

Action Concertée Incitative
MASSES DE DONNEES
Descriptif complet du projet

I - FICHE D'IDENTITE DU PROJET

Nom du Projet :

REMIX

Titre du Projet :

Mémoire Reconfigurable pour l'Indexation de Masses de Données

Type du Projet :

Projet de recherche	Projet de recherche multi-thématiques	Projet de recherche avec infrastructure	Autre
	XXXXXXXXXXXXXXXXXXXX		

Durée du projet : 36 mois

Description courte du Projet :

L'indexation est une technique couramment employée pour accélérer la recherche d'information au sein de volumes de données conséquents. Lorsque les tailles des index sont supérieures à la capacité mémoire centrale des machines, ils doivent être stockés sur disque. Dans ce cas, leur structure interne est entièrement conçue dans le but d'optimiser au maximum les accès (coûteux) aux disques, ceci afin de limiter les temps de réponse. De plus, leur organisation interne orientée disques contraint indirectement les algorithmes car ces derniers privilégient les traitements et parcours séquentiels, au détriment des accès directs, et souvent aléatoires, à l'information.

Cette ACI propose l'élaboration d'une mémoire vive *spécialisée* de très grande taille (plusieurs centaines de Giga octets, distribués sur plusieurs machines physiques) où réside le ou les index, afin de permettre l'émergence d'un point de vue radicalement différent sur l'indexation où la structure interne des index est résolument orientée mémoire (et non plus disque). L'objectif est de réviser les principes généraux des techniques d'indexation pour privilégier les accès directs aux données et les traitements parallèles. Dans ce cas, on autorise, par exemple, la redondance d'information à grande échelle ou encore l'utilisation de nombreuses structures annexes pré calculées, sans contrainte de taille, pourvu que l'ensemble participe à une accélération spectaculaire de la vitesse des recherches. Cette mémoire a la particularité d'employer des composants reconfigurables qui permettent, d'une part, d'adapter, *au niveau matériel*, la gestion de la mémoire aux caractéristiques propres de chaque type d'index et, d'autre part, d'offrir la possibilité d'implanter, *toujours au niveau matériel*, des algorithmes, spécifiques à chaque index, exhibant alors un fort parallélisme, contrebalançant leur extrême complexité avérée dans les architectures traditionnelles (fonctions de comparaison complexes par exemple).

Trois champs disciplinaires représentatifs serviront de support pour valider cette proposition : la génomique, la recherche d'images par le contenu et la recherche documentaire basée sur les textes et leurs structures.

Coordinateur du projet

Nom	Prénom	Laboratoire (sigle éventuel et nom complet)
LAVENIER	Dominique	IRISA, Symbiose, Rennes (Institut de Recherche en Informatique et Systèmes Aléatoires)

Action Concertée Incitative
MASSES DE DONNEES
Descriptif complet du projet

Organisme de rattachement financier pour le présent projet

INRIA

Equipes ou laboratoires partenaires

Symbiose, IRISA , Rennes

TexMex, IRISA, Rennes

Equipage, Valoria, Vannes

R2D2, IRISA, Rennes

Action Concertée Incitative
MASSES DE DONNEES
Descriptif complet du projet

II - PRESENTATION DETAILLEE DU PROJET

**A- IDENTIFICATION DU COORDINATEUR ET DES AUTRES
PARTENAIRES DU PROJET :**

A1- Coordinateur du Projet :

M. ou Mme. Prénom Nom	M. Dominique LAVENIER
Fonction	DR CNRS
Laboratoire (Nom complet et sigle le cas échéant)	Institut de Recherche en Informatique et Systèmes Aléatoires (IRISA)
Adresse	Campus de Beaulieu 35042 Rennes cedex
Téléphone	02 99 84 72 17
Fax	02 99 84 71 71
Mél	Dominique.Lavenier@irisa.fr

Action Concertée Incitative
MASSES DE DONNEES
Descriptif complet du projet

A2 (a) - Equipes ou laboratoires partenaires du Projet :

Identification de l'équipe ou du laboratoire

Equipe ou Laboratoire	Equipe Symbiose, IRISA
Adresse	Campus de Beaulieu 35042 Rennes cedex

Organisme de rattachement financier de l'équipe pour le présent projet

INRIA

Responsable du projet au sein de l'équipe ou du laboratoire

M. ou Mme. Prénom Nom	M. Dominique Lavenier
Fonction	DR CNRS
Téléphone	02 99 84 72 17
Fax	02 99 84 71 71
Mél	Dominique.Lavenier@irisa.fr

Autres membres de l'équipe participant au projet

Nom	Prénom	Poste statutaire	% du temps de recherche consacré au projet
NICOLAS	Jacques	CR INRIA	10 %
XXXXXX	Xxxxxx	Ingénieur à recruter	100 %
GIRAUD	Mathieu	Doctorant, ENS	30 %
RAIMBAULT	Frédéric	MCF, UBS, Vannes (délégation CNRS à l'IRISA)	50 %

Références :

- P. Guerdoux-Jamet, D. Lavenier, *SAMBA: Hardware Accelerator for Biological Sequence Comparison*, CABIOS, 13 (6), 1997
- C. Delamarche, P. Guerdoux-Jamet, R. Gras et J. Nicolas, *A symbolic-numeric approach to find patterns in genomes : Application to the translation initiation sites of E. coli*, Biochimie, 81, Elsevier, 1999
- F. Raimbault and D. Lavenier. ROOM, *Des machines reconfigurables orientées objet*, 8^{ième} Symposium en Architectures Nouvelles de Machines (Sympa), Tunis, 2002.
- D. Lavenier, *Speeding up genome computations with a systolic accelerator*, SIAM news, 31 (8) 1998.
- D. Lavenier, S. Derrien, S. Guyetant, S. Rubini, *A reconfigurable parallel disk system for filtering genomic banks*, ERSA'03, Las Vegas, USA, 2003.

Action Concertée Incitative
MASSES DE DONNEES
Descriptif complet du projet

A2 (b) - Equipes ou laboratoires partenaires du Projet :

Identification de l'équipe ou du laboratoire

Equipe ou Laboratoire	Equipe TexMex, IRISA
Adresse	Campus de Beaulieu 35042 Rennes cedex

Organisme de rattachement financier de l'équipe pour le présent projet

INRIA

Responsable du projet au sein de l'équipe ou du laboratoire

M. ou Mme. Prénom Nom	M. Laurent Amsaleg
Fonction	CR CNRS
Téléphone	02 99 84 74 44
Fax	02 99 84 71 71
Mél	Laurent.Amsaleg@irisa.fr

Autres membres de l'équipe participant au projet

Nom	Prénom	Poste statutaire	% du temps de recherche consacré au projet
BERTI-EQUILLE	Laure	MCF, univ. Rennes-1	20 %

Références :

L. Berti, *Quality-extended query processing for distributed sources*, International Workshop on Data Quality in Cooperative Information Systems, DQCIS'2003, Siena, Italy, Janvier 2003.

L. Amsaleg, P. Gros, *Content-based retrieval using local descriptors: problems and issues from a database perspective*, Pattern Analysis and Applications, 4(2/3):108-124, 2001.

L. Berti, F. Moussouni, A. Arcade, *Integration of biological data on Transcriptome*, Revue ISI-NIS, Numéro Spécial Interopérabilité et Intégration des Systèmes d'Information, Hermès, 6(3/2001):61-86, 2001.

L. Amsaleg, P. Gros, S.-A. Berrani, *A Robust Technique to Recognize Objects in Images, and the DB Problems it Raises*, Proc. of the Workshop on Multimedia Information Systems, Capri, Italie, Novembre 2001.

S. Berrani, L. Amsaleg, P. Gros, *Approximate k-nearest-neighbor searches : a new algorithm with probabilistic control of the precision*. Technical Report PI 1495, nov 2002, IRISA.

Action Concertée Incitative

MASSES DE DONNEES

Descriptif complet du projet

A2 (c) - Equipes ou laboratoires partenaires du Projet :

Identification de l'équipe ou du laboratoire

Equipe ou Laboratoire	Equipe R2D2, IRISA
Adresse	Campus de Beaulieu 35042 Rennes cedex

Organisme de rattachement financier de l'équipe pour le présent projet

INRIA

Responsable du projet au sein de l'équipe ou du laboratoire

M. ou Mme. Prénom Nom	M. François Charot
Fonction	CR INRIA
Téléphone	02 99 84 72 15
Fax	02 99 84 71 71
Mél	Francois.Charot@irisa.fr

Autres membres de l'équipe participant au projet

Nom	Prénom	Poste statutaire	% du temps de recherche consacré au projet
DERIEN	Steven	ATER, univ. Rennes-1	30 %
WAGNER	Charles	IR CNRS	20 %
XXXXXX	XXXXXX	Post-Doc à recruter	100 %

Références :

S. Derrien, *Techniques de partitionnement pour l'implantation de réseaux de processeurs sur FPGA*, thèse, université de Rennes 1, décembre 2002.

