AfIA

Association française
pour l'Intelligence Artificielle

Bulletin N° 97

Association française pour l'Intelligence Artificielle

AfIA

AfIA
Association française
pour l'Intelligence Artificielle

PRÉSENTATION DU BULLETIN

Le Bulletin de l'**A**ssociation **f**rançaise pour l'**I**ntelligence **A**rtificielle vise à fournir un cadre de discussions et d'échanges au sein de la communauté universitaire et industrielle. Ainsi, toutes les contributions, pour peu qu'elles aient un intérêt général pour l'ensemble des lecteurs, sont les bienvenues. En particulier, les annonces, les comptes rendus de conférences, les notes de lecture et les articles de débat sont très recherchés. Le Bulletin de l'AfIApublie également des dossiers plus substantiels sur différents thèmes liés à l'IA. Le comité de rédaction se réserve le droit de ne pas publier des contributions qu'il jugerait contraire à l'esprit du bulletin ou à sa politique éditoriale. En outre, les articles signés, de même que les contributions aux débats, reflètent le point de vue de leurs auteurs et n'engagent qu'eux-mêmes.

Édito

Le bulletin de ce trimestre est abondant et nous remercions tous les auteurs des articles pour leur travail. Le dossier du bulletin est consacré aux équipes académiques, nous avons eu le plaisir de recevoir les présentations de 10 équipes de recherche en lien avec l'intelligence artificielle. Dans ce bulletin, vous trouverez également le compte-rendu des événements reliant l'intelligence artificielle à une autre discipline : BioSS et IA est consacré aux travaux sur la représentation formelle et le raisonnement automatique appliqués à l'étude des systèmes vivants ; EIAH et IA est consacré aux liens entre le domaine des environnements interactifs pour l'apprentissage humain et l'IA ; IM-IA autour de l'Imagerie Médicale et l'IA. Nous relatons également le Forum Industriel et IA organisé par l'AFIA et la journée ROD qui est un événement lançant le groupe de travail Reasoning on Data du pré-GdR Aspects Formels et Algorithmiques de l'IA.

Ce bulletin contient aussi deux hommages à deux chercheurs en Intelligence Artificielle disparus ce trimestre : Alain COLMERAUER et Daniel KAYSER. Les articles ont été rédigés respectivement par Vincent RISCH et Adeline NAZARENKO, ils présentent le cheminement de ces deux hommes et l'impact de leurs travaux sur le développement de l'intelligence artificielle française et internationale.

Bonne lecture à tous !

Florence BANNAY,

AfIA
Association française
pour l'Intelligence Artificielle

SOMMAIRE
DU BULLETIN DE L'AfIA

AfIA
Association française
pour l'Intelligence Artificielle

Disparition d'Alain Colmerauer

Par **Vincent RISCH**
LSIS - CNRS UMR 7296
Aix–Marseille Université
vincent.risch@univ-amu.fr

Alain Colmerauer nous a quitté en mai dernier à l'âge de 76 ans, des suites d'un malaise soudain.

Professeur d'informatique à la Faculté des Sciences de Luminy (Université Aix-Marseille) depuis 1970, membre de l'Institut Universitaire de France à partir de 1999, il a profondément marqué l'essor de l'Intelligence Artificielle en France et dans le monde, par la création du langage Prolog, et par ses travaux novateurs dans le domaine des CSP.

Après des études à l'Institut de Mathématiques Appliquées de Grenoble, et la soutenance d'une thèse d'état sur les grammaires de précédence en 1967, la carrière d'Alain Colmerauer débute brillamment au cours d'un séjour de trois ans à l'Université de Montreal : à 28 ans, il y est nommé responsable du projet *TAUM*, destiné à la réalisation d'un des premiers prototypes de traduction automatique entre français et anglais. Il développe à cette époque les *systèmes-Q*, un formalisme de réécriture sur les arbres qui peut être vu comme l'ancêtre de Prolog.

De retour en France en 1970, Alain Colmerauer est nommé Professeur à la Faculté des Sciences de Luminy. Il s'entoure alors d'une équipe de quelques chercheurs avec lesquels il constitue le *Groupe d'Intelligence Artificielle*, au sein duquel il poursuit ses recherches sur le traitement automatique du langage naturel. C'est dans ce cadre qu'en 1972, faisant le lien avec les travaux de Robert Kowalsky dans le domaine de la démonstration automatique, il créée avec Philippe Roussel le premier interpréteur d'un langage de programmation de conception radicalement nouvelle : Prolog (pour *PROgrammation LOGique*). A vrai dire plus qu'un langage, Prolog représentera un paradigme permettant de penser la programmation d'une façon totalement différente, et une vraie révolution en Intelligence Artificielle.

