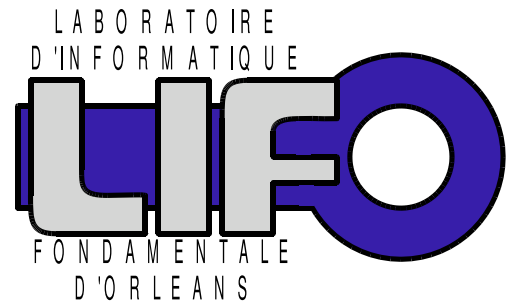




UNIVERSITE D'ORLEANS



4 rue Léonard de Vinci

BP 6759

F-45067 Orléans Cedex 2

FRANCE

<http://www.univ-orleans.fr/lifo>

Rapport de Recherche



FÉDÉRATION INFORMATIQUE CENTRE VAL DE LOIRE

ICVL

Projet Scientifique

Laboratoires de recherche :

LI : Laboratoire d'Informatique (EA 6300)

LIFO : Laboratoire d'Informatique Fondamentale d'Orléans (EA 4022)

Porteurs :

Agata Savary (LI)

Mirian Halfeld Ferrari Alves (LIFO)

8 juillet 2016



TABLE DES MATIÈRES

1	La fédération ICVL	4
1.1	ICVL : la rencontre de deux laboratoires	5
1.1.1	Les laboratoires	6
1.1.2	Les tutelles	8
1.2	Construction d'une fédération : aperçu de la méthode de travail . .	8
2	Les besoins sociétaux et les verrous scientifiques de nos recherches	12
2.1	Les besoins sociétaux	12
2.2	Les verrous scientifiques	13
3	Les axes de recherche	15
3.1	Axe A : Données massives, complexes ou hétérogènes	16
A1 :	Extraction des connaissances et reconnaissance des formes	16
A2 :	Données du Web	18
A3 :	Traitement automatique des langues	20
A4 :	Analyse et traitement d'images	21
3.2	Axe B : Algorithmique et optimisation	24
B1.	Algorithmique des graphes et complexité exponentielle . .	24
B2.	Ordonnancement et transport	25
3.3	Axe C : Sécurité, fiabilité et performances	28
C1.	Sécurité des systèmes	28
C2.	Programmes corrects et vérification de systèmes	29
C3.	Parallélisme et systèmes repartis	31
3.4	Vision globale d'une recherche complémentaire	35
4	Dynamique de coopération : actions, projets, ouverture internationale	38
5	ICVL : qualité et dynamisme dans la recherche	41
5.1	ICVL : motivation pour un plus grand investissement scientifique	41
5.2	ICVL : nouveaux horizons pour nos recherches	42
5.3	ICVL : l'union vers des objectifs internationaux	42
5.4	ICVL : vers une structuration transversale.	43

A	Détails sur les actions	45
A.1	Actions 2016	45
A.1.1	Algorithmique exponentielle et paramétrée en ordonnancement	45
A.1.2	Hybridation des techniques de clustering et d'extraction de motifs	46
A.1.3	Extraction de motifs séquentiels à grande échelle dans un environnement Hadoop / MapReduce	49
A.1.4	Similarité de trajectoires et optimisation	50
A.1.5	Masses de données dans une approche sémantique : consultation, maintenance, évolution	51
A.1.6	Techniques d'apprentissage pour le TALN	53
A.1.7	Analyse syntaxique (parsing) pour l'annotation de la temporalité	55
A.1.8	Expressions polylexicales et parsing	58
A.1.9	Éthique et risques des technologies numériques	59
B	Détails sur les projets	62
B.1	Projets passés et en cours	62
B.1.1	Information Technologies for Business Intelligence (IT4BI)	62
B.1.2	PARsing and Multi-word Expressions (PARSEME)	64
B.1.3	Syntactic Parsing and Multiword Expressions in French (ANR PARSEME-FR)	65
B.1.4	Performance, évolution et composition pour XML : modèles, algorithmes et systèmes (ANR Codex)	66
B.1.5	Big data on the cloud (BigTrend)	67
B.1.6	Fouille de données temporelles - Application à des données environnementales (FDTE)	68
B.1.7	Grands graphes : interrogation, fouille et analyse (GIRAFON)	69
B.1.8	Annotation temporelle (MSH TEMPORAL)	70
B.1.9	Outiller les données pour le développement des industries de la langue (IA ODIL)	71
B.1.10	Réseau Thématique Régional autour du risque (RTR Risque)	72
B.2	Projets soumis et en cours d'évaluation	74
B.2.1	Big Data Management and Analytics (BDMA)	74
B.2.2	Large Scale Anomaly Recognition (LASCAR)	75
B.2.3	Innovative RENaissance HERitage (I-RENHER)	76

B.2.4	SHAring GRAMmatical and LExical Resources (FP7 SHA-GRALER)	77
B.2.5	Discrimination prosodique automatisée et sémantique de l’oral (ANR SEMORAL)	78
B.2.6	Smart Loire	79
B.2.7	Analyse syntaxique des expressions figées (IA ASEF)	80
B.2.8	Annotation temporelle (IA ANNOTEP)	81
C	Exemples des facteurs de visibilité extérieure	82
D	Co-encadrements	93
E	Publications communes aux deux laboratoires	94
F	Publications co-signées par des membres de plusieurs équipes des laboratoires LIFO ou LI	98
G	Statuts	100
G.1	Statuts de la Fédération ICVL	100
H	Compte rendu de l’élection pour la direction de l’ICVL	106
H.1	Les élections pour la direction ICVL : candidature et procédure	106
H.2	Lettre de candidature	107
H.3	Résultats	109

LA FÉDÉRATION ICVL

La fédération Informatique Centre Val de Loire (ICVL) est issue de la rencontre des deux laboratoires de la Région Centre Val de Loire ayant comme ambition l'implantation d'une dynamique de collaboration pour l'enrichissement de leurs recherches en informatique.

Ce document présente le projet scientifique de cette fédération, décrivant les bases des collaborations des deux laboratoires, ainsi que la plus-value attendue par sa mise en place.

À titre d'introduction, nous rappelons les objectifs majeurs de ICVL ainsi que les informations de base concernant les laboratoires membres et leurs tutelles. Le travail de mise en place de notre fédération, initialisé en juin 2014, est résumé très brièvement dans ce chapitre. Nous évoquerons enfin, de manière sommaire, la méthodologie utilisée pour la construction du projet scientifique.

Il est à remarquer, dès maintenant, que tout notre effort dans la mise en place de ICVL porte déjà ses fruits dans la dynamisation scientifique de nos laboratoires. Des collaborations anciennes s'approfondissent et songent à des nouvelles perspectives, des forces complémentaires se mettent ensemble pour construire un pôle d'excellence dans *l'acquisition et le traitement efficaces, fiables et sécurisés de l'information*.

1.1 ICVL : la rencontre de deux laboratoires

La fédération ICVL réunit les deux laboratoires d'informatique de la région Centre Val de Loire, le LI (Laboratoire d'Informatique de l'Université de Tours) et le LIFO (Laboratoire d'Informatique Fondamentale d'Orléans). Cette fédération est le résultat d'une politique concertée entre les deux laboratoires, visant en priorité la dynamisation de la recherche en informatique dans notre région dans le respect de la diversité des thématiques de ses membres.

Objectifs de la Fédération

- valoriser et améliorer la visibilité des activités de recherche informatique en Région Centre ;
- susciter et favoriser des projets de recherche au niveau régional, national et international ;
- dynamiser l'activité de recherche au sein des deux Parties et encourager leur collaboration.

La fédération ICVL trouve ses fondements dans des liens tissés sur une décennie entre les deux laboratoires. Ces liens ont vocation à être pérennisés et amplifiés. Dans l'historique de collaboration scientifique citons, par exemple :

- les JIRC (Journées informatiques de la région Centre), une rencontre annuelle entre chercheurs des deux laboratoires ayant lieu depuis 2005 ;
- les PPF (Programmes Pluri-Formation) *Fouille de donnée en Région Centre* (2008) et *Géométries, Images, Communications*, à l'échelle de l'actuelle COMUE (2008) ;
- le projet *Complexité des méthodes exactes exponentielles pour des problèmes d'ordonnancement*, dans le cadre du GdR-CNRS RO (2009-2010) ;
- le projet *Vérification de Modules Composés*, financé par les conseils scientifiques de deux universités (2009-2010) ;
- le projet ANR DEFIS-Codex (cf. B.1.4), étendu au LIFO en 2009, permettant la continuation et l'extension des collaborations (2009-2012) ;

- l’action COST IC1207 PARSEME (cf. B.1.2), un réseaux scientifique européen dédié à la précision et l’efficacité des applications de Traitement Automatique des Langues (2013-2017) ;
- le projet ANR PARSEME-FR (cf. B.1.3), le spin-off français de l’action COST IC1207 PARSEME (2015-2019) ;
- le projet régional ODIL (2015-2016, cf. B.1.9) ;
- la co-organisation de l’École Jeunes Chercheu/r/se/s en Informatique Mathématique¹ à Orléans en 2015 ;
- le co-encadrement de deux thèses financées par une bourse nationale (2013-2016) et par une bourse régionale Orléans-Tours (2014-2017) ;
- le dépôt de plusieurs projets communs (dont 2 européens, 1 ANR et plusieurs projets régionaux, cf. section B) ;
- collaboration autour de la formation pour l’option *Informatique et sciences du numérique* de la terminale S au lycée.

1.1.1 Les laboratoires

Les deux laboratoires sont de tailles et de structures comparables. Ils ont reçu des évaluations comparables par l’AERES (note A). Leurs thématiques de recherche sont proches, non redondantes et souvent complémentaires. Géographiquement, ils sont répartis sur l’axe ligérien entre Orléans et Tours avec un site à Blois et un site à Bourges. Ce sont les seuls laboratoires de recherche en informatique de la région Centre Val de Loire. Ensemble ils comptabilisent environ 150 membres, repartis selon l’indication de la table 1.1.

Laboratoire	Direction	# Membres	# Permanents	# HDR
LI	Jean-Charles Billaut	75	47	22 (dont 14 PR)
LIFO	Jérôme Durand-Lose	68	42	18 (dont 13 PR)

TABLE 1.1 – Situation des laboratoires en septembre 2015

Le LI (EA 6300) compte² 47 permanents, 24 doctorants et 4 docteurs ou

1. <http://www.univ-orleans.fr/lifo/evenements/EJCM2015/>

2. Référence de janvier 2016.

post-doctorants. Le laboratoire est structuré en trois équipes, indiquées dans la table 1.2. Trois grands domaines d'applications fédèrent les activités du LI, à savoir : (i) santé et handicap, avec les problèmes d'aide technique au handicap physique, à l'autisme, au diagnostique et à la décision médicale, (ii) masses de données, avec des défis autour de la modélisation, la construction et l'exploitation de données, (iii) humanités numériques, avec des problématiques liées aux structures des bases de données, à la numérisation 3D, à la reconnaissance des formes.

Équipe LI	Thématiques
BdTIn	Bases de Données et Traitement des Langues Naturelles
OC	Ordonnancement et Conduite
RFAI	Reconnaissance des Formes et Analyse d'Images

TABLE 1.2 – Équipes du LI en 2015

Le LIFO (EA 4022) compte³ 42 permanents (dont 37 affiliés et 5 associés), 15 doctorants, 5 docteurs ou post-doctorants et 2 personnes administratives, 2 professeurs émérites et 2 personnes d'autres catégories. La table 1.3 montre les cinq équipes structurant le LIFO. Afin d'avoir une approche transversale, à une structuration formelle (les équipes) est ajouté une structuration souple : les thèmes transversaux. Trois thèmes sont mis en avant : (i) masse de données et calcul haute performance, (ii) modélisation et algorithmique, (iii) sécurité et sûreté.

Équipe LIFO	Thématiques
CA	Contraintes et Apprentissage
GAMoC	Graphes, Algorithmes et Modèles de Calcul
LMV	Logique, Modélisation et Vérification
Pamda	Parallélisme, systèmes distribués et bases de données
SDS	Sécurité et Distribution des Systèmes

TABLE 1.3 – Équipes du LIFO en 2015

3. Référence de janvier 2016.

1.1.2 Les tutelles

Au 1er janvier 2014, l'ENSI de Bourges a fusionné avec l'ENI Val-de-Loire à Blois pour devenir l'INSA Centre Val de Loire. Récemment, l'INSA est devenue tutelle du LI et du LIFO.

Ainsi, actuellement, le LI et le LIFO ont chacun deux tutelles : l'université d'accueil et l'INSA.

La fédération s'insère naturellement dans ce contexte, comptant avec trois tutelles :

- l'Université d'Orléans,
- l'Université François Rabelais de Tours,
- l'INSA Centre Val de Loire.

Il est également à noter que l'une des équipes du laboratoire LI, à savoir l'équipe Ordonnancement et conduite (OC), possède depuis le 1^{er} janvier 2012 le statut d'une *équipe de recherche labellisée* du Centre national de la recherche scientifique (ERL-CNRS 6305).

1.2 Construction d'une fédération : aperçu de la méthode de travail

Les synergies et les complémentarités des laboratoires LIFO et LI montrent qu'une fédération de ces deux unités apportera une valeur ajoutée aux activités de recherche menées par ces deux unités indépendamment. Cependant, son efficacité dépendra en grande partie de la motivation et de la participation des différents membres des deux laboratoires. C'est pourquoi la mise en place de la fédération a reposé sur une démarche largement consultative.

Tout d'abord le binôme de chefs de projet, nommé au printemps 2014 par les conseils des deux laboratoires, a organisé des réunions consultatives avec toutes les équipes de recherche, afin de recueillir leurs avis et attentes. A la base de ces éléments une enquête anonyme ouverte à tous les membres a été menée en automne 2014, avec des questions concernant :

- les **valeurs** communes à défendre au sein de la fédération,
- les **objectifs** de la fédération,
- les **craintes** vis-à-vis du projet,
- le périmètre d'**actions** à mener au sein de la fédération.

Le taux de participation à l'enquête a été de 46%. A la base des réponses il a été possible de classer les différents éléments dans l'ordre d'importance, d'analyser

les points de divergence et de convergence entre les deux laboratoires et d'identifier des risques éventuels pouvant menacer la réussite du projet. En particulier, les objectifs définis comme majeurs suite à cette enquête ont été introduits dans les statuts de la fédération.

En septembre-octobre 2014 le projet a été présenté aux trois tutelles. La réunion commune des deux conseils des laboratoires, ayant lieu en décembre 2014, a permis une discussion sur les résultats de l'enquête et sur le projet scientifique à construire.

Les responsables du projet ont également rencontré les directions des structures analogues, à savoir des fédérations NormaSTIC (Fédération Normande de Recherche en Sciences et Technologies de l'Information et de la Communication), MIRES (Mathématiques et leurs Interactions, Images et information numérique, Réseaux et Sécurité) et Denis Poisson (fédération en Mathématiques et Physique Théorique des Universités d'Orléans et de Tours). Ces échanges, ainsi que la lecture des documents fondateurs de ces structures, ont inspiré la rédaction des statuts de la fédération ICVL, soumis à une consultation large en janvier 2015 et validés par les Assemblées Générales du LIFO et du LI en septembre 2015.

La construction du projet scientifique de ICVL a été structurée autour de trois notions :

- **Axes transversaux** — il s'agit de larges thématiques englobant les compétences et les sensibilités scientifiques représentées par les deux laboratoires. Ils doivent servir à l'affichage externe, mais aussi à la communication et structuration interne. Un axe est souvent transversal à deux laboratoires, mais peut aussi correspondre à l'activité d'une ou plusieurs équipes d'un même laboratoire, l'un des objectifs étant que chaque membre de la fédération puisse s'identifier avec au moins un axe.
- **Actions** – elles correspondent à des collaborations sur de moyennes et longues durées, existant entre les deux laboratoires.
- **Initiatives** – ce sont des événements plus ponctuels qui visent la création ou le renforcement des actions (par exemple des stages co-encadrés, des journées thématiques, des séminaires, etc.). Les initiatives sont, dans la mesure du possible, financées par le budget de la fédération.

Un appel à actions et initiatives lancé en janvier 2015 a permis de définir **9 actions** (voir section 4) basées sur des collaborations nouvelles ou déjà existantes entre les deux laboratoires. Parallèlement aux actions, la fédération peut constituer des **groupes de travail** dédiés à des sujets spécifiques. Trois groupes de travail

créés jusqu'à présent concernent : (i) l'harmonisation des formations de Master informatique en Région Centre, (ii) le développement des liens avec la COMUE et la fédération MIREs, (iii) la préparation d'une demande d'accréditation auprès du CNRS. Notons également que le nom de la fédération ICVL résulte d'un appel à propositions et d'un vote direct.

La construction des **axes transversaux** de la fédération, présentés dans la section 3, résulte d'un effort important des membres des deux laboratoires. La méthodologie, qui se voulait ici encore largement consultative, a été inspirée d'un travail précédent d'analyse et de restructuration à l'échelle d'une des équipes du laboratoire LI. Cette méthodologie, présentée et validée lors des séminaires dédiés au sein du LIFO et du LI, a ensuite été mise en place en deux phases. Premièrement, une nouvelle enquête adressée à l'ensemble des membres a consisté à recueillir les contributions scientifiques majeures des deux laboratoires. Il s'agissait de fournir les informations concernant la nature de chaque contribution :

- résultats principaux obtenus,
- besoins sociétaux auxquels ces résultats répondent,
- verrous scientifiques auxquels on s'attaque,
- objets d'études sur lesquels porte la contribution,
- techniques scientifiques employées,
- l'importance et les aspects novateurs de la contribution,
- les indicateurs de succès, mesurés en terme de : (i) publications, (ii) réseautage et visibilité, (iii) utilisateurs externes des résultats,
- personnes et équipes internes impliquées dans la contribution,
- collaborations nationales et internationales liées à la contribution,
- pertinence par rapport à diverses politiques scientifiques (axes des laboratoires LIFO et LI, domaines privilégiés de la Région Centre, stratégie nationale française en lien avec Horizon 2020, défis interdisciplinaires du CNRS),
- perspectives et expertises recherchées pour la poursuite des travaux.

En réponse à cette enquête 59 contributions ont été soumises. Ces réponses ont ensuite été utilisées comme base de travail collaboratif lors de l'événement annuel majeur dans la vie des deux laboratoires, à savoir les Journées informatiques en Région Centre (JIRC)⁴ en mai 2015 à Bourges. Le travail a été mené au sein de quatre ateliers parallèles, dédiés aux thèmes suivants :

- besoins sociétaux,
- verrous scientifiques,

4. <http://www.univ-orleans.fr/lifo/evenements/JIRC2015/jirc/>

- techniques scientifiques,
- objets d'études.

Chaque atelier, regroupant une quinzaine de représentants des deux laboratoires et animé par un responsable, a étudié les 59 contributions soumises auparavant afin d'opérer des agrégations selon le critère en question. Ce travail a notamment résulté en la définition de quatre typologies dont les deux premières sont présentées dans la section 2. A la base de ces résultats, le binôme chef de projet a proposé la définition des axes transversaux, décrits dans la section 3 et soumis à la consultation auprès des membres.

En conclusion, le présent document est donc le fruit d'un travail largement consultatif et collaboratif de plusieurs dizaines de membres du LIFO et du LI, ce qui nous permet de croire en sa pertinence et utilité au projet de la fédération.

LES BESOINS SOCIÉTAUX ET LES VERROUS SCIENTIFIQUES DE NOS RECHERCHES

2.1 Les besoins sociétaux

Nos travaux de recherche s'organisent pour répondre essentiellement à cinq besoins sociétaux :

1. **Accès à l'information et création de connaissances.**

Un très grand panel de nos travaux de recherche répondent à ce besoin. Citons les travaux théoriques ou pratiques concernant la recherche, l'accès à, la visualisation, l'intégration, la comparaison ou l'analyse d'information diverse (structurée ou non structurée) en prenant, parfois, en compte la distribution des données pour proposer une manipulation efficace et sûre. Nous trouvons dans ce contexte non seulement des travaux du domaine de l'apprentissage, data mining, bases de données, traitement automatique des langues, vérification et reconnaissance d'image mais aussi du parallélisme et de l'optimisation.

2. **Services à la personne.**

Les travaux d'optimisation menant à des améliorations des plannings dans les hôpitaux, la proposition des systèmes d'aide à la communication des handicapés (comme la prédiction de mots) ou l'aide à la composition des services sont des exemples de nos recherches dans le cadre des services à la personne.

3. **Santé, bien-être, environnement et patrimoine.**

Ce besoin sociétal réunit nos travaux dans les domaines de : (i) la combinatoire (programmation linéaire, méta-heuristiques, etc.) où nous proposons des algorithmes et modèles pour la planification des tournées d'aide à domicile ou le déploiement des ambulances, (ii) la reconnaissance de formes

et des motifs qui s'appliquent à l'aide du diagnostic médical et à l'analyse archéologiques de tessons céramiques, (iii) la fouille de données temporelles appliquée à des données environnementales.

4. Transports, production et énergie : ville intelligente.

Dans ce cadre, nous nous penchons sur plusieurs thématiques en utilisant des techniques diverses. Parlons, par exemple, de nos travaux pour le traitement de masses de données issues des capteurs GPS, permettant une aide à la compréhension des déplacements de véhicules, piétons ou autres. Cela, sans oublier les algorithmes proposés pour la planification de la maintenance dans l'industrie de la production d'énergie ou les tournées de véhicules, illustre comment nos recherches répondent à ce besoin sociétal.