S. Derrien, A.C. Guillou, P. Quinton, T. Risset et Wagner, *Automatic Synthesis of efficient interfaces for compiled regular architectures*, Proc. System Architecture Modeling and Simulation Workshop, LNCS, Springer Verlag, 2002.

V. Messé et F. Charot, *A Flexible Code Generation Framework for the Design of Application Specific Programmable Processors*, Int. Workshop on Hardware/Software Codesign (CODES'99), May 1999.

F. Charot et V. Messé, *La compilation recyclable au service de la définition interactive d'ASIP*, Technique et Science Informatiques, 20, 2, février 2001.

S. Guyétant, S. Derrien, D. Lavenier, *Architecture parallèle pour la génomique*, SympA'8, 8ème Symposium en Architectures Nouvelles de Machines, Hamamet, Tunisie, 2002

Action Concertée Incitative

MASSES DE DONNEES

Descriptif complet du projet

A2 (d) - Equipes ou laboratoires partenaires du Projet :

Identification de l'équipe ou du laboratoire

Equipe ou Laboratoire	Equipe Equipage-APRIME, Valoria, UBS
Adresse	Campus de Tohanic, 56000 Vannes

Organisme de rattachement financier de l'équipe pour le présent projet

Université de Bretagne Sud

Responsable du projet au sein de l'équipe ou du laboratoire

M. ou Mme. Prénom Nom	M. Pierre-François Marteau
Fonction	Prof. UBS
Téléphone	02 97 01 72 21 (02 97 01 72 99)
Fax	02 97 01 72 00
Mél	Pierre-Francois.Marteau@univ-ubs.fr

Autres membres de l'équipe participant au projet

Nom	Prénom	Poste statutaire	% du temps de recherche consacré au projet
MENIER XXXXX	Gildas Xxxxx	MCF, UBS Doctorant à recruter	50 % 100 %

Références :

Marteau P.F. et Ménier G., *Alignement approximatif d'arbres pour la recherche d'information en contexte dans les données XML hétérogènes - Fusion d'information structurées et textuelles*, soumis à la revue Techniques et Sciences Informatiques, Hermès, 2003.

Ménier G., Marteau P.F., *Information retrieval in heterogeneous XML knowledge bases*, The 9th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, IEEE, 1-5 July, Annecy, France, 2002.

Ménier G., Marteau P.F., Le Saux E., Revault J., *SIRIUS : Un Moteur de Recherche d'Informations Contextuelles dans les Bases de Documents XML*, Journée Applications, Apprentissage et Acquisition de Connaissances à partir de Textes électroniques (A3CTE), 25-29 Juin, Grenoble, 2001

Marteau P.F., de Loupy C., Bellot P., El Bèze M., *Le Traitement Automatique du Langage Naturel Comme Outil d'Assistance à La Fonction d'Intelligence Economique : vers une Architecture " Push-Pull " d'accès à l'information*, revue Système et Sécurité, Sigre Editeur, Fevrier-Mars 1999.

Action Concertée Incitative

MASSES DE DONNEES

Descriptif complet du projet

B - DESCRIPTION DU PROJET

B1 – Objectifs et contexte :

Contexte

Ces dernières années, dans de nombreux domaines, on assiste à une très forte production de données en raison de l'accroissement conjugué des capacités de calcul et de stockage, à la multiplication des outils de calcul et à la démocratisation des systèmes de numérisation et d'acquisition de l'information. Le développement d'outils et de stratégies informatiques pour faire face à l'accumulation des données tout en facilitant l'extraction d'informations pertinentes est un problème d'actualité.

Dans ce contexte, une problématique générale est la recherche d'information par le contenu ; les trois champs disciplinaires suivants sont exemplaires des difficultés rencontrées pour maîtriser l'expansion de leur masses de données.

- En Génomique, une technique couramment employée est la comparaison d'une nouvelle séquence avec toutes celles contenues dans les banques de données afin d'en déduire des propriétés fonctionnelles. Depuis plus de 10 ans, la taille des banques double chaque année. Actuellement, cela représente quelques centaines de Giga octets, dans 10 ans on aura multiplié par 1000 ce volume.
- Dans le domaine de l'image, certaines applications recherchent, à partir d'une image exemple, celles qui, dans une base d'images, lui ressemblent le plus, visuellement parlant. Ce type de recherche par le contenu est notamment utilisé pour la reconnaissance de visage ou encore pour l'interrogation de photothèques (personnelles ou institutionnelles) lorsque l'on recherche des images similaires à celles que l'on possède. En plus des problèmes de volume cités ci-dessus, on fait face ici à la grande complexité de la comparaison entre les descriptions des images.
- Dans le domaine de la recherche documentaire textuelle, sur Internet ou dans des corpus privés, la comparaison de documents du point de vue de leur structure et/ou de leur contenu est une activité essentielle. La profusion des informations disponibles en ligne rend très difficile l'extraction de données pertinentes.

De façon générale, l'ensemble des applications nécessitant l'exploration de très grands volumes de données fait appel à des techniques d'indexation. Celles-ci, en organisant les données d'une manière particulière (et souvent spécifique aux données indexées et aux traitements qui leur sont appliqués), visent à confiner au maximum les recherches, limitant ainsi la quantité de données que l'on doit accéder et analyser pour répondre à une question. Sans index, les recherches ne peuvent qu'examiner exhaustivement les données. Indexer a donc pour conséquence de minimiser le temps mis par le système pour répondre aux requêtes. L'indexation est un sujet ancien, et la littérature scientifique comporte de très nombreuses solutions.

L'utilisation d'index s'est révélée particulièrement intéressante dès que la taille du corpus de données dans lequel se déroulent les recherches dépasse la taille de la mémoire primaire des machines. Dans ce cas, l'index est obligatoirement stocké sur disques, et la recherche ne peut accéder qu'à un sous-ensemble réduit de celui-ci. La lenteur des accès aux disques a un impact fondamental sur la conception de toutes les techniques d'index. D'abord, un index est conçu pour être le plus compact possible et donc pour occuper le moins d'espace sur le disque, afin de minimiser ce que l'on doit y lire. Ensuite, un index est conçu pour maximiser les accès séquentiels et éviter au maximum les accès aléatoires. Enfin, la quasi-totalité des index incorporent implicitement l'hypothèse que comparer des informations entre elles est bien plus économique que d'accéder au disque. Ces raisons font que le fonctionnement de tous les index est orienté "disque", c'est-à-dire que leur conception repose sur des notions de pages, de blocs, de pistes, de traitement en séquence de tous les enregistrements appartenant à une page qui a été chargée, etc.

Objectifs

Cette ACI propose l'élaboration d'une mémoire vive *spécialisée* de très grande taille (plusieurs centaines de Giga octets, distribués sur plusieurs machines physiques) où réside le ou les index, afin de permettre l'émergence d'un point de vue radicalement différent sur l'indexation où la structure interne des index serait orientée mémoire (et non plus disque). L'objectif est de réviser les principes généraux des techniques d'indexation pour privilégier les accès

Action Concertée Incitative

MASSES DE DONNEES

Descriptif complet du projet

directs aux données, les traitements parallèles, et où l'on autorise, par exemple, la redondance d'information à très large échelle ou encore l'usage de structures annexes pré-calculées, sans contraintes de taille, pourvu que l'ensemble participe à une accélération spectaculaire de la vitesse des recherches. Cette mémoire a la particularité d'employer des composants reconfigurables qui permettent, d'une part, en les configurant (à l'exécution) d'une façon précise, d'adapter, *au niveau matériel*, la gestion de la mémoire aux caractéristiques propres de chaque type d'index, et, d'autre part, d'offrir la possibilité d'implanter, *toujours au niveau matériel*, des algorithmes, spécifiques à chaque index, exhibant alors un fort parallélisme, contrebalançant leur extrême complexité avérée dans les architectures traditionnelles (l'alignement de deux séquences, l'évaluation de la similarité visuelle de deux images ou la comparaison de deux textes sont des opérations élémentaires d'une grande complexité).

La mémoire spécialisée envisagée ne se résume pas à mettre au point un dispositif qui étend très largement la mémoire vive d'un ordinateur standard. Dans ce cas, on souffrirait des problèmes classiques liés aux grandes mémoires : parce que le CPU est doté de différents niveaux de caches mémoire dont il faut préserver au maximum le contenu, les accès à la mémoire doivent être le moins aléatoires possible. Nous retrouverions, à une échelle différente, les contraintes liées à l'utilisation de disques.

Il s'agit plutôt de construire une mémoire annexe, - plus précisément une **mémoire d'index** - accédée par le processeur de la machine d'une manière spécifique, et qui possède comme caractéristique saillante d'assurer que l'accès à n'importe quelle adresse mémoire se fait avec un coût fixe, mettant à égalité les accès en séquences et les accès aléatoires. Le processeur accède à cette mémoire indirectement en lui soumettant des requêtes déclenchant l'exécution d'opérations élémentaires d'indexation qui lui renvoient les résultats attendus. Ces opérations élémentaires s'exécutent au sein de circuits reconfigurables où sont câblés des versions hautement parallèles des algorithmes les mettant en œuvre.

