques
  - e-mail
  - chat

rali

Idem pour Stockage et Transport

- Disques sont maintenant assez grands (» 10GB) pour conserver toutes nos archives personnelles
  - écrits
  - e-mails (envoyés et reçus)
  - photos
- via internet (même à la maison!!!)

rali

Problèmes engendrées par ces *solutions*...

- Recherche d'information
- Intégration de l'information
  - combinaison en tenant compte du contexte
- À tous les niveaux
  - PC, entreprise, Web

Cette problématique est qualifiée de

Partage d'information

rali

Méthodes de résolution tenant en compte la sémantique

- meta-data: auteur, titre, date...
- concepts traités par les documents
- relations entre ces concepts
  - et ceux d'autres documents
  - avec des connaissances générales
- problématique semblable à celle des images
  - format ou date
  - mais aussi contenu
    - photo satellite d'un endroit
    - photo de quelqu'un

rali

Métainformation
à traiter par la machine

- Comment rendre la sémantique disponible aux machines ?
- Comment exploiter ces informations pour chercher et intégrer l'information ?

- Besoin d'une base commune de sens entre les documents

rali

Ontologies

- Modèles formalisés de connaissances partagées dans un domaine
- Semblables aux Schémas de Base de données ou des modélisations UML sur des domaines restreints

