

Theoretische Informatik und neue Anwendungen



Le Web Sémantique – Technologies

 **Universität Trier**

Serge Linckels

Université du Luxembourg, FSTC, 4 novembre 2004



Vers la Crise du Web ?

Universität Trier

- Chacun « réalise » des pages Web
 - HTML hardcoding en Notepad
 - Conversion (Word, Excel...)
 - Avec un outil WYSIWYG (Dreamweaver, Frontpage...)
- Problèmes :
 - Aucune rigueur → efforts du W3C



- Problèmes sémantiques



Web Classique vs Web Sémantique

Universität Trier

- Service de l'Internet
- Existe depuis 1991
- Évolution peu contrôlée
- Basé sur divers langages
- Absence de sémantique
- Moteurs de recherche basés sur mots-clés

- Pas limité sur l'Internet
- Première idée en 1993
- Évolution planifiée et contrôlée
- Basé sur un langage commun
- Basé sur sémantique
- Moteurs de recherche sémantiques

Moteurs de recherche qui comprennent l'utilisateur

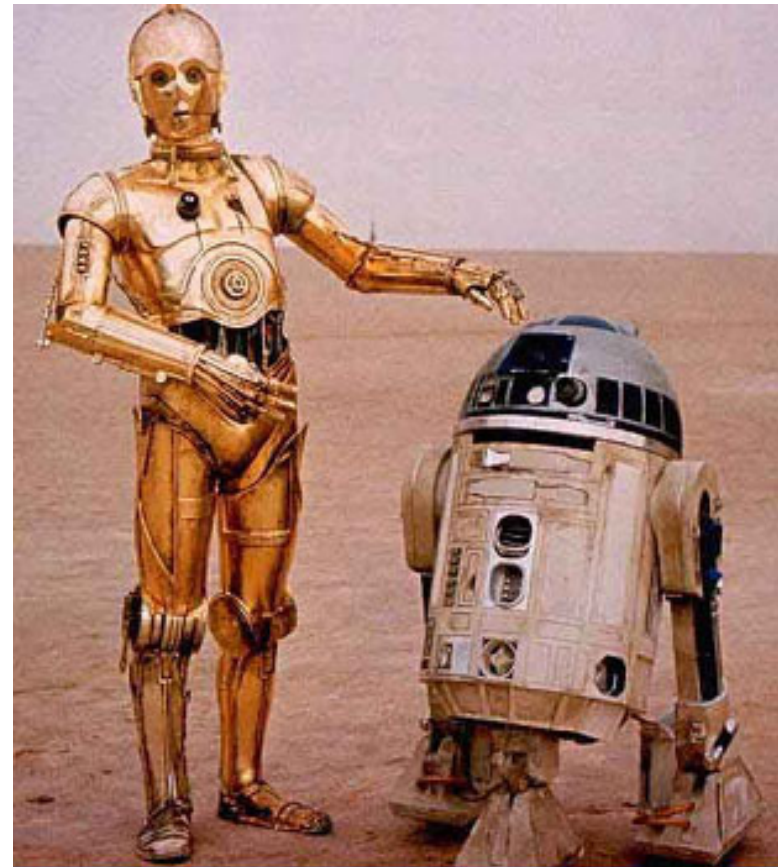
Moteurs de recherche intelligents → **Intelligence Artificielle**



Le Rêve d'Ordinateurs Intelligents

Universität Trier

- Sujet à la mode des années 70 et 80, p.ex. robotique
- Principes :
 - Une machine se comporte comme un humain
 - Une machine comprend un humain
- **Tentative échouée de rendre intelligentes les machines stupides**





Réflexions sur l'Intelligence

Universität Trier

- Qu'est-ce que l'intelligence ?
- Exemple de *Deep Blue* :
 - Super-Ordinateur de jeu d'échec d'IBM (10 billions d'opérations par seconde)
 - 1990 : Garry Kasparow proclame qu'aucun ordinateur ne va jamais le battre
 - 1997 : Première défaite pour l'humain
- « L'intelligence » des ordinateurs se définit par leur programmation
- *Le vrai problème n'est pas de savoir si les machines pensent, mais de savoir si les Hommes pensent*



