INSTITUT DE RECHERCHE EN INFORMATIQUE ET SYSTEMES ALEATOIRES
Campus universitaire de Beaulieu ● 263 Avenue du Général Leclerc - 35042 Rennes Cedex France
www.irisa.fr

OFFRE de THÈSE

Sujet de thèse : **Active Indexing Memory for Genomic Processing**

Description :

Context : Searching genomic database is a basic bioinformatics treatment. Given a query (a DNA sequence for example), the goal is to find all, or part of, the sequences having similarities. With the very fast biotechnology progresses, the size of the databases become gigantic and proportionally increase the response time.

To restrict time, one technique consists in indexing databases in such a way that a query refers only to a small portion of the data. However, this fast approach requires storage spaces which are much bigger than storing direct raw data. It also requires very fast memory access time which cannot be sustained by standard magnetic hard drives.

The ReMIX architecture, developed into the Symbiose Group, was a first tentative based on FLASH technology to propose a new indexing architecture. Experiments have demonstrated that this architecture was actually very fast. On the other hand, experiments have shown that the indexing step was a serious bottleneck, just because this step was performed outside ReMIX, on standard computers with limited memory capacity.

PhD Proposal : The goal of this PhD proposal is to push the limit of the ReMIX architecture by enhancing this concept with some auto-indexing capabilities. The main idea to investigate is to make the indexing process transparent for the user. Ideally, each time a new data is pushed to the database, an internal processing is triggered to restructure the index accordingly. To house such functionalities, the ReMIX architecture must be completely revisited and associated with advanced programming model such as MAP-REDUCE model, for example, to allow the indexing structures to be explicitly specified.

In a context where the rapid increasing of genomic data poses real challenges, the work of this thesis is to provide original solutions combining original architectural approaches, algorithms and technology.
http://www.irisa.fr/matisse/forms/sujets/2011/bioinformatique-symbiose

Département : D7 - Gestion des données et de la connaissance

Equipe : Bioinformatique – Symbiose ; http://www.irisa.fr/symbiose/

Directeur de thèse : Dominique Lavenier ; http://irisa.lavenier.net/

Encadrant(s) : Dominique Lavenier

Contact : lavenier@irisa.fr

Début des travaux : october 2011

Bibliographie : ReMIX web site: http://irisa.lavenier.net/reserach/remix.html