

# Ontologies et Web sémantique

---

## Cours 1: Introduction

Dr. TA Tuan Anh  
[ttanh@ciid.vast.vn](mailto:ttanh@ciid.vast.vn)

# Ontologies

---



# Système à base de connaissances

---

- ❑ Système informatique construit pour modéliser les compétences de résolution de problèmes des humains
  - doit fonctionner avec une base de connaissances sur un sujet donné
- ❑ Différence entre **données**, **informations**, **connaissances**
  - la donnée transporte l'information
  - l'information est extraite de la donnée
  - la connaissance permet de relier les informations
- ❑ La manipulation des connaissances relève de **l'intelligence artificielle**, qui permet de traiter des problèmes exponentiels, flous et/ou évolutifs
- ❑ Les systèmes experts « classiques » sont un cas particulier des systèmes à base de connaissances.

# Le cas de MYCIN

---

- ❑ MYCIN : un des premiers systèmes experts (développé à Stanford à partir de 1972)
- ❑ Objectif : diagnostique et traitement d'une infection bactérienne.
  - déterminer l'importance de l'infection
  - déterminer l'organisme responsable
  - identifier les médicaments possibles
  - choisir le meilleur traitement
  - raisonner à partir de données incomplètes
- ❑ MYCIN était un programme de recherche et n'a jamais été réellement utilisé à l'hôpital
  - incomplet,
  - difficile à évaluer,
  - mauvaise interface...
- ❑ Mais il a montré qu'on peut approcher un domaine d'expertise