La Sémantique, un Nouveau Début

Universität Trier

- Ne pas comprendre les données, mais le sens des données
- Exemple : Le contenu d'un livre est saisi. La machine sait retrouver la page où se trouve la réponse à un problème donné.
- **Les données sont structurées de sorte que les machines stupides peuvent les comprendre** 👍



Première Technologie : XML

Universität Trier

- Vers 1993 : HTML est insuffisant
 - Pas suffisamment d'éléments « tag »
 - Interprétations différentes suivant Browser
 - Pas de description sémantique
- Extension de HTML → XML (1996 par Jon Bosak)
- Caractéristiques :
 - Méta-Langage
 - Création de règles sur les données
 - Moyen indépendant de d'échange de données



XML Apporte de la Structure

Universität Trier

- Représentation classique (texte pur) :

Serge
Linckels
Les Générations Pascal

 Structure ?

- Représentation XML :

```
<Auteur>  
  <Nom>  
    <Prénom>Serge</Prénom>  
    <NomFam>Linckels</NomFam>  
  </Nom>  
  <AuteurDe>Les Générations Pascal</AuteurDe>  
</Auteur>
```

 Sémantique ?



Le Domaine de la Sémantique

Universität Trier

- Sémantique = Décrire le sens de quelque chose
 - Ajoute des données qui décrivent un contenu
 - Données sur des données = **Metadata**

- Application : Semantic Web
 - Décrire le contenu de pages web
 - Rend plus efficaces des moteurs de recherche

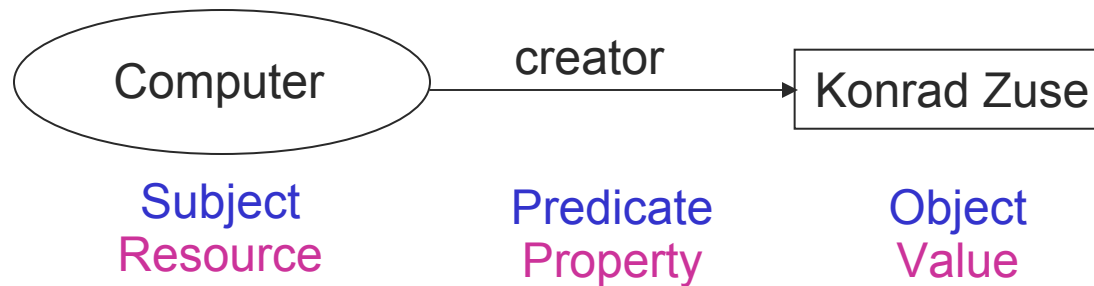
- Application : e-Commerce
 - Comprendre les souhaits des clients



Deuxième Technologie : RDF

Universität Trier

- Recommendation W3C
 - **RDF** (*Resource Description Framework*)
 - Introduit en 1998 pour construire le **Semantic Web**
- Exemple d'un modèle RDF (*directed graph*):





