

Présentation des projets financés au titre de l'édition 2008 du Programme DEFIS

ACRONYME et titre du projet	Page
ARTIS- Inversion articulatoire de la parole audiovisuelle pour la parole augmentée	2-3
CLASSEL- Classification croisée et sélection de modèle	4-5
COCQ- Codes Correcteurs quantiques	6-7
CODEX- Performance, évolution, et composition pour XML : modèles, algorithmes et systèmes	8-9
DECERT- Déduction certifiée	10-11
ECHANGE- Echantillonnage acoustique nouvelle génération	12-13
IDEA- Approches dynamiques des codes identifiants	14-15
NAFIT- Nouvelles formes algorithmiques de la théorie de l'information	16
NATIMAGES- Adaptivité pour la représentation des images naturelles et des textures	17-18
PARTOUT- Parallélisme partout	19
PHYL-ARIANE- Phylogénomique : Algorithmes et représentations intégrés pour l'analyse de l'évolution du vivant	20-21
QRAC- Algorithmes et complexités quantiques et probabilistes	22
SEQUOIA- Analyse syntaxique probabiliste à large couverture du français	23-24

Programme DEFIS

Edition 2008

Titre du projet

ARTIS- Inversion articulatoire de la parole audiovisuelle pour la parole augmentée

Résumé

L'inversion acoustique articulatoire, c'est-à-dire la récupération automatique des gestes articulatoires de la parole, représenterait une avancée stratégique en traitement automatique de la parole. En effet, cela ouvrirait la voie à de nouvelles applications : synthèse de la parole et têtes parlantes plus naturelles, pilotage automatique de têtes parlantes réalistes, retour d'efforts articulatoire pour l'apprentissage de langues étrangères, et plus généralement le traitement automatique de la parole. Les industries multimédia, des télécommunications, des interfaces homme machine, de l'assistance aux citoyens (personnes âgées, apprenants, handicapés...) sont donc concernées. Bien qu'il soit bien établi que les sujets humains exploitent l'origine articulatoire de la parole pour la perception comme pour la production, aucune solution opérationnelle n'a été apportée à l'inversion acoustique articulatoire. Nous sommes pourtant convaincus que les progrès techniques réalisés dans les domaines de l'imagerie médicale appliquée au conduit vocal, du suivi de capteurs électromagnétiques et de la vision par ordinateur appliquée au visage sont suffisants pour fournir des données articulatoires exploitables tant pour la conception que l'évaluation de méthodes d'inversion. Nous sommes aussi convaincus que les outils de simulation acoustique et articulatoire d'une part, de modélisation stochastique d'autre part, ont atteint un niveau de maturité suffisant. Il y a donc une double opportunité scientifique et technologique pour relever ce défi avec de bonnes chances de succès. L'objectif principal du projet ARTIS est de développer des outils d'inversion acoustique articulatoire, et de mettre en œuvre un système de retour d'efforts articulatoires, qui en inversant le signal de parole seul ou complété par des images du visage, génère de la parole augmentée : affichage des articulateurs habituellement cachés (langue, vélum) à l'aide d'une tête parlante virtuelle. Le développement des algorithmes d'inversion reposera sur des modèles physiques ou statistiques de la production de la parole.

Un point clé est le développement de modèles adaptables au locuteur. En effet, la plupart des travaux antérieurs fait appel à des modèles construits pour un seul locuteur ce qui représente une limitation forte.

La première étape consistera donc à acquérir des données

articulatoires statiques et dynamiques. Ensuite, des techniques de traitement d'images médicales et des algorithmes de recalage élastique seront élaborées afin d'obtenir des modèles articulatoires adaptables au locuteur à partir de données statiques (images IRM du conduit vocal). Les modèles dynamiques statistiques (correspondant aux données des capteurs électromagnétiques ou aux images ultrasonores) compléteront les modèles statiques. Deux systèmes d'inversion complémentaires seront développés: le premier à l'aide d'une approche d'analyse par synthèse, le second à l'aide de modèles articulatoires statistiques dynamiques. Les deux partageront les modèles articulatoires statiques et utiliseront des données spectrales standard en entrée. La troisième étape consistera à développer un prototype fournissant un retour d'efforts articulatoires en temps réel à partir de la parole et des images du locuteur. The retour sera présenté à l'aide de techniques de réalité augmentée.

Une évaluation sera conduite à chacune des étapes du projet. Le consortium ARTIS regroupe un ensemble de compétences remarquable en associant des équipes dotées d'un fonds théorique et d'une expérience pratiques dans les domaines de :

- la production de la parole, la modélisation stochastique, les modèles articulatoires, l'inversion acoustique-articulatoire,
- la vision par ordinateur, le traitement d'images et l'imagerie médicale. La force de ce projet est la combinaison d'expertises qui ont été acquises sur une longue période de temps. Le consortium a un niveau d'excellence internationale pour atteindre les objectifs de ce projet.

