

Internet et le web sémantique

Groupe 50

Vendredi, de 14h00 à 17h00 PK-4360 (cours)

Responsable(s) du cours

Nom du coordonnateur : MEMMI, Daniel**Nom de l'enseignant :** MEMMI, Daniel**Local :** PK-4430**Téléphone :** (514) 987-3000 #7939**Courriel :** memmi.daniel@uqam.ca**Site Web :** http://intra.info.uqam.ca/Members/memmi_d

Autre enseignant : VALTCHEV, Petko

Description du cours

Notations, langages et modèles du Web sémantique (cadre de description des ressources et ses schémas, la logique de description et les ontologies, les langages des requêtes, les langages de règles). Infrastructure du Web sémantique (couches du Web sémantique : données, schémas, ontologies, règles, etc.). Données sur le Web sémantique et Web des données. Principaux défis du Web sémantique : ingénierie d'ontologies, intégration des données. Usages typiques du Web sémantique : recherche sémantique, requêtes complexes sur les données.

Objectifs du cours

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 assiste en effet à la convergence d'évolutions techniques, d'usages sociaux et de nouveaux formalismes, dont nous essaierons de montrer l'évolution globale. On exposera 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.

Contenu du cours

Ce cours comporte 2 parties, assurée respectivement par D. Memmi et P. Valtchev:

1ère 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.

2ème partie : Web sémantique, formalismes et applications (P. Valtchev)

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.

Modalités d'évaluation

Deux contrôles, 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 (10 à 15 p.) sera aussi demandé.

Politique d'absence aux examens

L'autorisation de reprendre un examen en cas d'absence est de caractère exceptionnel. Pour obtenir un tel privilège, l'étudiant-e doit avoir des motifs sérieux et bien justifiés.

Il est de la responsabilité de l'étudiant-e de ne pas s'inscrire à des cours qui sont en conflit d'horaire, tant en ce qui concerne les séances de cours ou d'exercices que les examens. **De tels conflits d'horaire ne constituent pas un motif justifiant une demande d'examen de reprise.**

Dans le cas d'une absence pour raison médicale, l'étudiant-e doit joindre un certificat médical original et signé par le médecin décrivant la raison de l'absence à l'examen. Les dates d'invalidité doivent être clairement indiquées sur le certificat. Une vérification de la validité du certificat pourrait être faite. Dans le cas d'une absence pour une raison non médicale, l'étudiant-e doit fournir les documents originaux expliquant et justifiant l'absence à l'examen – par exemple, lettre de la Cour en cas de participation à un jury, copie du certificat de décès en cas de décès d'un proche, etc. Toute demande incomplète sera refusée. Si la direction du programme d'études de l'étudiant-e constate qu'un étudiant a un comportement récurrent d'absence aux examens, l'étudiant-e peut se voir refuser une reprise d'examen.

L'étudiant-e absent-e lors d'un examen doit, dans les cinq (5) jours ouvrables suivant la date de l'examen, présenter une demande de reprise en utilisant le formulaire prévu, disponible sur le site Web du département à l'adresse suivante : <http://info.uqam.ca/politiques/>

L'étudiant-e doit déposer le formulaire dûment complété au secrétariat de la direction de son programme d'études : PK-3150 pour les programmes de premier cycle, PK-4150 pour les programmes de cycles supérieurs. Pour plus de détails sur la politique d'absence aux examens du Département d'informatique, consultez le site web suivant : <http://info.uqam.ca/politiques>

Intégrité académique

PLAGIAT Règlement no 18 sur les infractions de nature académique. (extraits)

Tout acte de plagiat, fraude, copiage, tricherie ou falsification de document commis par une étudiante, un étudiant, de même que toute participation à ces actes ou tentative de les commettre, à l'occasion d'un examen ou d'un travail faisant l'objet d'une évaluation ou dans toute autre circonstance, constituent une infraction au sens de ce règlement.

La liste non limitative des infractions est définie comme suit :

- la substitution de personnes;
- l'utilisation totale ou partielle du texte d'autrui en la faisant passer pour sien ou sans indication de référence;
- la transmission d'un travail pour fins d'évaluation alors qu'il constitue essentiellement un travail qui a déjà été transmis pour fins d'évaluation académique à l'Université ou dans une autre institution d'enseignement, sauf avec l'accord préalable de l'enseignante, l'enseignant;
- l'obtention par vol, manoeuvre ou corruption de questions ou de réponses d'examen ou de tout autre document ou matériel non autorisés, ou encore d'une évaluation non méritée;
- la possession ou l'utilisation, avant ou pendant un examen, de tout document non autorisé;
- l'utilisation pendant un examen de la copie d'examen d'une autre personne;
- l'obtention de toute aide non autorisée, qu'elle soit collective ou individuelle;
- la falsification d'un document, notamment d'un document transmis par l'Université ou d'un document de l'Université transmis ou non à une tierce personne, quelles que soient les circonstances;
- la falsification de données de recherche dans un travail, notamment une thèse, un mémoire, un mémoire-crédation, un rapport de stage ou un rapport de recherche;
- Les sanctions reliées à ces infractions sont précisées à l'article 3 du Règlement no 18.

Les règlements concernant le plagiat seront strictement appliqués. Pour plus de renseignements, veuillez consulter les sites suivants : <http://www.sciences.uqam.ca/etudiants/integrite-academique.html> et <http://www.bibliotheques.uqam.ca/recherche/plagiat/index.html>

Médiagraphie

Partie 1

Castells, M. (1996). *The Rise of the Network Society*. Oxford: Blackwell.

Lazega, E. (1998). *Réseaux Sociaux et Structures Relationnelles*. Paris: PUF.

Tanenbaum, A. (2003). *Computer Networks*. NJ: Prentice Hall.

Wasserman, S., & Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.

Partie 2

VR ANTONIOU, G. et VAN HARMELEN, F. -- *A Semantic Web Primer* -- 2 ed, The MIT Press, 2008.

VC HITZLER, P., KRÖTZSCH, M. et RUDOLPH, S. -- *Foundations of Semantic Web Technologies* -- CRC Press/Chapman and Hall, 2009.

VC ALLEMANG, D. et HENDLER, J. -- *Semantic Web for the Working Ontologist* -- Morgan Kaufmann, 2008.

VC -- *D'autres références (selon les sujets abordés) seront remises durant la session (voir).*

A : article - C : comptes rendus - L : logiciel
S: Standard - U : uri - V : volume

C : complémentaire - O : Obligatoire - R : recommandé