Formaliser la Sémantique

Universität Trier

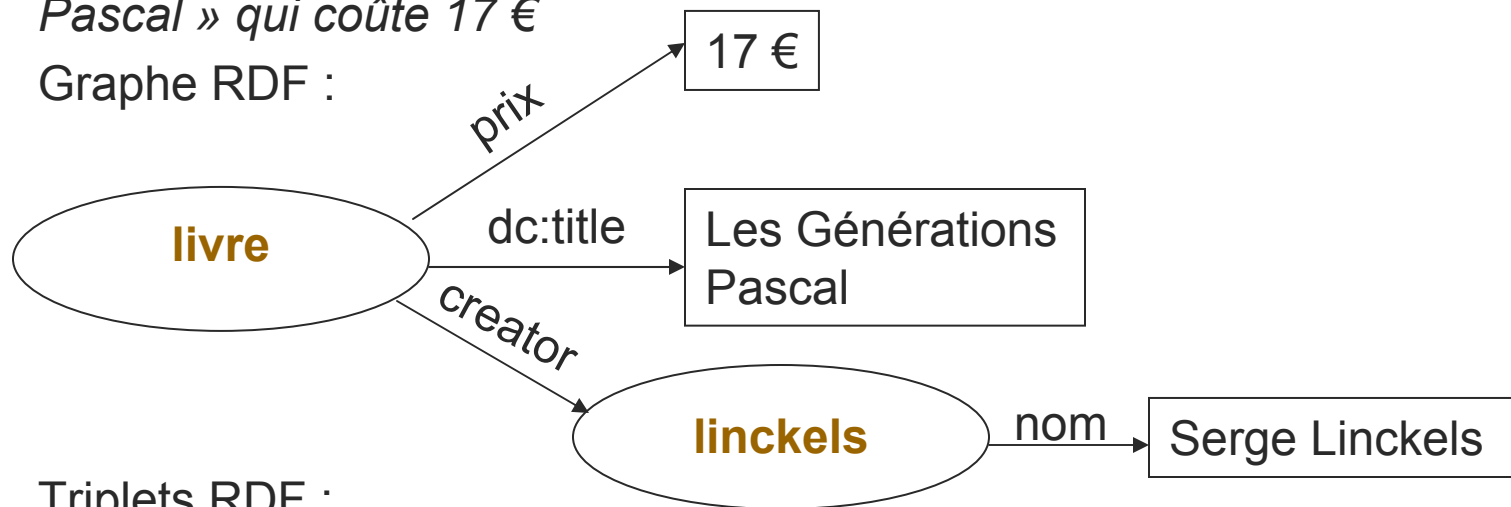
- RDF est un mécanisme pour exprimer des affirmations sur des **ressources**, de façon que des machines peuvent les interpréter
- Une affirmation (anglais : *statement*) est appelé **triplet** (anglais : *triple*)
- L'ensemble des triplets concernant une ressource décrit la sémantique de cette ressource



Exemple RDF

Universität Trier

- Affirmation : *Serge Linckels est l'auteur du livre « Les Générations Pascal » qui coûte 17 €*
- Graphe RDF :



- Triplets RDF :

Ressource	Propriété	Valeur
livre	prix	17 €
livre	dc:title	Les Générations Pascal
livre	creator	linckels
linckels	nom	Serge Linckels



Structure RDF

Universität Trier

- Triplet = (Ressource, Propriété, Valeur)
- Valeur = (Ressource | Littéral)
- Une ressource est décrite par une ou plusieurs autres ressources
Exemple : *La ressource **livre** a un créateur, la ressource **linckels***
- Une ressource est décrite par un ou plusieurs littéraux
Exemple : *La ressource **livre** a un titre « Les Générations Pascal » et un prix « 17 € »*



Le Problème des Synonymes

Universität Trier

- Pas de problème de synonymes
 - Serge Linckels a écrit le livre « Les Générations Pascal »
 - L'auteur du livre « Les Générations Pascal » est Serge Linckels
 - Le fameux Monsieur Linckels est l'écrivain de l'excellent manuel « Les Générations Pascal »



- Pas besoin d'un dictionnaire centralisé
 - Utilisation de *Namespaces*



Encodage

Universität Trier

- Plusieurs possibilités d'encodage (*serialization*), par exemple RDF/XML :

```
<?xml version="1.0"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:dc="http://purl.org/dc/elements/1.1/">
  <rdf:Description rdf:about="ISBN:2-495-10603-X">
    <prix>17 €</prix>
    <dc:title>Les Générations Pascal</dc:title>
    <dc:creator>
      <rdf:Description rdf:about="http://www.linckels.lu">
        <nom>Serge Linckels</nom>
      </rdf:Description>
    </dc:creator>
  </rdf:Description>
</rdf:RDF>
```