Partenaires

CNRS DR6 (partenaire coordinateur)
Grenoble INP
Université Paul Sabatier
CNRS DR1

Coordinateur

M. Yves Laprie– LORIA
Yves.Laprie@loria.fr

Aide de l'ANR

779 659 euros

Début et durée

Janvier 2009 - 42 mois

Référence

ANR-08-EMER-001

Programme DEFIS

Edition 2008

Titre du projet

CLASSEL- Classification croisée et sélection de modèle

Résumé

CLASSEL est un projet de recherche académique qui vise à développer des méthodes pour transformer des données en connaissances. Les données en question sont sous la forme d'une matrice individus-variables. Nous cherchons à comprendre comment construire de manière automatique, à partir de données, des groupes ou des hiérarchies d'individus définies simultanément à partir de groupes ou de hiérarchies de variables. Ces hiérarchies associant individus et variables sont ensuite exploitées pour compléter les données ou pour servir de base à la définition de terminologies ou de 'contextes'. C'est le problème de classification croisée. Lorsque qu'on attaque ce type de problème, une attention toute particulière doit être portée au problème fondamental du choix du nombre de groupe : c'est la question de la sélection de modèle. Nous abordons ces questions formellement dans un cadre statistique nouveau et particulièrement bien adapté.

Nous avons structuré notre projet autour de quatre tâches :

- l'étude de la classification croisée,
- l'étude du problème spécifique de la sélection de modèle,
- les questions algorithmiques liées notamment à notre volonté d'attaquer des grandes masses de données,
- les applications.

Nous proposons d'attaquer le problème de classification croisée formellement en utilisant une modélisation probabiliste. Notre projet vise à adapter ce type de modèle aux problèmes spécifiques de la classification croisée pour les problèmes de grande taille et d'adapter les algorithmes d'estimation du type EM. Le but principal de l'aspect Sélection de modèles de notre projet est le développement de nouvelles méthodes de mise en œuvre de sélection de modèles appliquées à l'apprentissage statistique. Nous comptons intégrer les points de vue issus des domaines de la fouille de données et d'apprentissage à la Statistique paramétrique classique en vue d'explorer de très grands ensembles de données. Nous visons à déterminer, au travers des idées de la sélection de modèle, quels prédicateurs ont le plus d'influence sur les résultats et à évaluer le degré d'incertitude de nos prévisions. La partie algorithmique vise à adapter les solutions proposées aux contraintes liées au passage à l'échelle. Elle a aussi pour objectif de tester les différentes solutions envisagées et de fournir à la communauté des

composants logiciels réutilisables. Les applications sont vues à la fois comme moteur et démonstrateurs de nos recherches. Ils sont en effet moteurs à travers les problèmes spécifiques posés. Ces applications concernent la segmentation marketing en collaboration avec l'université de Vienne (Autriche), les systèmes de recommandation à travers le challenge Netflix et la fouille de texte. Notre stratégie scientifique consiste à attaquer de front les questions fondamentales de la modélisation en apprentissage et de la sélection de modèle pour trouver des solutions en rupture avec l'existant. Pour atteindre cet objectif nous proposons de mettre en œuvre une approche décloisonnée mobilisant des chercheurs de différentes communautés STIC (statistiques, analyse de données, apprentissage et informatique) sur des applications concrètes liées à de grandes masses de données.

Le groupe projet associe logiquement des statisticiens (D. Fourdrinier du LITIS, des spécialistes de l'analyse des données comme G. Govaert d'Heudiasyc et M. Nadif du CRIP 5, des chercheurs de la communauté apprentissage S. Canu, A. Rakotomamonjy, G. Gasso (LITIS) et Y. Grandvalet (Heudiasyc) et des informaticiens comme F.X. Jollois du CRIP 5.

Partenaires

Université de Rouen (partenaire coordinateur)
CNRS DR18 (Heudiasyc)
Université Paris Descartes

Coordinateur

M. Dominique Fourdrinier – Université de Rouen / Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes
dominique.fourdrinier@univ-rouen.fr

Aide de l'ANR

318 146 euros

Début et durée

Janvier 2009 - 36 mois

Référence

ANR-08-EMER-002

Programme DEFIS

Edition 2008

Titre du projet

COCQ- Codes Correcteurs quantiques

Résumé

Ceci est une proposition de recherche fondamentale dans le domaine des codes correcteurs quantiques. Les applications possibles concernent à la fois la fiabilisation des mémoires dans le domaine des nanotechnologies, des calculs quantiques effectués dans un ordinateur quantique et des communications quantiques mises en œuvre lors de protocoles quantiques d'échange de clés.

L'objectif du projet est de généraliser au cadre quantique les constructions des codes classiques les plus efficaces. Ceci concernerait notamment les codes LDPC et les turbo-codes. Ils fourniraient une alternative très intéressante à la concaténation de petits codes quantiques utilisée dans les architectures actuelles de calcul quantique tolérants aux fautes. Ils offriraient notamment un gain comparable au gain observé dans le domaine des codes correcteurs classiques lorsque ces derniers ont commencé à remplacer les codes concaténés dans les schémas de codage. Des généralisations de codes LDPC au cas quantique ont déjà été proposées dans la littérature scientifique. Cependant, beaucoup de travail reste à faire en la matière. Il reste notamment à résoudre le problème fondamental de proposer des familles de tels codes qui à rendement donné aient une distance minimale non bornée. Ceci est essentiel pour atteindre des probabilités d'erreur suffisamment faibles. Nous nous proposons de réunir les compétences diverses de trois équipes de recherche pour résoudre ce problème. Les turbo-codes eux ont été généralisés au cadre quantique par un des participants de cette proposition de projet. Là aussi, beaucoup de travail reste à effectuer afin d'obtenir des familles de tels codes qui soient aussi performantes que leurs pendants classiques. Le problème à résoudre ici est de même nature que pour les codes LDPC, il s'agit ici de proposer des familles de tels codes avec une distance minimale non bornée. Ceci est un des objectifs à atteindre dans cette proposition de projet.