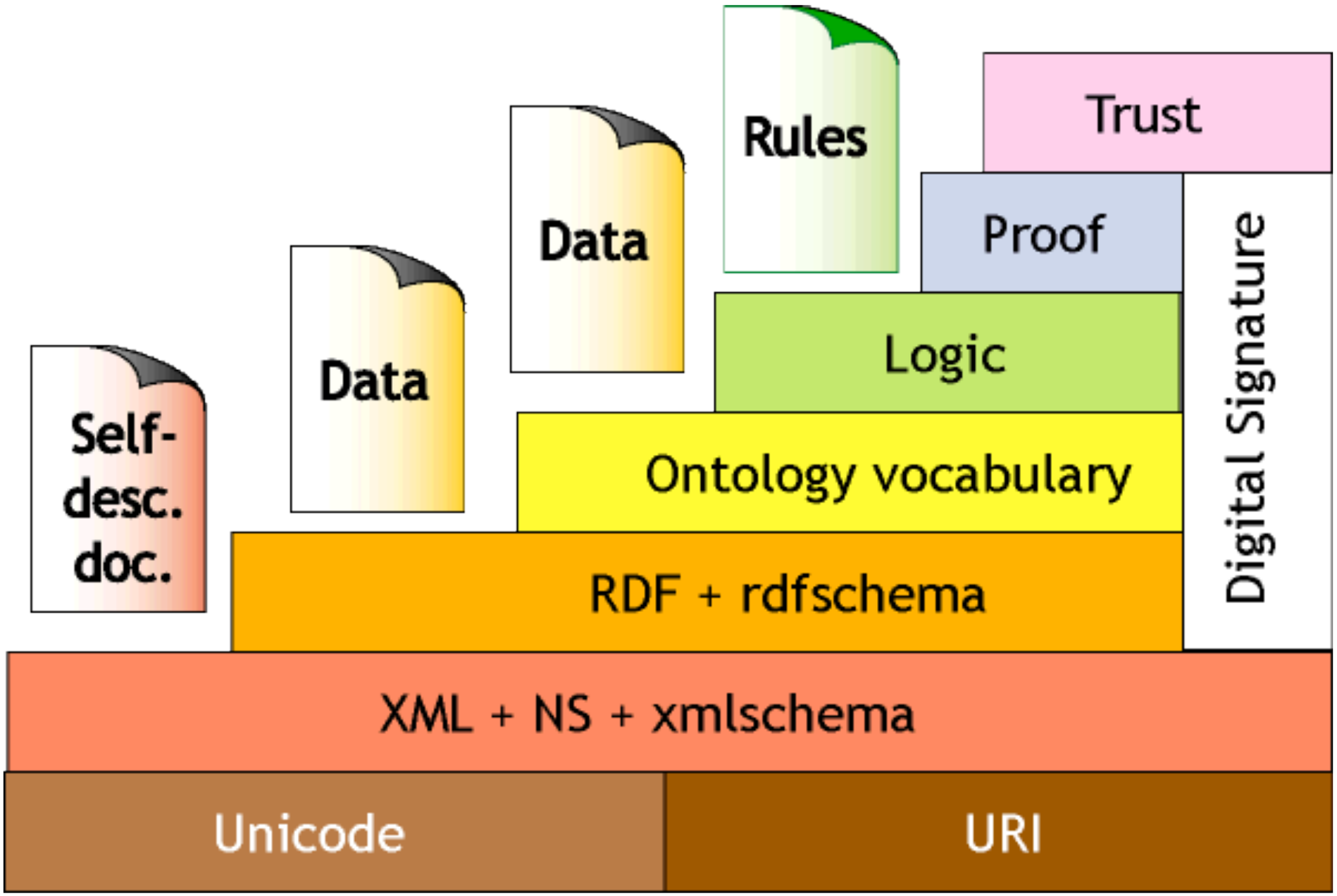
rali

Particularités des ontologies pour le Web

- Changement continuel (ajout, retrait)
  - difficile de déterminer l'information dont on veut tenir compte
- Représentations hétérogènes de contenus

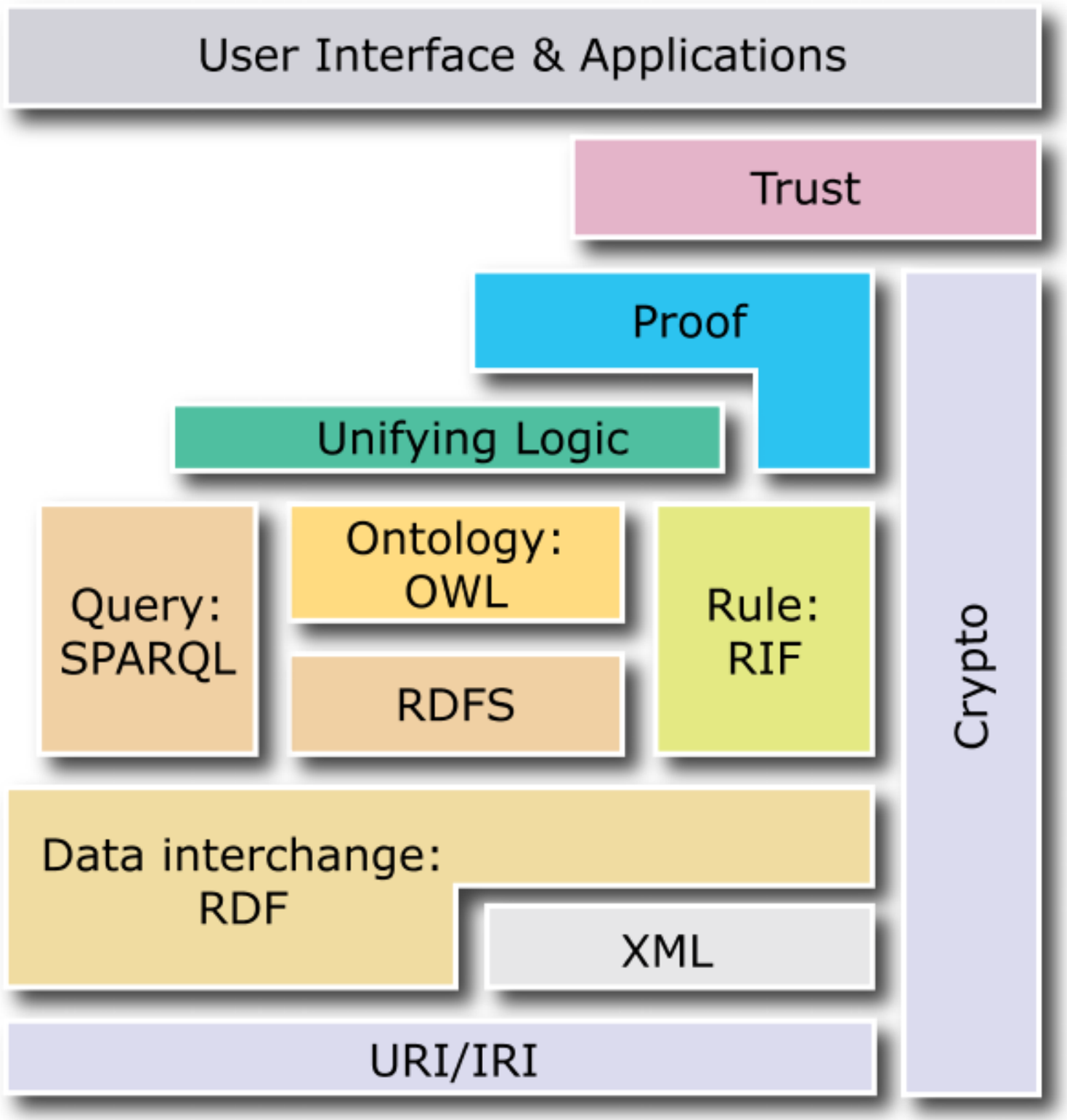
rali

Semantic Web Stack (Berners-Lee 2000)