Besoin de Règles

Universität Trier

- Rappel :
 - XML apporte une structure aux données
 - RDF apporte une syntaxe aux données
- **RDF Schema** permet de définir des règles comment utiliser les données
- Comment ?
 - Permet de créer un vocabulaire
 - Permet de décrire les objets et les propriétés des objets
 - Permet de créer une hiérarchie de classes



Éléments de RDF Schema

Universität Trier

- Les éléments en RDF Schema permettent de définir:
 - des classes (`rdfs:Class`)
 - des littéraux (`rdfs:Literal`)
 - des ensembles (`rdfs:Container`)
 - des types de données (`rdfs:Datatype`)
 - ...

- Exemple : Un livre est écrit par un auteur
 - Livre(titre, prix)
 - Auteur(nom)



Exemple : Définition des Classes

Universität Trier

■ La classe Livre :

```
<rdfs:Class rdf:about="http://exemple.net/elements#Livre">  
  <rdfs:isDefinedBy rdf:resource="http://exemple.net/elements#" />  
  <rdfs:subClassOf rdf:resource="http://www.w3.org/2000/01/rdf-schema#Resource" />  
  <rdfs:label>Livre</rdfs:label>  
  <rdfs:comment>Un livre écrit par un auteur</rdfs:comment>  
</rdfs:Class>
```

■ La classe Auteur :

```
<rdfs:Class rdf:about="http://exemple.net/elements#Auteur">  
  <rdfs:isDefinedBy rdf:resource="http://exemple.net/elements#" />  
  <rdfs:subClassOf rdf:resource="http://www.w3.org/2000/01/rdf-schema#Resource" />  
  <rdfs:label>Auteur</rdfs:label>  
  <rdfs:comment>Une personne qui a écrit (au moins) un livre</rdfs:comment>  
</rdfs:Class>
```



Exemple : Définition des Propriétés

Universität Trier

■ Propriétés pour la classe Livre :

```
<rdf:Property rdf:about="http://exemple.net/elements#titre">
  <rdfs:isDefinedBy rdf:resource="http://exemple.net/elements#" />
  <rdfs:domain rdf:resource="http://exemple.net/elements#Livre" />
  <rdfs:label>Le titre du livre</rdfs:label>
</rdf:Property>
```

```
<rdf:Property rdf:about="http://exemple.net/elements#prix">
  <rdfs:isDefinedBy rdf:resource="http://exemple.net/elements#" />
  <rdfs:domain rdf:resource="http://exemple.net/elements#Livre" />
  <rdfs:label>Le prix du livre en Euro</rdfs:label>
</rdf:Property>
```

■ Propriétés pour la classe Auteur :

```
<rdf:Property rdf:about="http://exemple.net/elements#nom">
  <rdfs:isDefinedBy rdf:resource="http://exemple.net/elements#" />
  <rdfs:domain rdf:resource="http://exemple.net/elements#Auteur" />
  <rdfs:label>Le nom (nom et prénom) de l'auteur</rdfs:label>
</rdf:Property>
```



Besoin d'un Langage plus Puissant

Universität Trier

- Rappel :
 - XML apporte une structure aux données
 - RDF(S) apporte une syntaxe aux données