Nous nous proposons de construire de tels codes en même temps que nous élaborons des algorithmes de décodage itératif qui leur soient adaptés. Les algorithmes de décodage itératif standards sont inappropriés ici, car dans le cas des codes LDPC quantiques par exemple, il y a nécessairement des cycles très courts dans leur graphe de Tanner, et ceux-ci ont besoin de subir un traitement spécial lors du décodage. Nous nous

proposons d'explorer deux techniques issues de la physique théorique pour traiter ce problème.

Enfin, nous combinerons aussi la recherche de bons codes LDPC et de turbo-codes avec l'estimation de la capacité des canaux quantiques sur lesquels ils sont censés opérer. Rappelons que cette capacité n'est pas connue pour la plupart des canaux intéressants, et nous nous proposons ici d'améliorer les bornes inférieures et supérieures existantes. Les méthodes que nous envisageons pour améliorer les bornes inférieures sont constructives et permettront de donner des familles de codes quantiques pouvant supporter un bruit plus élevé dans le canal.

Partenaires

INRIA Paris-Rocquencourt (partenaire coordinateur)
Université Bordeaux 1
Télécom Paris Tech
ENSEA

Coordinateur

M. Jean-Pierre Tillich – Centre de recherche INRIA Paris -
Rocquencourt / EPI SECRET
jean-pierre.tillich@inria.fr

Aide de l'ANR

409 770 euros

Début et durée

Janvier 2009 - 36 mois

Référence

ANR-08-EMER-003

Programme DEFIS

Edition 2008

Titre du projet

CODEX- Performance, évolution, et composition pour XML : modèles, algorithmes et systèmes

Résumé

Le développement de technologies centrées sur XML est sans précédent. De nombreux standards ont été ou sont actuellement définis par des organisations internationales telles que le W3C ou l'ISO. Cette intense activité de standardisation couvre des domaines aussi variés que l'interrogation, la représentation ou le codage d'informations, l'interaction, l'orchestration et la coordination de processus communiquant dans un environnement hautement distribué ou encore la validation ou la mise à jour de données.

Un deuxième aspect caractéristique du boom des technologies XML est la mise à disposition de pléthore d'outils gratuits ou de logiciels libres. Une telle abondance accroît drastiquement le nombre de développeurs potentiels comme cela n'a jamais été le cas de par le passé. Cette situation permet de multiples expérimentations et ouvre la voie à de réelles innovations. Enfin un dernier aspect, totalement nouveau, et sûrement moteur de cette immense activité réside dans l'étroite collaboration entre les mondes académiques et industriels tant au sein d'organisations de standardisation que de colloques internationaux. Cette synergie constitue un facteur de grande convergence entre les différents acteurs des technologies internet. Le programme de recherche que nous proposons a pour ambition de repousser les frontières technologiques en termes de gestion de données et de programmes au format XML selon trois directions corrélées. En premier lieu nous proposons de concevoir, étudier et implémenter des langages, algorithmes et prototypes pour une gestion de données XML à la fois expressive et hautement efficace. Un accent particulier sera mis sur l'obtention de systèmes XML massivement distribués. En second lieu, nous étudierons des modèles afin de décrire, contrôler et réagir au comportement dynamique inhérent à l'évolution des données et des schémas au cours du temps. Enfin nous définirons des théories, des modèles et des prototypes afin de permettre une composition et une interaction de services plus riche en termes d'expressivité tout en garantissant formellement des propriétés de correction pour de telles interactions.

Partenaires

INRIA Saclay-Ile de France (partenaire coordinateur)

INRIA Lille-Nord Europe
Université Paris-Sud LRI
CNRS DR2
INRIA Grenoble-Rhône Alpes
Université de Tours
Innovimax

Coordinateur

Mme. Ioana Manolescu-Goujot– INRIA Saclay - Île-de-France
ioana.manolescu@inria.fr