Cherchant à étendre les possibilités du premier Prolog, Alain Colmerauer définira ensuite une élégante algèbre permettant de raisonner en termes d'équations/inéquations sur des arbres infinis. Il élargira cette approche en dotant Prolog d'une assise arithmétique qui le conduira d'abord à inclure dans le courant des années 90 le traitement des *contraintes numériques*, puis à doter le langage d'une nouvelle algèbre sur les *intervalles*. Au delà de la création de Prolog, il apparaît ainsi aussi comme l'un des fondateurs du champs de recherches autour des CSP.

Ayant été particulièrement soucieux tout au long de sa vie d'une interaction entre recherche théorique et applications, Alain Colmerauer orientera pourtant ses derniers travaux vers des questions fondamentales de calculabilité, en proposant un nouveau cadre formel permettant l'évaluation des complexités de plusieurs programmes universels sur différentes machines.

Outre les contributions directes d'Alain Colmerauer dans le traitement des langues naturelles, de la démonstration automatique, et des CSP, il est remarquable de noter que certaines des solutions qu'il a apportées dans ces domaines ont eu une influence réelle sur des travaux menés dans des domaines connexes. Ainsi par exemple du mécanisme de *cut* de Prolog : intégré au langage pour des raisons pratiques d'effectivité, sa mise en oeuvre revient sur le plan théorique à éliminer des modèles, approche qui a participé de recherches originales autour du traitement de la *négation par échec* dans le champs de la Représentation des Connaissances et des ASP.

Alain Colmerauer a été récompensé par de nombreuses distinctions : Pomme d'Or du logiciel (1982) ; Prix Michel Monpetit (1985) de l'Académie des Sciences ; Chevalier de la légion d'honneur (1986) ; Membre associé AAAI (1991) ; Correspondant de l'Académie des Sciences.

Par sa compréhension profonde des problèmes scientifiques de sa discipline, et une approche résolument novatrice de ces problèmes, Alain Colmerauer marque de manière unique la jeune histoire de l'Intelligence Artificielle. Ceux qui ont eu la chance de le connaître retiennent d'Alain la limpidité de sa pensée, qu'illustrait notamment la recherche systématique de solutions d'une merveilleuse élégance. De cette élégance qu'il a su nous faire partager, qu'il soit remercié.

Disparition de Daniel Kayser

Par **Adeline NAZARENKO**
LIPN - CNRS UMR 7030
Universite Paris 13 – Sorbonne Paris Cité
adeline.nazarenko@lipn.univ-paris13.fr

Daniel Kayser est décédé en mai 2017 des suites d'une très longue maladie.

Il fut Professeur d'informatique à l'Université Paris 13. Avec quelques autres, il est de ceux qui ont créé le Laboratoire d'Informatique de Paris-Nord en 1986 et il en a été le premier directeur. Il a su y imprimer sa marque et en faire un laboratoire reconnu à la fois pour la qualité des recherches qui y sont menées et parce qu'il y fait bon vivre et travailler.

Au sein de l'Université Paris 13 et beaucoup plus largement, Daniel Kayser a oeuvré pour la mise en place de formations d'informatique exigeantes et cohérentes, à une époque où la discipline informatique était émergente et était souvent mal considérée. Il a perçu très tôt l'importance de l'informatique et a largement contribué à la faire reconnaître comme discipine scientifique autonome.

Pour autant, il a toujours été convaincu de l'importance du dialogue interdisciplinaire avec les mathématiques, la linguistique, les sciences cognitives, etc. Il a joué un rôle pionnier dans la constitution de ces dernières. De controverses en amitiés, il a échangé avec des scientifiques de toute discipline et de tout pays. Son travail de recherche l'a conduit aux frontières du traitement automatique des langues, de la représentation des connaissances et de l'intelligence artificielle et lui a valu une reconnaissance nationale et internationale dans ces domaines.

Daniel Kayser a contribué à former toute une génération de chercheurs. Par ses travaux, aux nombreux étudiants qu'il a encadrés et à tous les chercheurs plus ou moins jeunes qu'il a accompagnés, il a transmis des idées importantes dans ce domaine de l'intelligence artificielle auquel il n'a cessé de réfléchir : une certaine vision de la sémantique et du raisonnement.

En marge de ses cours et de ses recherches, Daniel Kayser a assumé de nombreuses responsabilités, au niveau local, national ou international. Il a répondu présent quand on avait besoin de lui mais il avait le souci de passer la main après quelques années et de rentrer dans le rang. Sa modestie, le souci des étudiants et l'attrait de la recherche l'ont sans cesse ramené sur le « terrain » et à ce qui lui tenait le plus à coeur : comprendre et transmettre.

