

## Chapitre II: le Web Sémantique

### II.2 Le web sémantique

# Plan du cours

- Introduction
- Le web visuel
- Le web syntaxique
- Le web sémantique
- **La sémantique**
- **Niveaux de la sémantique**
  - Le vocabulaire contrôlé
  - taxinomie
  - Thesaurus
  - Ontologie
- conclusion

# Introduction

- Le web actuel est composé essentiellement des page HTML qui est écrit en HTML (ensemble de balises contenues dans des pages web ).
- Les volumes d'information sur le web augmentent considérablement, ce qui renforce le besoin de partager, d'échanger et de réutiliser ces informations d'une manière efficace entre les utilisateurs.
- La plupart des informations se trouvant sur Internet sont effectivement lisibles par les hommes mais souvent difficilement interprétables par les machines.

# Le web visuel

- Dans le web visuel les machines(applications logicielles :qqs moteurs de recherche) ne comprennent plus les informations contenues dans les pages HTML .les mots clés (métadonnées en HTML balisé par <meta>) ne tiennent pas compte de la sémantique qui les lient entre eux.
  - Ainsi, une recherche de plusieurs mots ne donne pas nécessairement les bons résultats car le moteur ne tient pas compte de la sémantique associée à l'ensemble de ces mots
- Limites du HTML

# Le web visuel

- Pourquoi un moteur de recherche basé sur des mots clés ne détient pas forcément la sémantique générale associée à la requête utilisateur ?
- La raison principale est que la description des informations (les métadonnées) est limitée, ou alors non exploitée d'une manière plus "intelligente" par les machines

# Le web visuel

- C'est quoi une métadonnée ?
- Les métadonnées sont des données (structure de données) qui définissent ou décrivent d'autres données.
- Exemple :
  - Les métadonnées associées à un fichier :
  - Nom, titre, taille, auteur.....
- Les métadonnées sont un ensemble de paires :attribut-valeur qui fournissent des informations sur d'autres informations

# Le web syntaxique

- Le web syntaxique est un ensemble de documents non seulement pour but d'affichage mais également pour l'intégration et l'interopérabilité
- Pour comprendre les données, les machines ont besoins de métadonnées fournit en plus, XML est le langage le plus répondu pour ça
- XML est langage souple qui permet de définir de nouvelles balises personnelles afin de donner un sens aux données

# Le web syntaxique

- Exemple
- <etudiant>
- <id> 12</id>
- <nom> zian farid<nom>
- <specialité>informatique < specialité >
- </etudiant>
- aucune information ni fournie sur le sens de :Etudiant, Id ,nom, prenom, specialité

# Le web sémantique

## définitions

- Le Web sémantique est une extension du Web qui facilite l'automatisation du traitement des connaissances disponibles. C'est une extension du Web classique où les connaissances ne sont pas représentées dans une langue naturelle mais formalisées à l'aide de langages pouvant être interprétés par des machines

# Le web sémantique

## définitions

- ensemble de technologies visant à rendre le contenu des ressources du web accessible et utilisable par les programmes, agents logiciels et machines, grâce à un système de métadonnées formelles (données servant à décrire d'autres données), utilisant notamment la famille de langages développés par le World Wide web Consortium (W3C).

# Le web sémantique

## définitions

- La nouvelle génération du web –le web sémantique- a pour objectif de rendre possible le traitement de l'information web par les ordinateurs
- Le web sémantique a pour ambition de rendre les ressources web plus aisément accessibles et compréhensibles par les humains que par les machines grâce à la représentation sémantique de leurs contenus.

