

Annotations sémantiques et applications en recherche d'information

Infos pratiques

- > ECTS : 3.0
- > Nombre d'heures : 24.0
- > Période de l'année : Enseignement neuvième semestre
- > Méthodes d'enseignement : En présence
- > Forme d'enseignement : Cours magistral
- > Ouvert aux étudiants en échange : Oui
- > Composante : Philo, Info-Comm, Langages, Littératures & Arts du spectacle

Présentation

Ce cours exploite les savoirs et les données du rapprochement manifeste, ces dix dernières années, entre les domaines du TAL et de la Recherche d'Information. Il présentera ainsi des méthodes, modèles et applications qui s'intéressent à l'accès au contenu informationnel des textes et mobilisent un niveau d'analyse sémantique des textes. L'enjeu se situe bien sûr à l'aune d'une masse croissante de documents textuels de types très divers (depuis des fonds d'archives historiques numérisés jusqu'à des ensembles de pages web évolutives en passant par des articles scientifiques dans des domaines spécialisés comme la biologie par exemple) qu'il s'agit de saisir dans leur visée informationnelle au regard de certaines catégories sémantiques linguistiques. Les catégories sémantiques plus particulièrement explorées dans ce cours seront celles de la prise en charge énonciative, de la modalité, du temps et de l'aspect en montrant comment elles sont mobilisées dans des applications concrètes de recherche d'information, comme par exemple celles qui cherchent à identifier avec quelle distance énonciative ou avec quelle caractéristique émotionnelle des informations sont présentées dans un texte.

Objectifs

Comprendre les apports de la sémantique linguistique (lexicale et grammaticale) aux applications en recherche d'information

Évaluation

M3C en 2 sessions

- Régime standard session 1 – avec évaluation terminale (1 seule note) : Un projet par groupe
- Régime dérogatoire session 1 : Un projet par groupe
- Session 2 dite de rattrapage : Un projet par groupe

Compétences visées

Comprendre les apports, d'un point de vue méthodologique, de travaux menés en linguistique notamment dans le cadre de la linguistique énonciative, de la linguistique textuelle et de l'analyse du discours. Ouvrir à des pistes de réflexion sur l'accès au contenu informationnel des textes qui mobilisent ces travaux.

Bibliographie

- Battistelli D., 2011 : *Linguistique et recherche d'information : la problématique du temps*, Hermès, [Lavoisier](#), coll. Traitement de l'Information.
- CONDAMINES A. (ed), 2005 : *Sémantique et corpus*. Londres : Hermes
- Etienne A., Battistelli D. , Lecorvè G., 2020 : "L'expression des émotions dans les textes pour enfants : constitution d'un corpus annoté". In *Actes TALN'20 (Traitement automatique du langage naturel 2020)*.