rali

Semantic Web Stack- revised 2007



rali

RDF - Motivations

- metadata Web
- applications qui demandent des modèles ouverts d'information
- information traitable par des machines en dehors de leur environnement de création
- combinaison d'information
- traitement par des agents automatisés

rali

But de la conception

- Modèle simple de données
- Sémantique formelle et inférence prouvable
- Vocabulaire extensible
- Syntaxe à base de XML
- Support pour les types XML
- Permettre à n'importe qui d'énoncer des faits

pq implementer:
d'avoir les connaissence partagé
eviter imbiguité
modale capable de deduire

rali

RDF sur Web

- Représente de l'information (propriétés-valeurs) sur des ressources du WWW
- Vise la méta-information (e.g. titre, auteur, date de modification d'une page web)
- Identifie l'information via des
  - Uniform Resource Identifiers (URI)
  - Uniform Resource Name (URN)
    e.g : URN:ISBN:0-395-36341-1
- Information à être traitée par des applications plutôt que par des humains

Pourquoi un nouveau standard pour le Web Sémantique ?

Joshua Tauberer: http://www.rdfabout.com/intro/?section=1

- Sur le web sémantique, ce sont les ordinateurs qui *browsent:* cherchent les connaissances, les traitent et prennent action
- Web : plateforme décentralisée pour des
  - **présentations** distribuées (actuel)
  - **connaissances** distribuées (sémantique)

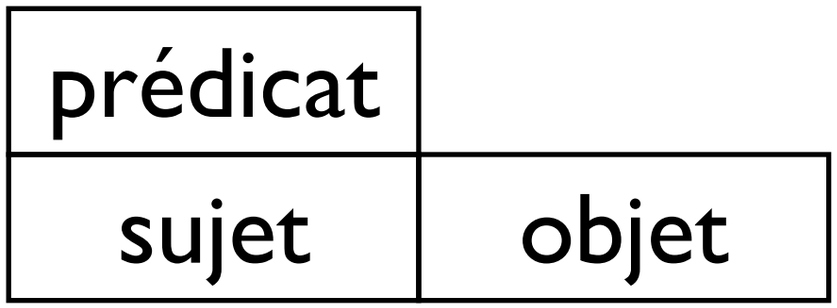
dans la web classique recherche boolean
mais dans la web semantique elle est capable de deduire des nouvelles information meme si ils sont pas codées

rali

Modèle à base de triplets

(sujet, prédicat, objet)
propriété *littéral*

prédicat(sujet,objet)




```
<rdf:Description rdf:about="sujet">
    <ex:predicate>
        <rdf:Description rdf:about="objet"/>
    </ex:predicate>
</rdf:Description>
```

RDF/XML

```
sujet ex:predicate objet.
```

Triplet - Turtle

Composantes

- Noeud
  - URI Reference (URIRef)
  - littéral
  - vide (nom local et arbitraire)
- Prédicat / propriété entre **deux** noeuds
  - URIRef
- Types de XML-Schema
- Littéral
  - identifié par sa représentation lexicale
  - peut être objet mais non sujet