5. Sécurité, sûreté et éthique.

Toutes nos recherches dans le cadre de la sécurité des systèmes, la protection de la vie privée, l'analyse des attaques, etc, ainsi que celles concernant le développement des programmes parallèles corrects, la fiabilité des logiciels, le renforcement de la confiance dans les infrastructures d'informatique en nuage, par leurs résultats théoriques ou appliqués, répondent à ce besoin sociétal. Plus récemment, l'intérêt porté aux problèmes éthiques liés aux nouvelles technologies motivent des travaux multidisciplinaires dans le but de proposer de nouvelles solutions pour la protection de la vie privée ou de données personnelles.

2.2 Les verrous scientifiques

Les verrous scientifiques que nos recherches tentent de lever peuvent être classés dans quatre groupes. Une contribution s'intéresse parfois à plusieurs de ces verrous, même si un verrou majeur est souvent identifié. Nos ateliers aux JIRC 2015 ont permis une large discussion sur ces points.

1. Intégration ou adaptation à l'environnement

Dans ce cadre, nous plaçons nos travaux qui proposent des méthodes : pour l'intégration et la prise en compte de l'utilisateur ; pour l'aide à l'interprétation, l'exploitation ou la visualisation des informations diverses ; pour l'intégration de contraintes (bruit, incomplétude) ; pour la dynamique ou l'évolution de l'environnement ; pour l'intégration de l'information. Nos résultats portent aussi sur des questions de temporalité et de passage à grande échelle (Big Data, multidimensionnalité).

2. Qualité, fiabilité et sécurité

Nos recherches s'efforcent de lever les verrous concernant la précision, fiabilité, confiance, ainsi que la qualité de données, programmes, processus et analyses. Nous sommes investis dans des thèmes concernant la vérification, la correction par construction et le parallélisme, tout en dédiant un volet de nos recherches aux défis de la sécurité.

3. Complexité et optimisation

Nous nous intéressons particulièrement aux problèmes combinatoires et d'optimisation et à la caractérisation des problèmes et de leurs classes de complexité. Nous sommes également motivés par les besoins d'efficacité et de performance des algorithmes, ainsi que par leurs implémentations.

4. Modélisation, représentation et génération

Nous proposons des solutions aux questions de représentation, modélisation et structuration de l'information, ainsi qu'aux interrogations sur l'interopérabilité. Nous sommes investis dans la génération automatique des modèles et de structures (par exemple, la génération automatique des architectures de déploiements prenant en compte les besoins fonctionnels et de sécurité) et dans la générique et composition des modèles, interfaces, architectures (par exemple, l'automatisation de la découverte et de la composition de services web sémantiques).

LES AXES DE RECHERCHE

En considérant en détail les thèmes des recherches des membres des deux laboratoires, nous avons établi trois axes transversaux de recherche. La transversalité ne se restreint pas au cadre de collaboration entre équipes, mais propose des associations plus fines, où les différents thèmes de chaque équipe sont considérés. Dans ce contexte, des axes concernant seulement des travaux du même laboratoire sont possibles. Cette classification a été faite en prenant en compte toutes les branches de nos recherches et englobe ainsi tous les membres de l'ICVL.

Dans la suite, nous nous efforçons pour donner une description large et précise des axes sans pourtant avoir la prétention de citer ou détailler les nombreuses contributions existantes. Il ne s'agit pas d'une liste exhaustive de nos recherches, mais une description concise des thématiques que nous essayons de faire avec justesse. Par surcroît, les axes représentent un premier rassemblement de nos forces vers des buts scientifiques communs et le premier pas vers la structuration d'un pôle de recherche pour *l'obtention et le traitement efficace, sûr et sécurisé de l'information*.

3.1 Axe A : Données massives, complexes ou hétérogènes

Équipes impliquées

BdTIn (LI), CA (LIFO), GAMoC (LIFO), LMV (LIFO), Pamda (LIFO), RFAI (LI)

Plusieurs problématiques sont soulevées par la grande quantité de données produites dans le monde. L'énorme complexité, dynamisme et hétérogénéité de ces données rendent très difficile le traitement et l'obtention d'information fiable et utile ainsi que leur interprétation et analyse. Cet axe s'intéresse à la proposition des solutions pour l'analyse et l'extraction de connaissances provenant de différentes sources de données, pour l'interrogation et la manipulation efficace et conviviale de données, pour la vérification de leur cohérence par rapport à des contraintes, ... Nous nous intéressons ainsi à des données plus ou moins complexes et structurées et qui peuvent être sémantiquement hétérogènes, incomplètes, incertaines et évolutives.

A1 : Extraction des connaissances et reconnaissance des formes

Membres permanents impliqués

Équipe BdTIn du LI : Thomas Devogele, Laurent Etienne, Arnaud Giacometti, Nicolas Labroche, Dominique Li, Patrick Marcel, Nizar Messai, Verónica Peralta, Arnaud Soulet

Équipe CA du LIFO : Sylvie Billot, Guillaume Cleuziou, Thi-Bich-Hanh Diep-Dao, Matthieu Exbrayat, Willy Lesaint, Lionel Martin, Frédéric Moal, Marcilio Pereira de Souto, Christel Vrain

Équipe RFAI du LI : Sabine Barrat, Fatma Bouali, Thierry Brouard, Hubert Cardot, Nicolas Monmarché, Nicolas Ragot, Jean-Yves Ramel, Romain Raviaux, Gilles Venturini

Un très grand nombre de nos travaux s'intéressent à l'extraction des connaissances, offrant un large panel de méthodes et perspectives d'interaction. Ainsi, en ce qui concerne le type de données traitées (plus au moins structurées, plus au

moins complexes, . . .) les travaux de ce sous-axe concernent non seulement des données dans les bases de données relationnelles, des données géolocalisées ou XML, et des réseaux sociaux, mais aussi des textes, images et vidéos et de la bio-informatique. Ces travaux englobent des technologies variées comme la fouille de données, OLAP (*On Line Analytical Processing*) ou la reconnaissance des formes, ainsi que des techniques diverses telles que statistiques, visualisation de données, apprentissage, etc. Dans ces recherches nous trouvons aussi un intérêt multiple pour la conception d'algorithmes spécialisés dans certains types de données, ainsi que par ceux complètement génériques. Une autre manière d'aborder nos nombreuses recherches en extraction des connaissances consiste à prendre en compte les méthodes numériques, symboliques ou hybrides que nous proposons.

Ainsi, d'un côté, nous pouvons citer des travaux s'intéressant aux méthodes numériques pour la classification non-supervisée où il est à remarquer un travail sur des modèles adaptés au traitement de données complexes (recouvrantes, multi-vues, images, . . .). Dans ce contexte, une nouvelle direction de recherche concerne la modélisation des *clustering* sous contraintes utilisateur en programmation par contraintes. Egalement, nous nous intéressons aux outils permettant de faire de la prévision et de l'identification de motifs (reconnaissance de formes) en particulier grâce aux méthodes d'apprentissage statistique (apprentissage profond, réseaux de neurones, machines à vecteur support, etc.). Une attention particulière est portée à l'intégration d'informations contextuelles, hétérogènes, et a priori, ainsi qu'à l'apprentissage incrémental et continu.

De l'autre côté nous trouvons les recherches concernant les méthodes symboliques comme, par exemple, l'extraction de motifs (extraction, stockage et exploration d'ensembles de motifs locaux), la définition de nouveaux motifs spatio-temporels, l'apprentissage relationnel statistique, les mesures de complexité de données (pour expliquer la difficulté de leur classification) et enfin les méthodes de caractérisation et mesure de similarité entre graphes.

Au delà de ces deux volets, des méthodes hybrides sont également source d'intérêt pour les recherches de ce sous-axe. Dans cette branche, d'ailleurs, des directions nouvelles, naissant de la complémentarité des travaux dans les deux branches citées auparavant, sont proposées (voir par exemple l'action qui propose l'hybridation des techniques de clustering et d'extraction des motifs, section A.1.2).

Nous pouvons classer dans ce volet certaines recherches en reconnaissance des formes basées sur l'hybridation et la collaboration d'outils et de méthodes : fusion numérique (boosting, combinaison de classificateur, . . .), fusion numérique-symbolique (théorie des croyances et logique floue, graphes et méthodes numériques),

intégration de l'expert dans les systèmes de reconnaissance et d'analyse (visual analytics).

Il est aussi à placer sous ce sous-axe, les propositions concernant les requêtes OLAP. Récemment un intérêt particulier a été porté à la personnalisation de ces requêtes. En effet, des travaux en direction de la définition des modèles de préférences et de la gestion de la qualité des données sont parmi les perspectives de ce sous-axe.

A2 : Données du Web

Membres permanents impliqués

Équipe BdTIn du LI : Béatrice Bouchou, Denis Maurel, Nizar Messai, Verónika Peralta, Yacine Sam, Agata Savary
Équipe GAMoC du LIFO : Anthony Perez
Équipe LMV du LIFO : Pierre Réty
Équipe Pamda du LIFO : Mostafa Bamha, Jacques Chabin, Mirian Halfeld Ferrari, Sophie Robert

Ce sous-axe se caractérise par les travaux sur des données XML (format privilégié pour la communication via le web), l'intégration de données, le web sémantique et les problématiques liées à l'intégration d'applications au travers la technologie des services web.

L'évolution des contraintes (de schéma ou d'intégrité) sur les documents XML a été le cadre de plusieurs contributions : la validation incrémentale ; la correction de documents par rapport à un schéma et des méthodes pour l'évolution des schémas XML (par exemple, la construction d'un schéma XML minimal incluant un ensemble de schémas initiaux).

L'interopérabilité, vue sous différentes facettes, a été la motivation pour des travaux non seulement sur l'intégration des contraintes XML (intégration des dépendances fonctionnelles) et sur l'intégration des ontologies (méthode générique de construction quasi-automatique d'un médiateur sémantique, ouvert et dynamique), mais aussi pour certains problèmes de composition de services (ou les défis principaux sont le nombre croissant de services et la variété des langages de description utilisés). Dans ce cadre, une nouvelle architecture pour la description et la découverte/composition de services web, tenant compte de la similarité sémantique entre services, a été proposée. Nos autres travaux se réfèrent

à la technologie de services web de type REST (Representational State Transfer), qui évolue vers un standard et donne accès à des ressources décrites et liées sémantiquement (via des RDF et des Linked Open Data). Nous nous penchons sur la configuration de réponses issues de plusieurs services REST, en prenant en compte le contexte et les profils utilisateurs. Le domaine d'application est la construction de parcours touristiques personnalisés.

De plus, l'intérêt par composition de services web a tissé des liens avec le sous-axe parallélisme (section C2.) – dans le cadre des travaux sur un modèle de composants pour les simulations scientifiques – ainsi qu'avec le sous-axe vérification (section C3.). Dans le premier cas, une méthode de composition visant aussi bien les applications de web services que les applications basées sur une programmation par composants a été proposée. Dans le second cas, nos travaux ont testé la compatibilité des schémas XML, utilisés par deux services web, via des algorithmes de comparaison approximative, modélisés par des grammaires d'arbres. Les perspectives vers la personnalisation de la composition des services sont à noter.

Ce sous-axe évolue actuellement vers la manipulation des données et contraintes sur les grands graphes où la sémantique des divers types de données est prise en compte. Par exemple, nous analysons les propriétés topologiques des graphes des réseaux sociaux afin d'y découvrir des "capitalistes sociaux" (personnes tentant d'augmenter leur visibilité en falsifiant certains paramètres). Nous nous intéressons également au développement d'un langage de requête déclaratif, à l'évolution des contraintes, aux méthodes de partitionnement des données, ainsi qu'aux algorithmes parallèles pour le traitement de grands graphes ou des requêtes complexes en utilisant le modèle MapReduce ou des variantes. Sur ce dernier point, il est essentiel de rappeler que, depuis un certain temps, nous nous intéressons aux algorithmes parallèles, capables d'assurer un équilibre des données, pour les requêtes sur bases de données relationnelles parallèles.

A3 : Traitement automatique des langues

Membres permanents impliqués

Équipe BdTln du LI : Jean-Yves Antoine, Béatrice Bouchou-Marhoff, Nathalie Friburger, Nicolas Labroche, Denis Maurel, Agata Savary, Arnaud Soulet
Équipe CA du LIFO : Sylvie Billot, Guillaume Cleuziou, Thi-Bich-Hanh Diep-Dao, Denis Duchier, Yannick Parmentier

Les intérêts de ce sous axe sont centrés surtout autour des ressources linguistiques, souvent multilingues, et couvrent divers aspects de leurs cycles de vie.

Premièrement nous développons des formalismes et des outils associés pour la modélisation fine des connaissances linguistiques, adaptée à l'intuition de l'utilisateur-linguiste. Il s'agit notamment des :

- grammaires de propriétés, que nous formalisons en théorie des modèles et dont nous modélisons les analyseurs sous forme de solveurs du problème de satisfaction de contraintes, CSP ;
- métagrammaires, qui sont des descriptions modulaires, expressives et concises, plus faciles à écrire et à maintenir qu'une grammaire, et dont des grammaires peuvent être générées automatiquement ;
- bases de connaissances pour la modélisation multilingue des entités nommées, avec représentation fine à la fois des relations qu'elles entretiennent au niveau indépendant de la langue et de leurs diverses réalisations de surface ;
- lexiques des expressions polylexicales, avec la prise en compte de langues à morphologie riche, et divers types de variation linguistique.

Plus récemment, nous nous intéressons aux entités polylexicales (noms composés, idiomes, expressions figées, etc.) en tant que défi majeur de la modélisation linguistique et nous tentons d'intégrer leur représentation à différents niveaux du traitement automatique des textes, en particulier dans l'analyse syntaxique.

Deuxièmement, nous développons des méthodes pour la production automatisée de ressources linguistiques, en particulier pour l'acquisition automatique semi-supervisée de taxonomies lexicales à partir de textes. Le cadre théorique de la pré-topologie est employé ici pour définir de nouveaux opérateurs adaptés à la modélisation des relations sémantiques entre termes.

Troisièmement, nous produisons des ressources concrètes telles que lexiques électroniques d'expressions polylexicales, taxonomies d'entités nommées et corpus, notamment oraux, annotés en entités nommées et en coréférence. Nous atta-

chons une attention particulière à l’accessibilité de ces ressources grâce à l’emploi des licences libres et des standards d’échange tels que LMF et TEI.

Finalement, nous proposons des outils de traitement de textes qui exploitent à des degrés divers les ressources linguistiques. Il s’agit des méthodes pour (i) la reconnaissance d’entités nommées à base de règles, d’apprentissage automatique ou hybrides, (ii) le clustering éphémère, c’est-à-dire l’organisation sémantique à la volée d’un ensemble de textes courts (ou snippets), (iii) la prédiction de mots sur un clavier virtuel pour les handicapés moteurs lourds.

A4 : Analyse et traitement d’images

Membres permanents impliqués

Équipe RFAI du LI : Thierry Brouard, Donatello Conte, Mathieu Delalandre, Moncef Hidane, Pascal Makris, Julien Mille, Julien Olivier, Jean-Yves Ramel, Gilles Venturini

Ce sous-axe se concentre sur les données spécifiques que sont les images et les vidéos. Nous nous intéressons en particulier aux méthodes d’analyse permettant de segmenter les images pour y retrouver des zones d’intérêts et pouvoir ensuite les identifier et les manipuler.

Pour cela nous travaillons du point de vue des traitements bas niveaux sur certaines méthodes de filtrage, de restauration et d’extraction de points d’intérêts. Nos méthodes de segmentation reposent quant à elles sur des méthodes interactives à base de graphes ainsi que sur les méthodes variationnelles (contours actifs, équations aux dérivées partielles), adossées à l’intégration d’information a priori ou bien de décision prise par des classificateurs. Récemment, nous nous intéressons également aux moyens d’effectuer ses traitements dans des systèmes temps réel.

Du point de vue applicatif, nous avons une expertise forte dans les domaines d’application liés au traitement des images de documents, à l’imagerie médicale (2D et 3D) et au suivi dans les vidéos.

Nous résumons maintenant le résultat de notre analyse sur le positionnement de l’axe en termes de verrous scientifiques à lever, besoins sociétaux auxquels il s’efforce de répondre, techniques scientifiques utilisées, ainsi qu’objets d’études

sur lesquels l'axe se focalise. Par la suite, nous évoquons la plus-value que le fédération ICVL apporte pour l'activité scientifique au sein de cet axe.

Positionnement de l'axe

Verrous scientifiques : les travaux de l'axe contribuent à lever les quatre verrous identifiés dans le chapitre 2.

Besoins sociétaux : les recherches de l'axe répondent, principalement, à *l'accès à l'information et création de connaissances*. Néanmoins, plusieurs travaux sont classés comme réponse aux besoins *services à la personne et transport, production et énergie, ville intelligente*, présentés dans le chapitre 2.

Techniques scientifiques : comme *techniques de traitement de données*, nous pouvons mentionner la classification, la visualisation, l'apprentissage statistique et symbolique, l'extraction de données. Les *techniques de modélisation* employées incluent la logique, les grammaires d'arbres et les graphes.

Objets d'études : données collectées ou enrichies, ainsi que programmes et langages.

La plus value de l'ICVL

Les activités en apprentissage et fouille de données au LI et au LIFO sont complémentaires, aussi bien par les méthodes que par les données étudiées. Cette complémentarité permet de répondre ensemble à des appels à projet sur des applications, comme par exemple le projet FDTE (cf. section B.1.6), où les différentes approches peuvent être confrontées et/ou combinées. Elle permet aussi de développer des méthodes originales fondées sur les compétences des différentes équipes, comme le montre par exemple, la proposition d'action action A.1.2.

L'expertise en web sémantique du LI associée à celle du groupe bases de données du LIFO rend possible l'ouverture d'une direction de recherche vers la consultation, manipulation et maintenance des grands volumes de données. Cette collaboration a été entamée par une bourse de thèse Orléans-Tours, le projet GIRAFON (cf. B.1.7) et l'action décrite dans la section A.1.5.

L'ICVL renforce les liens entre les chercheurs en TAL des deux laboratoires, partenaires actifs et complémentaires dans de nombreux projets (B.1.2, B.1.3, B.1.8, B.1.9, B.1.2, B.2.4, B.2.7, B.2.5, B.2.8) et actions (A.1.6, A.1.7 et A.1.8). Les perspectives des recherches associant le web sémantique (LI) et la fouille des texte pour extraction de taxonomie (LIFO) est une autre ouverture apportée par l'ICVL, et incluse dans la proposition de projet européen I-RENNER (cf. B.2.3).

Les extensions envisagées par le groupe fouille de données du LI seront possibles grâce aux compétences du LIFO en algorithmes parallèles et, notamment, en MapReduce. L'action A.1.3 et le projet GIRAFON (cf. B.1.7) en sont la preuve.

Finalement, une grande partie des compétences regroupées au sein de cet axe permet le soutien efficace au programme de formation en Master européen Erasmus Mundus IT4BI (cf. B.1.1) se déroulant à Blois depuis 2012.

3.2 Axe B : Algorithmique et optimisation

Équipes impliquées

GAMoC (LIFO), OC (LI), RFAI (LI)

La modélisation des problèmes de calcul et le développement des algorithmes optimaux pour leur solution sont les défis majeurs de l'informatique. Dans cet axe nous nous intéressons particulièrement aux graphes, qui sont parmi les modèles formels les plus répandus, ainsi qu'à la recherche des algorithmes optimaux associés pour la solution des problèmes souvent NP-difficiles. Par ailleurs, nous nous penchons sur l'optimisation des problèmes d'ordonnancement et de la conduite des systèmes sous différentes contraintes. Ces thématiques se reflètent dans les deux sous-axes tissant des liens autour de la théorie de la complexité.

B1. Algorithmique des graphes et complexité exponentielle

Permanents impliqués

Equipe OC du LI : Jean-Louis Bouquard, Christophe Lenté, Emmanuel Néron, Ameer Soukhal, Vincent T'Kindt

Equipe GAMoC du LIFO : Mathieu Liedloff, Anthony Perez, Henri Thuillier, Ioan Todinca, Jean-Marie Vanherpe

En algorithmique exacte on souhaite calculer une solution optimale à un problème d'optimisation NP-difficile ; les algorithmes conçus sont donc exponentiels, l'objectif étant néanmoins d'obtenir la meilleure complexité possible. Les travaux dans ce sous-axe s'intéressent à la fois à la complexité globale et à la complexité paramétrée. Dans ce premier cas il s'agit d'établir une borne de complexité globale dans le pire des cas. Dans le second la borne recherchée est plus fine car elle dépend d'un paramètre du problème.

L'une des approches développées dans ce sous-axe pour la résolution des problèmes difficiles porte sur les décompositions de graphes en petits morceaux, agencés selon des règles simples. Les algorithmes correspondant ont alors une complexité exponentielle non pas en la taille du graphe, mais en la taille des morceaux de la décomposition. Il s'agit des décompositions arborescentes ou bien des

décompositions linéaires optimales ou proches de l'optimum. De tels algorithmes exponentiels modérés sont ici appliqués notamment à la solution des problèmes à contraintes multiples dans le domaine de l'ordonnancement.

Nous nous intéressons également à la combinatoire des graphes cubiques (tels que chaque sommet a exactement trois voisins), liés au fameux théorème des 4 couleurs (tout graphe planaire est 4-coloriable). Il est équivalent de montrer que tout graphe planaire est 4 coloriable ou bien que tout graphe cubique planaire a ses arêtes coloriables en 3 couleurs. Même si le théorème des 4 couleurs est maintenant prouvé, il a donné lieu à de nombreuses conjectures, toujours ouvertes, sur les graphes cubiques (e.g. la conjecture de Fulkerson), où on cherche la couverture minimale en couplages parfaits d'un graphe cubique sans isthme (un graphe cubique 3-coloriable est optimal en ce sens). Ces conjectures sont des généralisations envisageables du théorème des 4 couleurs. Nous travaillons sur les partitions linéaires des graphes cubiques (partitions de l'ensemble d'arêtes forêts linéaires, dont les composantes connexes sont des chaînes) et sur une nouvelle notion de partition normale.