La constitution d'une mémoire d'index de plusieurs centaines de Giga octets, extensible suivant les besoins applicatifs, passe matériellement par une approche parallèle. Elle doit donc être répartie en plusieurs entités physiques, ce qui sous-entend la possibilité d'une distribution équitable de l'index sur chaque entité. Cette structuration parallèle, si elle impose un découpage préalable de l'index, présente néanmoins l'avantage de pouvoir conduire une recherche concurrente, et donc de contribuer à réduire les temps de réponse. Concrètement, la structure d'accueil d'une mémoire d'index répartie sera un cluster de PC.

Comparaison avec d'autres approches

Au début des années 1990, plusieurs travaux [Garcia-Molina 1992] se sont focalisés sur l'implantation des bases de données en mémoire centrale (référéncé sous le terme de MMDBS : *Main Memory Database Systems*). Leur objectif était d'accélérer les requêtes en mémorisant les données dans une grande mémoire vive en lieu et place des disques. Ces travaux se sont heurtés à deux problèmes majeurs. D'une part, il fallait maintenir en parallèle un système sophistiqué de tolérance aux pannes, de journalisation des transactions et de sauvegarde sur disque du fait du caractère non permanent et peu fiable des mémoires vives ; ce système fonctionnant en concurrence avec le moteur d'interrogation le ralentissait. Pour améliorer les performances, des mécanismes matériels *ad hoc* de sauvegarde étaient proposés ; de notre côté, le matériel reconfigurable que nous associons à la mémoire est directement destiné à l'accélération des traitements. D'autre part, un MMDBS impliquait soit d'avoir la maîtrise complète de la mise à jour de la base de données, soit de la dupliquer régulièrement ; au contraire, le système de mémoire que nous proposons ne contient que des références sur les données et s'adapte ainsi à n'importe quel système de gestion de base de données.

Plus récemment, des solutions à base de clusters de PC ont été développés pour supporter l'exécution des moteurs de recherche sur Internet. Par exemple, le système Google (<http://www.google.com/press/highlights.html>) repose sur un pool d'une dizaine de milliers de machines sous Linux [Wagner 2000]. A la différence du projet REMIX, ce genre de plate-forme est très spécialisée dans sa fonction (recherche de texte) et repose principalement sur des index inversés pour la recherche par mots clés [Brin 1998]; à notre connaissance il n'existe pas de projet d'adaptation de tel système pour une recherche mêlant structure et contenu, et plus encore sur des données images ou séquences génomiques. La même constatation s'applique aux choix d'entreprises commerciales comme AltaVista ou InfoSeek qui privilégient la technologie des multiprocesseurs symétriques (SMP) ; ces infrastructures n'apportent pas de solution nouvelle pour contourner le goulot d'étranglement des entrées-sorties sur disques [Lu 1998].

Action Concertée Incitative

MASSES DE DONNEES

Descriptif complet du projet

Contributions

Quatre équipes de recherche, d'horizon différents, participent à cette ACI multi-thématique. Les contributions attendues sont diverses et complémentaires :

- La recherche de nouveaux algorithmes d'indexation dans un contexte de support mémoire quasi illimité ;
- La création d'une plate-forme matérielle expérimentale originale, bâtie autour d'un cluster de PC où chaque nœud est enrichi d'une mémoire d'index reconfigurable;
- La mise en œuvre d'un environnement de programmation permettant de gérer simultanément le parallélisme et les aspects reconfigurables ;
- La validation dans trois domaines applicatifs représentatifs des masses de données : la génomique, la recherche d'images, et la recherche documentaire textuelle.

Action Concertée Incitative

MASSES DE DONNEES

Descriptif complet du projet

B2 – Description du projet :

REMIX : Mémoire Reconfigurable pour l'Indexation de Masses de Données

B2.1 – Introduction

Le projet de recherche multi-thématique proposé dans cette ACI vise :

1. à élaborer une mémoire reconfigurable de grande taille pour indexer des masses de données conséquentes ;
2. à repenser les algorithmes d'indexation dans le cadre de ce type de mémoire ;
3. à développer un environnement logiciel qui prenne à la fois en compte les aspects parallélisme et les aspects reconfiguration ;
4. à étudier la mise en œuvre, sur ce support, de trois applications émanant de trois domaines applicatifs différents.

Globalement, le projet peut être représenté par la figure 1. Au centre, il y a la plate-forme matérielle bâtie autour d'un cluster de PC où chaque PC est équipé d'une mémoire reconfigurable de 16 Giga octets dédiée exclusivement à l'indexation. Cet ensemble est inclut dans un environnement logiciel qui gère à la fois les aspects programmation, reconfiguration et exploitation du système. Enfin, trois domaines d'applications complètent ce projet : la génomique, l'image, et la fouille de données textuelles. Ces domaines ont été sélectionnés à la fois en fonction des quantités de données manipulées, et des compétences des équipes. Nous résumons ci-après leurs compétences ainsi que leur contribution à l'ACI.

- L'équipe **Symbiose** (projet INRIA de bio-informatique) possède un axe de recherche, baptisé *architecture et parallélisme*, qui a pour objectif de réduire les temps de calcul des traitements liés à la génomique. Les moyens mis en œuvre sont la parallélisation des traitements sur une grille dont les nœuds sont composés de super calculateurs (projet GénoGRID [Lavenier 2003], et la définition d'architectures spécialisées pour filtrer à grande vitesse les banques de données (projet RDISK [Lavenier 2003]). Un autre axe de recherche est la recherche de motifs complexes dans les banques de séquences. Cette recherche s'appuie sur une structure d'indexation construite autour d'arbres de suffixes [Knuth 1977]. Dans le cadre de l'ACI, les contributions de Symbiose seront doubles : (1) l'expertise en architecture et parallélisme, et (2) le domaine d'applications : la génomique. On visera, en particulier, l'implémentation parallèle des arbres de suffixes.
- L'équipe **R2D2** (projet IRISA) a pour problématique la conception de systèmes matériels spécialisés sur silicium dans le cadre de l'approche système monopuce (SoC). Plus spécifiquement, R2D2 cherche à réaliser des outils pour passer d'une spécification algorithmique à une mise en œuvre sur des plates-formes matérielles reconfigurables composées de blocs de type processeur (processeur de traitement de signal, processeur à jeux d'instructions spécifiques), et de blocs reconfigurables (à grain fin et à grain épais). Les travaux de l'équipe s'articulent selon trois axes : (1) le développement de méthodologies de synthèse d'accélérateurs matériels parallèles dédiés ; (2) l'exploration, l'estimation et le prototypage pour la conception de systèmes sur silicium ; (3) l'étude, en lien avec les domaines d'applications considérés et les contraintes de performance et faible consommation, de nouvelles architectures reconfigurables. Dans le cadre de l'ACI, R2D2 se concentrera sur la conception de la mémoire reconfigurable, ainsi que sur les aspects programmation et compilation.
- L'équipe **TexMex** (projet INRIA : techniques d'exploitation des données multimédias) étudie les différentes techniques d'exploitation de très larges corpus de données multimédias. Définition de schémas de description de documents, navigation dans des bases et recherche de documents similaires, indexation, analyse conjointe des différents médias composant un même document, aspects bases de données forment autant de thèmes qui font partie des axes de travail que cette équipe privilégie. Les chercheurs impliqués dans la proposition de projet (Laure Berti-Équille et Laurent Amsaleg) ont respectivement une forte expertise des problèmes liés à l'indexation de banques génomiques et aux aspects bases de données de l'indexation de bases d'images.

Action Concertée Incitative

MASSES DE DONNEES

Descriptif complet du projet

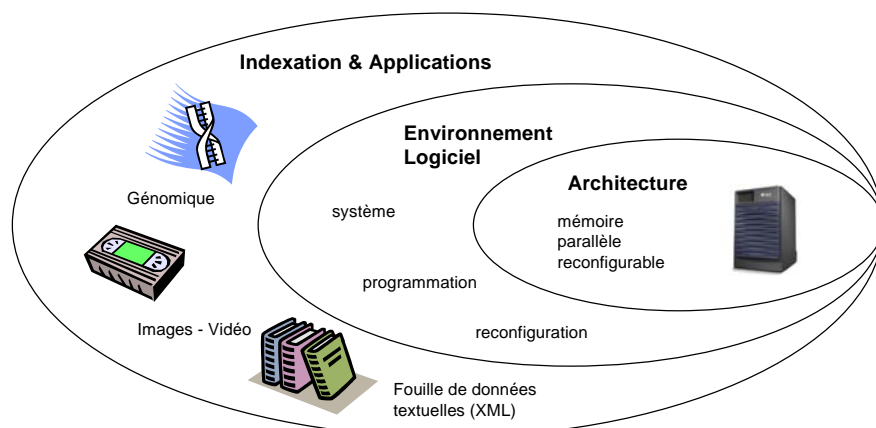


Figure 1 : Organisation de l'ACI REMIX. Trois niveaux sont considérés : l'architecture matérielle, l'environnement logiciel et les techniques d'indexation validées sur trois domaines d'applications.