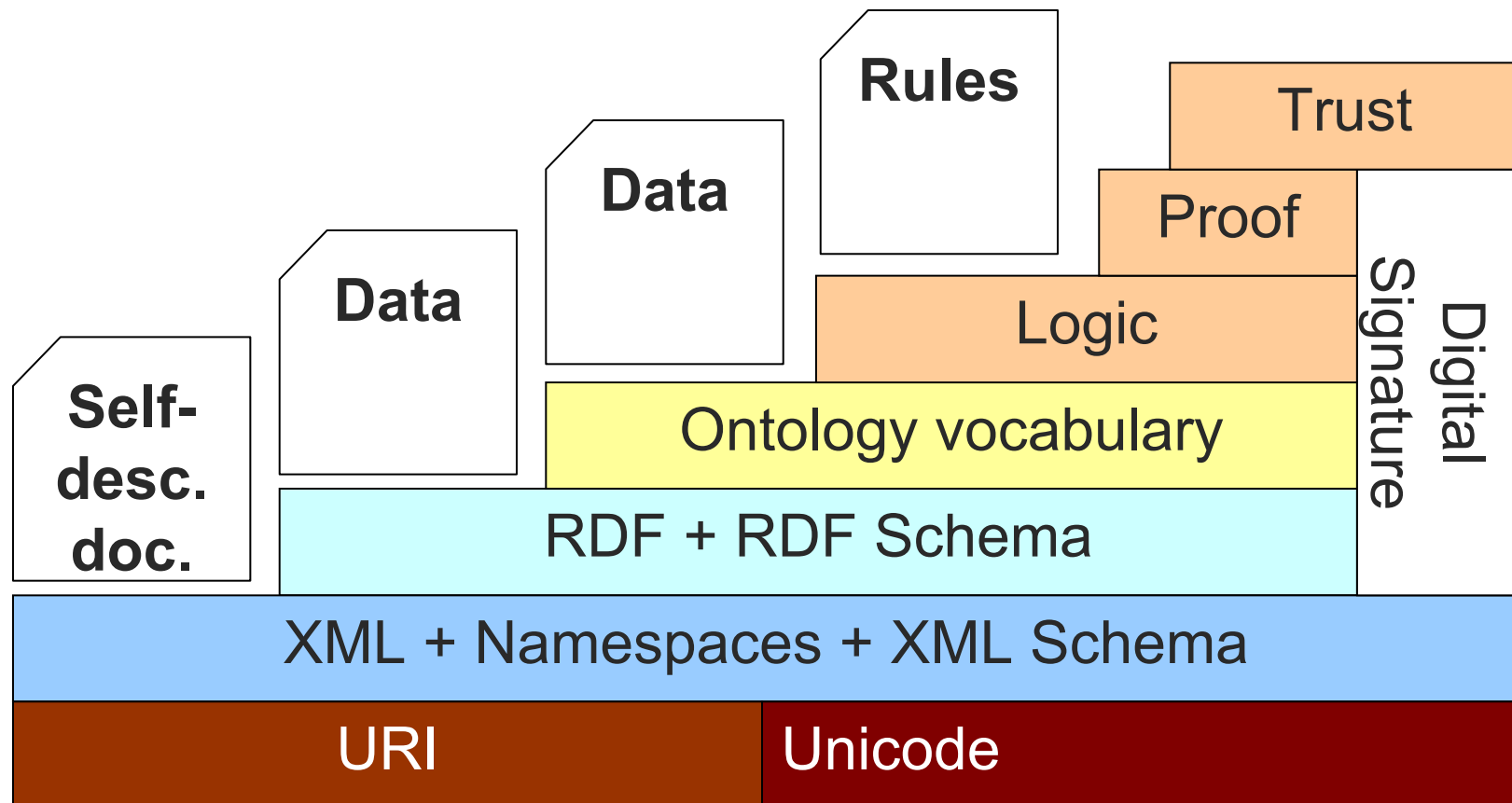
- Problèmes :
 - Pauvre pour exprimer des contraintes ou des règles (p.ex. restrictions → cardinalité, inégalité,...)
 - Impossible d'effectuer des raisonnements sur les données

→ Il faut un langage plus puissant à un niveau supérieur



Rappel de l'Architecture en Couches

Universität Trier





Notion d'Ontologie

Universität Trier

- Définition :

Ontologies are formal and consensual specifications of conceptualizations that provide a shared understanding of a domain, an understanding that can be communicated across people and application systems

(Fensel D., *Ontologies: A Silver Bullet for Knowledge Management and Electronic Commerce*, Springer, Berlin-Heidelberg-New York, 2004)