Aide de l'ANR

810 292 euros

Début et durée

Mars 2009 - 36 mois

Référence

ANR-08-EMER-004

Programme DEFIS

Edition 2008

Titre du projet

DECERT- Dédution certifiée

Résumé

Les niveaux les plus élevés des standards de développement logiciel, comme le niveau EAL7 des Critères Communs pour la certification de systèmes et des composants critiques, exigent l'utilisation de techniques formelles fondées sur des sémantiques mathématiques strictes. Ces techniques sont mécanisées par le biais d'outils déductifs qui permettent de démontrer des propriétés portant sur les logiciels. Notre vision à long terme est de contribuer à rendre l'utilisation de tels outils économiquement viable et, autant que possible, de rendre l'utilisation de méthodes formelles invisible aux développeurs et aux utilisateurs de systèmes. Afin que cette vision puisse se réaliser sur une échelle plus étendue, des composants automatisés et performants doivent interagir et forger un environnement de confiance pour le développement et/ou le déploiement de systèmes. L'objectif du projet DECERT est de concevoir une architecture pour des procédures de décision coopérant entre elles, avec une attention particulière portée à des fragments de l'arithmétique (arithmétiques bornée et non-bornée, sur les entiers et les réels, …) et à leur combinaison avec d'autres théories concernant les structures de données (listes, tableaux, ensembles, …). Pour garantir une confiance globale dans cette combinaison d'outils, les procédures de décision seront soit prouvées correctes à l'intérieur d'un assistant de preuve, soit produiront des certificats de résultat qui permettront à des outils extérieurs de vérifier la validité de leurs réponses. Nous définirons un format normalisé pour des témoins de preuves adapté aux procédures de décisions et leurs combinaisons. Ces témoins de preuve devront être produits sans trop de surcoût par les procédures de décisions et certifiés de manière efficace par les noyaux des assistants de preuve.

Nous proposons d'utiliser ces procédures dans plusieurs scénarii d'application provenant à la fois du milieu académique et industriel et qui demandent des garanties élevées de confiance. En développement prouvé de systèmes, des modèles de haut niveau sont généralement écrits dans des langages expressifs de modélisation telle la logique d'ordre supérieure, B ou TLA+ et dont les problèmes de vérification ne peuvent être entièrement automatisés. Dans ce contexte, des procédures de décision automatiques doivent interagir avec des assistants de preuve comme Coq ou Isabelle, sans que les assurances fortes de

correction associées à l'utilisation de ceux-ci soient compromises. Le deuxième scénario d'application que nous étudierons concerne l'utilisation de code téléchargé sur des plates-formes non dignes de confiance à travers le paradigme du code porteur de preuve (PCC : Proof Carrying Code). Afin de s'assurer que le code soit conforme à une politique de sûreté précise, il est assorti d'une preuve qui peut être générée par un outil de déduction automatique et qui est certifiée correcte sur le site hôte par un vérificateur de confiance. Dans les deux scénarii, la confiance repose sur un environnement de confiance fondé sur des certificats sous la forme de traces de preuve. En résumé, le projet se concentre sur les trois problèmes suivants :

1. Développer et implémenter des nouvelles et efficaces procédures de décision coopératives, en particulier pour des fragments de l'arithmétique.
2. Développer et standardiser une interface de sortie basée sur des certificats et des objets de preuve.
3. Intégrer les points 1 et 2 dans des assistants de preuve sceptiques, dans une architecture pour le code porteur de preuve et dans des outils de vérification tels que l'outil Rodin pour B et l'outil Frama-C du CEA pour la vérification de programmes C.

Partenaires

INRIA Rennes-Bretagne Atlantique (partenaire coordinateur)
INRIA Sophia-Antipolis-Méditerranée
INRIA Nancy-Grand Est
Université Paris Sud LRI
SYSTEREL
CEA

Coordinateur

M. Thomas Jensen – INRIA Rennes - Bretagne Atlantique
jensen@irisa.fr

Aide de l'ANR

787 694 euros

Début et durée

Janvier 2009 - 36 mois

Référence

ANR-08-EMER-005

Programme DEFIS

Edition 2008

Titre du projet

ECHANGE- Echantillonnage acoustique nouvelle génération

Résumé

Les signaux acoustiques sont porteurs d'une information très riche sur les phénomènes physiques complexes sous-jacents. Leur mesure et leur traitement est au cœur de nombreux domaines d'application, depuis les communications vocales et la production audiovisuelle (signaux audio phoniques) jusqu'au contrôle non destructif, la surveillance biomédicale, la veille sous-marine (ultrasons) ou l'exploration pétrolière (infrasons). Les capteurs acoustiques traditionnels (microphones, capteurs piézoélectriques et antennes de ces capteurs) réduisent le champ acoustique, objet spatio-temporel, à un ensemble de quelques séries temporelles, au risque de perdre une quantité significative d'information de nature spatiale. Le 'Compressed Sensing' propose une approche nouvelle pour acquérir des données complexes via un nombre limité de mesures, en exploitant des principes d'incertitude entre le domaine de mesure et un domaine où les données mesurées admettent une représentation parcimonieuse. L'objectif de ce projet est de mettre en place un cadre méthodologique et théorique, fondé sur les principes du compressed sensing, pour mesurer et traiter des champs acoustiques complexes via un nombre limité de capteurs acoustiques.

Le principal défi soulevé par ce projet est la nécessité d'aborder conjointement la définition l'optimisation et la validation : de fondations mathématiques solides pour la modélisation parcimonieuse de champs acoustiques, de dispositifs non-standard de captation acoustique tels que des antennes à disposition aléatoire ou, à plus long terme, des mesures optiques du champ le long de ligne ou des mesures sur des surfaces, d'algorithmes de décomposition parcimonieuse assez efficaces pour aborder des données d'un tel volume. Ce projet de recherche fondamentale rassemble les compétences complémentaires en acoustique, traitement du signal et mathématiques appliquées requises pour aborder avec succès ce défi.

