

Proposition de stage

Lieu :

**Unité de Recherche INRIA de Sophia Antipolis (06)
Equipe AxIS**

« Mise en correspondance des appels à projet et des thématiques de recherche INRIA »

Encadrants : Brigitte TROUSSE et Bernard SENACH

Durée : du 15/07/2006 au 31/12/2006

Indemnité de stage : entre 716 et 898€net/mois selon niveau d'études (minimum bac + 4)

L'objectif du stage est d'utiliser les technologies de fouille de données textuelles pour permettre aux équipes de l'Inria de répondre plus fréquemment à des appels à projets les concernant.

Actuellement, de nombreuses instances européennes, nationales et régionales lancent des appels à projets . Une veille transversale effectuée par différents services de l'INRIA (DIAS, REV...) est complétée par une veille locale effectuée en interne par les équipes de recherche de L'INRIA qui ont chacune leur propre réseau d'information, plus ou moins bien structuré.

Etant donné un appel à projet, il peut être difficile de déterminer quelles sont les équipes concernées : la mise en correspondance d'une thématique donnée, parfois formulée de façon assez floue, générale et ouverte avec une thématique de recherche souvent formulée de façon très technique suppose une bonne connaissance des travaux réalisés dans les différentes équipes de recherche. Le processus de mapping reste artisanal, de nombreuses inférences sont nécessaires et le poids des connaissances individuelles devient très lourd.

L'équipe AxIS souhaite mettre au point, à partir des techniques de fouilles de données textuelles, des mesures de distances facilitant la mise en correspondance des thématiques des appels à projet et celles des équipes de recherche de l'INRIA.

Travaux à réaliser :

Le travail consistera à :

- analyser les pages web élaborées par les différents projets de l'INRIA (pages Inria et pages équipes) ou rapports d'activité annuels (documents XML) pour en extraire des caractéristiques pouvant être utilisées dans le cadre du mapping avec les appels à projet
- analyser un corpus significatif d'appels pour en extraire les éléments pertinents pour la mise en correspondance avec les thématique de recherche des équipes
- définir une mesure de correspondance entre appel à projet et thématique de recherche

Au cours de ce travail, le stagiaire s'appuiera sur les travaux déjà réalisés dans AxIS appliqués à la classification des projets Inria soit à partir des rapports d'activités (documents XML) soit à partir de sites Web des projets.

Les étapes classiques de fouille de texte seront mises en œuvre (normalisation, segmentation, lemmatisation, numérisation, recherche de segments répétés ...). Les tableaux lexicaux supportant les statistiques descriptives discriminantes seront complétés par des analyses multi variées (AFC, ACP, ACM) permettant réduction des dimensions des données. La mise en relation d'une classification des thématiques de recherche avec des appels à projet qualifiés par les techniques de texte mining fera l'objet d'une validation empirique.