B2. Ordonnancement et transport

Permanents impliqués

Equipe OC du LI : Jean-Charles Billaut, Patrice Bonhomme, Jean-Louis Bouquard, Carl Esswein, Yannick Kergosien, Christophe Lenté, Patrick Martineau, Jorge Mendoza, Emmanuel Néron, Ameer Soukhal, Claudine Tacquard, Vincent T'Kindt
Equipe RFAI du LI : Nicolas Monmarché

Les problèmes d'ordonnancement et de transport sont des problèmes relevant de la Recherche Opérationnelle. Résoudre un problème d'ordonnancement, c'est prévoir l'exécution d'un ensemble de travaux dans le temps sur un ensemble de ressources de traitement dans le but d'optimiser un ou plusieurs critères (de coût, de durée, ...). Dans un problème de transport, il s'agit de planifier l'utilisation des ressources (camion, voiture, ...) qui permettent le transport d'un ensemble de travaux. Là encore, cela se fait dans le but d'optimiser un ou plusieurs critères de coût, de durée, etc.

Nous nous intéressons à l'étude et la résolution de ces problèmes, qu'ils apparaissent séparément ou conjointement, par les méthodes de la Recherche Opérationnelle.

Les problèmes d’ordonnancement et de transport mettent en œuvre des modèles complexes, difficiles, et possèdent de nombreuses applications dans les systèmes informatiques (HPC, Big Data), les systèmes de production (de bien, de services), les systèmes de santé (préparation de produits pour les chimiothérapies, gestion des blocs opératoires, transport de patients) ou les systèmes de mobilité et développement durable (systèmes de rechargement de véhicules électriques, calculs d’itinéraires touristiques).

Positionnement de l’axe

Verrous scientifiques : les travaux de l’axe contribuent à lever deux verrous : (i) la complexité et l’optimisation (plus précisément, la résolution des problèmes combinatoires ou intégrés et l’efficacité et performance des algorithmes et des implémentations), (ii) la modélisation et la structuration de l’information.

Besoins sociétaux : les recherches de l’axe répondent, principalement, aux besoins liés aux services à la personne dans la vie quotidienne, ainsi qu’au transport, production et énergie.

Techniques scientifiques : en optimisation, nous utilisons des méthodes exactes comme approchées sans garantie de performance (PL/PPC, heuristiques, métaheuristiques). Les graphes sont nos techniques de modélisation principales.

Les travaux de l’axe B offrent des perspectives de coopération inter-équipes et inter-laboratoires, ainsi que des liens possibles avec les axes A et C.

La plus value de l'ICVL

La fédération ICVL offre un cadre naturel pour le renforcement de la collaboration déjà existante entre les deux laboratoires autour de la thématique de l'algorithmique des graphes et complexité exponentielle. L'action A.1.1 témoigne de cette dynamique.

L'algorithmique des graphes offre de nombreuses perspectives applicatives dans les problèmes auxquels se consacrent l'axe A. Notons, en particulier l'algorithmique des grands graphes sur laquelle est axée l'activité récente du sous-axe A.2.

Les techniques de la recherche opérationnelle ont récemment été appliquées à des problèmes liés aux données massives et hétérogènes, qui sont au centre de l'intérêt de l'axe A. Il s'agit notamment de développer des algorithmes parallèles pour la résolution des problèmes d'indexation de graphes. Une thèse co-encadrée par les équipes RFAI et OC du LI est consacrée à ce sujet.

Les travaux sur l'ordonnancement dans les systèmes informatiques trouvent leurs applications dans : (i) le calcul de haute performance, strictement lié à la thématique des données massives (cf. axe A), (ii) l'amélioration des performances des systèmes d'exploitation et réseaux, ce qui complète à la fois thématique de la sécurité des systèmes d'exploitation (sous-axe C.1) et celle du parallélisme (sous-axe C.3).

3.3 Axe C : Sécurité, fiabilité et performances

Équipes impliquées

GAMoC (LIFO), LMV (LIFO), Pamda (LIFO), SDS (LIFO)

L'informatisation de la plupart des activités humaines et la demande croissante de travailler avec une quantité extraordinaire de données devenues accessibles imposent un investissement très important en sécurité (pour empêcher l'utilisation malveillante des informations sensibles), fiabilité (pour assurer un fonctionnement sans faille et précis) et performance (pour garantir l'efficacité des services). La maîtrise des éléments fondamentaux de ces trois domaines est d'une importance capitale dans la société moderne, non seulement pour des raisons économiques mais aussi pour le bien être des citoyens et des états.

L'activité scientifique de l'axe *sécurité, fiabilité et performance* se concentre sur les thématiques suivantes, que nous définissons comme des sous-axes ayant beaucoup d'interactions.

C1. Sécurité des systèmes

Permanents impliqués

Équipe SDS du LIFO : Pascal Berthomé, Christian Toinard, Jérémy Briffaut, Patrice Clemente, Jean-François Lalande, Benjamin Nguyen
Équipe LMV du LIFO : Frédéric Loulergue

Un des objectifs dans cette thématique a été de fournir un système d'exploitation garantissant une sécurité en profondeur, en contrôlant dynamiquement différents services (à degrés de fiabilité différents), co-existants au sein de ce système.

Dans ce cadre, les principaux résultats concernent la formalisation des propriétés qu'un système d'exploitation mandataire peut garantir, la mise en oeuvre de ces propriétés, ainsi que l'analyse des attaques observés sur un pot de miel sécurisé à haute interaction. Ces travaux ont évolué vers la sécurité des systèmes répartis (systèmes de calcul réparti, sécurité dans les *clouds*).

La sécurité logicielle fait aussi partie de cette thématique avec des travaux touchant la sécurité des langages de programmation, des applications embarquées et des applications web. Un point original de ces travaux concerne l'étude statique des programmes sous contrainte d'attaques. Finalement, la sécurité des applications mobiles, ainsi que la conception et le développement des techniques pour la protection de la vie privée et l'assistance à la personne, dans une démarche pluridisciplinaire, sont au coeur des préoccupations de ce groupe.

C2. Programmes corrects et vérification de systèmes

Permanents impliqués

Équipe LMV du LIFO : Siva Anantharaman, Nirina Andrianarivelo, Yohan Boichut, Wadoud Bousdira, Jean-Michel Couvreur, Frédéric Dabrowski, Khalil Djelloul, Ali Ed-Dbali, Frédéric Loulergue, Pierre Réty, Alexandre Tessier

Équipe Pamda du LIFO : Jacques Chabin

Ce sous-axe s'intéresse d'une part à la correction de programmes, et d'autre part à la vérification de systèmes.

La correction d'un programme peut s'entendre de deux façons : on peut vouloir vérifier certaines propriétés du programme, par exemple l'absence d'erreur à l'exécution ou l'absence d'interblocage pour les programmes parallèles, ou prouver qu'un programme implante correctement une spécification fonctionnelle. Dans les deux cas, ces propriétés peuvent être assurées par construction, ou a posteriori par vérification déductive. Nous explorons les quatre combinaisons possibles, chacune d'elles contribuant à la compréhension globale des langages de programmation étudiés. Par la conception de langages de programmation, en particulier pour le parallélisme, basée sur des sémantiques formelles, nous pouvons assurer par construction que tous les programmes du langage vérifient de bonnes propriétés. Nous concevons également des analyses statiques, soit incluses dans la conception du langage (par exemple du typage), soit qui permettent a posteriori d'assurer certaines propriétés des programmes (par exemple l'absence de *Data Race*). Il est extrêmement important que les sémantiques des langages de programmation et les analyses statiques soient formellement définies et que leurs propriétés soient prouvées. Notre outil de prédilection pour ce faire est l'assistant de preuve

Coq. Pour la correction de programmes par rapport à une spécification fonctionnelle, nous avons conçu et continuons à travailler sur un environnement, nommé SYDPACC pour la correction par construction de programmes fonctionnels parallèles dans l'assistant de preuve Coq. Nous prévoyons d'étendre SYDPACC pour la vérification de programmes appliqués au BigData, en particulier MapReduce. Nos travaux de vérification déductive a posteriori comprennent la conception de nouvelles logiques de programmes, la vérification avec Coq de passes de compilation pour des langages parallèles, et la vérification avec Frama-C de composants critiques de systèmes d'exploitation.

Les activités de vérification de systèmes sont des travaux sur la théorie, les algorithmes et les outils. Un des thèmes concerne la vérification de systèmes à nombre infini d'états, par sur-approximation d'automates d'arbres. Dans ce contexte, on s'intéresse à l'accessibilité d'états indésirables. Après des travaux sur les approximations régulières, nos recherches se concentrent sur les approximations non régulières et leurs applications à la vérification de systèmes parallèles et distribués. L'autre thème concerne la sémantique du parallélisme où nous nous intéressons à des systèmes tels que les automates communicants et les réseaux de Petri. La vérification de tels systèmes est difficile parce que la présence de composants concurrents augmente de façon importante le nombre des états. Nous proposons d'étudier des extensions des méthodes algorithmiques efficaces de vérification aux propriétés exprimables dans des logiques de la concurrence que sont les logiques « ordre partiel ». Les méthodes de vérification élaborées seront confrontées à la conception et l'analyse de systèmes communicants à ressources limitées tels que les réseaux de capteurs et de manière plus générale, les systèmes d'objets connectés.

La conception de langages de programmation parallèle est une activité commune avec le C3., et leur application avec l'Axe A (section 3.1). Des activités de vérification de systèmes et de programmes pour la sécurité sont réalisées avec le C1.

C3. Parallélisme et systèmes repartis

Permanents impliqués

Equipe GAMoC du LIFO : Florent Becker, Jérôme Durand-Lose, Nicolas Ollinger, Ioan Todinca

Équipe LMV du LIFO : Wadoud Bousdira, AbdelAli Ed-Dbali, Frédéric Loulergue

Équipe Pamda du LIFO : Mathias Bourgoïn, Sylvain Jubertie, Sébastien Limet, Emmanuel Melin, Sophie Robert

Ce sous-axe contribue à la modélisation, à la performance et à la facilité de la conception des applications par différentes directions de recherche concentrées sur le parallélisme et les systèmes repartis.

D’une part, nous nous penchons sur certains modèles non-conventionnels de calcul, pour la plupart parallèles ou distribués – machines à signaux et auto-assemblage de tuiles, machines de Turing à temps infini, automates cellulaires, pavages, etc. – ainsi que sur les aspects géométriques des interactions au sein de ces modèles. Ici la notion de signal (discret) est déplacée dans un cadre où le temps comme l’espace sont continus. Ce modèle sort de la théorie classique de la récursivité par la nature des positions et des dates, qui correspondent a priori à des nombres réels.

Par ailleurs, l’intégration des techniques issues des automates cellulaires avec des techniques d’algorithmique et combinatoire des graphes, nous permet de nous attaquer à des problèmes tels que bootstrapping percolation (identification de familles de graphes qui, avec une faible distribution initiale de cellules contaminées et une propagation par la règle majorité, contaminent presque sûrement toutes les cellules) et une restriction du modèle distribué CONGEST. Nous nous intéressons également au modèle de calcul *frugal* dans lequel les noeuds de calcul communiquent une seule fois au plus $O(\log n)$ bits d’information à un arbitre qui prend une décision.

Nous nous intéressons également à la conception d’outils et d’environnements pour le développement et la compilation vérifiée de programmes parallèles. Ainsi, ce sous-axe s’intéresse particulièrement au paradigme des squelettes algorithmiques avec la proposition des bibliothèques de squelettes algorithmiques efficaces pour C++ (OSL et SKELGIS). Rappelons qu’un squelette algorithmique peut être vu comme une fonction d’ordre supérieur qui saisit un modèle d’un al-

gorithme parallèle (pipeline, réduction parallèle, etc.). L'utilisateur construit un programme parallèle par la combinaison des appels aux squelettes disponibles.

En outre, le sous-axe s'intéresse aux questions concernant la parallélisation du traitement des informations SIG (système d'information géographique). Ce type de traitement met en œuvre de nombreuses données calculées ou relevées dans le temps et relatives à un espace dont la topologie est fortement liée à des informations géographiques par nature irrégulières. De ce fait, les modèles de programmation parallèle pour être efficaces doivent être adaptés à des structures de données fluctuantes en fonction d'informations extraites du SIG. Le passage à l'échelle est un des défis de nos recherches.

Dans ce sous-axe, nous plaçons aussi nos recherches concernant les algorithmes parallèles pour les simulations scientifiques interactives. L'intérêt du groupe porte sur la visualisation scientifique qui nécessite la composition d'applications à partir de nombreux codes parallèles différents avec un couplage plus ou moins fort, des instruments pour les interactions avec les utilisateurs, des codes de simulations synthétisant les modèles et enfin une sortie de visualisation et de rendu à haute performance. Dans ce cadre, nous trouvons, par exemple, la proposition d'un modèle de composants pour les simulations scientifiques interactives permettant d'exprimer des contraintes sur la cohérence des données.

Nous résumons maintenant le résultat de notre analyse sur le positionnement de l'axe en termes de verrous scientifiques à lever, besoins sociétaux auxquels il s'efforce de répondre, techniques scientifiques utilisées, ainsi qu'objets d'études sur lesquels l'axe se focalise.

Positionnement de l'axe

Verrous scientifiques : les travaux de l'axe contribuent à lever les quatre verrous identifiés dans la chapitre 2.

Besoins sociétaux : les recherches de l'axe répondent, principalement, à deux besoins (i) *sécurité, sûreté et éthique* et (ii) *accès à l'information et création de connaissances*, présentés dans la chapitre 2.

Techniques scientifiques : comme *techniques de programmation, vérification et sécurisation*, nous pouvons mentionner la vérification formelle, l'assistant de preuves, la programmation parallèle ou répartie, la programmation fonctionnelle, la cryptographie, la sécurisation et l'approches par squelettes. Les *techniques de modélisation*, employées incluent la logique, les grammaires d'arbres, les réseaux de Petri et les graphes.

Objets d'études : données collectées ou enrichies ainsi que programmes et langages.

Les travaux de l'axe C offrent des perspectives de coopération avec d'autres axes ainsi que des résultats pouvant être adaptés aux besoins des recherches de l'axe A.

La plus value de ICVL

La diversité des travaux liés à l'aide aux personnes handicapées (LI) et l'expertise du LIFO en terme de sécurité et sureté, permettrons la mise en place d'un travail commun sur le respect de la vie privée, intialisé avec l'action A.1.9.

Les connaissances du LIFO en parallélisme et systèmes repartis peuvent être mises à la contribution pour l'extraction de connaissances faite au LI. Le contact entre ces deux mondes ouvre de nouveaux horizons de recherche. En témoignent les actions des sections A.1.3 et A.1.4 et les possibilités existantes entre les travaux en fouille visuelle (LI) et ceux sur les algorithmes parallèles pour les simulations scientifiques interactives.

Des solutions de parallélisme structuré pourrons être proposées par le travail conjoint de deux équipes du LIFO (Pamda et LMV), en prenant en compte des besoins exprimés par les travaux dans l'axe A. Cette coopération est mise en oeuvre dans le projet GIRAFON (cf. B.1.7).

3.4 Vision globale d'une recherche complémentaire

La division thématique proposée par les axes n'est pas rigide, mais elle constitue une nouvelle manière de rassembler nos thèmes de recherche. Les axes, englobant des collaborations plus étroites ou naturelles ont, en outre, vocation à créer d'autres coopérations inter-thématiques.

Le tableau de la figure 3.1 synthétise notre enquête et le recensement des contributions et directions scientifiques de nos deux laboratoires. Il montre comment les axes définis par la fédération ICVL, contribuent à lever les verrous scientifiques identifiés, ainsi qu'à répondre aux besoins sociétaux mentionnés précédemment. La présence d'un axe dans une case du tableau indique qu'un bon pourcentage des travaux de l'axe se consacrent aux verrous scientifiques et aux besoins sociétaux représentés par la case. Néanmoins, notons que, dans certains cas, un axe apparaît de forme marginale dans une case (indiqué entre parenthèses, car, dans ce cas, l'axe y contribue avec relativement peu de travaux).

À partir de ce travail minutieux il nous est possible d'avancer, de façon synthétique, les premiers aspects qui illustrent la plus-value de l'ICVL. Nous avons, d'ailleurs, pendant la description des axes, évoqués les collaborations scientifiques qui semblent être la base de cette plus-value. Essayons maintenant de les résumer d'une manière plus schématique.

- l'ICVL rassemble, motive et développe la collaboration entre groupes de recherches complémentaires. Les exemples de ces collaborations, existantes ou en cours de construction, sont nombreux.
- Les trois sous-axes de l'axe A démontrent les liens entre le LI et le LIFO par des travaux liant, d'une part, le web sémantique, personnalisation de services, interrogation et évolution de bases de données et, d'autre part, la programmation par contraintes, les métagrammaires, la modélisation multilingue et les lexiques dans les travaux TAL. Il représente aussi la communion de diverses compétences en extraction de connaissances, constituant ainsi une grande force nationale dans ce domaine.
- L'axe C, centré au LIFO, représente une gamme de travaux qui, d'un côté, peut s'inspirer des besoins manifestes de l'axe A en vue d'atteindre des solutions efficaces et sûres et, d'autre part, peut appuyer l'axe A dans son passage à grande échelle.

- Les interactions en gestation entre les axes A et C sont nombreuses, allant de la fouille de données avec les algorithmes de visualisation scientifique aux solutions logicielles pour personnes handicapées et les questions de respect de la vie privée.
 - L'axe B gagne en rapprochant des travaux purement théoriques des recherches applicatives dans le monde de l'optimisation. Il représente sans doute un chantier intéressant pour de larges projets à venir.
- La réunion des forces du LI et du LIFO dans l'ICVL est fondamentale pour la création d'un groupe de référence dans le traitement de grandes masses de données.
 - L'ICVL réunit les forces des deux laboratoires dans la construction des projets régionaux, nationaux ou européens. Les projets (PARSEME, TEMPORAL, GIRAFON, ANNOTEP, PARSEME-FR, ODIL, I-RENHER, SEMORAL, etc., cf. section B), témoignent de cette dynamique, en accélération depuis 2013.
 - L'ICVL se construit dans la perspective de devenir un pôle de recherche dans le domaine de *l'acquisition et du traitement efficaces, fiables et sécurisés de l'information*.

		Verrous scientifiques			
		Intégration ou adaptation à l'environnement	Qualité, fiabilité et sécurité	Complexité et optimisation	Modélisation, représentation et génération
Besoins sociétaux	Accès à l'information et création de connaissances	A (1*B)	A, C	A, B, C	A (1*C)
	Services à la personne	A		B	A
	Santé, bien-être, environnement et patrimoine	A		B	B
	Transport, production et énergie : ville intelligente	A, B		B, A, C	B
	Sécurité, sûreté et éthique	C (1*A)	C (1*A)	C	C

Axes	Intitulé
A	Données massives, complexes ou hétérogènes
B	Algorithmique et optimisation
C	Sécurité, fiabilité et performances

FIGURE 3.1 – Tableau illustrant le croisement des axes thématiques avec les verrous scientifiques et les besoins sociétaux

DYNAMIQUE DE COOPÉRATION : ACTIONS, PROJETS, OUVERTURE INTERNATIONALE

Pour assurer une dynamique de coopération, la fédération veille et aide à la mise en place des actions. Ces actions correspondront à des collaborations sur de moyennes/longues durées, existant entre les deux laboratoires, impliquant des ensembles plus ou moins stables de personnes et dédiées à des sujets scientifiques plus précis. Même si une action est souvent inscrite dans un axe, elle peut aussi se constituer de la collaboration entre chercheurs de divers axes.

Pour 2016, neuf actions sont déclenchées. Il s'agit non seulement des travaux initiés récemment, mais aussi de la continuation des collaborations de longue date, parfois sur de nouvelles thématiques. Certaines actions seront appuyées par des projets (région, ANR, européens) soumis récemment (certains déjà retenus).

La liste des actions présentées aux JIRC'2015 est la suivante (voir annexe A.1 pour la description de chacune) :

1. Algorithmique exponentielle et paramétré en ordonnancement (axe B).
2. Hybridation des techniques de clustering et d'extraction de motifs (axe A).
3. Extraction de motifs séquentiels à grande échelle dans un environnement Hadoop/MapReduce (axes A et C).
4. Similarité de trajectoires et optimisation (axes A et C).
5. Masses de données dans une approche sémantique : consultation, maintenance, évolution (axe A).
6. Techniques d'apprentissage pour le TALN (axe A).
7. Analyse syntaxique (parsing) pour l'annotation de la temporalité (axe A).
8. Expressions polylexicales et parsing (axe A).
9. Éthique et risques des technologies numériques (axes A et C).

Dans chaque action , remarquons que nous citons seulement les permanents dont la participation est envisagée. Néanmoins, les actions comptent avec une importante participation de doctorants, ATER ou autres collaborateurs extérieurs à ICVL.

Le LI et le LIFO ont déjà été partenaires dans certains projets, mais la mise en place de ICVL contribue à l'augmentation de cette dynamique. Nous recensons les projets financés ci-dessous, résumés dans les annexes (annexe B.1), tous lancés dans l'esprit de collaboration de notre fédération :

1. Projet ANR Codex (2009-2012),
2. Réseau COST PARSEME (2013-2017),
3. Projet régional TEMPORAL (2013-2014),
4. Projet régional GIRAFON (2015-2017).
5. Projet ANR PARSEME-FR (2016-2020),
6. Projet régional ODIL (2016-2017).

