

# DIC938I

## INTERNET ET LE WEB SEMANTIQUE

Problématique : Ce cours étudie la variété des applications et usages qui se développent autour d'Internet et du Web en tant que réseaux à la fois techniques et sociaux. On exposera donc les bases techniques et informatiques de ce phénomène, mais on montrera aussi comment les outils sont appropriés par les utilisateurs. Dans la deuxième partie du cours, on traitera plus particulièrement le Web sémantique, ses bases théoriques et ses applications, car c'est actuellement un domaine de recherche très actif qui illustre bien les avancées et les questions suscitées par Internet.

1<sup>ère</sup> partie : réseaux d'information, structures et usages (D. Memmi)

### **1/ La société de l'information**

Place économique et sociale de l'information et conséquences actuelles. Formation, R&D et productivité. Information et connaissances ; connaissances tacites et explicites. Evolution vers une société en réseaux. Rôle croissant des technologies de l'information et de la communication dans cette évolution.

### **2/ Réseaux sociaux**

Importance pour la diffusion de l'information. Concepts de base : acteurs, relations et structures. Problématique, approches et disciplines pertinentes. Apports de la sociologie, de la théorie des graphes, de la physique statistique et de l'informatique. Travaux de recherche actuels et grandes tendances.

### **3/ Modèles de réseaux**

La sociologie structurale : enquêtes et modélisation. Modèles formels de la théorie des graphes et de la physique statistique. Notions de petit monde (*small world*) et de transitivité (*clustering*). Attachement préférentiel et lois de puissance. Confrontation avec divers réseaux réels : réseaux sociaux, réseaux techniques, graphe du Web.

### **4/ Internet et le Web**

Principes et bases techniques : commutation par paquets, organisation en couches physique, logique, virtuelle... Applications diverses, du courriel au Web. Principales étapes, fusion progressive avec le téléphone fixe et mobile. Services Web et calcul distribué. Structure et fonctions du Web. Synergie avec l'évolution sociale et économique.

### **5/ Usages d'Internet**

Principaux usages : courriel, clavardage, forums, transfert de fichiers, Web et services Web, blogs et wikis, téléphonie... Usages synchrones et asynchrones, texte, voix et vidéo. Réseaux sociaux et logiciels collaboratifs. Le Web 2.0 : tendances et grandes applications. Le réseau comme support de services.

### **6/ Recherche d'information**

Indexation et recherche. Le modèle vectoriel de représentation des documents. Mesures de distances. Les moteurs de recherche. Silence et bruit, problème d'évaluation des réponses. Structure du Web, PageRank et Google. Recherche collaborative de documents, systèmes de recommandation et expertise tacite. Les limites de l'approche textuelle lexicale.

### **7/ Contrôle des connaissances (1ère partie)**

### **8/ Bases technologiques du Web actuel**

Langage de balises HTML, motivation, structure et usage. Examen critique. Langage XML et données semi-structurées sur le Web. Langages de requêtes pour XML (XQuery, XPath). Définition de la structure des documents DTD vs XML Schéma. Présentation des documents (XSL, XHTML).

### **9/ Motivation et architecture du Web sémantique**

La vision originale du Web sémantique. Rôles complémentaires par rapport au Web « syntaxique ». Architecture en couches et principaux constituants. État actuel d'achèvement des couches. Notions initiales de base de connaissances et d'ontologies.

### **10/ Fondements de l'approche « sémantique »**

Représentation de connaissances structurées, réseaux sémantiques, frames. Logique en tant que modèle de connaissances et formalismes structurés à inspiration logique. Les logiques de descriptions, motivation et structure commune (syntaxe et sémantique). Diverses familles de logiques de description. Principaux outils : instanciation subsomption, mécanismes de requêtes. Algorithmes et complexité.

### **11/ Principaux formalismes et langages**

Le langage de description des ressources sur le Web RDF : syntaxe, sémantique, usage. Définition de vocabulaires pour RDF, le langage RDFS. Le langage de requêtes SPARQL. Représentation de haut niveau de granularité, les langages historiques (DAML, OIL, DAML+OIL). La couche ontologique du Web sémantique et langage d'ontologies pour le Web OWL. Les dialectes de OWL. La couche règles de raisonnement et le langage de règles SWRL.

### **12/ Ingénierie d'ontologies**

La modélisation conceptuelle et les multiples facettes d'une ontologie. Lexique, thesaurus, taxonomie, ontologie. Architecture d'une base de connaissances en OWL. Les activités du cycle de vie d'une ontologie : construction, évolution, versionnement. Outils d'aide à la construction d'ontologies. Les problématiques de fusion et d'alignement d'ontologies.

### **13/ Principaux usages et applications**

Annotation sémantique des documents, algorithmes et systèmes. Recherche d'information sémantique, précision et rappel sémantiques. Enseignement assisté par ordinateur sur le Web sémantique. Fouille de données pour et sur le Web sémantique. Applications actuelles du Web sémantique : FOAF, RSS, Semantic Wiki, Twine.

### **14/ Révision ou discussion complémentaire**

Séance de révision des concepts ou bien discussions supplémentaires.

### **15/ Contrôle des connaissances (2ème partie)**

#### Contrôle des connaissances

Deux contrôles sont prévus, c'est-à-dire un contrôle pour chaque partie du cours (50% chacun). Chaque contrôle se fera en salle, soit sous forme d'exposé oral (commentaire d'article ou synthèse ou projet), soit sous forme d'examen écrit, selon le nombre d'étudiants inscrits au cours. Dans le cas d'un exposé oral, un bref rapport écrit sur le même sujet (environ 10 p.) sera aussi demandé.