Daniel Kayser a toujours été à la fois enseignant et chercheur. C'était aussi un intellectuel : il n'avait peur ni des questions, ni du dialogue, ni même de la controverse mais il restait à l'écoute, curieux et à l'affût. En un mot, c'était un grand universitaire. C'était un esprit libre, exigeant, droit.

La maladie l'a soustrait brutalement il y a plus de 5 ans à la communauté scientifique à laquelle il tenait et pour qui il comptait tant.

Il a été présent pendant des années pour beaucoup d'entre nous, il nous a formés mais il nous a aussi laissé libres. C'était un universitaire, au sens le plus noble du terme.

Biographie universitaire

Pionnier français de l'Intelligence Artificielle et des sciences cognitives, spécialisé en représentation des connaissances pour la compréhension du langage naturel, Daniel Kayser a été Professeur à l'IUT d'Orsay de 1975 à 1984, puis Professeur à l'Université Paris Nord à dater d'octobre 1984. Membre fondateur du LIPN en 1986, il en est le premier directeur jusqu'en 1991. Il est aussi responsable de l'équipe Représentation des Connaissances et Langage Naturel de 1986 à 2002. Président de l'Association pour la Recherche Cognitive de 1992 à 1994 (il en a été l'instigateur avec Mario Borillo), directeur du Programme de Recherche Concertée « Intelligence Artificielle » de 1992 à 1995, c'est un acteur de la recherche influant et reconnu (membre d'honneur de l'Association Française d'Intelligence Artificielle en 2000, ECCAI Fellow en 2002). Il a été directeur de l'école doctorale de Galilée (1997-2001), responsable du département d'informatique de l'Institut Galilée (2002-2004), membre du conseil supérieur des universités (1986-87) et du comité national de la recherche scientifique (2003-2004) et il a assuré la présidence du comité scientifique disciplinaire n°1 de l'ANR de 2007 à 2009.

AfIA
Association française
pour l'Intelligence Artificielle

Dossier

« Présentation d'équipes de recherche Académiques »

Dossier réalisé par

Florence BANNAY
IRIT/ ADRIA
Université Paul Sabatier, Toulouse 3
florence.bannay@irit.fr

AfIA
Association française
pour l'Intelligence Artificielle

Un parcours des thèmes de l'IA au travers 10 équipes de recherche françaises

Par **Florence BANNAY**
IRIT/ADRIA
Université de Toulouse
bannay@irit.fr

Chaque année le dossier du bulletin de l'AfIA de juillet est consacré aux équipes de recherche académiques ou industrielles ayant des activités en IA. Cette année, c'est le tour des équipes académiques, nous avons donc invité les équipes qui le souhaitaient à présenter leurs recherches, les outils qu'elles développent et leurs projets. C'est une occasion pour les équipes de communiquer sur leurs activités, et au travers de ce bulletin de participer à faire le point sur la cartographie des recherches menées en France sur la thématique de l'IA. Nous présentons donc ici des équipes qui n'ont jamais été décrites dans le bulletin de l'AfIA ou des équipes dont la présentation avait été faite il y a plus de 5 ans.

Dix équipes de recherche académiques ont répondu positivement à notre sollicitation. Au travers de ces présentations, nous parcourons la France et ses grands laboratoires l'IRISA de Rennes est représenté par deux équipes : DRUID et SemLIS. Le LORIA de Nancy est également représenté par deux équipes : MultiSpeech et SYNALP. Les deux laboratoires de Toulouse l'IRIT et le LAAS sont représentés respectivement chacun par une équipe LILAC et ROC. En continuant dans le sud, le laboratoire LGI2P des Mines d'Alès est représenté par l'équipe KID. Le LIRMM de Montpellier par l'équipe ADVANSE. Le nouveau laboratoire LIS de l'Université d'Aix-Marseille est maintenant doté d'une nouvelle équipe consacrée aux approches logiques de l'intelligence artificielle et nommée LIRICA. Enfin la région parisienne est présente grâce à l'équipe CPU du laboratoire LIMSI à Orsay.

Cette diversité géographique des équipes reflète également une belle couverture des thèmes de l'intelligence artificielle de l'extraction des connaissances jusqu'à leur représentation en logique ou sous-forme de graphes et d'ontologies, du traitement automatique du langage naturel au raisonnement logique en passant par l'apprentissage profond ou les systèmes multi-agents...

Bonne visite guidée !

