

Projet IDEX Formations Innovantes INFOSHS Semantique Distributionnelle

| | |
|-------------------------------|--|
| Présentation | Selon Wittgenstein, le signifié d'un mot ou d'une expression est déterminé par son contexte: "... si je comprends la proposition, alors je connais la situation dont elle est la représentation..." (<i>Tractatus</i> , 4.021). Cette vision a une grande importance aujourd'hui dans l'analyse automatisée des textes. Dans ce cours on va étudier les techniques qu'ont été développés pour exploiter l'information contextuelle, en particulier l'analyse sémantique latente et les word embeddings. |
| Mode de validation | Une attestation de présence sera délivrée à chaque étudiant ayant suivi l'ensemble de la formation. Un certificat de réussite sera délivré aux étudiants ayant passé avec succès le QCM et l'épreuve de validation. |
| Programme détaillé | <ul style="list-style-type: none"> • Introduction à la sémantique distributionnelle • Latent Semantic Analysis • LSA pour la recherche d'information: Latent Semantic Indexing et Random Indexing • Latent Dirichlet Allocation • Word Embeddings et Semantic Vectors |
| Objectifs pédagogiques | Être capable d'appliquer les techniques de sémantique distributionnelle à des corpus de textes pour découvrir synonymes, antonymes, ou de toute façon des relations sémantiques entre mots différents. |
| Formateur(s) | Davide Buscaldi : maître de conférences à l'Université Paris 13, auteur de plus de 70 articles scientifiques publiés dans des conférences et journaux internationaux, avec plus de 10 ans d'expérience dans la recherche en TAL. Ses domaines d'intérêt sont la recherche d'information sémantique, les ontologies et la fouille de textes. |
| Pré-requis | Programmation en Python, utilisation de la librairie NLTK. |
| Public visé | Étudiants en Master 1, 2. Doctorants de 1ère année - Disciplines : informatique, langues, linguistique, littérature. S'adresse plus spécifiquement à des étudiants concernés par l'utilisation de corpus textuels. |
| Effectif | maximum 24 personnes |

Informations pratiques

| | |
|--|---|
| Email (contact pédagogique, pour toutes questions sur le contenu de la formation) | Email de contact : davide.buscaldi@lipn.univ-paris13.fr |
| Date(s) | Jeudi 30 juin : 9h-12h et 14h-17h Vendredi 1 juillet : 9h-12h et 14h-17h |
| Lieu | IUT de Villetaneuse, salles informatiques SHS ; se munir du présent document pour contrôle à l'entrée du bâtiment. |
| Durée de la formation | 2 jours (12h) |
| Nombre de jours validés par le Cfdip | |
| Lien vers l'inscription (obligatoire) | http://goo.gl/forms/hbOscM7Uiv |