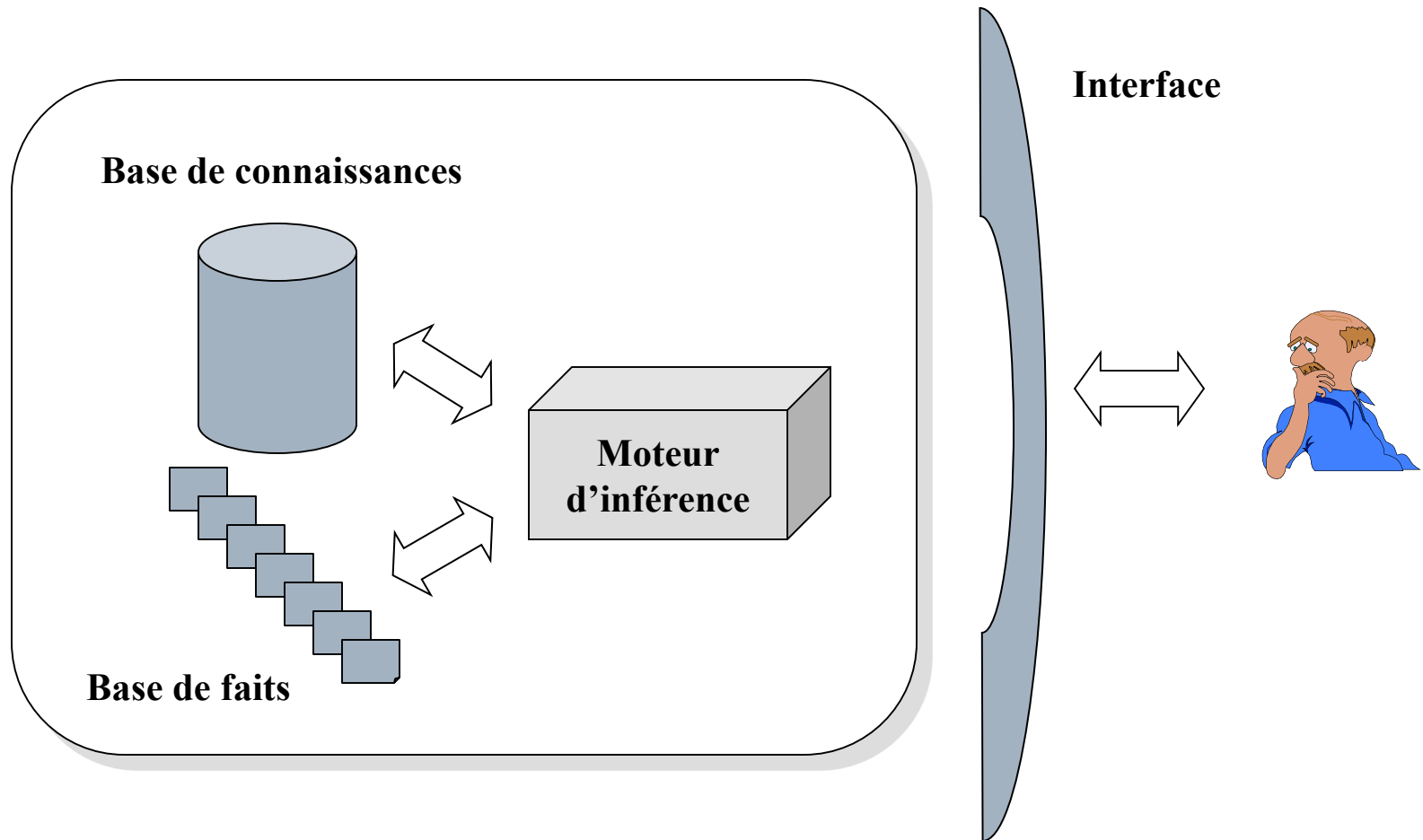
# La connaissance dans MYCIN

---

- La base de connaissances comporte des règles, e.g.,  
SI
  - la coloration de l'organisme est GRAM négatif
  - sa morphologie est un bâtonnet
  - il est aérobicALORS
  - il est vraisemblable (0,8) que l'organisme est un enterobacteriaceae
- On dispose donc d'un ensemble de règles de déductions
- Les inférences déduisent des faits à partir de faits initiaux et des règles. Deux approches de base :
  - partir de ce qu'on veut trouver et remonter vers les faits (chaînage arrière)
  - partir des faits et aller vers ce qu'on veut trouver (chaînage avant)

# Architecture du système

---



# Limites des SEs

---

- ❑ Peu de flexibilité
  - adaptés à un problème précis
  - dépendants du système d'inférence
- ❑ De grandes difficultés pour l'acquisition des connaissances
- ❑ Besoin de retravailler sur la modélisation de connaissances pour mieux séparer les connaissances stratégiques et celles du domaine
  - Connaissances stratégiques : expriment comment une tâche va être effectuée
  - Connaissances du domaine : concepts manipulés et leurs relations. « Ce qui est vrai » dans le domaine pour la tâche considérée

# Modèle conceptuel

---

- explicite la manière dont le système résout ainsi que les connaissances qu'il utilise
  - MC = Connaissances stratégiques + connaissances du domaine
- Un modèle est une abstraction du monde pour diminuer sa complexité, en se focalisant sur certains buts
  - conceptualiser entités du domaine et interactions
  - intermédiaire entre connaissances et code (i.e., bien structurer les connaissances du domaine, indépendamment de l'implémentation)
  - donc permettre d'accélérer l'acquisition de connaissances



# Modélisation des connaissances

---

- Modèle pour construire des bases de connaissances, quelques approches en exemple:
  - KADS/Common KADS – Université d'Amsterdam
  - “Components of Expertise” - Steels
  - “Generic Tasks” - Chandrasekaran
  - “Role-limiting Methods” - McDermott
  - Protégé - Musen, Stanford
  - TMDA - Motta

# Ingénierie des connaissances

---

- L'ingénierie des connaissances est l'étude des modèles symboliques formels, en fonction des usages qu'ils rendent possibles
  - Quels modèles prévoir en amont des bases de connaissances ?
  - Quels langages et formalismes pour décrire les modèles ?
  - Comment construire les modèles nécessaires à une application, selon quelles étapes et avec quels outils ?
  - Comment passer du modèle à l'application ?
  - Que devient le modèle en phase de maintenance ?
  - Autres utilisations du modèle : génération d'applications, etc.