AfIA
Association française
pour l'Intelligence Artificielle

ADVANSE : ADVanced Analytics for data SciencE

LIRMM
Université Montpellier-CNRS
https://www.lirmm.fr/recherche/equipes/advanse

Pascal PONCELET
poncelet@lirmm.fr
+33 4 67 41 85 00

Membres :

- Jérome AZÉ, PR
- Sandra BRINGAY, PR
- Dino IENCO, CR IRSTEA
- Pascal PONCELET, PR
- Pierre POMPIDOR, MCF
- Arnaud SALLABERRY, MCF
- Bilel MOULAHI, Post-doc
- Amin ABDAOUI, Doctorant
- Erick CUENCA, Doctorant
- Samiha FADLOUN, Doctorant
- Jessica PINAIRE, Doctorant
- Mike TAPI NZALI, Doctorant

Mots-clés

- Extraction et gestion de connaissances
- Fouille de données
- Motifs séquentiels
- Entrepôts de données
- Logique floue
- Ontologie
- Annotation automatique
- Bases de données

Thématique générale de l'équipe

Depuis sa création en 2014, l'équipe projet Advanse (Advanced ANalytics for Data SciencE) s'intéresse au processus d'extraction de connaissances. Ce processus interactif et itératif consiste à appliquer différentes étapes à un corpus de données : sélection des données, fusion de données, représentation des données, fouille des données, visualisation et validation des connaissances acquises. Cependant, étant donné la complexité des données liées aux nouvelles applications (structurées, semi-structurées, multidimensionnelles, qualitatives, quantitatives, textuelles, spatio-temporelles, approximatives, bruitées, etc.) et le fait que celles-ci soient nombreuses et évoluent rapidement au cours du temps, ce processus nécessite d'être adapté. L'objectif de l'équipe est de proposer de nouveaux algorithmes et méthodes pour faire face à ces nouveaux défis. Plus particulièrement, nous nous intéressons aux approches de fouille de données (extraction de motifs, apprentissage) et de visualisation d'informations (représentations visuelles, interactions). Nous portons une attention particulière à l'intégration de connaissances expertes au plus tôt dans ce processus, afin d'extraire des connaissances pertinentes pour le décideur. La visualisation a, quant à elle, non seulement pour but de rendre accessible les connaissances extraites par la fouille, mais aussi d'être intégrée aux systèmes de fouille comme support de l'extraction (Visual Analytics). L'équipe Advanse se focalise particulièrement sur les données liées à la santé et à l'environnement.

Projets marquants

Des motifs complexes...

Depuis de nombreuses années, l'équipe possède une expertise reconnue dans le domaine de l'extraction de motifs (notamment des motifs séquentiels). Nos travaux se sont poursuivis en abordant la réduction du nombre de motifs extraits. Menés dans le cadre de la thèse de Mickael Fabrègue [4] et du projet d'ANR Fresqueau, un nouvel algorithme d'extraction de motifs partiellement ordonnés clos [1] a été proposé. Ces motifs ont la particularité de pouvoir être représentés sous la forme d'un graphe facilement interprétable par des experts. Alors que les motifs séquentiels ne considèrent qu'une relation d'ordre, nous les avons étendus dans le cadre de la

1. Un motif partiellement ordonné est un motif permettant de représenter des ordres partiels entre ses événements. Ce motif est clos s'il est fréquent et s'il n'existe pas de super-motifs le contenant avec le même support.

thèse de Hugo Alastrista Salas [3], pour prendre en compte la dimension spatiale et obtenus des motifs dits « spatio-temporels ». Ces motifs permettent de déterminer au cours du temps les occurrences des événements et leur localisation pour, par exemple, mettre en évidence des corrélations. Une partie de ces travaux a notamment été menée avec l'institut National de Veille Sanitaire pour mieux appréhender les évolutions des épidémies de dengue. Les données spatio-temporelles peuvent également être considérées comme des trajectoires d'objets mobiles. Dans la thèse de Nhat Phan Hai [5], une nouvelle approche d'extraction unifiée de la plupart des trajectoires de la littérature a été proposée. Pour chacun de ces travaux, nous nous sommes intéressés à la recherche des motifs les plus pertinents, les top-k motifs, en fonction d'objectifs définis par les experts. Par exemple, via de nouvelles mesures pour les motifs partiellement ordonnés clos ou, pour l'extraction de trajectoires, d'une approche inspirée des Minimum Length Description. Ces travaux se sont poursuivis par l'application des approches d'extraction de motifs contextuels et de trajectoires sur des données de patients dans le cadre de la thèse de Jessica Pinaire en collaboration avec le CHU de Nîmes. L'objectif est d'utiliser ces motifs pour prédire l'évolution des populations de patients et les coûts associés et de recommander les soins les plus efficaces et les moins coûteux. Dans la définition de ses algorithmes, l'équipe attache une attention particulière à la généricité des propositions.

Outre les aspects liés à l'extraction de motifs, l'équipe s'est également intéressée à différentes techniques de fouilles de données supervisées ou non. Par exemple, dans le cas de la détection de cellules rares, les approches de détection d'*outliers* n'étant pas adaptées, nous avons proposé un algorithme de classification non supervisée permettant de retrouver des ensembles d'objets fortement connectés.