- Le groupe **APRIME** est l'un des projets de l'équipe **EQUIPAGE** du laboratoire VALORIA à l'UBS (Université de Bretagne Sud). Ce groupe s'intéresse aux mécanismes d'apprentissage susceptibles d'être mis en œuvre dans les processus d'indexation, de recherche et d'extraction d'information dans les grandes bases de données semi-structurées et distribuées XML. Les principes algorithmiques proposés pour gérer les structures d'index exploitent des modèles statistiques ou probabilistes, des modèles géométriques fondés sur des distances, des modèles à base de règles ainsi que des modèles de fusion de critères pour l'intégration des composantes numériques et symboliques (projet SIRIUS). Dans le cadre de l'ACI, la contribution d'APRIME ciblera principalement un domaine d'application : celui du traitement des grandes bases de données textuelles XML. La tâche principale traitée concernera l'adaptation des algorithmes d'indexation et de recherche en vue de leur implémentation sur l'architecture proposée. Nous nous intéresserons principalement aux algorithmes coûteux en temps de calcul, et tout particulièrement à un algorithme d'alignement approximatifs d'arbres de contexte [Ménier & Marteau 2002, 2003].

Ces quatre équipes de recherche possèdent des compétences variées, tant sur des aspects applicatifs (génomique, image, texte) liés à la recherche d'information dans de gros volumes de données, que sur des aspects plus génériques tels que l'architecture, le reconfigurable, le parallélisme, la compilation ou les environnements de programmation.

Au delà d'une réalisation commune, cette ACI offre aux équipes une opportunité réelle de poursuivre activement leurs propres axes de recherche et de valider concrètement diverses hypothèses sur un support *ad hoc*. La suite du texte détaille plus précisément le projet de l'ACI REMIX et, au travers de cette description, souligne les problématiques de recherches abordées par les équipes. Les problématiques de l'indexation sont d'abord exprimées. Puis l'architecture matérielle de la mémoire reconfigurable est présentée, suivie de l'environnement logiciel. Enfin, les trois domaines applicatifs qui serviront de support à cette sont ensuite décrits.

B2.2 – Repenser l'indexation

Equipes TexMex, Equipage et Symbiose

La communauté base de données s'est intéressée de longue date aux coûts liés au processus de recherche, en particulier lorsque la quantité de données à gérer est devenue telle qu'il était obligatoire de stocker cette masse d'information sur disques. Sont alors apparues les très nombreuses solutions d'indexation, le plus souvent arborescentes, visant à organiser les données d'une manière spécifique. Le but était ici à la fois de limiter la quantité de données à lire depuis les disques mais aussi à restreindre autant que possible la quantité de calculs à faire.

Action Concertée Incitative

MASSES DE DONNEES

Descriptif complet du projet

Les nombreux algorithmes proposés sont bien évidemment “ orientés disques ”, c'est-à-dire que leur fonctionnement se base sur les notions de blocs, de pistes et autres paramètres physiques des disques dont **on doit** tenir compte dans les transferts mémoire primaire mémoire secondaire, sans quoi il est impossible d'être performant. De plus, tous ces algorithmes visent (1) à exhiber un comportement qui soit **le plus séquentiel possible** afin de bénéficier de la rapidité des accès disques et (2) à éviter au maximum les accès disques aléatoires, terriblement lents. Finalement, ces algorithmes maximisent la compacité des structures en tentant de minimiser le nombre de pages disque sur lesquelles elles s'étendent, toujours pour limiter autant que possible les transferts entre mémoire principale et mémoire secondaire, goulet d'étranglement notoire.

L'architecture que propose le présent projet s'affranchit de toutes ces contraintes et permet d'aborder l'ensemble des problèmes de recherche par le contenu sous un angle radicalement différent. D'abord, la compacité des structures de données n'est pas un objectif premier : il semble plus intéressant, dans le contexte de cette ACI, de viser des structures permettant des accès aux données vraiment très rapides, même si cela se traduit par un accroissement de la taille des structures à gérer. Il est même envisageable d'introduire **volontairement** de la redondance de données (notion bannie des structures traditionnelles) dans un but accélérateur.

On peut ensuite repenser tous les schémas de fonctionnement des structures de données parce que la contrainte de la séquentialité des accès n'existe plus, le coût d'un accès aléatoire dans la mémoire d'index étant du même ordre que celui d'un accès en séquence. Il devient intéressant de privilégier les accès les plus directs possibles aux informations (au travers de pointeurs mémoire par exemple) au détriment de listes, traversées d'arbres, de sous-niveaux ou autres structures ensemblistes bien adaptées aux approches basées sur la notion de pages disques. Il est même possible d'envisager des modes de fonctionnement où l'on privilégie les opérations mémoires logiques (et, ou, ou exclusif, décalages) plutôt que de naviguer dans l'information.

Disposant d'une grande mémoire, il est finalement tout à fait envisageable de précalculer (hors ligne, avant toute exploitation par des recherches) certaines informations sur les données afin d'accélérer les recherches à venir, même si celles-ci sont volumineuses. On peut précalculer certaines données clés, permettant alors à la recherche en ligne d'en faire l'économie, ce qui diminue d'autant le temps pendant lequel un utilisateur attend une réponse.

Il est ainsi d'ores et déjà possible de discerner trois grands axes selon lesquels des innovations en recherche par le contenu peuvent se développer dans le contexte de cette ACI :

- repenser les structures d'index en les orientant mémoire (et en abandonnant les atavismes liés aux disques) ;
- précalculer et stocker en mémoire des informations permettant d'accélérer les traitements en lignes, critiques, en limitant le calculs effectivement effectués ;
- introduire de la redondance à des fins d'accélération.

B2.3 – Mémoire reconfigurable

Equipes R2D2 et Symbiose

La mémoire reconfigurable est le cœur du système. Elle doit supporter, entre autre, les deux contraintes suivantes :

1. contenir un index complet en mémoire vive, aussi volumineux soit-il ;
2. s'adapter aux mécanismes d'indexation selon la nature des masses de données ciblées : adressage spécialisé, taille des mots mémoire ajustée aux données, opérateurs de calcul dédiés à la nature des données, etc.

Une mémoire reconfigurable n'est pas seulement un espace de stockage, elle offre également des **services de calcul intégrés**. On peut, par exemple, imaginer que l'information stockée en mémoire est compressée et qu'un opérateur matériel câblé directement sur le port de données de la mémoire effectue une décompression à la volée.

L'architecture que nous proposons est illustrée par la figure 2. Une mémoire de grande taille (par exemple 16 Giga octets) est connectée via un composant reconfigurable à un ordinateur classique (un PC). Le rôle du composant reconfigurable est de supporter le dispositif matériel qui spécialise la mémoire d'index. Cette architecture diffère fondamentalement d'une machine standard “ gonflée ” en mémoire :

Action Concertée Incitative
MASSES DE DONNEES
Descriptif complet du projet

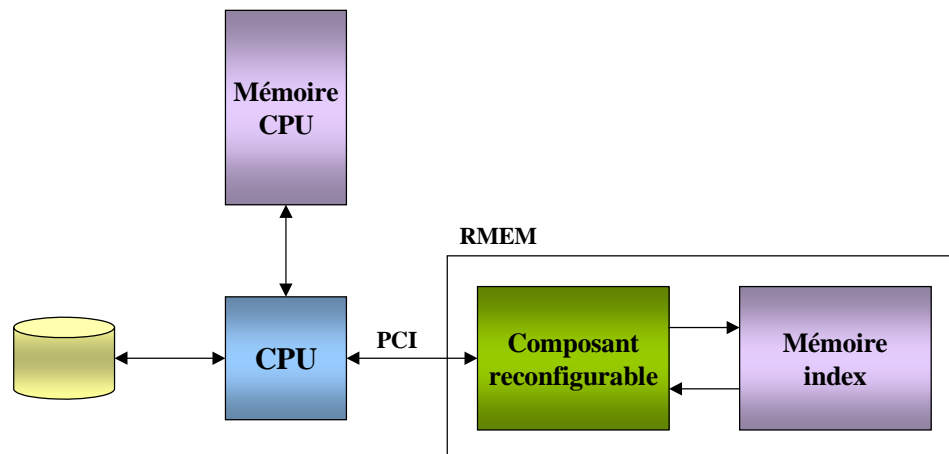


Figure 2 : Architecture d'une mémoire reconfigurable (RMEM) et connexion au processeur (CPU).