Un certain nombre de projets soumis non sélectionnés (cf. annexe B.2) ou en cours d'évaluation témoignent également de force de proposition de la fédération :

1. Projet européen SHAGRALER (soumis en 2011, non sélectionné),
2. Projet régional ASEF (soumis en 2012, non sélectionné),
3. Projet régional ANNOTEP (soumis en 2014 et 2015, non sélectionné),
4. Projet européen I-RENNER (soumis en 2015, non sélectionné),
5. Projet ANR SEMORAL (soumis en 2014 et en 2015, en cours d'évaluation),

L'ICVL n'a pas vocation à rester fermée entre le LI et le LIFO, la collaboration avec d'autres laboratoires, d'autres universités françaises et étrangères est un des ses objectifs le plus cher. Nous souhaitons, justement, que l'ICVL soit force de proposition et de réussite dans les collaborations vers l'extérieur. Ainsi, un de nos buts est de faire un recensement détaillé de toutes les collaborations déjà existantes pour mieux comprendre et investir dans leur renforcement et extension. L'annexe B.2.8 liste les principales collaborations de portée internationale et nationale ainsi que d'autres facteurs de visibilité extérieure de nos laboratoires. Ces collaborations sont diverses allant d'un Master Erasmus Mundus avec 4 pays et un réseau COST avec 33 pays à plusieurs projets internationaux, bi- ou tri- latéraux, d'échange ou de collaboration avec l'Amérique, l'Europe, l'Asie et l'Afrique. Cela est sans compter avec les réseaux nationaux et des collaborations individuelles de longue date. Comme facteur de visibilité, remarquons aussi

l'organisation d'un bon nombre de manifestations (congrès ou ateliers) de portée internationale ou nationale. Les projets nationaux incluent des partenariats assez variés : des laboratoires en Région Parisienne, Lyon, Grenoble, Bordeaux, Nancy, etc. L'ICVL souhaite encourager encore plus toutes les collaborations, en provoquant d'avantage d'échanges entre les chercheurs extérieurs et ceux membres de l'ICVL. Nous comprenons que c'est seulement via des partenariats sérieux et productifs que la fédération atteindra la dynamique scientifique souhaitée.

ICVL : QUALITÉ ET DYNAMISME DANS LA RECHERCHE

Ce chapitre clôt notre projet scientifique en rappelant l'importance de la fédération ICVL et en énonçant des possibilités d'avenir. Il ouvre aussi la discussion sur différentes méthodes et politiques pour le développement de l'ICVL.

5.1 ICVL : motivation pour un plus grand investissement scientifique

L'ICVL a l'ambition de devenir un instrument de motivation et dynamisation scientifique des membres de deux laboratoires. Nous encouragerons une participation constante des individus ou des groupes aux activités et rencontres scientifiques mis en place. Cette dynamique a commencé en 2014 avec l'appel aux actions. Les résultats sont très encourageants. Le montage de journées thématiques et l'appui aux développements prévus par les actions seront les premières mesures entreprises pour atteindre notre objectif.

Le travail de construction du projet scientifique a été extrêmement utile et enrichissant, permettant une ouverture vers de nouveaux horizons qui, nous l'espérons, s'annonce très riche en perspectives scientifiques. Notre démarche participative invite chaque membre à s'appropriier le projet de la fédération et à encourager ou élargir toute collaboration.

L'ICVL est ainsi un outil important pour le dynamisme de nos laboratoires de recherche, suscitant et favorisant des projets de recherche au niveau régional, national et international. La fédération est une étape essentielle pour la formation d'un pôle de référence dans *l'acquisition et le traitement efficaces, fiables et sécurisés de l'information*.

5.2 ICVL : nouveaux horizons pour nos recherches

Rappelons ici des aspects qui, déjà traités dans ce projet, illustrent la plus-value scientifique de l'ICVL. Tout d'abord l'organisation de l'ICVL en trois axes est le résultat d'une communion de forces nécessaire pour mener à bien toutes nos collaborations. La puissance interne des axes consiste, majoritairement, à faire émerger des recherches, possibles uniquement grâce à la mise en commun des compétences des deux laboratoires. Leur puissance externe repose sur l'offre d'un savoir faire qui peut s'adapter ou proposer des solutions à d'autres axes. L'ensemble de toutes ces collaborations permettra sans doute de donner des réponses pluridisciplinaires aux besoins sociétaux entrelacés entre ces différentes thématiques (figure 3.1).

En particulier, par la complémentarité des compétences du LI et du LIFO, l'ICVL sera : un noyau dans les domaines de l'extraction de connaissances ; un important appui aux chercheurs TAL, déjà très interactifs ; un instrument essentiel pour le développement de nos recherches vers la manipulation et traitement efficace de grandes masses de données, éventuellement liées par des informations sémantiques et un chantier ouvert au rapprochement des aspects théoriques et appliqués dans le domaine de l'optimisation.

Ce constat nous mène à conclure que la fédération ICVL est fondamentale pour la formation d'un pôle d'excellence dans *l'acquisition et le traitement efficaces, fiables et sécurisés de l'information*. Elle sera un centre des recherche capable de produire des résultats fondamentaux et appliqués levant les verrous scientifiques : intégration et adaptation à l'environnement ; qualité, fiabilité et sécurité ; complexité et optimisation ; modélisation et représentation de l'information et génération de modèles.

5.3 ICVL : l'union vers des objectifs internationaux

L'ICVL a déjà atteint une masse critiques nécessaire pour être force de proposition à l'échelle internationale. Les deux laboratoires jouent des rôles clefs dans une action européenne COST (cf. B.1.2) et ils ont été les initiateurs ou membres de deux propositions de projets européens (cf. B.2.4 et B.2.3). Ils ont accueilli des conférences et colloques internationaux à Blois et Orléans (CIAA-FSMNLP

2011¹, NWC 2011², UCNC 2012³, CSLP 2012⁴, FraDeCoP 2012 et 2013⁵, AM-LaGAP 2014⁶, WG 2014⁷, PaPDAS 2014⁸, CPDM 2014⁹, ETC 2015¹⁰). Le LI est partenaire d'un programme de Master Erasmus Mundus IT4BI (cf. B.1.1) en 2012-2018, dont le renouvellement sera demandé pour une nouvelle période de 6 ans. Le LIFO est informellement associé au LI dans la mise en place de ce cursus.

Dans le cadre de la fédération nous prévoyons la mise en place de mesures permettant encore d'augmenter la visibilité de nos recherches à l'échelle internationale. La définition de grands axes transversaux, décrite dans la section 3, en est le fondement. Elle permet la réflexion sur les expertises prioritaires à afficher lors du montage de nouveaux projets internationaux et des recherches de collaborations.

5.4 ICVL : vers une structuration transversale.

L'ICVL se propose aujourd'hui comme une fédération regroupant des forces pour valoriser et améliorer la visibilité et le dynamisme des recherches en informatique en Région Centre Val de Loire. Néanmoins, elle est aussi un banc d'essai vers d'autres possibilités de structuration. Par exemple, au moment de la rédaction de ce projet, l'ICVL a reçu de la part des deux laboratoires membres une demande de mener une étude sur une fusion LIFO/LI. L'ICVL donne donc une opportunité d'observer les possibilités de convergence de nos deux laboratoires vers des objectifs communs, pas exclusivement scientifiques. Une ouverture éventuelle vers d'autres laboratoires de la région ou de la COMUE dépendra aussi de la synergie développée par nos recherches conjointes au sein de l'ICVL. Cette fédération est donc une étape importante à vivre avant de répondre aux questions d'extension souvent avancées par nos tutelles.

Face à ces possibles scénarios, l'ICVL devra rester ouverte, lancer des débats, proposer, écouter, analyser diverses pistes, toujours dans le but d'améliorer la pro-

-
1. <http://ciaa-fsmnlp-2011.univ-tours.fr>
 2. <http://www.univ-orleans.fr/lifo/evenements/NWC2011/>
 3. <http://www.univ-orleans.fr/lifo/events/UCNC2012/>
 4. <http://www.univ-orleans.fr/lifo/events/CSLP2012/>
 5. <https://traclifo.univ-orleans.fr/PaPDAS/wiki/FraDeCoPP-2>
 6. <https://sites.google.com/site/amlagap/>
 7. <http://www.univ-orleans.fr/lifo/events/WG2014/>
 8. <http://traclifo.univ-orleans.fr/PaPDAS/wiki/workshop>
 9. <https://sites.google.com/site/journeecpdm/>
 10. <https://sites.google.com/site/etclustering/>

duction et la collaboration scientifique de ses membres, ainsi que de contribuer le plus activement possible à la formation à la recherche.

Comme officine d'étude d'une liaison plus étroite, l'ICVL pourra non seulement renforcer des mesures déjà en vogue, telles que l'encouragement et le soutien aux actions, aux projets et aux encadrements conjoints ; mais aussi ouvrir la discussion vers des ambitions plus controversées comme le montage d'équipes inter-laboratoires ou la mise à disposition, périodique et alternée, par les laboratoires, d'une allocation de recherche à un projet de thèse Orléans-Tours.

DÉTAILS SUR LES ACTIONS

A.1 Actions 2016

A.1.1 Algorithmique exponentielle et paramétrée en ordonnancement

Permanents dont la participation est envisagée

Christophe Lenté (LI-OC), Mathieu Liedloff (LIFO-GAMOC), Ameer Soukhal (LI-OC), Vincent T'kindt (LI-OC).

Sous-axe concerné : **B1**

Thématiques et objectifs

Sur un plan scientifique, un algorithme exponentiel est un algorithme exact résolvant un problème NP-difficile et dont la complexité (temporelle et/ou spatiale) au pire cas peut être calculée. Jusqu'aux années 2010, très peu de résultats étaient connus dans la littérature portant sur la théorie de l'ordonnancement : il existait des algorithmes exponentiels, mais pour des problèmes de graphe ou de décision. Nous avons proposés plusieurs algorithmes exponentiels pour des problèmes d'ordonnancement et enrichis ainsi, significativement, la littérature.

Historique des collaborations

Cette action s'inscrit dans le cadre d'une collaboration entre des chercheurs du LI et du LIFO depuis 2009. Plusieurs travaux de recherche autour de la mise au point d'algorithmes exponentiels en ordonnancement ont été développés et publiés dans des conférences et revues internationales.

Proposition de travail

Nous continuons à collaborer sur ce sujet et les perspectives sont nombreuses. Nous les détaillons selon deux axes :

- Mise au point d’algorithmes exponentiels pour des problèmes d’ordonnement de permutation. Ce travail s’articule autour de la thèse de doctorat de Lei Shang (LI, équipe OC) et est conduit en collaboration avec M. Liedloff du LIFO.
- Mise au point d’algorithmes paramétrés pour des problèmes d’ordonnement. Cette piste, complètement nouvelle vient en complément de la première. Il s’agira ici d’établir des algorithmes exactes dont la complexité au pire cas est fonction, non plus de la taille des instances, mais de la valeur d’un paramètre de l’instance.

A.1.2 Hybridation des techniques de clustering et d’extraction de motifs

Permanents dont la participation est envisagée

Sylvie Billot (LIFO-CA), Guillaume Cleuziou (LIFO-CA), Thi-Bich-Hanh Dao (LIFO-CA), Matthieu Exbrayat (LIFO-CA), Arnaud Giacometti (LI-BdTln), Nicolas Labroche (LI-BdTln), Arnaud Soulet (LI-BdTln), Marcilio de Souto (LIFO-CA), Christel Vrain (LIFO-CA)

Sous-axe concerné : A1

Thématiques et objectifs

L’objectif de cette action est une fertilisation croisée entre les techniques de clustering et celles d’extraction de motifs. Ces deux thématiques fortes en fouille de données se sont largement influencées ces dernières années dans la littérature. Au sein de la fédération ICVL, nous souhaitons mettre en place une collaboration nouvelle autour de l’hybridation de ces méthodes aussi bien de l’extraction de motifs vers le clustering que du clustering vers l’extraction de motifs pour conduire à de nouveaux algorithmes pour la fouille de données.

Historique des collaborations

Cette action s'inscrit dans une volonté commune d'initier une nouvelle thématique de recherche reposant sur les compétences du LI et du LIFO en clustering et recherche de motifs.

Proposition de travail

Les travaux portant sur l'interaction entre le clustering et l'extraction de motifs sont déjà présents au sein de la littérature. Nous proposons de les étendre selon deux directions complémentaires en observant tout d'abord ce que l'extraction de motifs peut apporter au clustering et ensuite ce que le clustering peut apporter à l'extraction de motifs.

Extraction de motifs pour le clustering Cette approche consiste soit à réutiliser les techniques algorithmiques pour extraire des clustering dans des sous-espaces distincts (subspace clustering), soit à exploiter directement les motifs extraits pour construire un clustering (clustering conceptuel).

Lorsque le nombre de dimensions décrivant les données devient très grand la qualité du clustering a souvent tendance à se dégrader. Dans ce contexte, l'idée clé du subspace *clustering* est de découvrir des *clustering* alternatifs en considérant des sous-espaces alternatifs. La recherche de ces sous-espaces alternatifs repose sur les techniques d'énumération de motifs correspondants aux dimensions de l'espace des données. Un autre apport des méthodes d'extraction de motifs pour le *clustering* est la combinaison de motifs pour construire une partition (parfois, avec recouvrement) des données. Ces approches de *clustering* conceptuel (aussi appelées, *bi-clustering* ou *co-clustering*) ont l'avantage de fournir une description en intention des clusters. Nos travaux dans ce cadre porteront sur les aspects sémantiques, visant une description efficace des clusters à partir des sous-espaces, mais également sur des aspects liés à la pertinence des méthodes de subspace *clustering* : comment évaluer l'intérêt d'un sous-espace et comment déterminer le nombre idéal de dimensions pour chaque cluster.

Nous explorerons enfin dans cette action une troisième voie autour du *clustering* multi-vues et multi-objectifs hybride : dans cette problématique nous considérerons un même jeu de données selon les deux aspects (vues ou objectifs) numériques et conceptuels dans le but de faire émerger de nouvelles solutions (voire des ensembles de solutions) réalisant un consensus sur les objectifs à la fois numériques (inertie ou vraisemblance) et conceptuels (définition intensionnelle et motif clos).

Enfin, dans le contexte actuel des grandes masses de données, les méthodes proposées devront être efficaces et pour le moins s’affranchir des problèmes classiques d’explosion combinatoire en recherche de motifs liés à l’exploration exhaustive des sous-ensembles de dimensions.

***clustering* pour l’extraction de motifs** À l’inverse, il est envisageable de tirer profit des approches de *clustering* pour améliorer les méthodes d’extraction de motifs. Cette voie a été moins explorée dans la littérature et souvent, le *clustering* est utilisé en simple post-traitement de l’extraction de motifs. Typiquement, les techniques de *clustering* sont utilisées pour regrouper les motifs similaires afin de réduire leurs redondances intrinsèques à l’approche. Cependant, nous pensons que le *clustering* pourrait être intégré plus en amont. Certaines propriétés du *clustering* devraient être intégrées dans le processus d’extraction de motifs pour en améliorer son fonctionnement. Par exemple, les méthodes d’extraction de motifs peu efficaces sur les données numériques pourraient exploiter une métrique à l’instar des méthodes de *clustering*.

Les thématiques de l’extraction de motifs et du *clustering*, présentes au sein de l’ICVL, sont propices à une collaboration nouvelle et immédiate. Plus particulièrement, nous souhaitons avancer sur les points suivants :

- Intégration d’une vue *conceptuelle* dans un processus de *clustering* multi-vues afin de faire émerger de nouvelles solutions consensuelles
- Intégration de la notion de métrique dans le support afin d’extraire des motifs dans des données numériques
- Amélioration de la pertinence et de la sémantique des approches de *sub-space clustering*
- Étude des interactions entre le *clustering* et d’autres formes de motifs

A.1.3 Extraction de motifs séquentiels à grande échelle dans un environnement Hadoop / MapReduce

Permanents dont la participation est envisagée

Mostafa Bamha (LIFO-PAMDA), Ali Ed-Dbali (LIFO-LMV), Arnaud Giacometti (LI-BdTln), Dominique Li (LI-BdTln), Patrick Martineau (LI-OC), Arnaud Soulet (LI-BdTln)

Sous-axes concernés : A1, B2, C2

Thématiques et objectifs

Depuis 25 ans, l'extraction de motifs intéressants joue un rôle très important dans le domaine de la fouille de données. Cependant celle-ci reste confrontée à des temps de traitement qui peuvent être importants sur de grandes masses de données. Dans ce contexte, différents algorithmes d'extraction de motifs fréquents ont été implantés dans un environnement Hadoop / MapReduce ou Spark. Néanmoins, la plupart des travaux existants portent sur l'extraction de motifs ensemblistes (itemsets) et très peu de travaux existent sur l'extraction de motifs plus complexes (séquences, graphes, etc.).

Dans ce cadre, l'objectif de cette action est de proposer de nouveaux algorithmes et outils pour l'extraction de motifs séquentiels sur de grandes masses de données en s'appuyant sur une complémentarité forte entre différents équipes de la fédération ICVL, à savoir :

- une expertise de l'équipe BdTln du LI dans le domaine de l'extraction de motifs locaux et de leur optimisation (mais dans un environnement non parallèle) ;
- une expertise des équipes PAMDA et LMV du LIFO et OC du LI dans les domaines du parallélisme et du calcul distribué, en particulier des plateformes de développement de type Hadoop / MapReduce.

Historique des collaborations

- Collaboration existante au LI entre les équipes BDTLN et OC sur cette thématique.
- Dépôt de projets : H2020 LASCAR (LI) et APR-IA GIRAFON (LI+LIFO, cf. B.1.7).

Proposition de travail

Dans un premier temps, les travaux vont se concentrer sur la conception et la preuve formelle de nouveaux algorithmes pour l'extraction de motifs séquentiels sur de grandes masses de données. Ensuite, des benchmarks devront être préparés pour évaluer et comparer les performances des différents algorithmes proposés, mais aussi comparer les forces et faiblesses des différentes variantes existantes du modèle Hadoop / MapReduce.

Enfin, les compétences de l'équipe OC devront permettre d'identifier comment pourraient être améliorés les mécanismes de répartition de charge du modèle Hadoop / MapReduce et de ses variantes, mais aussi si de nouvelles variantes pourraient être développées face aux contraintes particulières de l'extraction de motifs (la taille potentiellement exponentielle des espaces de recherche à parcourir).

Prospectives

L'action se concentrera en priorité sur l'extraction de motifs séquentiels. Mais elle pourrait ensuite être étendue au traitement et à l'exploration d'autres types de données ou à d'autres domaines d'application, telle que l'analyse et la fouille de grands graphes. Enfin, à moyen terme, les exigences de généricité et d'efficacité imposent le recours à des langages de requêtes pour la préparation des données à analyser. Dans ce cadre, des interactions avec l'action présentée dans la section A.1.5 pourraient être développées afin d'identifier les primitives nécessaires dans le développement d'un langage de requêtes pour l'interrogation et l'analyse des données distribuées.

A.1.4 Similarité de trajectoires et optimisation

Permanents dont la participation est envisagée

Thomas Devogele (LI-BdTln), Laurent Etienne (LI-BdTln), Emmanuel Melin (LIFO-PAMDA), Sophie Robert (LIFO-PAMDA)

Sous-axes concernés : A1, C3

Thématiques et objectifs

Afin de pouvoir traiter les masses de données issues des capteurs GPS, il est indispensable de disposer de mesures de similarité entre des trajectoires s'exécutant dans des temps raisonnables. Or, ce calcul de similarité est coûteux en termes d'extraction des bases de données des trajectoires et de calcul matriciel. Il est donc nécessaire de travailler à l'optimisation, la parallélisation des calculs et la distribution des données.

Historique des collaborations

Il s'agit d'une nouvelle collaboration.

Proposition de travail

L'équipe BdTln a une expertise sur la manipulation de gros volume de trajectoires d'objets mobiles. Elle se heurte à des problèmes de coût de calcul important lié à l'extraction des données issu de SGBD spatiale et à des calculs matriciels lourds. L'équipe PAMDA a une expertise la parallélisation de la visualisation de données géographiques. L'objectif de cette action est d'étudier et de réaliser un prototype de calcul parallèle de similarité de trajectoire.

Prospectives

Une première étude sera menée en janvier 2016 via un projet interne master 1 à Orléans (projet TER).

A.1.5 Masses de données dans une approche sémantique : consultation, maintenance, évolution

Thématiques et objectifs

Permanents dont la participation est envisagée

Mostafa Bamha (LIFO-PAMDA), Béatrice Bouchou-Markhoff (LI-BdTln), Jaques Chabin (LIFO-PAMDA), Mirian Halfeld Ferrari Alves (LIFO-PAMDA)

Sous-axes concernés : A2, C3

Les propositions concernant les langages de requêtes sur des grandes masses de données sont nombreuses, mais des questions concernant le pouvoir d'expression, l'efficacité de l'évaluation, et l'assurance de la cohérence des réponses obtenues sont encore source de débat. Plusieurs questions concernant la manipulation, la maintenance ou l'interrogation d'un grand volume de données distribuées restent ouvertes : de la représentation et du choix de langages aux techniques de partage et équilibrage dynamiques des données ou de la charge de travail.

Dans le cadre des masses de données du web sémantique, appelées *Linked Data* (ou *Global Giant Graph*), les requêtes sur ces données prennent naturellement en compte leur sémantique. Dans ce contexte l'action s'intéresse à proposer des outils pour l'interrogation efficace (comportant des opérations d'agrégations ou autres) et déclarative de grands volumes de données distribuées organisées via un schéma contenant des informations sémantiques. Des contraintes d'intégrité et la confiance selon la provenance sont à prendre en compte. L'action s'intéresse aussi à l'évolution et l'intégration des contraintes de schéma et d'intégrité.