Dans le cadre de ces travaux sur les motifs, la visualisation d'information est un aspect important et se focalise principalement sur la représentation visuelle et l'exploration de connaissances extraites. En parallèle de ces travaux, l'équipe a proposé de nouvelles méthodes de visualisation analytique (Visual Analytics). Les outils qui en résultent intègrent des méthodes d'analyse automatiques avec des méthodes de visualisation et d'interaction pour aider l'utilisateur à explorer les données et piloter les algorithmes d'analyse afin d'en extraire des connaissances. Par exemple, l'outil visuel HydroQual [2] a été conçu pour aider les experts à explorer dynamiquement les données afin d'en extraire automatiquement les motifs partiellement ordonnés clos en fonction de leur proximité géographique ou sémantique. Nous avons aussi proposé une nouvelle distance pour des séquences d'événements catégorisés, et une visualisation interactive permettant, d'une part de visualiser, mais aussi, de raffiner manuellement des clusters de séquences extraits automatiquement (https://www.youtube.com/watch?v=zbuNAG0oSMQ).

. . . aux réseaux

Nous nous sommes également intéressés au domaine des médias sociaux. L'objectif a été d'extraire de la connaissance à partir de ces données hétérogènes et notamment à partir de la structure des réseaux et des textes. Ces travaux ont été appliqués essentiellement dans le domaine de la santé. Ils ont été initiés dans le cadre du projet Parlons de nous, et poursuivis dans le cadre de plusieurs collaborations avec l'ICM et l'IMAG sur la thématique du cancer du sein, avec d'autres équipes du LIRMM dans le cadre de l'ANR SFIR, avec le LERASS-CERIC et le Département de virologie du CHU de Montpellier sur la thématique des nouvelles pratiques sexuelles à risque.

La construction et l'enrichissement automatique de ressources biomédicales (lexiques, ontologies, etc.) à partir de textes représente un élément clé pour l'analyse, la classification ou l'annotation de données textuelles. Ceci engendre des problèmes difficiles liés à l'extraction de la terminologie à partir de textes, la désambiguïsation lexicale et le peuplement d'ontologies biomédicales qui ont été étudié pendant la thèse de Juan Antonio Lossio [8], co-encadrée avec l'équipe SMILE, dans le cadre de l'ANR SIFR. Dans le cadre de la thèse de Mike Tapi NZali [7], nous avons mis en relation des connaissances avérées et des connaissances co-construites par les internautes. Pour cela, nous avons exploré des méthodes supervisées, semi-supervisées et non supervisées pour décrire les productions des patients

selon différentes dimensions (qui, quand, où et comment ?). Ces méthodes se sont avérées efficaces grâce à la construction originale et automatique d'une ressource mettant en relation le vocabulaire des patients et celui des professionnels de santé. Dans le cadre de la thèse d'Amine Abdaoui, nous avons produit un nouveau lexique dédié à l'analyse des sentiments en français qui a permis de positionner l'équipe en première place sur une des tâches du challenge DEFT 2016 et deuxième en 2017. L'équipe s'est également intéressée à la fiabilité des connaissances issues du Web Social. S'il est difficile d'empêcher les internautes de consulter des informations non pertinentes ou non fiables dans les médias sociaux, il est possible, en revanche, de concevoir des outils pour mettre en évidence des informations de qualité dans ces médias. Nous avons proposé différentes approches pour détecter automatiquement le rôle (expert, non expert) et la confiance que l'on peut accorder aux internautes dans les médias sociaux selon les thématiques et au cours du temps [1].

Pour finir, nous avons proposé des chaînes de traitements pour monitorer de manière globale la santé des populations. Nous avons combiné différentes méthodes pour capter les changements thématiques ou liés aux sentiments, brusques ou plus insidieux (e.g. votes entre différents classifieurs, détection de concepts *drift*, *active learning* pour prendre en compte les réactions des experts en charge du monitoring, etc.). Ici, les questions éthiques et juridiques seront centrales (e.g. vie privée, consentement). Le projet DontDoIt sur la détection des personnes à risque suicidaires en collaboration avec le département des Urgence Psychiatrique du CHU de Montpellier a également débuté avec le financement de post-doctorant UM de Bilel Boulahi.

Enfin, dans le cadre de la thèse de Samiha Fadloun, nous étudions les modes de visualisation et de navigation à l'intérieur d'un corpus de données dynamique et hétérogène. Le but est de fournir à des experts en maladie animale une plate-forme de surveillance du risque d'épidémiologie. Ce projet s'inscrit dans une approche de visualisation analytique. Nous sommes en train de mettre en place la plateforme dédiée. Elle inclut de nouveaux algorithmes de placement et de dessin d'éléments graphiques, assez génériques pour être étendus à d'autres applications.

