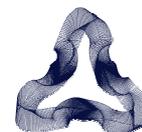


**Proposition de bourse de thèse CIFRE en informatique**  
**Algorithmes pour XML:**  
Masses de données, mémoire externe et performances extensibles.

---

## Innovimax

Innovimax SARL. Consulting, Training & XML Development  
<http://www.innovimax.fr>



## LACL

Laboratoire d'Algorithmique Complexité et Logique, Université de Paris-12  
<http://www.univ-paris12.fr/lacl>



**INNOVIMAX** est une jeune société spécialisée dans le conseil, la formation et de le développement de solutions logicielles autour de XML. Innovimax est membre du W3C, de l'ISO et de l'AFNOR et est très active dans ces instances sur les questions liées à la transformation XML ainsi que la validation de documents. Innovimax a comme client : Les Éditions Francis Lefebvre, La direction des Journaux Officiels (dépendant du premier ministre), l'OCDE, PTC, etc. Innovimax est membre du Pôle de compétitivité Cap Digital et participe activement à des salons internationaux.

Le **LACL** est un laboratoire public classé de catégorie A par le Ministère de la Recherche et dont une équipe poursuit des travaux sur la programmation haute-performance. Ces recherches ont été récompensées par un prix de thèse EADS, des financements ANR et DIGITEO Ile-de-France. Nos méthodes pour l'extensibilité des performances servent au calcul numérique, à l'apprentissage automatique, aux bases de données et à la vérification automatique de logiciel. Le LACL est membre du Pôle de compétitivité Sytem@tic, de l'Ecole Doctorale SIMME et dispose de ses propres moyens de calcul et de stockage parallèles.

Dans le cadre du développement d'Innovimax nous cherchons à approfondir nos recherches dans le cadre de l'algorithmique des transformations XML selon 3 axes principaux :

- 1) **Algorithmes de tri XML et Grouping** : le but est de trouver des classes d'algorithmes efficaces pour le traitement de très gros volume de fichiers ou de très gros fichiers, en prenant en compte les différentes problématiques. Fichier infini (log) ou fichier sur support en lecture bidirectionnelles possible (Filesystem). Une attention pourra être portée aux algorithmes d'approximation de tri ou de grouping dans un contexte d'utilisation de ressource minimale.
- 2) **Algorithmes de requête** : analyse conjointe des axes d'améliorations de requête : Les requêtes faites sur des fichiers XML peuvent avoir plusieurs axes d'améliorations selon leur complexité et leur degré de réutilisation : optimisation de la requête sans connaissance préalable du modèle de document, réécriture ou optimisation de la requête avec connaissance du modèle générique ou spécifique du document et enfin, changement de la structure du document et de la requête. Le but est d'analyser le seuil d'efficacité de ces différentes approches et proposer de nouvelles approches hybrides.
- 3) **Bidirectionnalité** : XML peut être en théorie traité de manière bidirectionnelle. Cela laisse un champ d'investigation quand aux techniques de traitements optimaux sur des fichiers de grands volumes avec des techniques BabyStep/GiantStep. On pourrait aussi coupler ce travail avec des techniques de projection de document c-à-d supprimer dans une représentation en mémoire intermédiaire les nœuds du document d'origine qui ne seront jamais utiles à l'évaluation d'une requête.

**FORMAT:** Recrutement selon les clauses d'une convention CIFRE<sup>1</sup> et inscription en thèse à l'Université Paris-12. Durée du contrat et des études 3 ans à partir de janvier 2008. Moyens de fonctionnement mis à disposition par innovimax et le LACL. Publication des résultats dans des colloques internationaux.

**CANDIDATURE :** Les candidats sont priés de fournir pour le 15 novembre 2007 un CV détaillé, un diplôme de niveau Master-2 et tout autre document utile. Des compétences en algorithmique haute performance ou XML seront un atout.

Mohamed Zergaoui  
Président Innovimax  
[innovimax@gmail.com](mailto:innovimax@gmail.com)

Gaétan Hains  
Directeur du LACL  
[gaetan@hains.org](mailto:gaetan@hains.org)

---

<sup>1</sup> [http://dr.education.fr/Alloc\\_doc/CIFRE.html](http://dr.education.fr/Alloc_doc/CIFRE.html)