# Modélisations à base d'ontologies

---

- Des SEs aux SBCs
  - De la simulation d'expert à la construction de modèles
- L'approche ici est de considérer la représentation des connaissances comme la construction de modèles dont la sémantique repose sur des ontologies et dont l'implantation repose sur la logique de systèmes symboliques
- Une ontologie est construite pour modéliser une partage de connaissances dans des SBCs
  - réutiliser des connaissances de domaines
  - partages, communications entre SBCs

# Notion d'ontologie

---

## PHILOSOPHIE

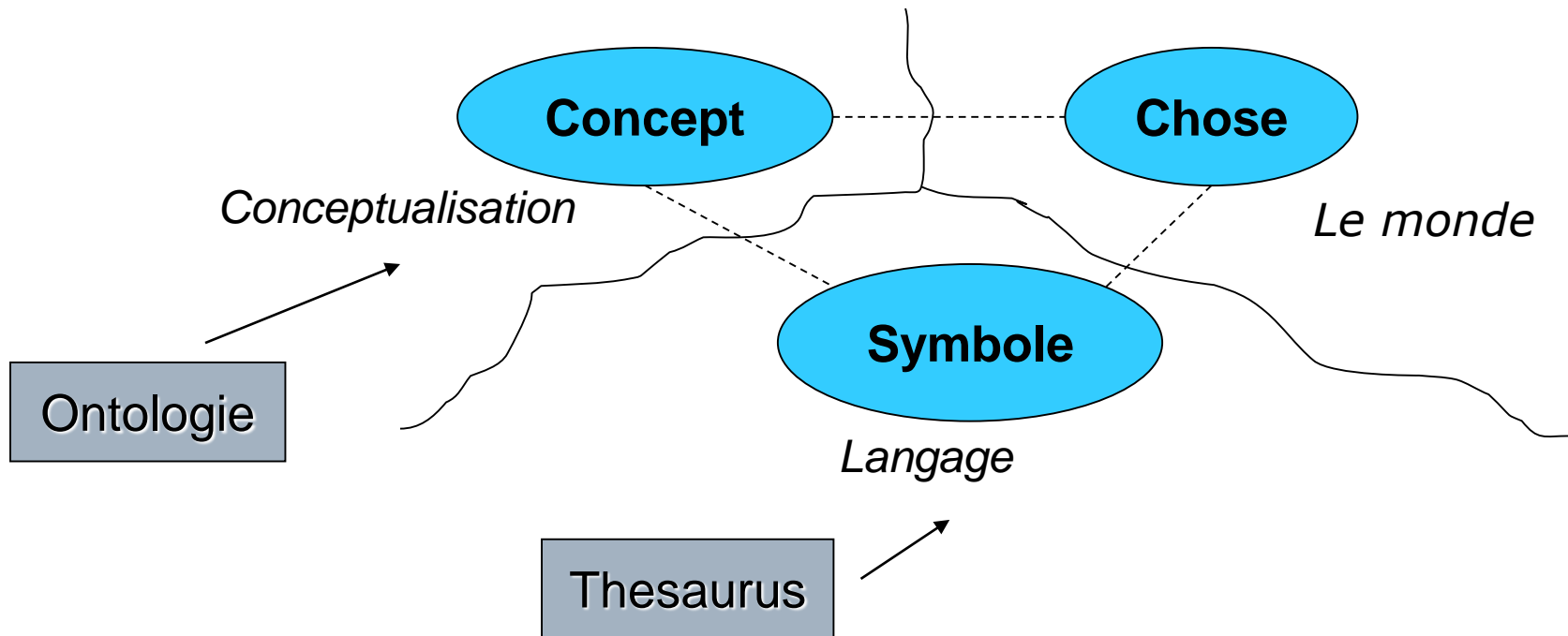
- Une notion pour décrire la science d'être (adressant la nature et l'organisation de la réalité)
  - quels types de choses existent ?
  - comment les choses devraient-elles relier ?