1. La mémoire (qui contient l'index) ne réside pas dans l'adressage du processeur. Par contre, le programme d'indexation réside dans la mémoire du CPU qui, lorsqu'il travaille sur l'index, lance des requêtes spécifiques à la mémoire d'index (par exemple, la recherche d'un motif sur une structure d'arbre).
2. La mémoire ne comporte pas de mécanismes de hiérarchie pour en accélérer l'accès. La notion de cache, par exemple, n'existe pas. L'indexation n'a donc pas à se préoccuper d'éventuelle localisation spatiale des données. Le coût (en temps) d'une lecture est constant, quelle que soit l'adresse pointée et quels que soient les accès antérieurs.
3. La mémoire est "augmentée" de ressources matérielles reconfigurables pour en spécialiser l'usage, tant d'un point de vue adressage, que traitement des données. On peut, par exemple, imaginer une sélection à la volée des données via un filtre matériel implanté directement en sortie des mémoires.

Concrètement, une mémoire reconfigurable (RMEM) comprend un composant reconfigurable auquel est rattaché plusieurs bancs mémoire. Les mémoires sont analogues à celles utilisées dans les PC, à savoir des barrettes de 512 Méga octets, voire un Giga octets, en technologie DDR SRAM à 266 MHz. Une organisation en mots de 256 bits, par exemple, délivre une bande passante mémoire de l'ordre de 8 Giga octets par seconde. L'ensemble mémoires + composant FPGA tient sur une carte standard et peut être connectée à un ordinateur par l'intermédiaire d'un *slot* PCI. La taille maximum de mémoire qui peut être implantée sur une seule carte est actuellement limitée par des aspects purement technologiques : les capacités des composants mémoire sont bornés ; on ne peut donc pas intégrer sur une carte unique, pour des raisons d'encombrement et de consommation électrique plusieurs centaines de Giga octets. A l'heure actuelle, 16 Giga octet semble une taille raisonnable.

Le composant reconfigurable est un circuit FPGA Virtex-II de la famille Xilinx [Xilinx,2002] possédant un grand nombre de broches d'entrées/sorties pour connecter l'ensemble des bancs. La technologie FPGA, développée depuis le milieu des années 80, permet de programmer en quelques centaines de milli secondes – et à volonté - une architecture matérielle spécifique pour un traitement donné. L'avantage, par rapport à un système micro-programmé est qu'une fonction ou une opération qui demande le déroulement de plusieurs dizaines d'instructions machine, peut être câblée et exécutée en un seul cycle d'horloge.

Ici, le composant FPGA joue plusieurs rôles. D'abord, il ajuste l'organisation des bancs mémoires en fonction de la structure de l'index. Ensuite, il contrôle l'adressage des mémoires par l'intermédiaires d'unités spécialisées qui, par exemple, peuvent être adaptées à un parcours d'arbre ou encore contenir des "look-up tables" spécifiques à un index. Enfin, les ressources reconfigurables peuvent avantageusement être mises à contribution pour effectuer directement des traitements sur les données dès la sortie de la mémoire d'index, tels que le calcul de distances entre

Action Concertée Incitative

MASSES DE DONNEES

Descriptif complet du projet

descripteurs d'images ou séquences génomiques. Actuellement, les capacités en ressources matérielles des composants FPGA de dernière génération offrent l'équivalent de plusieurs millions de portes pour des fréquences de fonctionnement de plusieurs centaines de MHz. En d'autres termes, la puissance de calcul que l'on adjoint à la mémoire peut représenter l'équivalent de plusieurs dizaines de Giga opérations arithmétiques, soit un ordre de grandeur supplémentaire par rapport à un microprocesseurs fonctionnant à quelques Giga Hz.

L'architecture RMEM que nous proposons est à rapprocher du concept de PAM (*Programmable Active Memory*) proposé par Jean Vuillemin dans le milieu des années 80 [Vuillemin, 1996]. Mais les réalisations physiques qui ont suivi se sont plutôt focalisées sur l'aspect "accélérateur de calcul", les mémoires étant plutôt utilisées comme stockage temporaire par rapport à un flux de données provenant de l'extérieur. Cette remarque est également vraie pour tous les projets de cartes reconfigurables, tant académiques que commerciaux, qui se sont succédés depuis 15 ans.

Pour atteindre des tailles d'index de plusieurs centaines de Giga octets, l'approche retenue consiste à "doper" chaque nœud d'un cluster de PC avec une telle mémoire, de manière à atteindre une capacité globale suffisante. Dans ce schéma, le cluster de PC offre d'abord une structure d'accueil relativement standard pour connecter un ensemble d'entités RMEM. Ensuite, c'est un support facilement extensible pour ajuster la mémoire d'indexation en fonction des besoins. Bien sûr, au niveau des applications, on fait implicitement les deux hypothèses suivantes :

1. une structure d'index peut être distribuée ;
2. une recherche parallèle ne pénalise pas les temps d'exécution.

Intuitivement, parmi les applications retenues, ces deux hypothèses sont viables, même s'il reste à mesurer sérieusement l'apport du parallélisme. En effet, en supposant une indexation parallèle possible, se pose le problème d'un équilibre équitablé entre les nœuds du cluster : la variabilité des temps de réponse est-il fonction de la nature des requêtes ? la manière dont est réparti l'index est-il optimum, etc.

B2.4 – Environnement logiciel

Equipes R2D2 et Symbiose

Le système d'indexation présenté dans le paragraphe précédent contient des éléments matériels qui font habituellement appel à des outils de développement très différents : les PC se programment dans un langage de programmation séquentiel (C, Java,...), le cluster de PC (les communications et les synchronisations) est géré à l'aide de bibliothèques de communication (MPI, RMI,...) et l'architecture du composant FPGA est spécifiée à l'aide d'un langage de description de matériel (VHDL) et d'outils de CAO.

Le modèle d'exécution

Compte tenu de la complexité de ce système il nous paraît primordial d'abstraire le fonctionnement des différents éléments matériels, en proposant un modèle d'exécution unifié. La figure 3 expose les différentes parties du modèle d'exécution du système d'indexation qui sont à prendre en charge par l'environnement de développement. L'idée générale repose sur le concept de processus séquentiel communiquant en associant un processus à chaque activité distincte du système. Tout les nœuds sont identiques tant du point de vue de leur matériel -- un processeur (en bas), un composant reconfigurable (au milieu) et de la mémoire (en haut) -- que du point de vue de leur logiciel ; seul le contenu de leur mémoire les différencie, chacune étant chargé avec une partie de l'index. A l'intérieur d'un nœud, un processus (*P0* sur la figure 3) est exécuté par le processeur du PC pour prendre en charge les aspects systèmes (configuration, chargement,...), les parties de l'application qui n'ont pas besoin de la mémoire d'index, les communications internes au cluster et celles avec l'extérieur. Dans la carte RMEM on trouve des processus spécifiques aux traitements (par exemple sur la figure *P2* et *P3*) et des processus spécifiques aux entrées-sorties avec le PC et la mémoire (par exemple sur la figure *P1* et *P4*).

Modèle de programmation

Pour atteindre un haut niveau de performance les différentes activités représentées figure 3 doivent être codées, placées et surtout synchronisées efficacement. Idéalement on aimerait pouvoir décrire une application d'indexation de manière conventionnelle (un algorithme séquentiel) et laisser aux outils de compilation le soin de répartir, de coordonner et de générer le code de l'ensemble des activités. Dans l'état actuel des outils et des connaissances nous pensons qu'il est plus réaliste de demander au programmeur de réaliser une partie du placement des activités

Action Concertée Incitative

MASSES DE DONNEES

Descriptif complet du projet

parallèles et de confier la synchronisation, la génération et l'optimisation de code à un outil *ad hoc*. Pour réaliser cet outil nous appuierons sur des travaux en cours dans l'équipe Symbiose autour de la conception d'accélérateurs programmables dans de la logique reconfigurable (ROOM) [Raimbault 2002] et sur les outils réalisés par l'équipe R2D2 pour la synthèse et l'optimisation de code recyclable (Calife) [Charot 2001]. Le modèle de programmation envisagé pour l'écriture d'un algorithme d'indexation sur l'ensemble de notre système d'indexation est celui d'un langage orienté objet, doté d'une extension parallèle pour l'expression de la répartition des données, associé à une bibliothèque de composants optimisés (processeurs ou opérateurs spécialisés) pour chaque domaine d'application.