Toutes ces directions font le sujet d'une recherche en deux niveaux : front-end (langages de requêtes, algorithmes pour cohérence et intégration, etc) et back-end (algorithmes parallèles pour des requêtes complexes, distribution de données en fonction des types de requêtes plus fréquents).

Historique des collaborations

Des collaborations dans le passé dans le domaine de XML (mise à jour, validation de contraintes) et actuellement dans le cadre de la thèse Orléans-Tours.

Proposition de travail

Les collaborations ont déjà commencé à travers :

- Thèse Orléans-Tours : proposition d'un langage d'interrogation type datalog sur des données distribuées, algorithmes d'évaluation de requêtes prenant en compte la cohérence et la confiance de données. Participation au projet APR-IA, soumis récemment, et au projet européen H2020 I-RENNER (cf. B.2.3).
- Des perspectives de collaborations internationales existent avec, en particulier, le Brésil.
- Les premiers travaux se concentrent sur la question d'interrogation d'un graphe de données distribuées contraint par des règles d'intégrité qui doivent être vérifiées dans la production des résultats.

Prospectives

Des interactions sont déjà envisageables dans deux cadres :

- Création d'un langage de requêtes comme outil pour la préparation des données à analyser via des algorithmes de data mining, essentiels dans la découverte de connaissances sur de grands volumes données. Une interaction avec les actions concernant le data mining est donc nécessaire dans cette perspective.
- Efficacité des algorithmes d'évaluation des requêtes basés sur des environnements d'exécution distribués accompagnés des bibliothèques de programmation parallèle. L'interaction avec les collègues travaillant sur des algorithmes parallèles est donc souhaitable.

A.1.6 Techniques d'apprentissage pour le TALN

Permanents dont la participation est envisagée

Jean-Yves Antoine (LI-BdTIn), Denys Duchier (LIFO-CA), Nicolas Labroche (LI-BdTIn), Yannick Parmentier (LIFO-CA), Agata Savary (LI-BdTIn)

Sous-axes concernés : **A1, A2**

Thématiques et objectifs

L'équipe CA du LIFO développe des recherches sur des techniques d'apprentissage automatique qui sont de plus en plus utilisées en TAL. De son côté, certains chercheurs en TAL de l'équipe BdTIn s'intéressent également à l'utilisation de ces techniques, sans en faire un axe de recherche spécifique.

Les chercheurs en fouille de données et classification menées au sein de l'équipe BdTIn du LI ont par ailleurs également travaillé sur des données textuelles (thèse en fouille de texte de Damien Nouvel, collaboration actuelle avec l'entreprise Eloquenz sur la détection d'auteurs sur blogs). Une action LI/LIFO permettra d'approfondir ces recherches sur les techniques de TAL centrées sur les données, et ce d'autant plus qu'un des domaines d'excellence de l'équipe BdTIn réside dans la constitution de corpus qui servent précisément de données d'apprentissage dans le domaine.

Du point de vue du LIFO, l'intérêt d'un travail avec le LI sur le sujet réside avant tout sur l'étude de l'adaptation des modèles appris à un autre type de données (passage du langage écrit au langage oral dans notre cas) et de réfléchir aux liens entre traits d'apprentissage pertinents pour la linguistique et ceux qui font sens pour les techniques de classification. Plusieurs domaines applicatifs développés dans nos laboratoires peuvent être concernés par cette thématique : détection des entités nommées, détection d'auteurs, résolution de la coréférence.

Historique des collaborations

À ce jour, il n'existe pas de collaboration entre le LI et le LIFO sur ce sujet spécifique, mais les chercheurs en TAL de deux laboratoires ont beaucoup de contact. Collaboration avec le laboratoire LaTTiCe (Isabelle Tellier, ancien membre LIFO, et Frédéric Landragin),.

Proposition de travail

Le travail prévu dans le cadre de cette action pour 2015 se focalise donc sur la résolution des coréférences. En collaboration avec le laboratoire LLL (porteurs : Jean-Yves Antoine pour le LI, Emmanuel Schang pour le LLL), nous avons constitué dans le cadre du projet régional ANCOR le plus grand corpus de parole transcrite annoté en relations de coréférence. Ce corpus a déjà servi de base d'apprentissage au laboratoire LaTTiCe (Isabelle Tellier, Frédéric Landragin) pour l'apprentissage d'un petit système de résolution à base de classifieurs multiples (CROC) réalisé en utilisant la plate-forme Weka. Weka est une plate-forme générique intégrant un très grand nombre de techniques d'apprentissage pour la classification. Dans le cadre de cette action, nous souhaiterions utiliser une plate-forme dédiée spécifiquement à la résolution des co-référence (BART), qui intègre également plusieurs modèles d'apprentissage et que le LI est en train d'adapter au français. L'idée est ensuite de comparer les résultats obtenus à ceux du système CROC.

D'un point de vue scientifique, cette action revêt plusieurs objectifs :

- Comparaison et optimisation des différentes techniques de classification proposées par BART,
- Etude des relations entre traits linguistiques observés en corpus et traits utiles à l'apprentissage,
- Etude éventuelle des questions d'adaptation de modèles d'un registre oral à un autre (interviews présentes dans le sous-corpus ANCOR-ESLO et parole

très interactive présente dans les sous-corpus ANCOR-OTG et ANCOR-UBS).

- D'un point de vue budgétaire, cette action repose sur une demande de financement de stage de Master. Le système CROC a été précisément développé dans le même cadre (stage de Master d'Adèle Desoyer) et sur un sujet très proche, ce qui un indicateur de réussite de cette action.

Prospectives

Deux perspectives de recherche à plus long terme sont clairement visées par cette action : le rapprochement de nos recherches en classification et plus généralement en apprentissage. Dans le cadre d'une application au TAL, nous avons déjà identifié les domaines applicatifs suivants :

- détection des entités nommées ;
- détection d'auteur.

Dans le cas du travail spécifiquement prévu sur la résolution des coréférences, la question du dépôt d'un projet ANR sur la question en collaborations avec le LaTTiCe est clairement envisageable. Ceci dépendra de la réussite du LaTTiCe d'un dépôt de projet proche cette année mais avec d'autres partenaires (le LI est associé à cette demande comme sous-traitant sur la partie Corpus) : projet DEMOCRAT actuellement en seconde phase d'évaluation.

A.1.7 Analyse syntaxique (parsing) pour l'annotation de la temporalité

Permanents dont la participation est envisagée

Jean-Yves Antoine (LI-BdTIn), Denys Duchier (LIFO-CA), Yannick Parmentier (LIFO-CA), Agata Savary (LI-BdTIn)

Sous-axe concerné : A3

Thématiques et objectifs

Cette action a pour objectif de poursuivre un travail déjà entamé sur la constitution d'un corpus de parole transcrite enrichi par une annotation en arbres syntaxiques. Ceci dans la perspective d'une annotation ultérieure en temporalité, pour

laquelle nous avons besoin d'un tel corpus arboré (treebank). Elle mobilisera pour cela l'expérience du LI (équipe BdTln) en matière d'analyse et de représentation de la temporalité, ainsi que celle du LIFO (équipe CA) en matière de parsing syntaxique.

Les thématiques de recherche proposées ici font actuellement l'objet de deux soumissions de projet :

1. Dépôt d'une demande de post-doctorat centrée sur l'annotation en temporalité dans le cadre de l'APR-IA de la région centre.
2. Participation du LIFO (et à un degré moindre du LI) à une demande de financement de projet portée par le LLL dans le cadre de l'APR-IA. Cette demande est liée au corpus ESLO et comportera une sous-tâche dédiée à l'annotation en arbres syntaxiques du corpus. Le LI interviendrait précisément sur la question de l'utilisation en temporalité.

Cette proposition d'action a pour objet de formaliser ces collaborations. Dans le cas où l'un des projets mentionnés ci-dessus serait accepté, l'action proposée à la fédération ne ferait plus l'objet de demande de financement, mais plutôt de labélisation.

Historique des collaborations

Ce travail a été initié dans le cadre du projet TEMPORAL (cf. B.1.8), une action de recherche soutenu par le MSH Val de Loire sur l'annotation de la temporalité. Ce petit projet collaboratif a regroupé dans un premier temps des chercheurs du LI et du LLL. Il a été porté par Emmanuel Schang (LLL) et Jean-Yves Antoine (LI) puis piloté scientifiquement par Anaïs Lefeuvre (LI). Les travaux des différents participants à l'action ont consisté à réfléchir à l'adaptation d'une norme ISO de représentation de la temporalité, TimeML, à la parole transcrite. Elle a conduit à l'issue du projet à la proposition d'un enrichissement de la norme TimeML consistant principalement à annoter la temporalité non plus au niveau des mots (unité lexicales) mais des nœuds d'arbres syntaxiques construits sur le corpus. Cette proposition a été validée par une communication dans TALN'2014.

Anaïs Lefeuvre et Jean-Yves Antoine ont par ailleurs rejoint le groupe de normalisation de l'AFNOR miroir du groupe ISO TC37/SC4 pour précisément porter cette demande de modification au niveau international. Cette évolution pose la question de la constitution de corpus arborés de parole transcrite. Le LIFO et le LI travaillant précisant de concert sur la question du parsing (thèse de Jakub Waszcuk co-encadrée par Yannick Parmentier pour le LI et Agata Savary pour

le LIFO ; pilotage par les mêmes encadrants de l'action COST PARSEME, cf. B.1.2), les laboratoires LI, LIFO et LLL ont décidé de poursuivre la collaboration initiée par TEMPORAL sur la temporalité en y intégrant la question de l'annotation syntaxique nécessitée par notre proposition de format de représentation alternatif à TimeML. Le LIFO a ainsi déjà participé à une réunion scientifique post-TEMPORAL.

Proposition de travail

Comme précisé plus haut, cette action de recherche fait l'objet de deux dépôts de demande de financement sur les deux problématiques scientifiques que soulève cette collaboration : l'annotation syntaxique de corpus de parole transcrite d'une part, et son utilisation pour la représentation de la temporalité d'autre part.

L'action que nous proposons dans le cadre de la fédération a pour objectif d'assurer la réalisation d'une avancée minimale mais absolument nécessaire à la poursuite de la réflexion dans le cas d'une non-acceptation sur les appels à projets précités : l'annotation en arbre syntaxique de corpus de parole transcrite. Elle consistera donc au financement d'un stage de Master centré sur l'adaptation au langage oral de parseurs syntaxiques ainsi qu'à la production d'un petit corpus pilote arboré grâce à ces outils (annotation automatique puis révision manuelle). Cette étude sera menée au choix sur les parseurs développés par le LIFO ou sur l'adaptation de parseurs disponibles en open source.

Prospectives

Deux perspectives de recherche à plus long terme sont clairement visées par cette action :

- annotation de la temporalité avec un paradigme de représentation enrichi ;
- constitution d'un large corpus de parole annoté en arbre syntaxique.

Cette action sur l'analyse syntaxique entre en interaction avec l'action *Expressions polylexicales et parsing* (cf. A.1.8) portée par les équipes BdTln du LI et CA du LIFO.

A.1.8 Expressions polylexicales et parsing

Permanents dont la participation est envisagée

Yannick Parmentier (LIFO-CA), Agata Savary (LI-BdTIn), Denis Mau-
rel (LI-BdTIn), Jean-Yves Antoine (LI-BdTIn), Béatrice Bouchou-
Markhoff (LI-BdTIn), Denys Duchier (LIFO-CA), Nicolas Labroche
(LI-BdTIn), Nathalie Friburger (LI-BdTIn)

Sous-axe concerné : **A3**

Thématiques et objectifs

Les expressions polylexicales (e.g. merle blanc, Banque centrale européenne, chérir l'espoir, prendre le taureau par les cornes) sont l'un des défis majeur dans le traitement automatique des langues, et plus particulièrement pour le parsing syntaxique et sémantique, car elles possèdent des propriétés imprévisibles à différents niveau de traitement linguistique (lexique, morphologie, syntaxe, sémantique, pragmatique). Les objectifs de cette action de la fédération ICVL est inscrite pleinement dans le réseau européen COST IC1207 PARSEME (cf. B.1.2), et vise la prise en compte des expression polylexicales dans le parsing profond, dans un contexte multilingue. D'autre part, nous souhaitons établir des liens entre le parsing des expressions polylexicales et d'autres compétences présentes au sein des équipes de la fédération. Ainsi, deux sous-objectifs (exprimés notamment dans la proposition de projet ANR PARSEME-FR, cf. B.1.3) concernent :

- le rapprochement du domaine de la reconnaissance d'entités nommées polylexicales à celui du parsing profond,
- le rapprochement du domaine de entity linking et Linked Open Data (incluant beaucoup d'entités polylexicales) au parsing syntaxique et sémantique.

Historique des collaborations

- Action COST IC1207 PARSEME (www.parseme.eu) – en cours depuis 2013 (jusqu'à 2017), réseau de 180 membres de 30 pays ; Yannick Parmentier et Agata Savary sont à l'origine de cette action. Yannick, Jakub et Agata sont membres de son Steering Committee.
- Thèse de doctorat de Jakub Waszczuk - co-encadrée par Yannick Parmentier et Agata Savary. Participation de Denys Duchier et Nicolas Labroche au

comité de pilotage de cette thèse.

- Trois dépôts de projets :
 - PARSEME-FR (cf. B.1.3)- proposition de projet ANR, le spin-off français de l'action COST IC1207 PARSEME, démarré en 2016,
 - SHAGRALER (cf. B.2.4) – proposition de projet européen STREP (appel FP7-ICT- 2011-SME-DCL), soumise en 2011 (non sélectionnée),
 - ASEF (cf. B.2.7) – proposition de projet régional de recherche d'initiative académique (APR-IA 2012), soumise en 2012 (non sélectionnée).

Proposition de travail

- Poursuite du co-encadrement de la thèse de Jakub, dédiée au parsing symbolique des expressions polylexicales en français (et polonais).
- Poursuite de la coordination du réseau COST IC1207 PARSEME (avec 3-4 réunions du réseau par an).
- Collaboration étroite avec d'autres chercheurs français, experts du parsing probabiliste et symbolique (laboratoires ALPAGE Paris, LIF Aix-Marseille, LIGM Paris-Est), si le projet ANR PARSEME-FR est sélectionné pour un financement.

Prospectives

Interactions possibles avec d'autres actions : action TAL parsing et apprentissage automatique.

A.1.9 Éthique et risques des technologies numériques

Permanents dont la participation est envisagée

Christian Toinard (LIFO-SDS), Jean-Yves Antoine (LI-BdTln), Pascal Berthomé (LIFO-SDS), Yannick Kergosien (LI-OC), Patrick Martineau (LI-OC), Benjamin Nguyen (LIFO-SDS),
Sous-axes concernés : A3, B2, C1

Thématiques et objectifs

Cette action a pour objectif d'entamer un travail de réflexion et d'expérimentation éthiques sur les risques potentiellement liés à l'usage et au développement des nouvelles technologies. Elle s'appuiera sur une typologique de risques qui a déjà été élaborée dans le cadre d'une collaboration entre certains participants à l'action, mais également des chercheurs extérieurs. Cette réflexion éthique a déjà été menée entre les participants dans deux domaines de recherche principaux :

- aides techniques aux personnes handicapées et dépendantes ;
- traitement Automatique des Langues Naturelles.

Nous aimerions étendre cette analyse à la question du respect de la vie privée liées aux systèmes d'information liés à la santé, aux réseaux et informatique distribuée. En ce sens, l'équipe SDS du LIFO dépose actuellement un projet ANR multidisciplinaires avec des juristes et des économistes.

Cette action a toutefois pour ambition à terme de couvrir tous les champs de recherche de nos deux laboratoires.

Historique des collaborations

Ce travail a été initié dans le domaine de l'aide au handicap par Christian Toinard (LIFO), Anaïs Lefeuvre (ATER au LI en 2014-2015) et Jean-Yves Antoine (LI). Il a donné lieu à publication commune et à la participation au RTR (Réseau Thématique Régional) Risques porté par le LIFO (cf. projet B.1.10). Ce RTR se focalise sur le domaine du handicap, qui ne fait pour cette raison par l'objet de demande dans la cadre de cette action. Cette dernière doit précisément servir à diffuser et généraliser le travail réalisé au sein de nos deux laboratoires.

Proposition de travail

Cette action concerne l'analyse éthique conséquentialiste des applications actuellement menées au sein du LIFO Bourges dans le domaine des assistances à la personne dépendante et la protection de la vie privée. L'idée est de faire conduire, sous la forme d'un comité éthique ad hoc, une analyse en termes de facteurs de risque, vulnérabilité et criticité des applications développées par le LIFO. Elle impliquera des collègues des deux laboratoires, une démarche éthique ne pouvant être conduite par les seuls chercheurs développant la recherche concernée.

- Analyse en facteur de risque et vulnérabilité – réflexion éthique basée sur la typologie définie par (Lefeuvre et al. 2015) et consultation d'experts.

L'action nécessitera donc la mise en place de réunion avec frais de mission afférents.

- Analyse en criticité – Une analyse en criticité nécessite de mesurer des études expérimentales en termes de sévérité d'impact et de prévalence des technologies concernées. Elle nécessite le recrutement d'un stagiaire qui mènera précisément ce travail sous la responsabilité du comité éthique ad hoc.

Prospectives

Cette action, de nature transversale, porte une demande de réflexion épistémologique sur nos actions de recherche. Elle pourrait donner lieu à la mise en place d'un comité éthique local qui serait par exemple saisi pour avis consultatif sur toute réponse de projet, demande de bourse doctorale etc. proposée au sein de la fédération, tout en soutenant des recherches spécifiques dans le domaine ainsi qu'une réflexion éthique sur certains thèmes (par exemple, dans le cadre des travaux pour l'aide au handicap, risque sociaux, etc).

DÉTAILS SUR LES PROJETS

En plus des actions présentées dans la section précédente, des projets internationaux, nationaux et régionaux incluant les laboratoires LIFO et LI témoignent également de la dynamique croissante de leur collaboration. Dans cette section nous présentons quelques uns de ces projets, passés, courant et soumis. Il s'agit en particulier des initiatives qui apportent un effet de défragmentation (inter-labo ou inter-axe) au sein de la fédération.

B.1 Projets passés et en cours

B.1.1 Information Technologies for Business Intelligence (IT4BI)

Fiche du projet

Statut : en cours

Dates : 2012-2018

Consortium : Université Libre de Bruxelles (Belgique), Université François Rabelais Tours (LI), Universitat Politècnica de Catalunya (Espagne), Ecole CentraleSupélec (France), Technische Universität Berlin (Allemagne) ; participation informelle du LIFO

Coordinateur : Esteban Zimányi (ULB)

Financement : programme européen Erasmus Mundus

Permanents impliqués

Béatrice Bouchou-Markhoff (LI-BdTIn), Thomas Devogele (LI-BdTIn), Arnaud Giacometti (LI-BdTIn), Mirian Halfeld Ferrari (LIFO-Pamda), Nicolas Labroche (LI-BdTIn), Dominique Li (LI-BdTIn), Patrick Marcel (LI-BdTIn), Nizar Messai (LI-BdTIn), Verónika Peralta (LI-BdTIn), Agata Savary (LI-BdTIn), Arnaud Soulet (LI-BdTIn)

Sous-axes concernés : **A1, A2, A3**

Il s'agit d'une offre de formation au niveau Master, et un réseau européen d'universités et entreprises qui l'implémentent, financé par la Commission Européenne dans le cadre du programme Erasmus Mundus, visant à améliorer la qualité de l'enseignement supérieur européen et à renforcer la compréhension interculturelle. Il s'adresse à des étudiants du monde entier, en offrant un système de bourses européennes, mais aussi des inscriptions auto-financées.

Les objectifs de IT4BI incluent la formation de cadres européennes et mondiales dans le domaine de l'aide à la décision, basée sur des acquis de recherche et innovation dans les domaines d'entrepôts de données, fouille de données, analyse de contenu, gestion des processus métiers, analyse visuelle de données, systèmes d'information, science de données, science de décision et génie logiciel. Le coordinateur local de ce programme est Patrick Marcel.

Tous les ans, le LI à accueille à Blois une promotion de plus d'une vingtaine d'étudiants étrangers dans ce cadre, pour leur deuxième semestre de la 1^{re} année du Master. Le LIFO est informellement associé à ce programme par la participation aux enseignements liés aux spécialités de recherche communes aux deux laboratoires.

Le renouvellement de ce programme sera bientôt demandé pour une nouvelle période de 6 ans et prévoit une extension de la participation des membres du LIFO dans la mise en place des programmes d'enseignement.

B.1.2 PARsing and Multi-word Expressions (PARSEME)

Fiche du projet

Statut : en cours
Dates : 2013-2017
Consortium : 30 pays, 200 membres
Coordinateur : Agata Savary (LI)
Financement : réseau européen COST

Permanents impliqués

Yannick Parmentier (LIFO-CA), Agata Savary (LI-BdTln)
Sous-axe concerné : A3

PARSEME¹ est un réseau scientifique européen dans le domaine de traitement automatique des langues, axé sur les liens entre les expressions polylexicales (idiomes, mots composés, entités nommées polylexicales, etc.) et l'analyse syntaxique (parsing). Il est financé par l'action COST IC1207 et ses objectifs incluent : (i) de mettre le multilinguisme au cœur d'études linguistiques et technologiques, (ii) d'établir un réseau de recherche en TAL prenant en compte de nombreuses langues, ainsi que divers cadres théoriques et méthodologiques, (iii) de réconcilier la précision linguistique et l'efficacité computationnelle dans les applications du TAL. Le réseau compte environ 200 membres de 30 pays, travaillant sur 29 langues et 6 dialectes, représentant plusieurs théories linguistiques (HPSG, LFG, TAG, etc.), différentes méthodologies de parsing (statistique, symbolique, hybride) et d'autres applications (traduction automatique, recherche d'informations, etc.).