## INGÉNIERIE DES CONNAISSANCES

- Une ontologie est une spécification explicite pour la conceptualisation d'un domaine
  - = ensemble des objets reconnus comme existant dans le domaine
  - construire une ontologie c'est aussi décider de la manière d'être et d'exister des objets

# Triangle du sens

---



# Ontologies dans l'IC

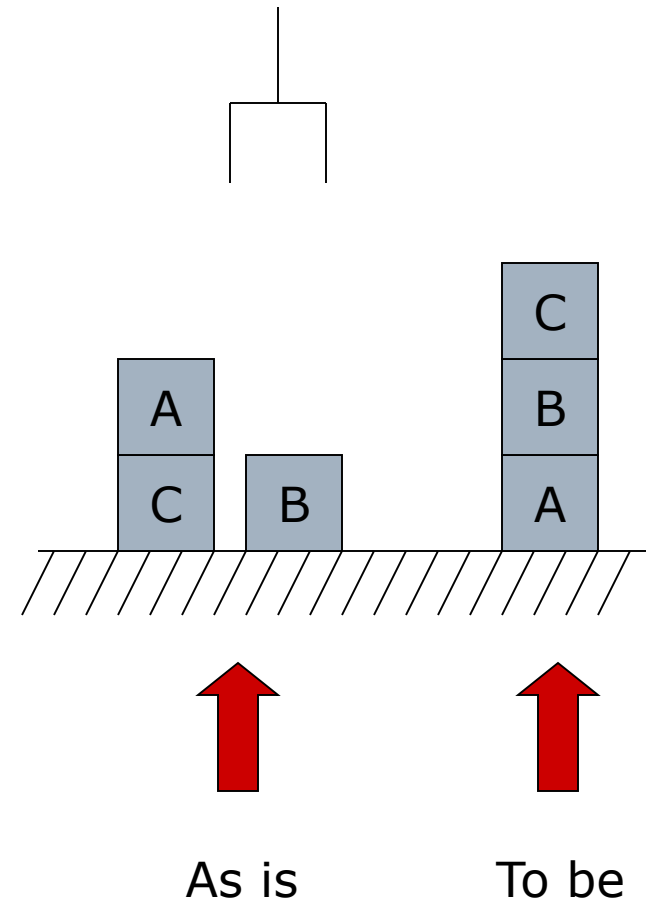
---

- ❑ Définit un vocabulaire commun pour les chercheurs qui doivent partager les informations dans un domaine
- ❑ Elles incluent des définitions machine-interprétables des concepts de base dans le domaine et des relations parmi eux
- ❑ La construction d'une ontologie a pour but de partager et de réutiliser des connaissances d'un domaine
  - permet la compréhension commune du domaine sans ambiguïté
  - comme un moyen de communication et de raisonnement
    - ❑ entre personnes (peut être informelle)
    - ❑ entre agents (doit être formelle)

# Exemple d'ontologie

---

- ❑ Concepts et Relations
  - Boit A, B, C
  - $\text{Sur}(X, Y)$ : X sur Y
  - $\text{Main}(X)$ : X dans la main
  - $\text{SurTable}(X)$  X sur la table
- ❑ Autre solution
  - Boit A, B, C; Table T
  - $\text{Sur}(X, Y)$ : X sur Y
  - Contraints pour la relation sur?



# Ontologies vs. bases de connaissances

---

- ❑ Ontologie = conceptualisation en classes génériques, relations et règles
  - les connaissances pertinentes pour un domaine d'application
- ❑ Base de connaissances = ontologie + instances (a.k.a. modèle du domaine)