Outre l'aspect analyse des médias sociaux, l'équipe s'intéresse également à l'étude de réseaux modélisés à l'aide de graphes. Les travaux menés ont des applications dans de nombreux domaines comme la biologie (réseaux de cellules, réseaux métaboliques, ...), la sociologie (réseaux sociaux), chimie (molécules et interactions), image satellitaires (réseau de segments adjacents), données ouvertes (réseau RDF). Outre le fait que ces graphes sont évolutifs et de plus en plus volumineux, ils possèdent la particularité de pouvoir comporter différents attributs aussi bien sur les sommets que sur les arêtes. Dans ce cadre, nous abordons les différentes étapes du processus d'extraction.

Tout d'abord, les graphes doivent être stockés et l'élaboration de modèles et de structures dédiées sont nécessaires pour le stockage et l'accès aux informations. Pendant la thèse de Vijay Ingallali, nous avons abordé la problématique d'interrogation de bases de données de graphes en proposant des méthodes efficaces d'indexation pour améliorer les performances des requêtes sur des graphes multivariés [2]. Nous avons proposé un système d'indexation de la base et de requêtage efficace pour rechercher les occurrences d'un graphe dans la base (isomorphisme de graphe). Par la suite, nous avons montré que les données RDF peuvent être considérées comme des graphes multivariés et que la problématique peut être transformée par bijection au problème d'homomorphisme de sous-graphes multivariés. Nous développons actuellement des algorithmes pour extraire des sous-graphes fréquents à partir de ces données et proposons de nouvelles techniques d'élagage pour parcourir l'espace de recherche ainsi que de nouvelles mesures de support plus adaptées.

L'analyse de réseaux comprend aussi d'autres aspects. Dans un premier temps nous nous sommes intéressés à la structure des graphes, et en particulier des graphes petits-mondes et sans échelle. Nous

2. Un graphe multivarié est un graphe dont les arêtes et les sommets sont munis d'une ou plusieurs propriétés qualitatives et/ou quantitatives.

nous sommes particulièrement focalisés sur la génération de tels graphes pour fournir des données de *benchmarking*. Nous avons ainsi proposé deux algorithmes différents de génération de ces graphes. Dans le cadre de la collaboration avec l'Université de Calabria (visite de Roberto Interdonato) nos recherches se sont intéressées à la détection de communautés locales dans des graphes multi-couches (différents types d'arêtes entre les sommets) et avons proposé différents algorithmes pour extraire automatiquement ces communautés.

Étant donné la volumétrie des graphes manipulés, l'un des aspects importants est la visualisation. Nous nous sommes intéressés à la visualisation de graphes évoluant au cours du temps. Les méthodes proposées permettent ainsi d'explorer interactivement des graphes dynamiques comportant des dizaines de milliers de sommets. En partant de l'observation d'André Skupin selon laquelle les formes des cartes géographiques sont plus facilement mémorisables que des formes plus abstraites, nous avons proposé un algorithme de visualisation de hiérarchies (arbres) dynamiques sous forme de cartes géographiques. Plus récemment, nous nous sommes focalisés sur les graphes multi-couches. Nous avons proposé un algorithme de « bundling » d'arêtes[3] permettant de séparer les différents types d'arêtes. Comme le paradigme noeud-lien n'est pas forcément le plus adapté à ce type de graphe, nous avons aussi proposé une méthode basée sur des simples diagrammes. Enfin, les ContactTrees sont des approches égocentriques originales permettant d'explorer les relations d'un acteur d'un réseau tout en représentant un grand nombre d'attributs. Précédemment, nous avons mis en évidence que la combinaison d'analyse automatique et de visualisations interactives (Visual Analytics) était une approche utile pour extraire des connaissances. Dans ce cadre, deux autres exemples sont en lien avec l'étude des réseaux. Tout d'abord nous avons proposé une méthode de visualisation analytique pour la détection de cellules rares [6]. Ces travaux se poursuivent actuellement dans la thèse d'Erick Cuenca où nous proposons, via différentes visualisations interactives, une exploration de gros graphes multi-variés. Le but est d'arriver à caractériser les différentes zones de ces graphes grâce à l'identification des motifs qu'elles comportent.

Bibliographie

Références

[1] Amine Abdaoui, Jérôme Azé, Sandra Bringay, and Pascal Poncelet. Collaborative content-based method for estimating user reputation in online forums. In Jianyong Wang, Wojciech Cellary, Dingding Wang, Hua Wang, Shu-Ching Chen, Tao Li, and Yanchun Zhang, editors, *Web Information Systems Engineering - WISE 2015 - 16th International Conference, Miami, FL, USA, November 1-3, 2015, Proceedings, Part II*, volume 9419 of *Lecture Notes in Computer Science*, pages 292–299. Springer, 2015.