L'activité scientifique du réseau est structurée en 4 groupes de travail :

1. Interface lexique-grammaire.
2. Techniques du parsing pour les expressions polylexicales.
3. Traitement statistique, hybride et multilingue d'expressions polylexicales.
4. Annotation d'expressions polylexicales dans des corpus arborés.

1. www.parseme.eu

Les outils principaux du réseau sont des réunions scientifiques bi-annuelles, des missions individuelles de courte durée, des ateliers et des écoles d'été/hiver.

Deux membres de la fédération ICVL appartiennent au comité de pilotage du réseau : Yannick Parmentier est le leader du groupe de travail 2, tandis que Agata Savary joue le rôle de la présidente du comité d'administration. Ils sont également co-encadrants d'une thèse de doctorat s'inscrivant au coeur de la thématique du réseau. Notons que cette thématique trouve son instanciation nationale dans le projet B.1.3, PARSEME-FR. Elle coïncide également directement avec l'action A.1.8 et elle entretient des liens plus ou moins forts avec les actions A.1.7 (via le parsing) et A.1.6 (via les entités nommées polylexicales).

B.1.3 Syntactic Parsing and Multiword Expressions in French (ANR PARSEME-FR)

Fiche du projet

Statut : sélectionné

Dates : 2016-2019

Consortium : LIGM (Paris-Est), ALPAGE-INRIA (Paris-Diderot), LIF (Aix-Marseille), LIFO, LI

Coordinateur : Matthieu Constant (LIGM, Paris-Est)

Financement : ANR

Permanents impliqués

Jean-Yves Antoine (LI-BdTIn), Béatrice Bouchou-Markhoff (LI-BdTIn), Denys Duchier (LIFO-CA), Denis Maurel (LI-BdTIn), Yannick Parmentier (LIFO-CA), Agata Savary (LI-BdTIn)

Sous-axes concernés : **A2, A3**

Ce projet ANR est un spin-off national de l'action COST PARSEME (cf. projet B.1.2). Il a pour objectifs de : (i) faire avancer l'état de l'art sur la représentation des expressions polylexicales dans les ressources linguistiques du français (lexiques, corpus arborés, Linked Open Data), au niveau syntaxique, sémantique et pragmatique, (ii) intégrer le traitement des expressions polylexicales dans le parsing syntaxique et sémantique.

Les membres du consortium, mis à part le LIFO et le LI, sont : (i) le Laboratoire d'informatique Gaspard-Monge (LIGM) de l'Université Paris-Est Marne-la-Vallée, (ii) ALPAGE, une équipe d'INRIA à l'Université Paris-Diderot, (iii) le Laboratoire d'Informatique Fondamentale (LIF) de l'Université Aix-Marseille. Yannick Parmentier et Agata Savary sont coordinateurs locaux de ce projet pour le LIFO et le LI.

Ce projet, tout comme l'action COST PARSEME, est liés aux Actions A.1.8, A.1.7 et A.1.6.

B.1.4 Performance, évolution et composition pour XML : modèles, algorithmes et systèmes (ANR Codex)

Fiche du projet

Statut : terminé

Dates : 2009-2012

Consortium : INRIA (Saclay), INRIA (Lille), Paris-Sud Sorbonne, PPS Paris 7, INRIA Grenoble, LIF (Marseille), LI

Coordinateur : Ioana Manolescu (INRIA Saclay)

Financement : ANR

Permanents impliqués

Béatrice Bouchou-Markhoff (LI-BdTIn), Mirian Halfeld Ferrari (LIFO-Pamda), Agata Savary (LI-BdTIn)

Sous-axe concerné : **A2**

Les objectifs de ce projet ANR ont été tournés vers des traitements efficaces de données XML, en prenant en compte leur dynamique, évolution, interopérabilité, composition et distribution. Mirian Halfeld Ferrari a d'abord été le coordinateur local pour le LI, puis, ayant été recrutée en tant que professeur au LIFO, elle a permis d'inclure son nouveau laboratoire d'accueil dans le consortium de ce projet.

B.1.5 Big data on the cloud (BigTrend)

Fiche du projet

Statut : Terminé

Dates : 2013-2014

Consortium : Groupe Cyrès (France), LI, Renault Innovation

Coordinateur : Christophe Cerqueira (Groupe Cyrès)

Financement : européen FEDER

Permanents impliqués

Arnaud Giacometti (LI-BdTIn), Dominique Li (LI-BdTIn), Patrick Marcel (LI-BdTIn), Patrick Martineau (LI-OC), Arnaud Soulet (LI-BdTIn)

Sous-axes concernés : **A1, B2**

La finalité du projet a été d'élaborer de nouvelles méthodes, de nouveaux langages naturels, de haut niveau, compréhensible par des opérationnels non-informaticien. Ceci afin de permettre d'enrichir l'offre de services de Cyrès au travers d'innovations et de méthodes qui sont encore inexplorées dans les entreprises. Il a permis de proposer de nouvelles solutions à des clients (en l'occurrence Renault innovation) comme la mise en place de modèles prédictifs ou le traitement temps réel de grandes quantités d'information. Techniquement, il s'agissait de permettre une fouille de données sur un très grand volume de données, dans un mode Cloud Public ou Cloud Privé, de façon rapide (temps réel ou quasi temps réel) et simple pour tout utilisateur. Le système devrait être totalement modulable pour charger, exploiter et sécuriser les données des entreprises dans le cloud.

B.1.6 Fouille de données temporelles - Application à des données environnementales (FDTE)

Fiche du projet

Statut : terminé

Dates : 2012-2014

Consortium : LIFO, LI, ISTO (Université d'Orléans), MAPMO (Université d'Orléans)

Coordinateur : Christelle Vrain (LIFO)

Financement : régional

Permanents impliqués

Thi-Bich-Hanh DIEP-DAO (LIFO-CA), Fatma Bouali (LI-RFAI), Matthieu Exbrayat (LIFO-CA), Lionel Martin (LIFO-CA), Gilles Venturini (LI-RFAI), Christel Vrain (LIFO-CA)

Sous-axe concerné : **A1**

Les données temporelles concernent de nombreuses problématiques de la société car elles sont utilisées pour représenter, modéliser ou prédire l'évolution de phénomènes dynamiques dans le temps, phénomènes que l'on rencontre à tous les niveaux de la société (démographie, finance, environnement, etc). Ces données sont l'objet depuis toujours de travaux en informatique et en mathématiques avec des applications importantes, notamment en Région Centre. Les travaux développés ont porté sur des données environnementales (tourbières de La Guette en Région Centre et de Frasnès dans le Jura). Ce projet a visé à adapter ou développer des méthodes innovantes permettant la mise au point de modèles quantitatifs et qualitatifs à différentes granularités temporelles.

B.1.7 Grands graphes : interrogation, fouille et analyse (GIRAFON)

Fiche du projet

Statut : en cours

Dates : 2015-2017

Consortium : LIFO, LI

Coordinateur : Frédéric Loulergue (LIFO)

Financement : régional, Recherche d'initiative académique (APR-IA)

Permanents impliqués

Mostafa Bamha (LIFO-Pamda), Pascal Berthomé (LIFO-SDS), Béatrice Bouchou-Markhoff (LI-BdTIn), Wadoud Bousdira (LIFO-LMV), Jacques Chabin (LIFO-Pamda), Arnaud Giacometti (LI-BdTIn), Ali Ed-Dbali (LIFO-LMV), Mirian Halfeld Ferrari Alves (LIFO-Pamda), Dominique Li (LI-BdTIn), Sébastien Limet (LIFO-Pamda), Frédéric Loulergue (LIFO-LMV), Benjamin Nguyen (LIFO-SDS), Sophie Robert (LIFO-Pamda), Arnaud Soulet (LI-BdTIn), Christel Vrain (LIFO-CA)

Sous-axes concernés : **A1, A2, C1, C3**

L'objectif du projet GIRAFON est la conception d'algorithmes génériques de fouilles de grands graphes s'appuyant sur un langage de requêtes riche et des environnements d'exécution distribués accompagnés de bibliothèques de programmation dédiées, tout en prenant en compte la protection de la vie privée des individus. Cet objectif se décline en quatre tâches qui mèneront au développement de prototypes logiciels servant de preuve de concept de la pertinence des approches proposées, et qui seront évalués dans des environnements informatiques répartis :

1. Algorithmes pour l'analyse et la fouille de grands graphes.
2. Langages de requêtes sur les grands graphes.
3. Programmation parallèle structurée sur les grands graphes.
4. Protection de la vie privée.

GIRAFON est un projet fédérateur qui met en collaboration divers chercheurs des axes A et C. Ce projet a été retenu et son démarrage est prévu pour novembre 2015. Frédéric Loulergue est le coordinateur scientifique.

B.1.8 Annotation temporelle (MSH TEMPORAL)

Fiche du projet

Statut : terminé

Dates : 2013-2014

Consortium : LLL, LI, en association avec LIFO

Coordinateur : Emmanuel Schang (LLL)

Financement : régional, MSH-VdL 2013-II

Permanents impliqués

Jean-Yves Antoine (LI-BdTIn), Denis Maurel (LI-BdTIn), Yannick Parmentier (LIFO-CA), Agata Savary (LI-BdTIn)

Sous-axe concerné : **A3**

Ce projet a visé l'élaboration de la méthodologie pour la construction d'un corpus numérique oral annoté en référence et coréférence temporelle. Ses résultats directs ont consisté en des propositions de modifications de la norme d'annotation temporelle ISO TimeML. La mise en place de la nouvelle méthodologie, basée notamment sur des sorties d'un analyseur syntaxique, font objet des programmes scientifiques des projets régionaux ANNOTEP (cf. B.2.8) et ODIL (cf. B.1.9).

B.1.9 Outiller les données pour le développement des industries de la langue (IA ODIL)

Fiche du projet

Statut : en cours

Dates : 2016-2017

Consortium : LI, LIFO, LLL (Orléans)

Coordinateur : Jean-Yves Antoine (LI)

Financement : régional, Recherche d'initiative académique (APR-IA)

Permanents impliqués

Jean-Yves Antoine (LI-BdTIn), Denys Duchier (LIFO-CA), Denis Maurel (LI-BdTIn), Yannick Parmentier (LIFO-CA), Agata Savary (LI-BdTIn)

Sous-axe concerné : **A3**

Cette proposition de projet régional fait suite à une autre proposition nommée ANNOTEP (cf. B.2.8), non-sélectionnée en 2015. Son objectif est de concevoir un corpus représentatif de la variété des pratiques linguistiques, transcrit et annoté, afin de documenter les analyses linguistiques et assurer un transfert technologique pour les industries de la langue. L'une des originalités de la proposition est d'annoter un corpus en informations sémantiques, notamment temporelles, en se basant sur les résultats de l'analyse syntaxique. Le point de départ est ici le corpus de parole ESLO (résultat d'une large enquête socio-linguistique à Orléans en 1968), dont le Laboratoire Ligérien de Linguistique de l'Université d'Orléans est le dépositaire. Des liens forts existent entre ce projet et l'Action A.1.6.

B.1.10 Réseau Thématique Régional autour du risque (RTR Risque)

Fiche du projet

Statut : en cours

Dates : 2014–2018

Consortium : INSA Centre Val de Loire, Université de Tours (LI), Université d'Orléans (LIFO), Centre Mutualiste de rééducation et réadaptation fonctionnelles de Kerpape, CHU Garche, Adapt Centre

Coordinateur : Christian Toinard (LIFO-SDS)

Financement : régional

Permanents impliqués

Jean-Yves Antoine (LI-BdTln), Yannick Kergosien (LI-OC), Benjamin Nguyen (LIFO-SDS), Patrick Martineau (LI-OC), Christian Toinard (LIFO-SDS)

Sous-axes concernés : A3, B2, C1

L'objectif de ce réseau est de s'intéresser de façon pluridisciplinaire à l'apport mais aussi aux effets délétères (iatrogènes) de l'usage d'outils informatiques (orthèses numériques par exemple) pour la prise en charge du handicap, mettre en place un réseau (médecins, psychologues, soignants, éducateurs spécialisés, usagers, associations, centres spécialisés, agences de santé, philosophes, juristes, économistes, etc...) de gens qui souhaitent mesurer l'impact d'outils informatiques sur la qualité des soins. Le but est de mettre un aspect éthique au centre de la démarche qui tend à vouloir améliorer la prise en charge du handicap avec des approches numériques.

Les retombées sont la mise en place d'un réseau réellement pluridisciplinaire avec des gens qui arrivent à coopérer en ayant les mêmes préoccupations qui sont de mettre l'éthique au centre de la démarche d'informatisation de la prise en charge du handicap. Ce réseau doit montrer des premiers résultats d'études, qui sont mises en place au sein d'établissements spécialisés de rééducation, pour évaluer l'efficacité des approches et aussi définir une méthodologie d'analyse de risque et de réflexion éthique.

Une partie de ces travaux, impliquant notamment les chercheurs du LIFO et du LI, concernent les questions d'une meilleure identification des contraintes de fonctionnement des systèmes sécurisés de manière à mieux modéliser le problème d'optimisation à résoudre.

B.2 Projets soumis et en cours d'évaluation

B.2.1 Big Data Management and Analytics (BDMA)

Fiche du projet

Statut : soumis en 2016

Consortium : Université Libre de Bruxelles (Belgique), Université François Rabelais Tours (LI), Universitat Politècnica de Catalunya (Espagne), Technische Universität Berlin (Allemagne); Technische Universiteit Eindhoven (Pays-Bas); participation informelle du LIFO

Coordinateur : Esteban Zimányi (ULB)

Financement : européen, H2020 Erasmus+ (EAC/A04/2015)

Permanents impliqués

Jean-Yves Antoine (LI-BdTIn), Béatrice Bouchou-Markhoff (LI-BdTIn), Thomas Devogele (LI-BdTIn), Laurent Etienne (LI-BdTIn), Arnaud Giacometti (LI-BdTIn), Mirian Halfeld Ferrari (LIFO-Pamda), Nicolas Labroche (LI-BdTIn), Dominique Li (LI-BdTIn), Patrick Marcel (LI-BdTIn), Nizar Messai (LI-BdTIn), Verónica Peralta (LI-BdTIn), Agata Savary (LI-BdTIn), Arnaud Soulet (LI-BdTIn)

Sous-axes concernés : **A1, A2, A3**

C'est une proposition de programme international au niveau Master qui serait une évolution du Master Erasmus Mundus actuel IT4BI (voir projet B.1.1, p. 62) financé par la Commission Européenne.

Il s'agit d'un programme d'excellence dans les domaines de Business Intelligence (BI) et Big Data (BD), orienté vers des étudiants du monde entier, en offrant un système de bourses européennes, mais aussi des inscriptions auto-financées. Il est divisé en 4 semestres, chacun proposé par un partenaire différent. La spécialisation du semestre 3 proposée à Blois, intitulée *Content and Usage Analytics*, est axée sur la précision de l'analyse de donnée hétérogènes, tout en visant trois parmi les aspects principaux des Big Data : Value, Variability and Veracity.

Le coordinateur local de ce programme est Patrick Marcel. Le LIFO y est informellement associé par la participation aux enseignements liés aux spécialités de recherche communes aux deux laboratoires.

B.2.2 Large Scale Anomaly Recognition (LASCAR)

Fiche du projet

Statut : soumis en 2015, non-sélectionné

Consortium : LI, entreprise Cyres Group (France), LIST (Luxembourg Institute of Science and Technology) , entreprise L-1 Identity Solutions AG (Allemagne), entreprise NordPay (Espagne)

Coordinateur : Christophe CERQUEIRA (Cyres)

Financement : européen H2020, appel ICT-16-2015 (Big data – re-search)

Permanents impliqués

Arnaud Giacometti (LI-BdTIn), Dominique Li (LI-BdTIn), Patrick Martineau (LI-OC), Arnaud Soulet (LI-BdTIn), Gilles Venturini (LI-FRAI)

Sous-axes concernés : **A1, B2**

L'objectif principal du projet a été de créer un outil générique d'analyse et de fouille visuelle de données qui soit : (i) adaptable à un large spectre d'applications, (ii) utilisable par des non-experts du domaine de l'analyse de données, (iii) capable de traiter à la fois de gros volumes de données et des ensembles de données hétérogènes. L'outil développé devait être testé sur deux grands domaines applications :

- l'analyse de transactions bancaires pour détecter les comportements non usuels, à risque et frauduleux,
- l'analyse de vidéos de surveillance pour détecter automatiquement des tentatives d'intrusion.

Les défis scientifiques incluaient :

- un langage déclaratif de haut niveau pour analyser de très gros volumes de données,

- nouveaux ordonnanceurs de tâches pour exécuter des requêtes de fouille de données sur des environnements de type Hadoop/MapReduce,
- solutions d’analyse visuelle de données passant à l’échelle,
- nouvelles méthodes de détection de comportements anormaux et frauduleux (dans des transactions bancaires ou flux vidéo).

B.2.3 Innovative RENaissance HERitage (I-RENNHER)

Fiche du projet

Statut : soumis en 2015, non sélectionné.

Consortium : Centre d’études supérieures de la renaissance (CESR, Université de Tours), LI (avec le LIFO en tant que partenaire associé), Université de Poitiers, BulkyPix (France), CYRES, pôle Ingensi (France), Università degli studi di Milano (Italie), Museo Galileo (Italie), Antenna International (Royaume Uni), Pixel Ratio (Espagne), Alhambra, Generalife y Palacio de Carlos V (Espagne), Uniwersytet Kardynała Stefana Wyszyńskiego (Pologne), Uniwersytet Warszawski (Pologne), Instytut Badań Literackich PAN (Pologne), Max-Planck-Institut für Wissenschaftsgeschichte (Allemagne)

Coordinateur : Benoist Pierre (CESR Tours)

Financement : européen, H2020-Reflective-6-2015 Innovation ecosystems of digital cultural assets

Permanents impliqués

Béatrice Bouchou-Markhoff (LI-BdTin), Guillaume Cleuziou (LIFO-CA), Jacques Chabin (LIFO-Pamda), Mirian Halfeld Ferrari Alves (LIFO-Pamda), Frédéric Louergue (LIFO-LMV), Gilles Venturini (LI-RFAI), Christel Vrain (LIFO-CA)

Sous-axes concernés : **A1, A2**

I-RENNHER est un projet de recherche scientifique, soumis en mai 2015, répondant à l’appel à projets européen Horizon 2020, Reflective 2015 - *Innovation ecosystems of digital cultural assets*. L’objectif de I-RENNHER est la création d’un

système intelligent d'information dédié à la compréhension globale du patrimoine renaissant européen.

Le projet prévoit le développement d'une plateforme de données scientifiques, hétérogènes et partagées sur la Renaissance. Grâce aux ontologies, ce système complexe mettra en lien des connaissances scientifiques issues de domaines disciplinaires divers. Ce réseau de relations inédites permettra l'élaboration de nouveaux savoirs au profit d'une nouvelle compréhension du patrimoine historique européen, favorisant le développement de l'économie touristique. La diffusion de ces nouveaux savoirs est prévue via une application numérique intelligente connectée à la plateforme de données hétérogènes et capable d'en extraire les informations circonstanciées en fonction des critères sélectionnés.

Le projet sera dirigé par le Centre d'études supérieures de la Renaissance (CESR-UFRT / CNRS). La fédération ICVL s'est associée à ce projet, s'intéressant aux tâches concernant, par exemple, l'interrogation cohérente des données en prenant en compte leur provenance, l'évolution de schéma et contraintes, l'extraction automatique ou semi-automatique d'un thésaurus de type WordNet...

B.2.4 SHAring GRAMmatical and LEXical Resources (FP7 SHAGRALER)

Fiche du projet

Statut : soumis en 2011, non sélectionné

Consortium : LIFO, LI, Tagmatica (Paris), IPIPAN (Varsovie), POLENG (Poznań), Universität Düsseldorf

Coordinateur : Yannick Parmentier (LIFO)

Financement : européen, FP7-ICT-2011-SME-DCL

Permanents impliqués

Thi-Bich-Hanh Dao (LIFO-CA), Denys Duchier (LIFO-CA), Nathalie Friburger (LI-BdTIn), Denis Maurel (LI-BdTIn), Yannick Parmentier (LIFO-CA), Agata Savary (LI-BdTIn)

Sous-axe concerné : **A3**

Cette proposition de projet européen a été un premier pas vers une collaboration plus étroite entre le LIFO et le LI dans le domaine du TAL et des ressources linguistiques. Elle avait pour objectifs de explorer des méthodes de production semi-automatique de ressources telles que méta-grammaires, grammaires locales et lexiques électroniques, en vue d'applications telles que traduction automatique et extraction d'information.

B.2.5 Discrimination prosodique automatisée et sémantique de l'oral (ANR SEMORAL)

Fiche du projet

Statut : soumis en 2014 et en 2015, non sélectionné

Consortium : LLL (Orléans), LIFO, LI, Prisme-Iraus (Orléans), Modyco (UMR, Paris-Ouest), FoReLL (UPRES, Poitiers)

Coordinateur : François Nemo (LLL Orléans)

Financement : ANR

Permanents impliqués

Jean-Yves Antoine (LI-BdTIn), Denis Maurel (LI-BdTIn), Yannick Parmentier (LIFO-CA), Agata Savary (LI-BdTIn)

Sous-axe concerné : **A3**

Cette proposition de projet ANR fédère des experts pluridisciplinaires dans les domaines du traitement du signal, traitement automatique des langues et linguistique. Son objectif principal est de faciliter l'emploi des données prosodiques facilement et massivement accessibles (par exemple grâce à la reconnaissance automatique de la parole intégrée dans les smartphones) pour le traitement automatique de la sémantique de la parole spontanée. L'un des points d'entrée du projet est le corpus ANCOR (cf. Action A.1.6), le plus grand corpus oral dans le monde annoté pour la coréférence, créé en 2011-2013 dans le cadre d'un projet régional incluant le LI et le LLL (Laboratoire Ligérien de Linguistique de l'Université d'Orléans, partenaire privilégié des projets en TAL à la fois du LIFO et du LI, depuis de nombreuses années).