Partenaires

INRIA Rennes-Bretagne Atlantique (partenaire coordinateur)
UPMC Paris 6
UPMC Paris 6 (IJLRA)

Coordinateur M. Rémi Gribonval– INRIA Rennes - Bretagne Atlantique

Aide de l'ANR 629 733 euros

Début et durée Janvier 2009 - 36 mois

Référence ANR-08-EMER-006

Programme DEFIS

Edition 2008

Titre du projet

IDEA- Approches dynamiques des codes identifiants

Résumé

Les problèmes de reconstruction sont centraux en combinatoire. Nous pouvons par exemple mentionner la question de déterminer, parmi un ensemble de vecteurs binaires, un sous-ensemble minimal dont la somme (ou exclusif) est égale à un vecteur donné. Dans le cadre de la théorie des codes correcteurs d'erreurs, cette question est équivalente au problème de décodage. Considérons maintenant le problème analogue de reconstruction d'un ensemble de vecteurs dont la somme (ou inclusif) est égale à un vecteur donné. En termes d'ensembles, il s'agit de reconstruire une collection de sous-ensembles d'un ensemble donné à partir de leur union. Ce problème possède un grand nombre d'applications dans différent domaine : tatouage numérique, tests groupés (dans lesquelles des échantillons sont testés en pools, et où le but est de déterminer les échantillons ayant une propriété donnée). Les familles de sous-ensembles ayant de telles propriétés de reconstruction sont connues sous la dénomination de codes super imposés, et ont été étudiés dans différentes communautés, parfois de façon indépendante. Nous nous intéressons au cas où l'ensemble considéré a des propriétés structurelles additionnelles, dans le sens où c'est l'ensemble des sommets d'un certain graphe. Sélectionnons un certain sous-ensemble de ces sommets, qui va jouer un rôle particulier, et que nous appelons code. Considérons alors la collection formée des voisins de sommets du code. Un tel code ayant les propriétés de reconstruction susmentionnées est appelé un code identifiant.

Cette notion a été introduite en 1998 par Karpovsky et al. afin de modéliser un problème de détection de défaillances dans des réseaux multiprocesseurs. Dans ce cadre, on suppose que chacun des processeurs du sous-ensemble correspondant au code a la capacité de tester s'il y a un processeur défectueux dans son voisinage. L'objectif est de déterminer l'ensemble des processeurs défectueux (ou de certifier qu'aucun processeur ne l'est) à partir des réponses fournies par les processeurs du code. Les liens avec la théorie des codes super imposés a été établi récemment, et donne des perspectives pour des développements futurs ainsi que pour des applications nouvelles. Ce problème relève clairement de la théorie des graphes, qui paraît donc naturellement être un outil approprié pour étudier ces codes. Ses liens avec les codes super imposés suggèrent de

plus que les techniques spécifiques de la théorie des codes peuvent être pertinentes pour explorer ces questions, et donner des perspectives originales sur le problème. Nos objectifs concernent l'investigation de versions dynamiques de ces codes, et ce selon deux angles. Premièrement, nous allons supposer que le graphe est donné de façon dynamique, et nous allons construire un code identifiant efficace étape par étape, au fur et à mesure que le graphe nous est donné. Cette approche est complètement nouvelle dans le cadre des codes identifiants, et semble de plus particulièrement pertinente dans des perspectives applicatives (concernant les graphes scale-free). L'aspect dynamique prend en particulier en compte l'ajout ou la suppression de sommets ou d'arêtes, et conduit alors à l'étude de la robustesse des codes vis-à-vis des changements de structures. Deuxièmement, nous allons développer une approche adaptative de ces codes, dans le sens où les sommets peuvent être testés séquentiellement, et où la séquence de sommets à interroger va être construite au fur et à mesure, en fonction des réponses des sommets déjà interrogés. Ceci conduit à la détermination d'une stratégie d'identification, et donc à un arbre de décision, plutôt qu'à la détermination d'un code statique. Nous nous intéresserons à des questions structurelles aussi bien qu'algorithmiques (complexité et détermination d'algorithmes efficaces). Le développement d'outils appropriés passera par l'étude de versions statiques de ces codes dans des versions généralisées.