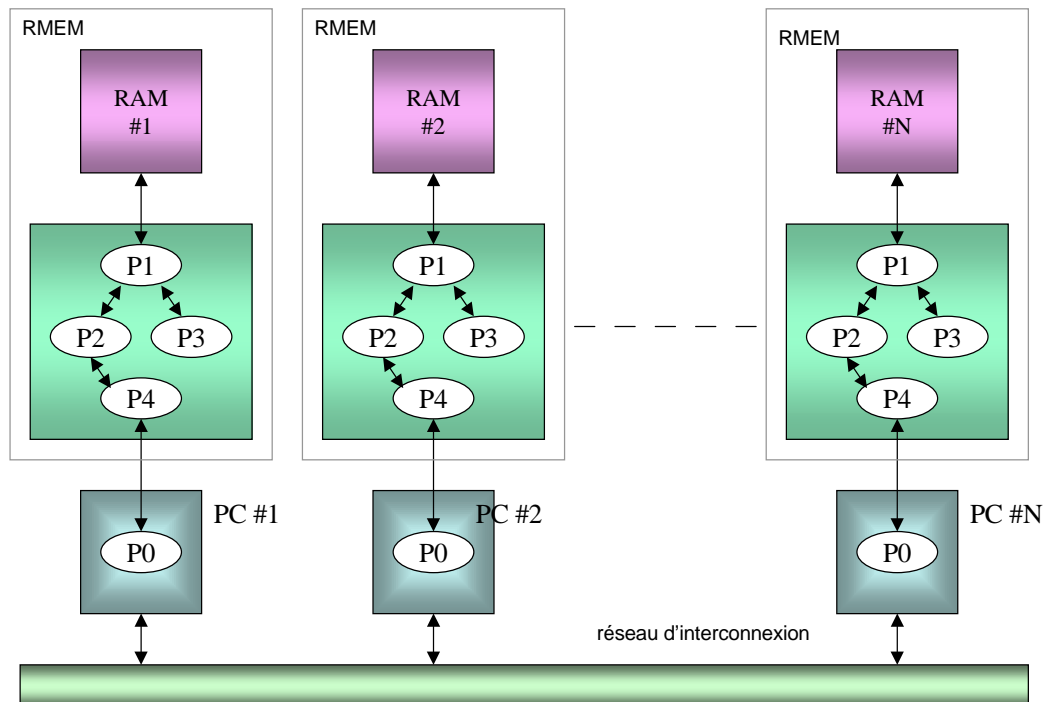


Figure 3 : modèle d'exécution du système d'indexation

Compilation

A- La spécialisation du matériel

Avec l'évolution de la densité des circuits VLSI, il est possible d'implanter dans un composant FPGA des cœurs de processeurs, soit sous la forme d'un bloc pré-caractérisé (ou cœur matériel), soit sous la forme d'un bloc utilisant les ressources du composant FPGA (on parle alors de cœur logiciel). Les principaux fournisseurs de circuit FPGA proposent aujourd'hui de telles solutions dans l'optique d'une conception conjointe matérielle-logicielle : les parties de l'application non critiques, ou dominées par du contrôle complexe, sont décrites par du code exécutés par des cœurs de processeurs logiciels tandis que les parties de calcul intensif sont implantées sous la forme de blocs spécialisés ; ces derniers étant produits à l'aide d'outils de synthèse d'architecture (outil MMAAlpha, ou plate-forme BSS développée dans R2D2). Mais il peut être également judicieux de recourir à des approches qui visent à concevoir des processeurs à jeu d'instructions spécifiques (ASIP), dimensionnés pour répondre exactement aux besoins des applications visées. Lorsque le processeur utilisé est un cœur logiciel, un environnement de développement logiciel (compilateur, simulateur) lui est généralement associé, il peut être utilisé si la qualité du code produit est jugée satisfaisante. Si les orientations poussent à définir un cœur de processeur de type ASIP, l'environnement Calife/Armor développé dans R2D2 peut être employé pour construire un optimiseur de code ainsi que les outils de simulation associés.

Action Concertée Incitative

MASSES DE DONNEES

Descriptif complet du projet

B - La distribution des tâches dans un composant FPGA

Pour faciliter l'implantation d'un ensemble de processus communiquant dans un composant reconfigurable, l'équipe Symbiose a développé le concept de ROOM (*Reconfigurable Object Oriented Machine*). Le principe [Raimbault, 2002] est de décrire l'algorithme à implanter dans le FPGA dans un sous-ensemble de Java et de produire le code des processus en associant un processus par classe d'objet. De cette manière le programmeur guide la distribution des tâches en découpant son algorithme en classe mais toutes les communications et les synchronisations sont générées automatiquement par le compilateur. De plus, l'encapsulation sous la forme de classe et leur implantation indépendante permet de composer les parties de l'application exécutées par des cœurs de processeurs logiciels, les parties réalisées par des opérateurs matériels spécialisés, et les parties confiées à des ASIP. Le concept de ROOM et son compilateur sont directement exploitables dans le cadre du projet REMIX pour produire le code des processeurs qui seront implanté dans le composant FPGA de chaque carte RMEM.

C - La répartition des données

L'environnement de programmation d'une carte RMEM doit être étendu pour prendre en compte la répartition de la mémoire d'indexation sur tout un cluster de PC. Sous l'hypothèse que la mémoire est uniformément répartie et que les traitements sont identiques sur tous les nœuds du système d'indexation, le paradigme de parallélisme par les données constitue une solution séduisante. Du point de vue algorithmique il s'agit de demander au programmeur de dissocier les traitements effectués sur un PC de ceux effectués sur sa carte RMEM. Du point de vue compilation il s'agit d'étendre le concept de ROOM en fournissant une extension data-parallèle à son langage de programmation. Pour cela on pourra s'inspirer des nombreux travaux [Briot][Guerraoui95] visant à étendre les langages orientés objet pour la programmation de systèmes parallèles.

B2.5 – Domaines d'applications

Cette dernière partie passe en revue les trois champs disciplinaires qui, aujourd'hui, sont confrontés à une recherche d'information à partir d'une masse de données en constante augmentation : la génomique, l'image et les données textuelles. Les trois domaines sont maîtrisés par les équipes Symbiose, TexMex et Equipage et exploitent des techniques d'indexation diverses qui, après avoir été repensées dans le cadre d'une mémoire d'indexation de très grande taille, devront être validées sur la plate-forme REMIX.

Recherche de motifs dans les génomes

Equipe Symbiose

Depuis quelques années, le séquençage massif entrepris par de nombreux laboratoires conduit à une explosion des données génomiques (le volume de données double tous les ans). Parmi celles-ci on compte, à l'heure actuelle, pas moins de 150 génomes entièrement séquencés, et plus de 350 en cours de séquençage. Possédant ces génomes, les biologistes cherchent maintenant à comprendre leur structure par le biais d'outils bio-informatique. La recherche de motifs fait parti des interrogations de base lorsqu'il s'agit d'appréhender un génome dans son ensemble. Pour ce faire, le génome doit être stocké dans un format facilitant l'accès à ce type d'information. L'arbre des suffixes est, à cet égard, particulièrement intéressant pour découvrir des motifs spécifiés sous forme d'expressions régulières ou de grammaires algébriques [Gras 1995].

Si l'arbre des suffixes facilite la localisation des motifs, son volume, par contre, équivaut à 12 fois la taille des données primaires. A titre indicatif, la taille du génome humain est de 3×10^9 nucléotides. Un génome complet peut cependant être découpé en segments qui se chevauchent et auxquels on attribue un arbre des suffixes. Une recherche peut alors traiter en parallèle ces différents arbres et fusionner les résultats. Nos premières investigations montrent qu'une fragmentation d'un chromosome humain, par exemple, en morceaux de taille relativement restreinte a même tendance à accélérer le processus de recherche sur une machine séquentielle. La parallélisation n'en sera que plus efficace !

Dans cette application, outre le fait de pouvoir disposer d'une indexation complète en mémoire, la structure reconfigurable peut notablement accélérer les temps de traitement. En effet, un motif complexe est souvent partitionné en requêtes élémentaires sur l'arbre des suffixes, requêtes qui produisent chacune une liste d'indices qui doit être confrontée avec les autres listes pour ne retenir que les ensembles d'indices correspondant au motif recherché. Ce post traitement est coûteux lorsque les listes sont longues. L'espace reconfigurable peut alors être avantageusement utilisé pour câbler des unités spécialisées traitant ces données à la volée, dès leur sortie mémoire.

Action Concertée Incitative

MASSES DE DONNEES

Descriptif complet du projet

Indexation des grandes banques de données de séquences

Equipe TexMeX

Il existe de nombreux algorithmes pour l'indexation des données dans les systèmes de recherche d'information et dans les bases de données. Parmi les techniques d'indexation, certaines ont été appliquées aux données génomiques pour la recherche intensive, rapide et efficace d'alignements locaux entre séquences. Sur de très grands volumes de données, l'avantage des approches basées sur l'indexation est de réduire significativement les temps d'accès aux données par rapport aux systèmes classiques basés sur une recherche séquentielle exhaustive comme le sont les logiciels BLAST et FASTA, par exemple.

Relativement peu de travaux ont été menés dans le domaine de l'indexation des séquences génomiques principalement du fait de la taille importante des index à construire et à la nécessité de leur appliquer des techniques de compression adaptées. Néanmoins, à partir des principaux schémas d'indexation proposés pour des données génomiques : RAMDB [Fondra, 2000], FLASH [Califano, 1993] basés sur une structure de type table de hachage ; CAFE [Williams, 2002] basé sur des techniques d'indexation selon le modèle vectoriel (fichiers inversés), M-tree [Ciaccia, 1997] et plus proches voisins [Chen, 1997], nos travaux de recherche visent à proposer des algorithmes d'indexation qui prennent en compte parallèlement les différents niveaux structurels d'une même séquence (séquence ADN, séquence protéique, structures secondaire et tertiaire) [Dulucq, 2001], afin de réaliser une recherche simultanée sur ces différents niveaux en optimisant l'utilisation des tables d'index selon la précision recherchée.