rali

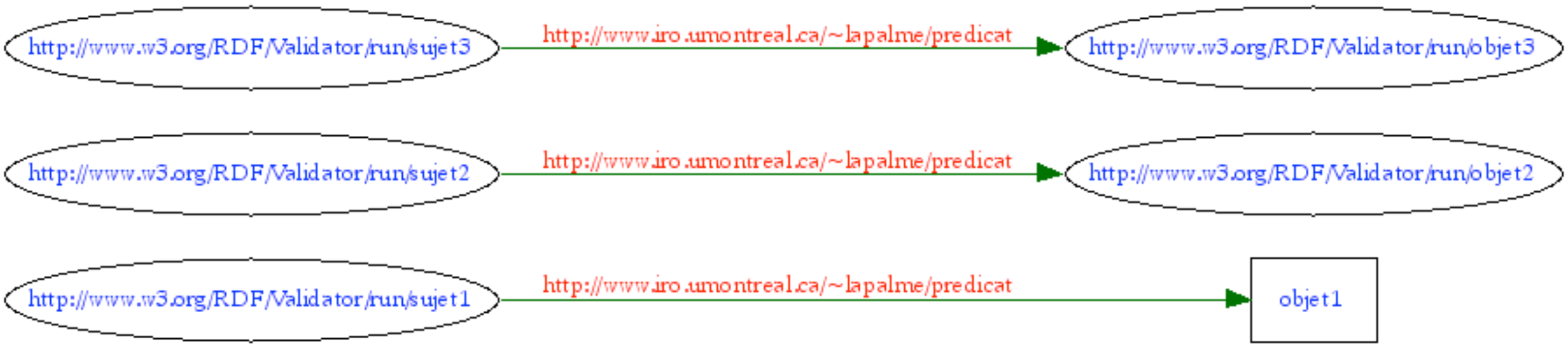
Encodages XML

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:gl="http://www.iro.umontreal.ca/~lapalme/">
   <!-- gl:predicat(sujet1,"objet1") -->
   <rdf:Description rdf:about="sujet1">
         <gl:predicat>objet1</gl:predicat>
   </rdf:Description>

   <!-- gl:predicat(sujet2,objet2) -->
   <rdf:Description rdf:about="sujet2">
      <gl:predicat>
         <rdf:Description rdf:about="objet2"/>
      </gl:predicat>
   </rdf:Description>

   <!-- gl:predicat(sujet3,objet3) -->
   <rdf:Description rdf:about="sujet3">
      <gl:predicat rdf:resource="abjet3"/>
   </rdf:Description>

</rdf:RDF>
```

rali

Graphes et triplets résultants

http://www.w3.org/RDF/Validator/




```
http://www.w3.org/RDF/Validator/run/sujet1
   http://www.iro.umontreal.ca/~lapalme/predicat
   "objet1"

http://www.w3.org/RDF/Validator/run/sujet2
   http://www.iro.umontreal.ca/~lapalme/predicat
   http://www.w3.org/RDF/Validator/run/objet2

http://www.w3.org/RDF/Validator/run/sujet3
   http://www.iro.umontreal.ca/~lapalme/predicat
   http://www.w3.org/RDF/Validator/run/objet3
```

rali

Encodage *Turtle*

Terse RDF Triple Language

```
@prefix ex:  <file:///Users/glapalme/Desktop/GeneveFevMars2010/RDF/> .
@prefix gl:  <http://www.iro.umontreal.ca/~lapalme/> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