Partenaires

Université de Bordeaux 1 (partenaire coordinateur)
CNRS DR11

Coordinateur

M. Ralf Klasing – Laboratoire Bordelais de Recherche en
Informatique
klasing@labri.fr

Aide de l'ANR

242 757 euros

Début et durée

Janvier 2009 - 36 mois

Référence

ANR-08-EMER-007

Programme DEFIS

Edition 2008

Titre du projet	NAFIT- Nouvelles formes algorithmiques de la théorie de l'information
Résumé	<p>La théorie algorithmique de l'information (ou complexité de Kolmogorov) a été conçue comme une théorie de l'information issue de la notion d'algorithme. Fondée dans les années 60 mais principalement développée dans les années 70 à 80, cette théorie est devenue célèbre par la beauté et la simplicité de ses raisonnements. Elle a eu des applications directes et indirectes dans des domaines aussi variés que la combinatoire pour lesquels elle a permis de développer des outils puissants, pour l'analyse des algorithmes pour lesquels elle permet des raisonnements compréhensibles dans des situations complexes, mais aussi dans des domaines aussi divers que la sociologie ou l'économie (nous n'en sommes pas spécialistes). Un fort regain d'intérêt pour cette théorie se développe actuellement (et peut être vérifié dans les grandes conférences d'informatique théorique) dans des modèles un peu différents, plus adaptés à l'algorithmique moderne qui traite rarement de la résolution par un programme d'un problème à une entrée mais traite plutôt de flux de données, d'informatique interactive et d'informatique "online". Nous nous proposons dans ce projet d'étudier la théorie des flux d'information en la liant aux notions d'aléatoire, d'imprédictabilité, et de complexité d'interaction. Notre étude a aussi l'intérêt de faire un lien avec la théorie de l'information de Shannon, la théorie des probabilités et c'est un sujet très moderne que de rapprocher les deux domaines des algorithmiques et des probabilités. Dans ce rapprochement, nous nous situons clairement en informatique théorique.</p>
Partenaires	CNRS DR12 (partenaire coordinateur) CNRS DR1
Coordinateur	M. Alexander Shen– Laboratoire d'Informatique Fondamentale de Marseille Alexander.Shen@lif.univ-mrs.fr
Aide de l'ANR	320 209 euros
Début et durée	Janvier 2009 - 48 mois
Référence	ANR-08-EMER-008

Programme DEFIS

Edition 2008

Titre du projet

NATIMAGES- Adaptivité pour la représentation des images naturelles et des textures

Résumé

L'équipe de NatImages s'attaque au problème difficile de l'extraction d'informations dans les images naturelles et les textures. Nous considérons plus particulièrement trois applications: le tatouage compressible, l'imagerie optique du cortex et l'imagerie astrophysique. Ces trois applications nécessitent une modélisation adaptative des images pour pouvoir extraire l'information intéressante depuis les données observées (en l'occurrence les tatouages, les signes d'activité corticale et les sources astrophysiques). Le projet NatImages se concentre à la fois sur les méthodes parcimonieuses et les méthodes variationnelles pour extraire les structures complexes des images et des textures. La géométrie que l'on rencontre dans les vidéos à tatouer, l'imagerie corticale et l'imagerie astrophysique ne correspond pas à des concepts simples comme des contours ou des textures oscillantes. Cette géométrie complexe nécessite l'utilisation d'outils adaptatifs qui étendent les méthodes utilisées de manière intensive en traitement des images: la régularisation variationnelle (avec des modèles de fonctions à variation bornée) et les décompositions parcimonieuses (dans les bases d'ondelettes et au delà). Le projet NatImages est une chance unique de développer de nouveaux outils adaptatifs pour capturer la géométrie des images naturelles et des textures. Les méthodes théoriques développées par NatImages permettront de continuer des avancées récentes dans les domaines des décompositions parcimonieuses et des énergies variationnelles. La performance des décompositions parcimonieuses sera mesurée à l'aide de modèles statistiques adaptatifs. Les régularisations variationnelles seront étendues grâce à des énergies à variation bornée variant spatialement. Les nouveaux outils développés par NatImages exploiteront de nouveaux dictionnaires de "texturelets" ainsi que des espaces de Hilbert adaptatifs et des espaces de Banach non-locaux. Les dictionnaires de "texturelets" correspondent à des décompositions adaptatives qui peuvent être optimisées à la volée sur une image donnée ou bien apprises depuis un ensemble de données. Les espaces de Hilbert adaptatifs et les espaces non-locaux sont optimisés pour respecter la géométrie locale et non-locale des textures. De nouveaux algorithmes seront développés pour résoudre efficacement les problèmes d'optimisation inhérents à ces régularisations parcimonieuses et

variationnelles. Ces algorithmes rapides seront inspirés par des idées récemment proposées dans l'Analyse en Composantes Morphologiques et en optimisation convexe non lisse. Les progrès effectués par NatImages dans chaque domaine (méthodes, outils et algorithmes) auront un impact immédiat sur les performances dans chacune des applications considérées par NatImages: détection de tatouages, extraction d'activité corticale et traitement d'images astrophysiques hyperspectrales.

Partenaires CNRS DR1 (partenaire coordinateur)
CEA
CNRS DR19
ENS Cachan

Coordinateur M. Gabriel Peyré – CEREMADE, Université Paris-Dauphine
gabriel.peyre@ceremade.dauphine.fr

