

PROPOSITION DE STAGE EN COURS D'ETUDES

Référence : **DTIM-2012-09**
(à rappeler dans toute correspondance)

Lieu : Palaiseau

Département/Dir./Serv. :
DTIM / SIF

Tél. : 01 80 38 65 65

Responsable du stage : Valentina Dragos

Email. : valentina.dragos@onera.fr

DESCRIPTION DU STAGE

Domaine d'étude : Ingénierie des connaissances

Type de stage Fin d'études bac+5 Master 2 recherche Bac+2 à bac+4

Intitulé : Mesures de similarité sémantique pour la cotation des informations

Sujet : Le stage aborde la problématique de la qualité des informations produites par plusieurs sources dans un environnement hétérogène. Dans un tel environnement des sources distinctes fournissent des informations ayant des formats variés (données vidéo, capteur ou fournies par des opérateurs humains), et dont les contenus peuvent être corrélés. Des lors, ces informations peuvent s'avérer incohérentes, des contradictions apparaissant au fil du temps, ou redondantes, car la même information peut être décrite différemment par des sources distinctes. L'utilisation des informations ainsi recueillies doit être soumise à une étape d'analyse préalable afin d'évaluer leur qualité : on appelle ce processus cotation des informations. Si la littérature propose des solutions satisfaisantes pour le traitement des données numériques, la prise en compte des données symboliques s'avère déficitaire. L'exploitation des données textuelles, telles que les descriptions des scènes vidéo ou les messages de renseignement d'origine humaine représente aujourd'hui une problématique de recherche émergente.

Les données symboliques étant difficilement traitables, nous avons abordé cette problématique selon un point de vue sémantique. Une top-ontologie du domaine a été construite permettant le développement des algorithmes de cotation des informations capables d'exploiter la sémantique des informations analysées. L'objectif du stage est de définir des mesures pour estimer la similarité sémantique des informations appartenant à une collection de données. Par la suite, ces mesures se trouveront au cœur des algorithmes de cotation des informations.

Une première expérimentation sera réalisée en exploitant un corpus de messages de renseignement d'origine humaine, permettant de valider les mesures proposées.

Est-il possible d'envisager un travail en binôme ? Non

Méthodes à mettre en oeuvre :

- | | |
|---|---|
| <input checked="" type="checkbox"/> Recherche théorique | <input checked="" type="checkbox"/> Travail de synthèse |
| <input checked="" type="checkbox"/> Recherche appliquée | <input checked="" type="checkbox"/> Travail de documentation |
| <input type="checkbox"/> Recherche expérimentale | <input checked="" type="checkbox"/> Participation à une réalisation |

Possibilité de prolongation en thèse : Oui

Durée du stage : Minimum : 4 mois Maximum : 6 mois

Période souhaitée :

PROFIL DU STAGIAIRE

Connaissances et niveau requis :

Java, OWI, Protégé

Ecoles ou établissements souhaités :

Université, Grandes Ecoles