ex:sujet1 gl:predicat "objet1" .
ex:sujet2 gl:predicat ex:objet2 .
ex:sujet3 gl:predicat ex:objet3 .
```

Comparaison RDF et XML

- RDF est construit *par-dessus* XML
- XML est un arbre, RDF ensemble de triplets
- XML est ordonné, RDF sans ordre
- RDF permet de ne comprendre qu'un sous-ensemble
- RDF est plus facile à interroger
- XML est syntaxique, RDF est sémantique
- RDF permet de *déduire*
- RDF et Ontologies permettent aux programmes de déduire
- RDF permet de s'abstraire de la syntaxe des documents
- RDF est plus compliqué et demande une planification
- Pas tous les projets ont besoin de RDF...

rali

Outils de traitement RDF

- Java : Jena (issu des travaux du HP Labs UK)
- C : Redland avec interface pour Perl, PHP, Python, Ruby
- Gestion efficace de triplets
- Semantic Web Development Tools
  - http://www.w3.org/2001/sw/wiki/Tools

rali

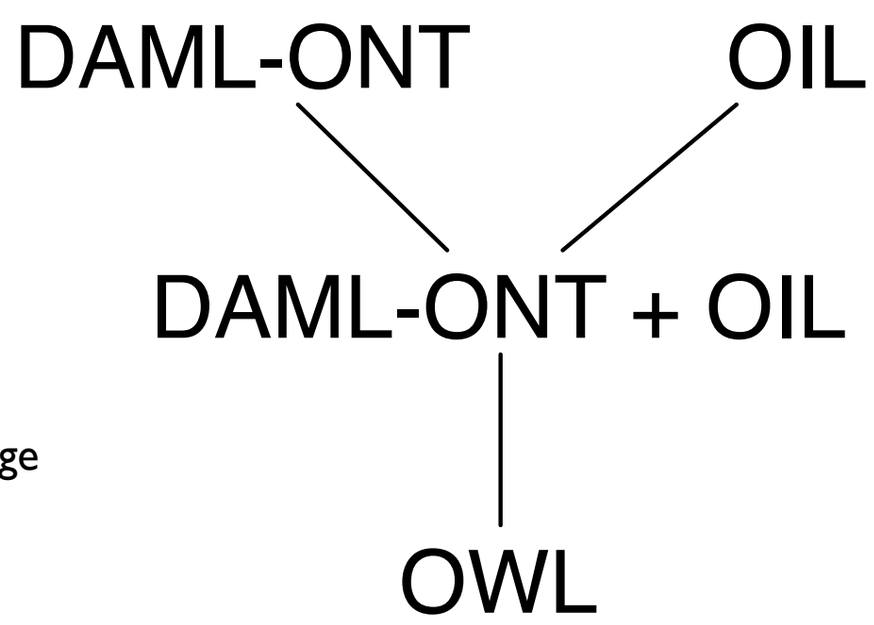
Avantages des ontologies *structurées*

- vérification de consistance
- complétion de l'information
- interopérabilité
- support à
  - validation et au test
  - configuration
  - recherche coopérative et structurée
- exploiter
  - la généralisation
  - la spécialisation

rali

OWL motivations

- expressivité de RDF et RDFS est limitée
  - RDF : prédicats binaires
  - RDFS : hiérarchie de sous-classes et de propriétés
- besoin d'exprimer des relations plus riches



DAML: DARPA Agent Markup Language

rali

Exigences pour un langage d'ontologie

- Syntaxe bien définie (RDF/XML)
- Sémantique formelle (*Description Logic)*
  - permet de raisonner sur
    - appartenance à une classe
    - équivalence de classe
  - permet de vérifier
    - consistance de l'ontologie
    - possibilités de relations non voulues
  - classification automatique des instances

rali

OWL : Web Ontology Language

- Représentation de connaissances riches et complexes à propos de
  - *choses*
  - groupes de *choses*
  - relations entre *choses*
- Basé sur une logique *calculatoire* permettant
  - vérifier la consistance des connaissances
  - expliciter des connaissances implicites
- Documents OWL peuvent être liés entre eux

rali

Principes

- Langage **déclaratif** pour exprimer des ontologies
- **Ce n'est pas**
  - un langage de schéma
    - peut pas forcer l'apparition de certaines informations
  - un modèle de base de données
    - monde ouvert plutôt que fermé : une information manquante peut être vraie

une BD peut toutefois servir d'infrastructure pour conserver l'ontologie

rali

Modélisation des connaissances

- **Axiomes** : énoncés de base supposés vrais
- **Entités** : référents aux objets du monde
- **Expressions** : combinaisons d'entités pour former des descriptions complexes à partir de formes de base

Le résultat de la modélisation est appelé **ontologie**

rali

Représentation de connaissances

- Énoncés de base
  - *il pleut*
  - *tout homme est mortel*
- Conséquences des énoncés
  - un énoncé *a* est vrai si d'autres *A* le sont
  - *A* entraîne (*entails*) *a*
  - *A* est *consistant* s'il y a une situation où tous ses énoncés sont vrais
  - *A* est *inconsistant* si on ne peut trouver de situation où tous ses énoncés sont vrais
- Sémantique formelle définit les états pour lesquels un ensemble d'énoncés sont vrais

rali

Raisonnement à partir des axiomes

- Calcul automatique des conséquences d'un ensemble d'axiomes
- Outils sont appelés *reasoners*
- Pas toujours facile à contrôler ou à en comprendre les résultats

rali

Outils

- OWL-API : interface Java
- Editeurs d'ontologie
  - Protégé, SWOOP
  - TopBraid Composer (Commercial)
- Raisonneurs
  - Fact++ (Manchester)
  - Pellet
  - RacerPro

rali

État actuel du Web Sémantique

- Encore en mouvement
- Standards encore en évolution
- Commence à entrer dans les moeurs
- *Retro-ingénierie* difficile

rali

SKOS+FOAF

- réseaux d'information distribués et extensibles à cause de leur base RDF
- FOAF
  - approche évolutive pour l'extension d'information
  - facilement extensible par n'importe qui
- SKOS
  - approche plus organisée
  - standardisé par un comité du W3C

rali

RDFa

attributs XHTML pour supporter RDF

- expression de données structurées dans un langage de balisage
- rendre le texte et les liens HTML accessibles aux machines
- ne pas répéter de contenu
- éviter d'avoir à distribuer séparément le contenu lisible par la machine
- règles de traitement pour produire des triplets RDF à partir du XHTML+RDFa

rali

Bibliographie

- G. Lapalme, *Looking at the Forest instead of the Trees*, tutoriel XML
- D. Allemang et J. Hendler, *Semantic Web for the Working Ontologist*, Morgan Kaufmann, 2008.
- G. Antoniou et F van Harmelen, *A Semantic Web Primer*, MIT Press, 2009, 2nd ed.
- Pascal Hitzler, Markus Krötzsch, Sebastian Rudolph, *Foundations of Semantic Web Technologies*, Chapman & Hall/CRC, 2009.
- http://www.w3.org/2001/sw/

rali