Aide de l'ANR 307 278 euros

Début et durée Janvier 2009 - 48 mois

Référence ANR-08-EMER-009

Programme DEFIS

Edition 2008

Titre du projet	
Partout	PARTOUT- Parallélisme partout
Résumé	<p>Le but du projet PARTOUT est d'étudier, à partir d'un point de vue langage de programmation, l'impact de la globalisation du parallélisme qui couvre maintenant tout le spectre de l'informatique, en partant des architectures multicores, en passant par les multiprocesseurs et les systèmes distribués, jusqu'aux applications déployées sur le Web. La question centrale est: Comment programmer le parallélisme sous-jacent aussi bien aux multicores qu'au Web ? Nous pensons que les solutions logicielles que nous proposons seront bénéfiques aux techniques de programmation concurrentes, parallèles et distribuées, ainsi qu'à la programmation Web.</p> <p>Sur le plan scientifique, nous mettons l'accent sur les fondements du parallélisme, et plus particulièrement sur les trois aspects suivants: expressivité des langages, efficacité d'implémentation et sûreté de l'exécution. Le but de ce projet de recherche fondamentale est de progresser sur ces trois aspects et d'atteindre une meilleure compréhension des liens existant entre eux.</p>
Partenaires	INRIA Sophia-Antipolis Méditerranée (partenaire coordinateur) CNAM Université Paris Sud
Coordinateur	M. Frederic Boussinot– INRIA Sophia Antipolis-Méditerranée frederic.boussinot@sophia.inria.fr
Aide de l'ANR	403 832 euros
Début et durée	Janvier 2009 - 48 mois
Référence	ANR-08-EMER-010

Programme DEFIS

Edition 2008

Titre du projet

PHYL-ARIANE- Phylogénomique : Algorithmes et représentations intégrés pour l'analyse de l'évolution du vivant

Résumé

Mots-clefs : algorithmes, combinatoire des arbres, super arbres, modèles combinatoires et probabilistes, données haut-débit et multi-échelles, évolution des espèces, analyse de gènes et de génomes, visualisation de collections d'arbres. Le but de ce projet est de proposer des algorithmes pour répondre au défi informatique que posent les données haut-débit générées par les projets de séquençage de génomes.

Le contexte du projet est celui de la phylogénomique, un domaine émergent consistant à inférer des connaissances sur l'évolution des gènes et l'histoire des organismes vivants depuis les séquences biologiques issues de génomes complets. L'arbre est généralement le modèle utilisé pour représenter l'histoire évolutive des espèces. Une grande partie de cet arbre, dont notamment les zones les plus anciennes, sont encore à élucider (64% selon le NCBI, dont les relations de parenté entre grands groupes bactériens) et ceci pour au moins deux raisons :

- Un manque de données « complètes » : le 1er génome complet a été séquencé il y a seulement dix ans, et le soucis d'avoir une véritable représentation de la diversité du vivant dans les projets génomes est très récent ;

- Mais surtout un manque de méthodes capables de prendre en compte la complexité des données de façon fidèle : actuellement, on reconstruit d'abord un arbre de gène décrivant les relations de parenté entre séquences apparentées à un gène ; l'arbre du vivant est ensuite estimé à partir d'un ensemble d'arbres de gènes. Cependant, les histoires de gènes et de génomes sont bien plus complexes que le simple modèle arboré et sont engendrées par un processus à plusieurs échelles : des micro-événements comme les substitutions agissent au niveau de chaque site et des macro-événements comme les duplications ou les transferts de gènes agissent au niveau d'un génome ou entre génomes. Elucider les relations des grands groupes du Vivant depuis les données haut-débit récemment obtenues nécessite une avancée méthodologique importante dans la prise en compte de la complexité des événements évolutifs affectant les gènes. C'est ce que propose de réaliser le projet Phyl-ARIANE, sur la base de modèles et d'algorithmes de nature combinatoire et/ou probabiliste, prenant à la fois en compte les micro- et les macro-événements. Une autre source d'originalité se situe dans l'élucidation conjointe des histoires évolutives des

gènes et de l'arbre du vivant. Il nous semble fondamental de ne pas dissocier l'estimation de ces phénomènes responsables chacun d'une partie du signal porté par les données. Enfin, ce projet effectuera aussi un saut qualitatif dans le domaine de la représentation de collections d'arbres et de la navigation de connaissances sur les phénomènes évolutifs.

Nous proposons des algorithmes d'exploration automatique de collections de gènes, mettant en lumière les points communs entre gènes, et mettant en regard les histoires de gènes et d'espèces, afin de faciliter la production de nouvelles connaissances dans les études futures. Les progrès réalisés dans ce projet sur la combinatoire des arbres, les algorithmes de réconciliation d'arbres et la représentation d'arbres bénéficieront aux nombreux domaines où ce modèle est employé : classification hiérarchique, fouille de données, analyse d'images, chimie, etc. Par ailleurs, l'application d'algorithmes de complexité paramétrique à des données de très grandes tailles contribuera à populariser cette technique relativement récente de l'informatique fondamentale. Les connaissances sur l'Evolution résultant de ce projet permettront de nombreuses avancées à court et moyen terme en sciences de la vie : inférence automatique de la fonction des gènes, détection de gènes impliqués dans l'interaction entre plusieurs organismes interdépendants pour l'agronomie, meilleure compréhension du contexte génomique lié aux transferts de gènes, et compréhension des mécanismes responsables de la transmission entre bactéries de la résistance aux antibiotiques.