[2] Pierre Accorsi, Nathalie Lalande, Mickaël Fabrègue, Agnès Braud, Pascal Poncelet, Arnaud Sallaberry, Sandra Bringay, Maguelonne Teisseire, Flavie Cernesson, and Florence Le Ber. Hydroqual : Visual analysis of river water quality. In Min Chen, David S. Ebert, and Chris North, editors, *2014 IEEE Conference on Visual Analytics Science and Technology, VAST 2014, Paris, France, October 25-31, 2014*, pages 123–132. IEEE Computer Society, 2014.

[3] Hugo Alatrista-Salas, Sandra Bringay, Frédéric Flouvat, Nazha Selmaoui-Folcher, and Maguelonne Teisseire. Spatio-sequential patterns mining : Beyond the boundaries. *Intell. Data Anal.*, 20(2) :293–316, 2016.

[4] Mickaël Fabrègue, Agnès Braud, Sandra Bringay, Florence Le Ber, and Maguelonne Teisseire. Mining closed partially ordered patterns, a new optimized algorithm. *Knowl.-Based Syst.*, 79 :68–79, 2015.

[5] NhatHai Phan, Pascal Poncelet, and Maguelonne Teisseire. All in one : Mining multiple movement patterns. *International Journal of Information Technology and Decision Making*, 15(5) :1115–1156, 2016.

3. Le « bundling » d'arêtes est une méthode de réduction d'encombrement de dessins de graphes consistant à regrouper les arêtes similaires, i.e. les arêtes spatialement proches.

[6] Eniko Szekely, Arnaud Sallaberry, Faraz Zaidi, and Pascal Poncelet. A graph-based method for detecting rare events : Identifying pathologic cells. *IEEE Computer Graphics and Applications*, 35(3) :65–73, 2015.

[7] Mike Donald Tapi-Nzali, Jérôme Azé, Sandra Bringay, Christian Lavergne, Caroline Mollevi, and Thomas Opitz. Formalisation semi-automatique d'un vocabulaire patient/médecin dédié au cancer du sein. *Revue d'Intelligence Artificielle*, 30(5) :533–556, 2016.

[8] Juan Antonio Lossio Ventura, Clement Jonquet, Mathieu Roche, and Maguelonne Teisseire. Biomedical term extraction : overview and a new methodology. *Inf. Retr. Journal*, 19(1-2) :59–99, 2016.

CPU : Cognition, Perception, Usages

Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur
CNRS, Orsay
https://www.limsi.fr/fr/recherche/cpu

Nicolas SABOURET
Nicolas.Sabouret@limsi.fr
+33 1 69 85 81 02

Jean-Claude MARTIN
Jean-Claude.Martin@limsi.fr
+33 1 69 85 80 00

Membres impliqués : [4]

- Jean-Claude MARTIN, PR
- Nicolas SABOURET, PR
- Vincent BOCCARA, MCF
- Céline CLAVEL, MCF
- Virginie DEMULIER, MCF
- Élise PRIGENT, MCF
- Mathieu JÉGOU, Post-doc
- Lauriane HUGUET, Doctorant
- Lydia OULD OUALI, Doctorant
- Alya YACOUBI, Doctorant
- Guillayme DEMARY, Doctorant

Mots-clés

- Cognition Humaine et Artificielle
- Interaction Affective Multimodale
- Modélisation des Variabilités Intra- et Inter-individuelles
- Conception et Usages

Thématique générale de l'équipe

L'équipe CPU est une équipe pluridisciplinaire regroupant des chercheurs en informatique, en psychologie-sociale et en ergonomie-cognitive. Elle développe des recherches sur la cognition humaine et la conception, l'évaluation et les usages des interactions homme-machine. En particulier, elle propose des modèles informatiques de la cognition humaine et de l'activité humaine qui sont utilisés dans des applications allant de la formation à l'étude du comportement.

Les principales activités en Intelligence Artificielle de l'équipe CPU sont donc situées dans le domaine de la simulation de l'humain. Les thématiques développées sont celles de l'informatique affective et de la simulation multi-agents.

Informatique Affective

L'Informatique Affective regroupe l'ensemble des travaux en Sciences de l'information (intelligence artificielle, robotique interactive, interaction homme-machine) et en Sciences Humaines et Sociales qui s'intéressent à la reconnaissance automatique, à la synthèse par et/ou à la modélisation informatique des phénomènes affectifs (émotions, personnalité, relations sociales).