La taille d'un index où se superposent plusieurs niveaux d'information représente forcément un volume d'information conséquent, mais non rédhibitoire dans le cadre d'une plate-forme matérielle comme REMIX. Aussi, dans le cadre de cette ACI nous visons la mise en œuvre d'algorithmes d'indexation parallèles par heuristiques selon plusieurs niveaux structurels. Les diverses implémentations donneront lieu à une étude de performance comparant les techniques existantes à celles que nous proposons.

Recherche d'images par le contenu

Equipe TexMex

La recherche d'images par le contenu au sein de grandes bases de données est un processus notoirement coûteux. Ces coûts ont grossièrement deux origines. Il y a d'abord les coûts liés au traitement des images, fixes ou vidéo, afin d'analyser leur signal et en déduire une description à même de refléter leur contenu. Un exemple souvent mis en avant est le calcul d'histogrammes de couleur sur les images (mais bien d'autres schémas de description existent, souvent beaucoup plus complexes). De façon générale, les descriptions sont en général des vecteurs numériques définissant des points dans des espaces de grandes dimensions. La similarité visuelle de deux images est supposée proportionnelle à la similarité de leurs descriptions. Extraire de chaque image une ou plusieurs descriptions est un processus complexe, cher, et dont le coût devient tangible (voire handicapant) lorsque l'on doit décrire des milliers d'heures de vidéo ou des milliers de photos.

La seconde source de coûts est, quant à elle, liée au processus de recherche d'images similaires. Après avoir calculé une description de l'image (ou de la séquence) requête, il faut rechercher dans la base s'il existe des images ou séquences ressemblantes, au sens visuel. Là, la similarité de contenus se traduit fréquemment par une recherche de plus proches voisins **via** des calculs de distance souvent Euclidiens dans l'espace de description. L'origine du coût très important de la recherche est d'une part lié à la masse de données qu'il faut examiner et aux calculs de distance, très nombreux.

De très nombreuses stratégies d'indexation ont été proposées dans la littérature. Les algorithmes d'indexation multimédia se classent en deux grandes catégories : ceux qui regroupent les données en fonction de leur proximité (*data-partitionning index methods*) [Guttman 1984, White 1996], et ceux qui découpent *a priori* l'espace multidimensionnel et qui ensuite stockent les données selon ce découpage (*space-partitionning index methods*) [Nievergelt 1984, Henrich 1998].

Ces index sont très clairement orientés disque, et l'examen détaillé de leur fonctionnement montre qu'ils s'appuient sur des notions comme le facteur de remplissage des pages, l'éclatement et l'équilibrage de pages disque saturées, etc. Ils sont aussi conçus pour être très compacts. Il n'est donc pas possible d'en reprendre le principe ni de les plonger tels quels dans un contexte orienté mémoire vive. Il faut au contraire concevoir de nouvelles stratégies utilisant pleinement les accès directs, les pointeurs. Les recherches que nous entreprendrons dans le contexte de cette ACI ont ainsi pour objectif de concevoir une nouvelle structure de données, orientée mémoire, destinée à

Action Concertée Incitative

MASSES DE DONNEES

Descriptif complet du projet

indexer une grande quantité de descripteurs d'images, et supportant notamment des recherches de plus proches voisins.

Il faut aussi tenter de s'appuyer sur des outils issus de pré calculs sur les données, outils onéreux dans un contexte disque mais puissants dans un contexte mémoire. Un exemple typique est l'utilisation du principe de l'inégalité triangulaire afin d'éviter, à l'exécution, le calcul de nombreuses distances entre descripteurs. Ce principe demande de pré calculer, hors ligne et avant toute interrogation, toutes les distances entre toutes les descriptions de toutes les images. Cette table de distances, très volumineuse, permet des gains de temps très rapide car il est possible de savoir, grâce à une unique comparaison (simplement comparer deux valeurs entières ou réelles) qu'il est inutile de calculer la distance (processus complexe) entre un vecteur requête et un descripteur de la base. Dans un contexte habituel, petites mémoires et disques, cette table pose de nombreuses difficultés. Ici, dans le contexte de cette ACI, les contraintes de tailles n'existent plus, et cette table permet des accélérations de traitements spectaculaires. D'autres principes peuvent être appliqués de manière analogue. Aussi, notre but est d'inventer, grâce à cette ACI, de nouvelles structures annexes, bâties à partir de pré-calculs éventuellement complexes sur les données, occupant certainement beaucoup de mémoire, mais accélérant fortement les traitements.

Données textuelles – XML

Equipe APRIME

Il existe bon nombre d'approches qui se sont historiquement intéressées à la question de l'interrogation de banques de données semi-structurées. Certaines relèvent du modèle relationnel [Codd, 1983] : par exemple [Stonebraker et al.], d'autres mettent en exergue les modèles classiques de recherche et d'information (modèle booléen, vectoriel, etc.) [Salton et McGill 1983, Frakes et Baeza-Yates 1992], d'autres proposent des modèles d'hyperdocument [Conklin 1987], ou encore des modèles orientés objets [Kim et Lochovski 1989, Cattell 1991]. Aucune de ces approches initiales ne permettent de gérer de manière coordonnée les éléments de structures et les données non structurées. Plus récemment, de nouvelles approches ont été développées pour traiter cette question de la recherche d'information en contexte structurel, ce type de recherche pouvant tenir compte plus ou moins précisément de la localisation dans la structure du document de l'occurrence de l'information non structurée recherchée. Les travaux de Baeza-Yates [Baeza-yates et al., 1997], de Sung Hyon Myaeng [Sung Hyon Myaeng et al., 1998] ou encore de Bremer [Bremer et Gertz, 2000] et Wolf [Wolff et al., 1999] visent tout particulièrement le développement d'algorithmes de recherche d'information semi-structurée efficaces basés sur une exploitation des contextes structurels d'occurrence des éléments informationnels et sur des heuristiques de classement des réponses. Ces approches introduisent des algèbres sur des ensembles d'éléments structurels hiérarchiques qui correspondent à la structure arborescente de la base documentaire traitée. La nécessité d'une connaissance parfaite de la structure arborescente de la base documentaire pour exploiter la richesse des fonctions de recherche constitue la principale limite rencontrée par ces dernières approches.

Dans cette dynamique, nous avons proposé et développé [Ménier & Marteau 2001,2002] [Marteau & Ménier 2003] des principes d'indexation et de recherche d'information dans des bases documentaires XML. La structure des documents XML varie en général d'un document à l'autre. Elle est implicitement véhiculée par le document lui-même et pas systématiquement déterminée par une organisation structurelle globale stable (DTD). Pour cette raison, les documents XML sont considérés plutôt comme des documents semi-structurés, par opposition aux documents structurés pour lesquels aucune irrégularité structurelle n'existe [Abiteboul, 1997].

Les principes que nous développons mettent en œuvre des mécanismes de recherche qui exploitent simultanément les éléments des structures documentaires indexées et les contenus documentaires non structurés. L'innovation par rapport aux dernières approches référencées repose sur l'introduction de mécanismes d'alignement approximatif des éléments structurels indexés. Ces mécanismes ont pour objectifs d'assouplir les critères de recherche portant sur le contexte structuré pour permettre à l'utilisateur d'interroger une bases d'information en faisant appel à une connaissance nulle, incomplète, imprécise ou totale de la structure de la base documentaire exploitée.

Dans le contexte de l'exploitation de grandes masses de données, la question de l'efficacité des algorithmes d'indexation et de recherche se pose en termes de temps de calcul, temps d'accès disque (I/O) mais aussi de gestion de l'espace mémoire.

Une des solutions envisageables pour tenter de gérer ces difficultés est l'exploitation d'architectures matérielles spécialisées. Les travaux préliminaires sur le thème de l'optimisation et de la parallélisation des bases de données XML sont très récents et assez rares : on peut citer par exemple [Yasuo & al., 2000], [Lü & al., 2002].

Action Concertée Incitative

MASSES DE DONNEES

Descriptif complet du projet

Nous nous proposons dans le cadre de l'ACI d'étudier l'adaptation de nos mécanismes d'indexation à une architecture matérielle spécialisée. La recherche d'opérateurs primitifs intervenants dans le processus d'indexation ou de recherche d'information permet de cerner les besoins calculatoires d'une telle architecture. Dans le cas de la gestion d'une grande quantité d'information, le parallélisme de traitement ne se limite pas nécessairement à la seule accélération matérielle des opérations élémentaires de recherche : il est nécessaire de mettre en évidence les indépendances intrinsèques de la base pour exploiter en parallèle les informations jugées pertinentes (filtrage, élagage etc..)