Partenaires Université Montpellier 2 LIRMM (partenaire coordinateur)
Université Montpellier 2 ISEM
CNRS DR7

Coordinateur M. Vincent Berry – Université Montpellier 2
vberry@lirmm.fr

Aide de l'ANR 350 625 euros

Début et durée Janvier 2009 - 36 mois

Référence ANR-08-EMER-011

Programme DEFIS

Edition 2008

Titre du projet **QRAC- Algorithmes et complexités quantiques et probabilistes**

Résumé

Notre projet consiste à réunir les forces des experts des deux communautés du calcul probabiliste et quantique. Notre objectif principal consiste à contribuer significativement dans les domaines du calcul probabiliste et quantique, mais aussi, en réunissant nos compétences, à assurer un transfert de techniques du calcul probabiliste au calcul quantique, et vice-versa.

Pour le calcul probabiliste, nous nous concentrons essentiellement sur des modèles de masses de données dont l'accès est restreint, et pour lesquels les techniques probabilistes et parfois d'approximation sont des outils fondamentaux. Nous couvrirons les domaines du test approché de propriétés (Property Testing), des algorithmes de flots (Streaming Algorithms), des algorithmes en ligne (Online Algorithms), et de la théorie algorithmique des jeux (Algorithmic Game Theory). Pour le calcul quantique, nous continuerons de développer des algorithmes, en particulier pour le problème du sous-groupe caché, les problèmes de vecteurs dans les réseaux, mais aussi sur des algorithmes élaborés à partir des marches quantiques. Nous étudierons aussi les bornes inférieures pour la complexité en requêtes quantiques, la théorie de l'information quantique et la communication quantique. A l'intersection des modèles de calcul quantique et probabiliste, notre objectif est d'approfondir l'étude des codes localement décodables, la complexité de la communication, et les méthodes de bornes inférieures pour la taille des circuits.

Partenaires

Université Paris Sud 11 (partenaire coordinateur)

Coordinateur

M. Frédéric Magniez – Université Paris Sud Orsay : Paris XI
magniez@lri.fr

Aide de l'ANR

420 000 euros

Début et durée

Janvier 2009 - 48 mois

Référence

ANR-08-EMER-012

Programme DEFIS

Edition 2008

Titre du projet

SEQUOIA- Analyse syntaxique probabiliste à large couverture du français

Résumé

Le problème auquel on se propose de s'attaquer dans ce projet est celui de l'analyse syntaxique automatique probabiliste à large couverture du français. L'analyse syntaxique automatique consiste à produire, à partir d'une phrase d'une langue donnée, toutes les analyses syntaxiques de cette dernière. Un analyseur est probabiliste s'il associe de plus à toute analyse produite, une probabilité, calculée à l'aide d'un modèle probabiliste. Un analyseur est dit à large couverture s'il est capable d'analyser un échantillon significatif des phrases de la langue étudiée. L'analyse syntaxique constitue généralement une étape préliminaire à une analyse du sens de la phrase et intéresse à ce titre plusieurs applications relevant du traitement automatique des langues, telles que l'extraction d'informations ou la traduction automatique. Les raisons pour lesquelles nous nous intéressons à ce problème sont de trois ordres :

1 - L'explosion de la diffusion de contenus textuels numériques, notamment sur l'internet pose de manière cruciale le problème du traitement de leur contenu. Les techniques actuelles mises en œuvre pour le traitement du contenu textuel font généralement l'économie d'une analyse syntaxique complète des phrases du texte à traiter. La raison principale est l'absence d'analyseurs syntaxiques possédant la maturité suffisante pour être intégrés dans une application de grande ampleur. Les conséquences sont un traitement moins précis de l'information.

2 - L'analyse syntaxique probabiliste a connu, depuis une quinzaine d'années un développement spectaculaire, qui a permis une amélioration significative des performances des analyseurs. Les développements ont surtout été réalisés aux Etats-Unis et portent sur l'anglais. La France accuse de ce point de vue un retard important, tant au niveau de la recherche qu'au niveau de la réalisation d'analyseurs pour le français.

3- Les équipes impliquées dans le présent projet sont à la pointe de la recherche dans leurs domaines respectifs (traitement automatique des langues, modélisation probabiliste de la langue, analyse syntaxique et apprentissage automatique). Leur collaboration permettra l'émergence d'une équipe originale dans le paysage de la recherche française. Elle sera de plus particulièrement bien armée pour s'attaquer à un problème de cette ampleur. Ce projet relève par certains aspects de la recherche fondamentale et par d'autres de la recherche

appliquée. Il permettra d'une part une meilleure visibilité de la France sur ce thème de recherche et, d'autre part, de fournir à la communauté du traitement automatique du français un analyseur de performances au moins égales à l'état de l'art international. La stratégie générale adoptée pour ce projet consiste à produire un analyseur état de l'art pour le français, en recourant aux techniques qui ont prouvé leur efficacité pour l'anglais puis, à l'issue de l'analyse des résultats obtenus, aller au delà de l'état de l'art en reposant sur les compétences complémentaires des équipes partenaires.

Partenaires

Université de la Méditerranée Aix Marseille II (partenaire coordinateur)
INRIA Paris Rocquencourt
Université de la Sorbonne Paris 4
Université d'Avignon et Pays du Vaucluse

Coordinateur

M. Alexis Nasr – Laboratoire d'Informatique Fondamentale
alexis.nasr@lif.univ-mrs.fr

Aide de l'ANR

623 157 euros

Début et durée

Janvier 2009 - 36 mois

Référence

ANR-08-EMER-013