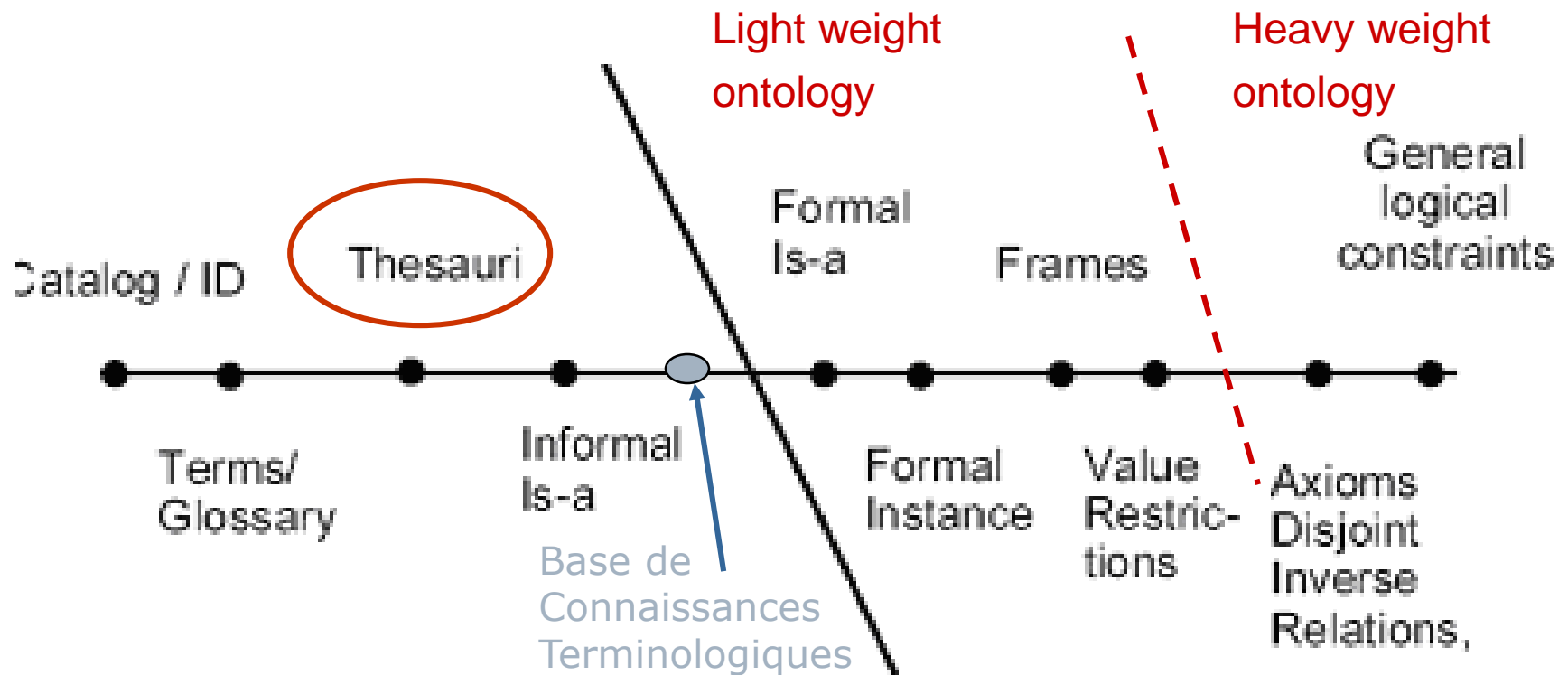
# Une classification des ontologies

---

- ❑ Ontologies de représentation
  - définit un ensemble de primitives de représentation
  - exemple : la «frame ontology» d'ONTOLINGUA
- ❑ Ontologies génériques (*Dont le «haut» est réutilisable*)
- ❑ Ontologies du domaine
- ❑ Ontologies d'une MRP
  - le rôle joué par chaque concept dans une méthode particulière est rendu explicite (ontologies de tâche)
- ❑ Ontologies d'application
  - voir une ontologie d'application comme une double spécialisation : d'une ontologie du domaine et d'une ontologie de méthode

# Ontologies et autres ressources terminologiques

---



# Ontologie vs. thesaurus

---

- ❑ Corpus, taxonomie et thesaurus
  - Corpus = un dictionnaire linguistique
  - Taxonomie = une hiérarchie de vocabulaire
  - Thesaurus = un ensemble de vocabulaire avec relations «is\_a», «synonyme», «voir\_aussi»,...
- ❑ Ontologie
  - Contenu = une taxinomie des concepts, des « rôles »
  - Décrite dans un langage de représentation de connaissances et exploitée par un système informatique
    - ❑ Possibilité de comparer et de classer des concepts
    - ❑ Capacité générative
    - ❑ Inférences

# Ontologie vs. thesaurus

---

## A simple taxonomy

- Vehicle
  - ground vehicle
    - motor car
      - 4 or more wheel car
        - car
        - truck
      - motor bike
    - train
  - ship
  - air craft

(a) A simple classification.