# Le web sémantique

## définitions

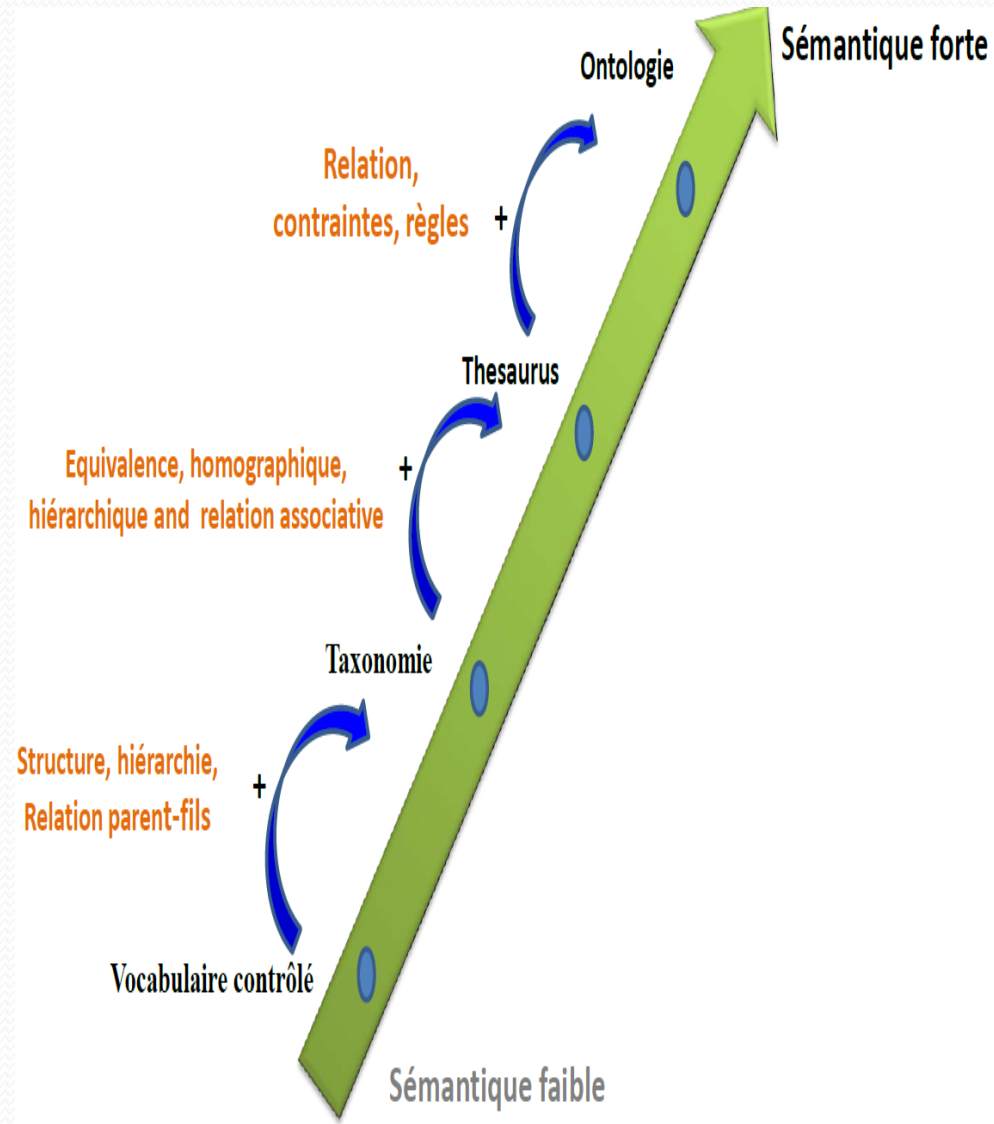
- Il faut fournir aux ordinateurs des informations qu'ils peuvent comprendre.
- Le web sémantiques doit combiner les technologies du web avec les formalismes de représentation de connaissances

# La sémantique

- La sémantique et l'étude de sens des signes, telles que les mots et les termes .

# Niveaux de la sémantique

- Il existe 4 représentations qui peuvent être utilisés pour organiser et modéliser les termes et les concepts sémantiquement :
  - Le vocabulaire contrôlé
  - taxinomie
  - Thesaurus
  - ontologie



# Niveaux de la sémantique

## 1. Le vocabulaire contrôlé

- Le niveau sémantique le plus bas
- Une liste de termes (mots, phrases, notation) énumérés explicitement
- Tout terme dans le vocabulaire doit avoir une définition non ambiguë et non redondante
- En métadonnées, utilisé dans un but de classification

# Niveaux de la sémantique

## 1. Le vocabulaire contrôlé

### Exemple :

Amazon.com a le vocabulaire contrôlé (voir tableau ci-dessous) qui peut être sélectionné par les utilisateurs afin de réaliser une recherche de produits.