L'exploitation d'architectures parallèles n'a pas donc pas seulement comme objectif l'accélération des accès aux informations : la gestion de masses importantes de données nécessite également une organisation répartie des informations. Nous nous proposons d'étudier ces questions pour proposer des solutions adaptées à la gestion de masse importante de données textuelles semi-structurées.

B2.6 – Bibliographie

Abiteboul S., *Querying semi-structured data*, In F. Afrati and Ph. Kolaitis, editors, Proc. of the 6th Int. Conf. on Database Theory (ICDT), Lecture Notes in Computer Science 1186, pages 1-18. Springer, January 1997.

Baeza-Yates R. and Navarro G., *Proximal Nodes: A Model to query document databases by contents and structure*, ACM Transactions on Information Systems 15 (4), October 1997, pp. 401-435., 1997.

Brin S., Page L., *The anatomy of a large-scale hypertextual web search engine*, 7th International WWW Conference, 1998.

Califano, I. Rigoutsos, *FLASH : Fast look-up algorithm for string homology*, Proc. Int. Conf. Intelligent Systems for molecular biology, p. 56-64, 1993.

Chen W., Aberer K., *Efficient Querying on Genomic Databases by using Metric Space Indexing techniques*, Proc. of 8th Int. Conf. And Workshop on Database en Expert Systems Application DEXA'97, 1997.

Charot F., Messé V., *La compilation recible au service de la définition interactive d'ASIP*, Technique et Science Informatiques, 20, 2, février 2001.

Ciaccia P., Patella M., Zezula P., *M-tree: an efficient access method for similarity search in metric spaces*, Proc. of VLDB'97, 1997.

Codd E., *A relational model of data for large shared data banks*, CACM 26, 1 (Jan.). In Milestones of Research, 1983.

Conklin J., *Hypertext : an introduction and survey*, IEEE Computer 10, 9 (Sept.), 17-41, 1987.

Dulucq S., Tichit L., *A propos de la comparaison de structures secondaires d'ARN*, Actes des journées JOBIM, p. 125-133, 2001.

Fondrat C, Dessen P., *A rapid Access Motif Database (RAMdb) with a search algorithm for the retrieval patterns in nucleic acids or protein databanks*, Computer Applications Biosciences, vol.11, no.3, p.273-279, 2000

Frakes W. and Baeza-Yates R. Eds., *Information Retrieval : Data Structures and Algorithms*, Printice-Hall, 1992.

Garcia-Molina H., *Main Memory Database Systems: An overview*, IEEE Transactions on Knowledge and Data Engineering, vol. 4, no. 6, 1992.

Gras R, Nicolas J., *FOREST : browser of repeats in huge sequences*, The third International Conference on Computational Biology Intelligent Systems for Molecular Biology 95, Cambridge, 1995

Guttman A., *R-Trees: A Dynamic Index Structure for Spatial Searching*, Proceedings of the ACM SIGMOD International Conference on Management of Data, Boston, Massachusetts, 1984.

Action Concertée Incitative

MASSES DE DONNEES

Descriptif complet du projet

Henrich A., *The LSD^h-Tree: An Access Structure for Feature Vectors*, Proceedings of the Fourteenth ICDE International Conference on Data Engineering, February 23--27, Orlando, Florida, USA, 1998.

Kim W. and Lochovski F., Eds. *Object-Oriented Concepts, Databases and Applications*, Addison-Wesley,

Knuth D.E., Morris J.H., and Pratt V.R., *Fast pattern matching in strings*. SIAM J.Comput, 6(2):323--350, 1977.

Lavenier D. et al., GénoGRID : an experimental grid for genomic applications, HealthGrid 2003, Lyon, January 16-17, 2003

Lavenier D., Derrien S., Guyetant S., Rubini S., *A reconfigurable parallel disk system for filtering genomic banks*, ERSA'03, Las Vegas, USA, 2003.

Lu Z., McKinley K., Cahoon B., *The Hardware/Software Balancing Act for Information Retrieval on Symetric Multiprocessors*, EuroPar 1998.

Lü K., Zhu Y., Sun W., *Parallel Processing XML Documents*, International Database Engineering and Applications Symposium (IDEAS'02), July 17 - 19, Edmonton, Canada, 2002

Marteau P.F. et Ménier G., *Alignement approximatif d'arbres pour la recherche d'information en contexte dans les données XML hétérogènes - Fusion d'information structurées et textuelles*, soumis à la revue Techniques et Sciences Informatiques, Hermes, 2003.

Ménier G., Marteau P.F., *Information retrieval in heterogeneous XML knowledge bases*, The 9th International Conference on Information Processing and Magement of Uncertainty in Knowledge-Based Systems , IEEE, 1-5 July, Annecy, France, 2002.

Nievergelt J., Hinterberger H., Sevcik K., *The Grid File: An Adaptable, Symmetric Multikey File Structure*, ACM Transactions On Database Systems, vol 9, n 1, 1984.

Raimbault F., Lavenier D., *ROOM, Des machines reconfigurables orientées objet*, 8^{ième} Symposium en Architectures Nouvelles de Machines (Sympa), Tunis, 2002.

Rubini S., Lavenier D., *Les Architectures Reconfigurables*, Calculateurs Parallèles, vol. 9, no. 1, 1997

Salton G. and Buckley C., *Term-weighting approaches in automatic text retrieval*, Information Processing and Management, 24, 513-523, 1988.

Stonebraker M., Stettner H., Lynn N., Kalash J. and Guttman A., *Document processing in a relational database system*, ACM TOIS 1,2 (Apr.), 143-158, 1983.

Sung Hyon Myaeng and Dong-Hyun Jang, *A flexible model for retrieval of sgml documents*, SIGIR98, 1998.

Vuillemin J., Bertin P., Roncin D., Shand M. , Touati H. et Boucard P., *Programmable Active Memories: the Coming of Age*, IEEE Trans. on VLSI, vol. 4, no. 1, 1996

Yasuo Y., Nobuyuki I., Isao N., *High-performance XML Storage/Retrieval System*, FUJITSU Sci. Tech. J.,36,2, pp.185-192 (Dec.), 2000

Wagner M., *Google Bets The Farm On Linux*, <http://www.internetwk.com/lead/lead060100.htm>, June 2000.

White D., Jain R., *Similarity Indexing with the SS-tree*, Proceedings of the Twelfth ICDE International Conference on Data Engineering, February 26--March 1, New Orleans, Louisiana, 1996

Williams H.E., Zobel J., *Indexing and retrieval for genomic databases*, IEEE transactions on knowledge and data engineering, vol.14, n° 1, January/February 2002.

Wolff J. E., Flörke H., Cremers A., *XPRES: a Ranking Approach to Retrieval on Structured Documents*, technical report , University of Bonn, 1999

Action Concertée Incitative
MASSES DE DONNEES
Descriptif complet du projet

B3 – Résultats attendus :

Résultats finaux :

- Une plate-forme fonctionnelle REMIX d'une capacité mémoire de 256 Giga octets intégrée dans un cluster de 16 PC possédant chacun une carte RMEM de 16 Giga octets
- Un environnement de programmation adéquate intégrant les aspects parallélisation et reconfiguration.
- La définition de nouvelles stratégies d'indexation dans l'optique d'un tel support matériel
- Le portage d'au moins trois applications émanant des domaines de la génomique, de l'image et des bases de données textuelles XML
- Une étude comparative de l'apport de REMIX par rapport aux techniques d'indexation usuelles

A notre connaissance, il n'existe aucun projet similaire combinant à la fois les aspects architecture spécialisées et techniques d'indexation. C'est un projet multi-thématique original et exploratoire qui vise à lever le verrou de l'indexation de masse de données en agissant simultanément et de manière complémentaire sur deux leviers : l'architecture et l'algorithmique de l'indexation.

Echéancier : projet sur 3 ans

- **0 – 6 mois**
 - **étude**
 - cahier des charges d'une carte RMEM
 - définition de l'environnement de programmation
 - spécification des applications
- **6 – 12 mois**
 - **étude et développement**
 - fabrication 2 cartes RMEM
 - développement environnement de programmation
 - développement de nouvelles techniques d'indexation dans le cadre de REMIX
- **12 – 18 mois**
 - **test**
 - carte RMEM – modification éventuelle du cahier des charges
 - mise au point de l'environnement de programmation
 - **réalisation d'un démonstrateur**
 - 2 cartes RMEM
 - premier portage d'une ou 2 applications d'indexation
 - validation du concept
 - **remise du rapport intermédiaire**
- **18 – 24**
 - **passage à l'échelle**
 - achat cluster
 - fabrication d'une série de 16 cartes RMEM
 - assemblage de la plate-forme REMIX
- **24 – 36**
 - **test grandeur nature**
 - portage des applications : génomique – image – document semi-structuré
 - étude comparative
 - **rapport de fin d'étude**