B.2.6 Smart Loire

Fiche du projet

Statut : soumis en 2015

Consortium : LI, INSA CVL, CESR (Centre d'Etudes Supérieures de la Renaissance), LoireXplorer (blog tourisme), Agence Départementale du Tourisme de Touraine

Coordinateur : Thomas Devogele (LI)

Financement : régional, Recherche d'initiative académique (APR-IA)

Permanents impliqués

Thomas Devogele (LI-BdTIn), Laurent Etienne (LI-BdTIn), Nizar Messaï (LI-BdTIn), Yacine Sam (LI-BdTIn), Jorge Mendoza (LI-OC), Yannick Kergosien (LI-BdTIn)

Sous-axes concernés : **A1, B2**

La région Centre-Val de Loire constitue un site unique en France pour le tourisme grâce à la richesse et la diversité de ses composantes. La finalité du projet Smart Loire est de fournir aux visiteurs et aux acteurs du tourisme des services web. Leur généricité doit permettre de les intégrer aussi bien à un site web proposant des itinéraires à vélo dans la région qu'à une application mobile d'un office de tourisme local. Des développements dans le cadre de l'application mobile Caravel², déjà opérationnelle, sont prévus afin de valider ces services innovants. Du point de vue de la recherche en informatique, trois verrous scientifiques sont visés : la composition de services web (*mashup*) touristiques, l'optimisation d'itinéraire et l'analyse des trajectoires des visiteurs.

2. <http://www.caravel-app.fr/>

B.2.7 Analyse syntaxique des expressions figées (IA ASEF)

Fiche du projet

Statut : soumis en 2012, non sélectionné

Consortium : LIFO, LI

Coordinateur : Yannick Parmentier (LIFO)

Financement : régional, Recherche d'initiative académique (APR-IA)

Permanents impliqués

Thi-Bich-Hanh Dao (LIFO-CA), Denys Duchier (LIFO-CA), Nathalie Friburger (LI-BdTIn), Denis Maurel (LI-BdTIn), Yannick Parmentier (LIFO-CA), Agata Savary (LI-BdTIn)

Sous-axe concerné : **A3**

Cette proposition de projet régional, même si non sélectionnée, a été l'un des précurseurs de la collaboration LIFO-LI dans le domaine des expressions poly-lexicales, qui constitue actuellement un pilier des recherches communes de la fédération via les projets COST PARSEME (cf. B.1.2) et ANR PARSEME-FR (cf. B.1.3), ainsi l'Action A.1.8. Les objectifs du projet étaient : (i) de proposer un modèle de description aussi général que possible des expressions figées (intégrant des informations morphologiques, syntaxiques et sémantiques), (ii) de l'appliquer au français et éventuellement à d'autres langues (enrichissant ainsi les ressources linguistiques produites par le LI et le LIFO), et (iii) de le valider en mettant en œuvre des algorithmes d'analyse syntaxique sur des énoncés intégrant des expressions figées (e.g. corpus issus du web, etc.).

B.2.8 Annotation temporelle (IA ANNOTEP)

Fiche du projet

Statut : soumis en 2014 et en 2015, non sélectionné

Consortium : LI en association avec LIFO, LLL (Orléans), Modyco (Paris Ouest)

Coordinateur : Jean-Yves Antoine (LI)

Financement : régional, Recherche d'initiative académique (APR-IA)

Permanents impliqués

Jean-Yves Antoine (LI-BdTIn), Denys Duchier (LIFO-CA), Nathalie Friburger (LI-BdTIn), Denis Maurel (LI-BdTIn), Yannick Parmentier (LIFO-CA), Agata Savary (LI-BdTIn)

Sous-axe concerné : **A3**

Cette proposition de projet régional a été axée autour des acquis du projet TEMPORAL (cf. B.1.8), qui consistaient en l'évolution de la méthodologie de l'annotation des événements et des relations temporelles en corpus vers des structures arborescentes. L'objectif de ce projet a été de fournir librement la communauté le plus grand corpus francophone annoté en référence temporelle. Par sa taille (100 000 mots) ce corpus devait constituer une ressource de premier plan au niveau international et répondre aux besoins des méthodes d'apprentissage automatique utilisées par le TAL (lien avec l'Action A.1.7).

EXEMPLES DES FACTEURS DE VISIBILITÉ EXTÉRIEURE

Dénomination de l'initiative	Coordinateur local	Dates	Budget (euros)	Type	Portée	Rôle de l'équipe LIFO/LI	Ampleur	Partenaires principaux	Axes
Master Erasmus Mundus J1481	Patrick Marcel	2012-2018		co-diplôme	internationale	participation	4 pays, 10 pays associés, recrutement mondial	Université Libre de Bruxelles, Belgique; Université polytechnique de Catalogne, Espagne; Technische Universität Berlin, Allemagne	A1,A2,A3
Agata Smuty 1207 PARRSENE	Agata Smuty, Yannick Paumier	2013-2017	700 K	réseau financé	internationale	coordination	33 pays, 200 chercheurs	Pologne, Allemagne, Norvège, Grèce	A3
Prole Sove (Egile)	Denis Maurer	2004-2005	5,5 K	Projet filantier	internationale	coordination	2 pays	Université de Belgrade, Serbie	A3
Polonium (Egile)	Agata Smuty	2007-2008	5,74 K	Projets échanges internationaux	internationale	coordination	2 pays	Académie polonaise des sciences, Pologne	A3

Dénomination de l'initiative	Coordinateur local	Dates	Budget (euros)	Type	Portée	Rôle de l'équipe LIFO/LI	Ampleur	Partenaires principaux	Axes
Projet ANR franco-japonais PafDPA5	Frédéric Louérgue	2011-2014		projet financé	internationale	coordination	2 pays	Kocchi University of Technology, Japon	C3
Projet ANR Jeune Chercheur PONOSaD	Yannick Kergozien	2013-2017	131.104 K	projet financé	internationale	coordination	2 pays, 3 partenaires	CYRRELT, Montréal, Canada; Université Laval - Québec, Canada	B2
Projet CAPES-COFECUB MAP	Marcílio de Souto	-2014		projet financé	internationale	coordination	2 pays	Université de Sao Paulo, Brésil	A1
projet ECOS-COMICT C12E05	Nicolas Ollinger	2013-2015		projet financé	internationale	participation	2 pays	US Nîce et Univ. Adolfo B1 Ibanez Chili	B1
Projet franco-indien JFCPARR	Nicolas Ragot	2012-2016	1 M	projet financé	internationale	coordination	2 pays	ISI Calcutta, Inde	A1, A4

Dénomination de l'initiative	Coordinateur local	Dates	Budget (euros)	Type	Portée	Rôle de l'équipe LIFO/LI	Ampleur	Partenaires principaux	Axes
Projet Paradite	Jean-Yves Ramel	2010-2012	70 K	Google Award	internationale	coordination	Partenaires français, financement international	CESR Tours	A1, A4
Projet STIC-AMISUD CLEVER	Miriam Haffeld Ferrari	2011-2012	31 K	projet financé	internationale	coordination locale	3 pays, 4 partenaires	Universidade Federal do Rio Grande do Norte (UFRRN), Brésil; Facultad de Ingeniería de la República de Uruguay (INCO), Laboratoire d'Informatique de Grenoble	A2
Projet STIC-AMISUD SWANS	Miriam Haffeld Ferrari	2013-2014	13 K	projet financé	internationale	coordination locale	3 pays, 4 partenaires	Universidade Federal do Rio Grande do Norte (UFRRN), Brésil; Facultad de Ingeniería de la República de Uruguay (INCO), Laboratoire d'Informatique de Grenoble	A2

Dénomination de l'initiative	Coordinateur local	Dates	Budget (euros)	Type	Portée	Rôle de l'équipe LIFO/LI	Ampleur	Partenaires principaux	Axes
CAPPERIS Collaborative Action for the Protection of Privacy Rights in the Information Society	Benjamin Nguyen	fin en 2017		Project Lab PMRIA	Internationale	Coordinateur d'une action transversale	2 pays, 7 partenaires	France, Belgique	C1
Livre collaboratif: Constraints and Language, Cambridge Scholar Publishing, 2014.	Denis Duchier			livre collaboratif	internationale		Philippe Blache and Henning Christiansen and Denis Duchier and Verónica Dafil and Jørgen Villadsen (Ed.)		A3
Newsletter du TC-15 (Graphes en Reconnaissance de formes)	Jean-Yves Ramel, Donatello Conte			groupeement de recherche	internationale	participation	Beaucoup de destinataires		A1,A4

Dénomination de l'initiative	Coordinateur local	Dates	Budget (euros)	Type	Portée	Rôle de l'équipe LIFO/LI	Ampleur	Partenaires principaux	Axes
Centre d'Excellence africain en Mathématiques, Informatique et Technologies de l'information et de la Communication (CEA-MITIC)	Béatrice BOUCHOU-MARRKHOF (membre du Conseil Scientifique de CEA-MITIC)	2015-2018	5 M	Collaboration de longue durée avec financement par la Banque Mondiale	internationale	partenaire	Sénégal, Burkina Faso, Cameroun, Côte d'Ivoire, Gambie, Nigeria, Tunisie	UGB de Saint Louis du Sénégal	A2
Atelier sur la Protection de la Vie Privée	Benjamin Nguyen	2015	20K	congrès	francophone	Organisateur		France	C1
Colloque HLPF (High-Level Parallel Programming and Applications)	Frédéric Loulrye	Depuis 2013		congrès	internationale	coordination			C3
Organisation du workshop DRAS (Docupment Analysis Systems)	Jean-Yves Ramel	01/04/2014		congrès	internationale	coordination	5 pays	DPK1 (Allemagne), Osaka (Japon), ISI Calcutta (Inde)	A1, A4

Dénomination de l'initiative	Coordinateur local	Dates	Budget (euros)	Type	Portée	Rôle de l'équipe LIFO/LI	Ampleur	Partenaires principaux	Axes
colloque STACS (Symposium on Theoretical Aspects of Computer Science)	Nicolas Ollinger	2016		conférence	internationale	coordination			B1
colloque WG (Workshop on Graph- Theoretic Concepts in Computer Science)	Jean Tóth	2014		conférence	internationale	coordination			B1
Covert Channels	Jean-François Lafont			Collaboration + dépôt de projet STUDYUM	Internationale	coordination	3	FR, Italie, Pologne	C1
Collaboration bilatérale de longue durée avec l'Université Gaston Berger de Saint Louis (Sénégal)	Arnaud Giacometti	depuis 2000	Certelles ARUF, l'ambassa de de France au Sénégal	co-tutelle de thèse, double- diplôme de master	internationale	coordination	2 pays; 4 co-tutelles de thèse; 4 stages de master	Université Gaston Berger de Saint Louis, Sénégal	A1, A2

Dénomination de l'initiative	Coordinateur local	Dates	Budget (euros)	Type	Portée	Rôle de l'équipe LIFO/LI	Ampleur	Partenaires principaux	Axes
Ensemble de collaborations avec le Magreb	Denis Maurel	total de 8 mois financés depuis 2008	séjour+voyage	stages doctoraux, coencadrement de master	internationale	coordination	2 pays, 6 stages de doctorant, une certitude de master et une codirection de thèse	Université de Sfax, Tunisie et université de Saida, Algérie	A3
Constraints and Data Mining	C. Yvain - T.B.H. Dao	2015-2018		workshop, tutoriel e-EGC	nationale			Collaborateurs internationaux : J. Davidson, T. Guns	A1
projet ANR AGAPE	Jean Todiñica	2008-2012	220 K	projet financé	nationale	coordination	3 partenaires	LIFO, INRIA Sophia Antipolis, LIRMM	B1
Projet ANR DIGDOC	Jean-Yves Ramel	2011 - 2014	175 K	projet financé	nationale	participation	5 partenaires	LIRIS (Rouen), LIRIS (Lyon), Labri (Bordeaux), La Rochelle, Bibliothèque Nationale de France	A1, A4
Projet ANR FyNamo	Sophie Robert	2008-2011	121 K	projet financé	nationale	participation	4 partenaires	LBT Laboratoire de Biologie Théorique (CNRS, Paris) CEA INRIA Grenoble	C3

Dénomination de l'initiative	Coordinateur local	Dates	Budget (euros)	Type	Portée	Rôle de l'équipe LIFO/LI	Ampleur	Partenaires principaux	Axes
projet ANR GraphEn	Jean Toffin	2015-2019		projet financé	nationale	participation	3 partenaires	LPTA Metz	B1
Projet Exaviz	Sophie Robert	2011-2016	136 K	projet financé	nationale	participation	5 partenaires	LBT Laboratoire de Biologie Théorique (CNRS, Paris) LJMSJ INRA Grenoble CEMHTI	C3
Extraction et Gestion des Connaissances (EGC)	Gilles Venturini			groupe de travail	nationale	coordination			A1, A4
Groupe de Recherche en Communication Ecrite	Jean-Yves Ramel			groupe de travail	nationale	participation	2 pays	CVC (Barcelone), LIRIS (Lyon), Labri (Bordeaux, ES), La Rochelle, IRISA (Rennes), LIPADE (Paris), LORIA (Nancy), Telecoms ParisTech	A1, A4

Dénomination de l'initiative	Coordinateur local	Dates	Budget (euros)	Type	Portée	Rôle de l'équipe LIFO/LI	Ampleur	Partenaires principaux	Axes
Groupe de travail LAMHRA du GDR GPL	Frédéric Louérgue	dépuis 2008		groupe de travail	nationale	coordination			C3
GT SDA2 du GDR JM	Nicolas Ollinger	dépuis 2015		groupe de travail	nationale	coordination			B1
rencontres JGA 2015	Mathieu Liedloff	2015		groupe de travail	nationale	coordination			B1
Comité d'animation GDR JA	Christel Vrain	2016-2017		GDR	national				A1
ACTE LYN Travailité Longitudinale Hybride Unitaire Appliquée aux emballages de Grande Consommation	Patrice Clemente	2016-2019	1,2M dont 127K pour le LIFO	FCU	nationale	Coordonnateur académique	9 dont 3 académiques et 6 industriels	FR	C1

Dénomination de l'initiative	Coordinateur local	Dates	Budget (euros)	Type	Portée	Rôle de l'équipe LIFO/LI	Ampleur	Partenaires principaux	Axes
Projet PRACE MFBSP		2013-2014		special interest group	nationale	participation	1 partenaire	Université Paris-Est Crétail (UP-EC)	C2, C3
Réseau Thématique Régional sur le Risque Médical	Christian Touard	Depuis 2014		RTR	Régionale (CVL/ Bretagne)	Coordinateur	E1 + E2FO		C1
Valorisation des Données Personnelles	Benjamin Nguyen	fin en 2016	150K	Projet Jdex Paris Saclay	Régionale (île de France)	Participatif extérieur	6 dont 2 industriels		C1

CO-ENCADREMENTS

Thèses de doctorat

- Than Binh Nguyen
 - Titre : Vers une assistance intelligente à l’exploitation des services et données du web
 - Encadrants : Béatrice Bouchou-Markhof (LI-BdTIn) et Mirian Halfeld Ferrari Alves (LIFO-Pamda)
 - Dates : débutée en septembre 2014
 - Financement : bourse Orléans-Tours
- Jakub Waszczuk
 - Titre : Analyse syntaxique d’unités polylexicales
 - Encadrants : Agata Savary (LI-BdTIn) et Yannick Parmentier (LIFO-CA)
 - Dates : débutée en octobre 2013
 - Financement : bourse ministérielle

PUBLICATIONS COMMUNES AUX DEUX LABORATOIRES

- [1] Joshua Amavi, Béatrice Bouchou, and Agata Savary. On correcting XML documents with respect to a schema. *Comput. J.*, 57(5) :639–674, 2014.
- [2] Jean-Yves Antoine, Marie-Elisabeth Labat, Anaïs Lefeuvre, and Christian Toinard. Vers une méthode de maîtrise des risques dans l’informatisation de l’aide au handicap. In *Envirorisk’2014, Le forum de la gestion des risques technologiques, naturels et sanitaires*, Gestion des risques naturels, technologiques et sanitaires - Forum Envirorisk’2014, page 9 pages, Bourges, France, November 2014. Cépaduès.
- [3] Cheikh Ba, Umberto Souza da Costa, Mírian Halfeld Ferrari, Rémy Ferré, Martin A. Musicante, Verónica Peralta, and Sophie Robert. Preference-driven refinement of service compositions. In *CLOSER 2014 - Proceedings of the 4th International Conference on Cloud Computing and Services Science, Barcelona, Spain, April 3-5, 2014.*, pages 268–275, 2014.
- [4] Gabriel Bergounioux, Emmanuel Schang, Denis Maurel, Agata Savary, Jérôme Durand-Lose, and Yannick Parmentier. L’ordinateur et les langues. *Covalences*, 82 :14–15, 2012.
- [5] Béatrice Bouchou, Mirian Halfeld Ferrari Alves, and Maria Adriana Vidi-gal de Lima. A grammarware for the incremental validation of integrity constraints on XML documents under multiple updates. *T. Large-Scale Data- and Knowledge-Centered Systems*, 6 :167–197, 2012.
- [6] Béatrice Bouchou, Mirian Halfeld Ferrari Alves, and Maria Adriana Lima. Attribute Grammar for XML Integrity Constraint Validation. In Abdelka-der Hameurlain, Stephen W. Liddle, Klaus-Dieter Schewe, and Xiaofang Zhou, editors, *Database and Expert Systems Applications - 22nd International Conference, DEXA 2011*, volume 6860, pages 94–109, Toulouse, France, August 2011. Springer.

- [7] Iris Eshkol-Taravella, Denis Maurel, Isabelle Tellier, Nathalie Friburger, and Samer Taalab. Annoter ESLO1 par des cascades de transducteurs et par apprentissage automatique. In *Journée d'étude de l'Atala : Annoter les corpus oraux.*, Paris, France, 2011.
- [8] Christophe Lenté, Mathieu Liedloff, Emmanuel Néron, Ameer Soukhal, and Vincent T'Kindt. Complexité d'algorithmes exponentiels : application au domaine de l'ordonnancement. In *ROADEF'2010 : 11ième congrès de la Société Française de Recherche Opérationnelle et d'Aide à la Décision*, Toulouse, France, February 2010.
- [9] Christophe Lenté, Mathieu Liedloff, Ameer Soukhal, and Vincent T'Kindt. Algorithmes Exponentiels Pour Des Problèmes D'Ordonnancement à Une Machine Et Machines Parallèles. In *12ième congrès de la Société Française de Recherche Opérationnelle et d'Aide à la Décision (ROADEF 2011)*, France, 2011.
- [10] Christophe Lenté, Mathieu Liedloff, Ameer Soukhal, and Vincent T'Kindt. Sort & Search Exponential-Time Algorithms for Scheduling Problems,. In *Proceedings of Multidisciplinary International Conference on Scheduling : Theory and Applications (MISTA 2011)*, pages pp. 523–528, United States, 2011.
- [11] Christophe Lenté, Mathieu Liedloff, Ameer Soukhal, and Vincent T'Kindt. Complexité au pire des cas d'algorithmes exponentiels pour des problèmes de séquençement. In *13ième congrès de la Société Française de Recherche Opérationnelle et d'Aide à la Décision (ROADEF 2012)*, France, 2012.
- [12] Christophe Lenté, Mathieu Liedloff, Ameer Soukhal, and Vincent T'Kindt. Généralisation de la méthode Trier et Chercher : application à des problèmes à machines parallèles. In *13ième congrès de la Société Française de Recherche Opérationnelle et d'Aide à la Décision (ROADEF 2012)*, France, 2012.
- [13] Christophe Lenté, Mathieu Liedloff, Ameer Soukhal, and Vincent T'Kindt. Scheduling parallel machines with exponential algorithms. In *International Workshop on Approximation, Parameterized and EXact algorithms (APEX 2012)*, France, 2012.
- [14] Christophe Lenté, Mathieu Liedloff, Ameer Soukhal, and Vincent T'Kindt. Scheduling parallel machines with exponential algorithms. In *13th International Conference on Project Management and Scheduling (PMS 2012)*, pages pp. 203–206, Belgium, 2012.

