

Titre : Développement d'une interface de navigation à la LIS pour un outil de fouille de textes

Lieu : IRISA - Rennes

Contact : Peggy Cellier <cellier@irisa.fr> et Sébastien Ferré <ferre@irisa.fr>

Durée : 3-4 mois

Rémunération : 554,4 € / mois

Contexte :

Ce stage est proposé et financé dans le cadre plus large d'un projet de recherche (ANR Hybride) réunissant plusieurs laboratoires dont : GREYC (Caen) - LIPN (Paris 13) - IRISA (Rennes), INSERM (Paris). Ce projet a pour objectif général d'extraire des informations à partir de textes biologiques et médicaux sur les maladies orphelines (dites aussi maladies rares) pour apporter une aide automatique à l'enrichissement d'une base de connaissances, notamment du site d'information Orphanet dédié à ce type de maladies. L'autre objectif fondamental est combiner des méthodes de fouille de données avec des méthodes de traitement automatique de la langue. Pour une présentation générale du projet et de quelques résultats en particulier, on pourra consulter le site dédié : <http://hybride.loria.fr/>.

Sujet :

SDMC est un outil de fouille de motifs séquentiels : <https://sdmc.greyc.fr>. Cet outil est destiné à tous types de public. Il est utilisé dans le cadre du projet Hybride mais aussi plus largement sur différents types de textes (par exemple pour l'analyse stylistique de textes). Son interface a ainsi été conçue pour être utilisable par des non informaticiens : il suffit de rentrer un texte, éventuellement des paramètres de fouille et l'outil extrait automatiquement des motifs séquentiels du texte. L'inconvénient de la fouille de motifs est le nombre de motifs qui peut être très important. Des outils de manipulation sont nécessaires pour permettre à un utilisateur de mieux appréhender et sélectionner les motifs qui vont l'intéresser.

L'objectif du stage est ainsi de développer une interface de navigation à la LIS (Logical Information System : <http://www.irisa.fr/LIS/software-fr/>), pour naviguer dans les motifs. En particulier, il est important de pouvoir connecter les motifs avec le texte original. Le développement sera fait avec des technologies web, en particulier PHP.

Compétences requises :

- autonomie
- bonne compétence en développement, en particulier en PHP, programmation web
- connaissances en architecture client/serveur