- Explication :

Une ontologie est tout ce qu'il faut pour que des humains et des machines peuvent décrire, comprendre et raisonner sur un sujet donnée



Langages Ontologiques

Universität Trier

- Diverses spécifications et recommandations récentes :
 - **OWL** (*Web Ontology Language*) : recommandation W3C qui se base sur RDF(S)
 - **OIL** (*Ontology Interface Layer*) : essai d'améliorer RDF Schema
 - **DAML** (*DARPA Agent Markup Language*) : langage crée pour enrichir XML et RDF
 - **DAML+OIL** : alternative à OWL qui se base sur XML et OIL, plutôt que sur RDF(S)
 - **RDFS(FA)** (*Fixed layered metamodeling Architecture*) : proposition pour éliminer les lacunes et failles dans RDF Schema par un langage plus évolué



Troisième Technologie : OWL

Universität Trier

- Permet de définir des ontologies
- Permet de définir des règles et une syntaxe pour une ontologie donnée (mieux que RDFS)
- Permet de définir des restrictions sur des objets et propriétés (p.e. cardinalité, valeur,...)
- Permet de raisonner sur des données



Exemple de Restriction en OWL

Universität Trier

- Exemple :
Tout auteur a écrit au moins un article (notion de cardinalité)

```
<owl:Class rdf:about="http://exemple.net/elements#Auteur">  
  <rdfs:subClassOf rdf:resource="http://www.w3.org/2000/01/rdf-schema#Resource" />  
  <rdfs:subClassOf>  
    <owl:Restriction>  
      <owl:minCardinality rdf:datatype="&xsd;nonNegativeInteger">1</owl:minCardinality>  
    </owl:Restriction>  
  </rdfs:subClassOf>  
</owl:Class>
```



Raisonner sur les Connaissances

Universität Trier

- 3 types différents d'OWL
 - OWL Lite : très limité dans les restrictions
 - **OWL DL** : plus de liberté dans les restrictions et de raisonnement
 - OWL Full : pas de limitation mais pas de garanties de calculabilité

- OWL DL : *Description Logic*
 - Basé sur la logique de prédicats du premier ordre
 - Permet d'effectuer des raisonnements dans un temps fini
 - Cependant, certaines limitations d'expressivité



Notions de Raisonnement

Universität Trier

1. Raisonnement transitif : $P(x,y) \wedge P(y,z) \rightarrow P(x,z)$
 - Règle : Tous les humains ont 2 yeux.
 - Fait : Jean est un humain
 - Conclusion calculée : Jean a 2 yeux.

2. Raisonnement symétrique : $P(x,y) \leftrightarrow P(y,x)$
 - Le livre « Les Générations Pascal » fut créé par « Serge Linckels »
 - Donc, « Serge Linckels » a écrit le livre « Les Générations Pascal »



Notions de Raisonnement

Universität Trier

3. Raisonnement fonctionnel : $P(x,y) \wedge P(x,z) \rightarrow x = z$
4. Raisonnement inverse : $P1(x,y) \rightarrow P2(y,x)$
5. Raisonnement fonctionnel inverse : $P(y,x) \wedge P(z,x) \rightarrow y = z$



Conclusion sur le Semantic Web

Universität Trier

- Le Semantic Web :
 - Est une vision d'un nouveau web ; le SW n'existe pas (encore)
 - Plusieurs technologies ont été et vont être créées et améliorées en vu de construire le SW

- Aujourd'hui :
 - Des technologies du SW sont utilisées pour construire des pages WWW ou pour concevoir des applications
 - C'est un domaine de recherche actuel



Prochaine Séance

Universität Trier

- jeudi, 16 décembre (11:45 h – 13:00 h)
 - Présentation des avancées dans le domaine du traitement de la langue naturelle sur machine
 - Présentation de l'outil d'éducation CHESt
 - moteur de recherche sémantique
 - interaction avec l'utilisateur sur base de la langue naturelle