## An ontology of vehicle

- Vehicle world
  - type
    - ground vehicle
    - ship
    - air craft
  - function
    - to carry persons
    - to protect persons
  - attribute
    - power
    - size
  - machinery
    - engine
    - body
  - traffic system
  - ...

(b) Vehicle world ontology.

# Représentation des ontologies : historique

---

- ❑ Réseaux sémantiques (Brachman, Levêques)
- ❑ Logique du 1er ordre : CycL, KIF
- ❑ Frames : Frame Logic, Ontolingua
- ❑ Logiques de description
- ❑ Graphes conceptuels de Sowa
- ❑ Standards : DAML (Drapa) et OIL (Europe), OWL (W3C), ...

# Web sémantique et autre vision

---



# Qu'est-ce c'est le web ?

---

## Qu'est-ce c'est le web ?

- Le web est inventé par [Tim Berners-Lee](#)
- Un espace d'informations global construit à partir de :
  - HTTP (comment transférer des données de ressource)
    - autres protocoles: FTP, SMTP,...
  - URI (comment identifier des ressources)
    - deux types : URN (ressources non-adressables) et URL (ressources adressables)
  - HTML (comment formater des documents pour présentation)
    - autres formats multimédia : image, vidéo, audio,...

# Le Web actuel

---

- Un lieu où les machines présentent les documents et les gens interprètent et font lier entre eux (i.e., web syntaxique)
- Problèmes du web syntaxique
  - Rechercher des documents
    - Comment peut-on utiliser Google pour trouver tous les articles écrits par Tim Berners-Lee sur le web sémantique ?
  - Extraire des informations à partir de documents
    - Quel est le prix d'un livre donné ?
  - Combiner des informations pour obtenir des connaissances
    - Quels sont les vendeurs qui vendent souvent des livres avec prix bon marché ?



# Web sémantique est la solution

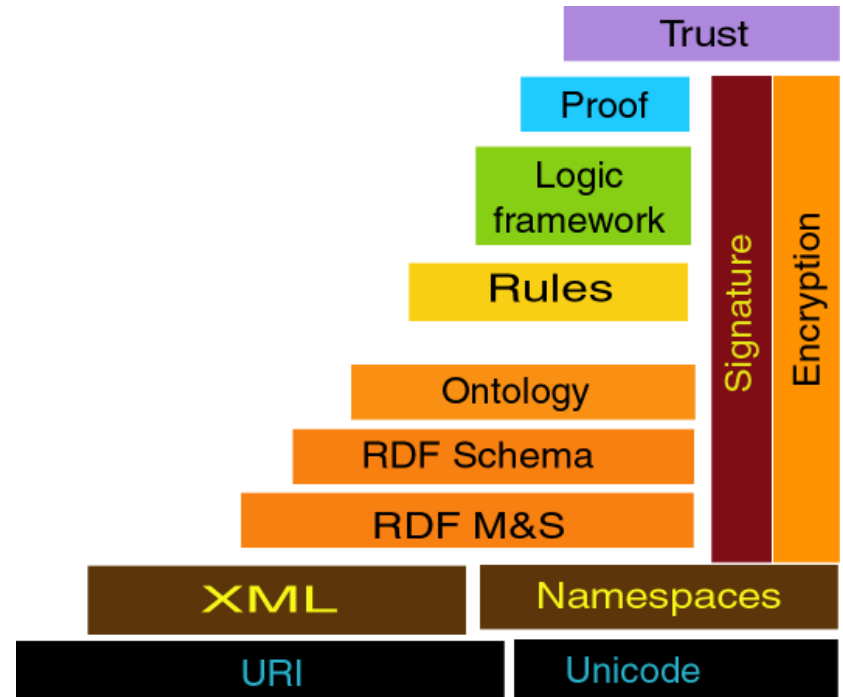
---

- Au lieu de publier des documents en langages naturels
  - les publier sous forme de données manipulables par machine
  - Utiliser des termes compréhensibles par machine pour représenter les informations
- Le Web Sémantique doit rendre les ressources accessibles pour les traitements automatiques en
  - Etendant les balises de visualisation avec les balises sémantiques
  - Utiliser des ontologies pour fournir un vocabulaire commun qui donne la sémantique des ressources
  - Besoin d'avoir des langages d'ontologie communs avant que nous pourrions partager les sémantiques