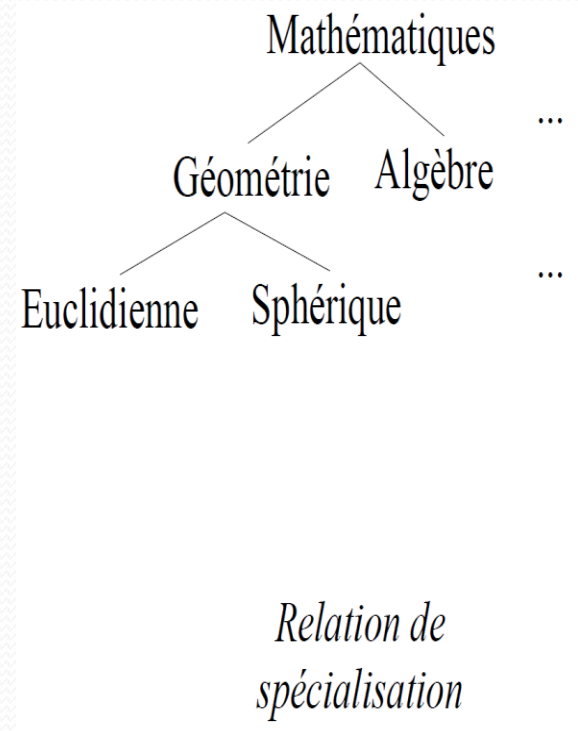
| Books           | Electronics       | Travel                |
|-----------------|-------------------|-----------------------|
| Popular Music   | Camera & Photo    | Cell Phones & Service |
| Music Downloads | Software          | Outlet                |
| Classical Music | Tools & Hardware  | Auctions              |
| DVD             | Office Products   | zShops                |
| Yellow Pages    | Outdoor Living    | Medical Supplies      |
| Restaurants     | Kitchen           | Indust. Supplies      |
| Movie Showtimes | Jewelry & Watches | Automotive            |
| .....           | ...               | ....                  |

Le principal objectif du vocabulaire contrôlé est d'empêcher les utilisateurs de définir leurs propres termes qui peuvent être ambigus, dénués de sens ou mal orthographiés

# Niveaux de la sémantique

## La taxinomie

- Une classification basée sur un thème ou un sujet. cette technique classifie le vocabulaire contrôlé en hiérarchie



- Une taxinomie classifie les termes en hiérarchie ou sous forme d'un arbre mettant en évidence leur relations avec les autres termes

# Niveaux de la sémantique

## La taxinomie

- Les premières utilisations des taxinomies furent en biologie pour classifier les animaux et les plantes selon leur relation naturelle
- Les types de relation y existe sont:
  - « Est sous classe de »
  - « Est super classe de »
- L'utilisateur et l'ordinateur peut comprendre la sémantique des mots en analysant les relation existant entre eux

# Niveaux de la sémantique

## Le thesaurus

- Un thesaurus est une collection de termes d'un vocabulaire contrôlé reliés entre eux par des relations conceptuelles .
- Les thesaurus sont des extensions des taxinomie .En effet, les termes sont reliés par des relations hiérarchiques et par d'autres relations et déclarations

# Niveaux de la sémantique

## Le thesaurus

- Les types de relation
- Equivalence : annonce que deux terme  $t_1, t_2$  ont le même sens
- Homographique : deux terme  $t_1, t_2$  ont la même orthographe mais ayant deux sens différents  $t_2$
- Hiérarchique: annonce une relation de type « super classe de » entre deux terme  $t_1$ ,
- associative pour....

# Niveaux de la sémantique

## Les ontologies

- Niveau sémantique plus haut
- Similaire à un thesaurus mais avec des **relations très riches** en sémantique entre les termes et les attributs. En plus des **règles strictes** qui guide la spécification des termes et des relations
- Une ontologie définit tout le vocabulaire d'un domaine (sujet) : la **définition** des entités(objets) ,ses attributs , et les relation entre eux