- [15] Christophe Lenté, Mathieu Liedloff, Ameer Soukhal, and Vincent T'Kindt. On an extension of the Sort & Search method with application to scheduling theory. *Theoretical Computer Science*, 511 :13–22, 2013.
- [16] Christophe Lenté, Mathieu Liedloff, Ameer Soukhal, and Vincent T'Kindt. Exponential Algorithms for Scheduling Problems. Technical report, 2014.
- [17] Christophe Lenté, Mathieu Liedloff, Vincent T'Kindt, and Ioan Todinca. Algorithmes modérément exponentiels pour problèmes NP-difficiles. In Nicolas Ollinger, editor, *Informatique Mathématique*, collection Alpha, pages 47–85. CNRS Éditions, March 2015.
- [18] Christophe Lenté, Matthieu Liedloff, Ameer Soukhal, and Vincent T'Kindt. Exponential Algorithms for the Two Parallel Machine Scheduling Problem. In *INFORMS Annual Meeting*, pages –, Charlotte, United States, November 2011.
- [19] Christophe Lenté, Matthieu Liedloff, Ameer Soukhal, and Vincent T'Kindt. Exponential-time algorithms for scheduling problems. In *10th Workshop on Model and Algorithms for Planning and Scheduling Problems*, pages –, Nymburk, Czech Republic, June 2011.
- [20] Mathieu Liedloff, Christophe Lenté, Ameer Soukhal, and Vincent T'Kindt. Exponential-time algorithms for scheduling problems. In *Workshop on Models and Algorithms for Planning and Scheduling Problems (MAPSP 2011)*, Czech Republic, 2011.
- [21] Agata Savary, Manfred Sailer, Yannick Parmentier, Michael Rosner, Victoria Rosén, Adam Przepiórkowski, Cvetana Krstev, Veronika Vincze, Beata Wójtowicz, Miriam Butt, Gyri Smørdal Losnegaard, Carla Parra Escartín, Jakub Waszczuk, Matthieu Constant, Petya Osenova, and Federico Sangati. PARSEME – PARSing and Multiword Expressions within a European multilingual network. Submitted to Language & Technology Conference (LTC'15), Poznań, November 27-29, 2015, 2015.
- [22] Lei Shang, Christophe Lenté, Mathieu Liedloff, and Vincent T'Kindt. An exponential dynamic programming algorithm for the 3-machine flowshop scheduling problem to minimize the makespan. In *7th Multidisciplinary International Conference on Scheduling : Theory and Applications (MISTA 2015)*, MISTA 2015 proceedings, pages 755–758, Prague, Czech Republic, August 2015.
- [23] Vincent T'Kindt, Christophe Lenté, and Mathieu Liedloff. A Study of worst-case complexity for parallel machine scheduling problems based on an ex-

tension of the Sort & Search method. In *Workshop on Models and Algorithms for Planning and Scheduling Problems (MAPSP 2013)*, France, 2013.

PUBLICATIONS CO-SIGNÉES PAR DES MEMBRES DE PLUSIEURS ÉQUIPES DES LABORATOIRES LIFO ou LI

- [1] Zeina Abu-Aisheh, Romain Raveaux, Jean-Yves Ramel, and Patrick Martineau. An Exact Graph Edit Distance Algorithm for Solving Pattern Recognition Problems. In *4th International Conference on Pattern Recognition Applications and Methods 2015*, Lisbon, Portugal, January 2015.
- [2] Pascal Berthomé, Jean-François Lalande, and Vincent Levorato. Implementation of exponential and parametrized algorithms in the AGAPE project. Technical report, January 2012.
- [3] Yohan Boichut, Jacques Chabin, and Pierre Réty. Towards More Precise Rewriting Approximations. In *Proceedings of Language and Automata Theory and Applications (LATA)*, volume LNCS 8977, Nice, France, March 2015. Springer.
- [4] Yohan Boichut, Thi-Bich-Hanh Dao, and Valérie Murat. Characterizing Conclusive Approximations by Logical Formulae. In Giorgio Delzanno and Igor Potapov, editors, *Reachability Problems 2011*, volume LNCS 6945 of *Reachability Problems (RP 2011)*, Gênes, Italy, September 2011.
- [5] Denys Duchier, Jérôme Durand-Lose, and Maxime Senot. Fractal parallelism : Solving sat in bounded space and time. In Kyung-Yong Chwa Cheong Otfried and Kunsoo Park, editors, *ISAAC '10*, number 6506 in LNCS, pages 279–290, South Korea, December 2010. Springer.
- [6] Denys Duchier, Jérôme Durand-Lose, and Maxime Senot. Massively Parallel Automata in Euclidean Space-Time. In Jacob Beal, Olivier Michel, and Antoine Spicher, editors, *IEEE 4th International Conference on Self-Adaptive and Self-Organizing Systems Workshops (SASOW '10) : Spatial*

- Computing Workshop (SCW '10)*, pages 104–109, Budapest, Hungary, 2010. IEEE Computer Society.
- [7] Denys Duchier, Jérôme Durand-Lose, and Maxime Senot. Computing with signals : a generic and modular signal machine for satisfiability problems. In *Workshop New Worlds of Computation (NWC '11)*, Orléans, France, May 2011.
 - [8] Denys Duchier, Jérôme Durand-Lose, and Maxime Senot. Solving Q-SAT in bounded space and time by geometrical computation. In Hristo Ganchev, Benedikt Löwe, Dag Normann, Ivan Soskov, and Mariya Soskova, editors, *Models of computability in context, 7th Int. Conf. Computability in Europe*, pages 76–86, Sofia, Bulgaria, June 2011. St. Kliment Ohridski University Press, Sofia University.
 - [9] Denys Duchier, Jérôme Durand-Lose, and Maxime Senot. Computing in the fractal cloud : modular generic solvers for SAT and Q-SAT variants. In S.B. Cooper M. Agrawal and A. Li, editors, *Theory and Applications of Models of Computation (TAMC 2012)*, volume 7287 of *LNCS*, pages 435–447, Beijing, China, May 2012. Springer.
 - [10] Asma Guesmi, Patrice Clemente, Frédéric Loulergue, and Pascal Berthomé. Cloud Resources Placement Based on Functional and Non-Functional Requirements. In *SECRYPT*, Colmar, France, July 2015. SCITEPRESS.

STATUTS

G.1 Statuts de la Fédération ICVL

Version approuvée par les assemblées générales du LI (10 septembre 2015) et du LIFO (21 septembre 2015)

Dénomination et composition

Article 1. La fédération ICVL (Informatique Centre Val de Loire) est une fédération de recherche composée des entités suivantes :

- le LI (Laboratoire d’Informatique), équipe d’accueil EA 6300 sous la double tutelle de l’Université François Rabelais de Tours et de l’INSA Centre Val de Loire et
 - le LIFO (Laboratoire d’Informatique Fondamentale d’Orléans), équipe d’accueil EA4022 sous la double tutelle de l’Université d’Orléans et de l’INSA Centre Val de Loire
- ci-après désignées collectivement par *Parties* et individuellement par *Partie*.

Article 2. Une convention, signée entre les tutelles des deux laboratoires, complète et assure l’application de ces statuts. La durée de la convention établit la durée des instances définies dans ces statuts.

Objectifs

Article 3. Les Parties ont décidé de s’associer pour former une fédération de recherche qui se veut un élément dynamique de promotion des synergies scientifiques

et de structuration du potentiel recherche de la Région Centre dans le domaine de l'informatique.

La fédération ICVL défendra les valeurs telles que la qualité de la recherche scientifique et de l'enseignement supérieur, la promotion de jeunes chercheurs, la transparence des procédures et la démocratie dans la prise de décision.

ICVL a comme missions principales de :

- valoriser et améliorer la visibilité des activités de recherche informatique en Région Centre ;
- susciter et favoriser des projets de recherche au niveau régional, national et international ;
- dynamiser l'activité de recherche au sein des deux Parties et encourager leur collaboration.

Membres

Article 4. Les membres de la fédération ICVL sont les membres pleins des Parties, définis selon les critères de chaque Partie. Chaque année, les Parties informent la fédération des éventuels modifications dans la liste de leurs membres pleins.

Direction

Article 5. La direction de la fédération est assurée par une *équipe de direction* composée d'un directeur et d'un directeur adjoint pour une durée égale à celle de la convention citée dans l' **Article 2.** (limitée à 5 ans si la durée dépasse 5 ans). Le directeur et le directeur adjoint changent leurs rôles à la mi-mandat. Une limite d'appartenance à l'équipe de direction est fixée à 10 ans. Cette limite peut être prolongée sous condition d'interruption d'au moins 5 ans.

Article 6. Le directeur et le directeur adjoint ne doivent pas être rattachés à la même Partie et ils doivent être des membres de la fédération dans l' **Article 4.**

Article 7. Les fonctions de directeur et directeur adjoint de la fédération sont non cumulables avec toute autre fonction de direction liée à la recherche dans leurs entités de rattachement (directeur ou directeur adjoint des laboratoires, directeur de ITP (Institut Thématique Pluridisciplinaire), responsable d'équipe dans un laboratoire, directeur ou directeur adjoint d'une autre fédération, membre de l'équipe de direction d'une des tutelles...).

Article 8. L'équipe de direction est nommée conjointement par les tutelles des Parties sur une proposition des Parties. La proposition des Parties est faite en deux étapes :

- le directeur et le directeur adjoint sont élus sur une même liste par les membres de la fédération via un suffrage direct à majorité absolue, à un ou deux tours,
- les conseils de laboratoires des Parties valident le résultat de cette élection.

Le veto peut être prononcé par un des conseils cités ci-dessus qui, dans ce cas, présente son argumentation et propose, dans un délai de 15 jours, une nouvelle liste qui est remise à l'approbation des autres conseils de laboratoires des Parties.

Le mandat d'une nouvelle équipe de direction commence au début de chaque convention (**Article 2.**).

Article 9. En cas de départ d'un des membres de l'équipe de direction avant la fin de son mandat, le membre restant propose un nouveau membre pour composer l'équipe de direction dans un délai d'un mois. Le partage des rôles du directeur et du directeur-adjoint est établi conjointement pour la durée restante du mandat en cours. Cette nouvelle équipe doit être validée dans les conditions établies dans l'**Article 8.**

Article 10. En cas de départ conjoint de deux membres de l'équipe de direction avant la fin de leur mandat, une nouvelle équipe de direction est nommée dans les conditions établies dans l'**Article 8.**

- pour la durée restante du mandat en cours si celle-ci est supérieure à la durée d'un demi-mandat,
- pour la durée restante du mandat en cours et le mandat suivant dans le cas contraire.

Article 11. Le directeur met en oeuvre la politique de la fédération, organise l'animation scientifique et l'accès à l'information. Il assure l'utilisation des fonds communs. Il publie annuellement le rapport moral et financier de la fédération. Ce rapport est voté par les assemblées générales des Parties, selon un calendrier proposé par les Parties et en accord avec la direction de la fédération. Au terme de chaque convention (**Article 2.**) il rédige un rapport d'activité qu'il transmet à chacune des Parties et il organise les élections pour la nouvelle équipe de direction.

Article 12. Le directeur adjoint assiste le directeur dans la gestion et l'organisation de la fédération et le remplace en cas de besoin.

Conseil de fédération

Article 13. Il est institué un conseil de fédération dont la composition est la suivante :

- le directeur de la fédération, Président,
- le directeur adjoint de la fédération,
- les directeurs des Parties (ou un de leurs directeurs-adjoints ou un représentant désigné pour la représentation de la direction de chaque Partie),
- 8 membres, représentants des chercheurs et enseignants chercheurs, 4 de chaque Partie, désignés par les conseils de laboratoires et représentant la diversité des thématiques de recherche des Parties,
- 2 représentants des personnels BIATTS, un pour chaque Partie, choisis par l'ensemble des personnels BIATTS,
- 2 représentants des doctorants, un pour chaque Partie, choisis par l'ensemble des doctorants.

Le conseil de fédération comprendra au moins cinq représentants de chaque collège (collège B : maîtres de conférences, chargés de recherche et assimilés ; collège A : professeurs, directeurs de recherche et assimilés). Il comprendra aussi au moins un représentant de chaque tutelle. Son renouvellement se fait au début de chaque convention (**Article 2.**).

Article 14. Le conseil de fédération est une instance de concertation qui élabore la politique de la fédération. Il est tenu informé des décisions de la direction. Il se réunit au moins deux fois par an, à l'initiative de l'équipe de direction.

Article 15. Ce conseil est consulté sur l'état, le programme et l'orientation des recherches, les moyens budgétaires à demander par la fédération et la répartition de ceux qui lui sont alloués. Il assiste l'équipe de direction de la fédération dans l'organisation des activités communes, la négociation et la répartition des ressources de la fédération. Il élabore le projet scientifique de la fédération et l'éventuelle mise en place ou modification des axes de recherche prioritaires. Il prépare le renouvellement de la convention mentionnée dans **Article 2.**

Article 16. L'équipe de direction peut faire appel aux membres du conseil de fédération pour aider dans la mise en place des actions d'animation de la recherche établies par la fédération ainsi que les consulter, éventuellement par voie électronique, pour certaines prises de décision.

Article 17. Le président peut inviter toute personne extérieure à titre d'expert sur un point de l'ordre du jour. Tout membre de la fédération peut demander au

président du conseil de fédération l'inscription d'un sujet particulier dans l'ordre de jour d'une séance du conseil de fédération.

Article 18. Tout membre de la fédération peut assister à une séance du conseil de fédération.

Comité scientifique

Article 19. Il est institué auprès de la fédération un comité scientifique ainsi composé :

- l'équipe de direction de la fédération,
- les directeurs des Parties ou ses représentants,
- un représentant de chaque tutelle
- 4 personnalités scientifiques extérieures aux Parties, nommées d'un commun accord entre l'ensemble des Parties, sur proposition du comité de direction.

Le président du comité scientifique est élu lors de la première réunion de ce comité parmi les membres extérieurs sur proposition de l'équipe de direction de la fédération.

Article 20. Le comité scientifique de la fédération se réunit au moins deux fois pendant la durée de la convention quinquennale mentionnée dans l'**Article 2.** sur convocation de son Président ou, à défaut de ce dernier, sur convocation du directeur de la fédération. À cette occasion un rapport d'activité est présenté par le directeur de la fédération.

Article 21. Le comité scientifique de la fédération a pour mission :

- d'examiner les résultats obtenus et la répartition des moyens fédéraux,
- de discuter des orientations, des axes et de proposer des infléchissements,
- de rendre un avis sur le renouvellement de la fédération.

Moyens

Article 22. Les dotations financières octroyées par les tutelles sur une base récurrente, ainsi que les contributions de chaque Partie forment le budget propre de la fédération.

Divers

Article 23. Tout vote nominatif est un vote à bulletins secrets.

Article 24. Toute modification des présents statuts doit être approuvée par les assemblées générales des Parties.

Article 25. Il est possible de mettre fin au mandat du directeur ou du directeur adjoint de la fédération par l'obtention d'au moins deux tiers des voix des membres pleins dans chacune des assemblées générales extraordinaires des Parties.

COMPTE RENDU DE L'ÉLECTION POUR LA DIRECTION DE L'ICVL

Vous trouverez dans ce chapitre les extraits des emails faisant part des différentes étapes de la procédure de l'élection de la direction de l'ICVL.

H.1 Les élections pour la direction ICVL : candidature et procédure

Dans l'email NEWS ICVL OCTOBRE 2015

APPEL A CANDIDATURE - DIRECTION ICVL

Nous vous rappelons que l'appel à candidature pour la direction ICVL est ouvert. Les élections sont prévues pour fin novembre 2015, et la prise des fonctions de l'équipe élue pour février 2016. Yannick Parmentier (LIFO) et Nicolas Ragot (LI) ont eu la gentillesse d'accepter de coordonner ces élections. Actuellement, avec les conseils scientifiques, nous préparons la liste des électeurs. L'élection sera faite par voie électronique, Yannick ayant proposé l'utilisation du système helios (système open-source développé dans le cadre d'une thèse du MIT), dont une instance est hébergée en ligne et utilisable gratuitement : <https://vote.heliosvoting.org/> Comme les statuts prévoient un vote à majorité absolue pour la direction, ce choix convient bien.

H.2 Lettre de candidature

Envoyée par email aux deux laboratoires le 05 novembre 2015

CANDIDATURE POUR LA DIRECTION DE LA FEDERATION INFORMATIQUE CENTRE VAL DE LOIRE (ICVL)

Par la présente lettre, nous répondons à l'appel à candidature pour la direction de la fédération ICVL (Informatique Centre Val de Loire). Nous, Agata Savary et Mirian Halfeld Ferrari Alves, nous proposons en tant qu'équipe de direction.

Nous croyons que la fédération ICVL offre une chance d'apporter une vraie valeur ajoutée aux activités de recherche effectuées par nos deux laboratoires. Cependant sa réussite dépendra surtout de la motivation et de la participation de ses membres. C'est en partant de ce principe que nous avons oeuvré depuis le mois de juin 2014, en tant que binôme chef de projet, pour la mise en place largement consultative de cette fédération. En automne 2014, nous avons notamment mené une enquête auprès des membres du LIFO et du LI sur les valeurs communes à défendre au sein de la fédération, sur ses objectifs, sur son périmètre d'actions, comme sur les craintes vis-à-vis de ce projet. En janvier 2015, nous avons lancé un appel aux actions transversales et aux initiatives pour le soutien de ces actions, ainsi qu'aux propositions de nom pour la fédération. En mai 2015, nous avons animé les Journées Informatiques en Région Centre, avec des ateliers dédiés à l'analyse des besoins sociétaux, verrous scientifiques, techniques et objets d'études liés aux activités scientifiques des deux laboratoires. Finalement, suite aux résultats de ces ateliers, nous avons rédigé, en octobre dernier, la proposition de projet scientifique de la fédération.

En tant que candidates à la direction, nous avons l'intention de poursuivre le principe de développement et de gestion collaboratifs de la fédération. Nous avons l'ambition de lancer un débat commun sur la future politique de recherche, qui abordera notamment les moyens que l'ICVL souhaite se donner afin de poursuivre les objectifs documentés dans les statuts. Parmi des moyens possibles et déjà évoqués – mais à discuter, ensemble, en profondeur - se trouveraient, par exemple, la poursuite de structuration en actions et initiatives, la recherche de financements communs, des bourses de thèse dédiées aux collaborations inter-laboratoires, des formations au montage de projets européens, une politique commune de recrutements (PR, MCF, ATER,...), une contribution au dossier de demande d'affiliation au CNRS, ou encore la promotion d'éventuelles équipes inter-laboratoires ou des masters conjoints. Il serait également intéressant de se poser

la question de la sélection de thèmes de compétences prioritaires que la fédération souhaite afficher vers l'extérieur. D'autres éléments de cette politique commune pourront certainement être suggérés et débattus grâce à la participation active des membres de l'ICVL. En résumé, nous proposons que l'ICVL soit non seulement le cadre privilégié pour discuter des politiques scientifiques communes, mais aussi, parfois, un terrain d'essai pour de nouvelles stratégies à adopter. Assurer le débat et travailler vers le consensus est l'un de nos buts majeurs.

Parmi d'autres mesures nécessaires pour la poursuite du montage de la fédération, nous envisageons notamment : (i) de finaliser la rédaction du projet scientifique, (ii) de rédiger la convention, (iii) de mettre en place le conseil de la fédération et son comité scientifique, (iv) de créer toutes les procédures administratives nécessaires mentionnées dans les statuts, (v) de développer des moyens de communication (site web, etc.). Dans le cas où nous serions élues, nous avons la volonté de poursuivre un véritable travail en binôme, avec un partage équilibré des tâches, décisions et initiatives à prendre par chacune de nous. Nous proposons de prendre nos fonctions dans l'ordre suivant : Agata Savary en tant que directrice de la fédération et Mirian Halfeld Ferrari Alves en tant que vice-directrice, pour la première moitié du mandat. Mirian Halfeld Ferrari Alves en tant que directrice et Agata Savary en tant que vice-directrice pour la deuxième moitié du mandat. Mirian Halfeld Ferrari Alves est actuellement responsable de l'équipe Pamda du Laboratoire LIFO, mais elle s'engage à quitter cette fonction en cas d'élection à la direction de l'ICVL, au plus tard le 22 janvier 2016. Agata Savary n'exerce aucune fonction de direction de recherche au sein du laboratoire LI. Nous vous remercions par avance pour tout l'intérêt porté à notre candidature et pour la confiance de ceux qui voteront en notre faveur. Dans le cas où d'autres candidats seraient élus, nous serons bien sûr prêtes à collaborer pour transmettre l'expérience acquise lors du montage du projet ICVL.

Amicalement,

Agata Savary

Mirian Halfeld Ferrari Alves

H.3 Résultats

- Ci-dessous vous trouverez l'email envoyé par Yannick Parmentier au LIFO et Nicolas Ragot au LI le 9 décembre 2015 avec le compte rendu des élections ICVL.
- Les résultats ont été validés par le conseil du LI le 7 janvier 2016 et le conseil du LIFO le 08 janvier 2016.

Bonjour à tous,

le 2nd tour du vote pour la direction de la Fédération Informatique en Centre-Val-de-Loire est à présent clos.

Les résultats sont les suivants :

- * nombre d'inscrits: 148 (81 pour le LI, 67 pour le LIFO)
- * nombre de votants: 60
- * nombre de voix pour la candidature Halfeld / Savary : 57
- * nombre de bulletin blancs : 3

Félicitations à Agata et Mirian.

Leur nomination à la tête de la Fédération ICVL doit encore être validée en conseils de laboratoire du LI et LIFO courant janvier.

Merci à tous pour votre attention,
bonne fin de semaine

Yannick

On Mon, Dec 07, 2015 at 10:09:30AM +0100, Yannick Parmentier wrote:

> Bonjour à tous,

>

> suite aux résultats du vote de la direction de la fédération ICVL de
> la semaine passée (voir ci-dessous pour rappel), un 2nd tour a lieu à
> compter de ce jour et jusqu'à mercredi prochain 18h.

>

> Comme la dernière fois, vous avez du recevoir votre bulletin de vote
> électronique par email (contenant vos identifiants pour accéder au
> site de vote). Si tel n'était pas le cas, merci de me contacter.
>
> Bonne semaine à tous,
>
> Yannick
>
> PS: Nicolas, peux tu faire suivre ce message aux membres du LI,
> merci.
>
>
> On Wed, Dec 02, 2015 at 06:17:14PM +0100, Yannick Parmentier wrote:
>> Voici les résultats du vote pour la direction de la Fédération ICVL:
>>
>> * nombre d'inscrits: 148 (81 pour le LI, 67 pour le LIFO)
>> * nombre de votants: 70
>>
>> * nombre de voix pour la candidature Halfeld / Savary : 67
>> * nombre de bulletin blancs : 3
>>
>> Conformément au règlement du vote, la majorité absolue n'ayant pas été
>> atteinte, un second tour sera organisé la semaine prochaine de lundi à
>> mercredi selon les mêmes modalités.
>>
>> Merci pour votre attention,
>> bonne soirée
>>
>> Yannick
>>