# Langages pour le Web Sémantique

---

- ❑ XML est un langage de représentation de données
- ❑ RDF est un modèle de base pour fournir des meta-données
- ❑ RDFS fournit des primitives permettant de définir les concepts ontologiques
- ❑ OWL fournit une couche ontologique plus fine
- ❑ Autres langages peut être étendu de l'OWL pour fournir les couches plus hautes.



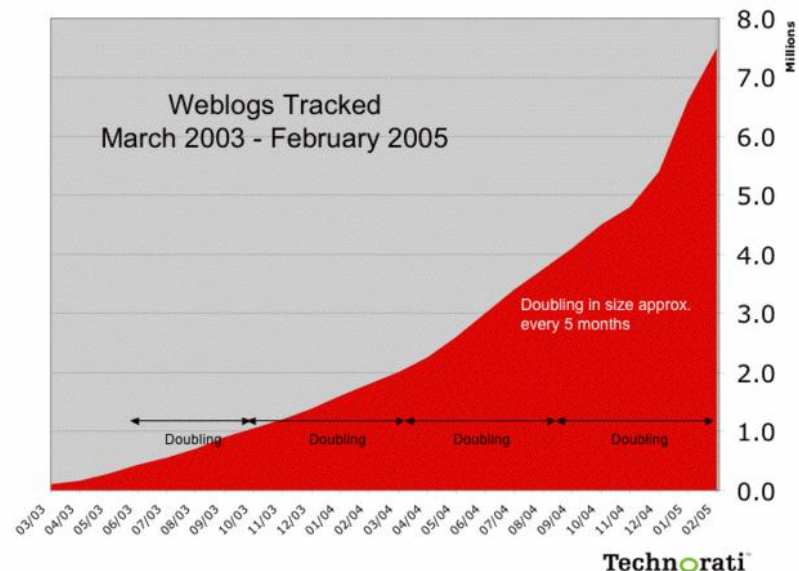
# La vue sociale du Web

---

- ❑ Logiciels sociaux
  - facilitant la communication de groupe, la construction et la solidification de liens sociaux, le travail collaboratif, le jeu à plusieurs, la création collective, organisés autour des outils en ligne
  - Outils pour communication en ligne
    - ❑ Chat
    - ❑ Forum
    - ❑ Blog (weblog)
    - ❑ Wiki
    - ❑ Réseaux sociaux
    - ❑ Etc.

# Blog/weblog

- ❑ Explosion de weblogs dans l'usage du Web (multiplié par 2 tous les 5 mois)
- ❑ Blog (né de la contraction de contraction de web log), est un site web ou les entrées sont organisées sous forme de journal (journal intime, journal de bord,...), souvent dans un ordre chronologique inverse
- ❑ Plusieurs types de blog
  - photoblog, travelblog, ...



# Réseaux sociaux

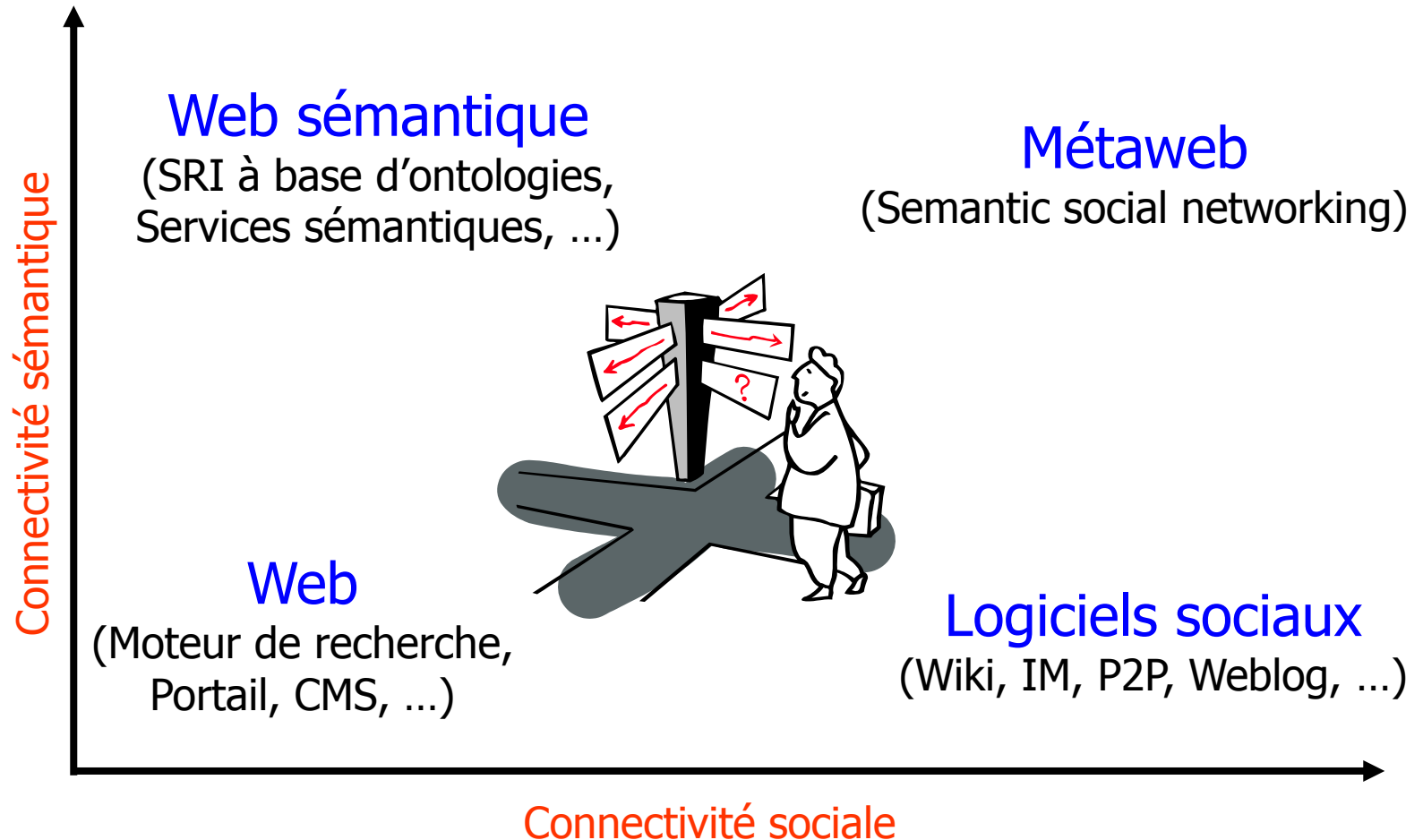
---

- Un réseau social est une structure composée de noeuds et d'arcs qui représentent généralement des relations entre individus/organismes.
- Pour mettre les personnes **en relation mutuelle** pour les actes de
  - partage,
  - échange,
  - communication,
  - collaboration, etc.
- Les réseaux sociaux en ligne peuvent aider chacun des participants à **développer ses relations**
  - E.g., Facebook, Twitter, Linkin,...



# Le futur du Web [Spivack, 2004]

---



# Matériels

---

- En ligne
  - Cours "Introduction à l'ontologie en Intelligence Artificielle" de Fabien Gandon, <http://www-sop.inria.fr/acacia/cours/essi2005/>
  - Cours "Semantic Web" de Jos de Bruijn, <http://www.debruijn.net/teaching/swt/>
  - Cours "Description Logic" de Enrico Franconi, <http://www.inf.unibz.it/~franconi/dl/course/>
  - Livre "A semantic Web primer" de Grigoris Antoniou and Frank van Harmelen, <http://www.ics.forth.gr/isl/swprimer/>
  - Documentation de W3C, <http://www.w3.org/2001/sw/>
  - et beaucoup d'autres, ...
- Grand remerciement aux auteurs !