Les recherches conduites par l'équipe CPU dans

4. Seules sont représentées les personnes impliquées dans la thématique IA.

cet axe concernent l'étude et la modélisation des composantes cognitives dans les agents artificiels et chez les humains, notamment en interaction avec les systèmes informatiques. Nous nous intéressons à la dimension affective dans ces processus, par opposition à d'autres approches plus classiques qui s'intéressent à l'aspect rationnel du comportement humain.

Nos travaux poursuivent ici plusieurs objectifs :

- Comprendre les processus cognitifs, affectifs et sociaux sous-jacents aux comportements des individus dans des situations données (par exemple dans une situation d'apprentissage, dans une situation de jeu ou dans une situation de crise) ;
- Comprendre comment l'humain intègre les informations multimodales affichées par d'autres humains ou par des agents conversationnels animés ;
- Simuler les processus de décision et de communication qui définissent le comportement humain et prendre en compte, dans ces processus, la part non-rationnelle (par exemple les erreurs de communication provoquées par le stress ou l'influence d'une émotion sur une prise de décision).
- Modéliser, concevoir et évaluer ces agents conversationnels animés capables d'exprimer des émotions, des attitudes sociales, des personnalités variées et crédibles.

Les techniques d'IA utilisées sont à la fois issues de la représentation des connaissances et de l'apprentissage automatique. Ainsi, l'équipe CPU a développé depuis plusieurs années la plate-forme d'agents conversationnels MARC [2] qui intègre un modèle d'évaluation cognitive basé sur la théorie CPM de Klaus Scherer [11]. Ce modèle est construit à partir d'un ensemble de règles sémantiques appliquées à la situation d'interaction pour construire une expression faciale d'émotion. Par exemple : est-ce que la situation est surprenante, est-ce qu'elle est conforme à mes buts, est-ce qu'elle est contrôlable, *etc.* Les modèles de raisonnement automatique que nous construisons dans ce type d'architecture sont basés sur de la logique floue ou de la logique modale. Nous avons, par exemple, proposé un modèle général de théorie de l'esprit basé sur la logique BDI [1].

Nous utilisons aussi largement des techniques basées sur l'apprentissage automatique à partir de corpus multi-modaux pour entraîner des systèmes de reconnaissance de comportements sociaux. Ainsi, dans la thèse de Tom Giraud [7], nous avons étudié les caractéristiques socio-émotionnelles telles que l'alexythimie [5], l'anxiété ou encore la gestion du stress dans le cadre d'interactions sociales du type entretien d'embauche ou interaction corporelle avec un coach de sport virtuel.

Enfin, le comportement des agents conversationnels peut être aussi construit à partir de corpus d'interaction. Dans la thèse de Caroline Faur [5, 4], nous avons proposé un modèle informatique de la personnalité basé sur la théorie du Regulatory Focus [8].

Simulation Multi-Agents

La simulation multi-agents intervient dans deux aspects de nos recherches en intelligence artificielle. Tout d'abord, nous nous appuyons sur la simulation multi-agents pour étudier les phénomènes collectifs. Un des objectifs de cet axe est de développer des modèles informatiques de la théorie de l'esprit pour intégrer la représentation de l'autre (système ou humain) dans la prise de décision, le raisonnement, l'adaptation du comportement et pour mieux contrôler l'interaction. Ainsi, dans la thèse de Lauriane Huguet (2015-18), nous proposons des modèles informatiques de prise de décision et de communication en situation de stress pour une équipe de secouristes virtuels. Dans la thèse de Guillaume Demary (2015-18), nous étudions l'impact du style de followership sur le comportement et la capacité des leaders d'équipes médicales à en détecter les indices non-verbaux chez des subordonnés virtuels. Ces deux modèles servent de support pour la formation de leaders d'équipe médicale d'urgence dans le cadre du projet ANR VICTEAMS.

Nous nous intéressons aussi à la simulation multi-agents de l'activité humaine, par exemple pour l'étude et la réduction de la consommation énergétique. Ainsi, dans la thèse de Thomas Huraux (2012-2015) réalisée en co-encadrement avec EDF, nous avons proposé un modèle qui permet de regrouper au sein d'une même simulation in-

5. Difficulté à identifier et exprimer des émotions.

formatique des experts issus de domaines scientifiques très différents (ergonomie, énergétique, thermique...). La méthodologie adoptée combine une approche de simulation participative avec des données statistiques sur les populations pour générer des diagrammes d'activité et les coupler avec des données de consommation (Figure 1). Ainsi, nous avons proposé une méthode de génération de populations synthétiques pour la simulation multi-agents qui s'appuie sur les données statistiques de l'INSEE sur la population française et sur les « enquêtes emploi-du-temps » [10].

Projets marquants

Les principales thématiques de recherche en Intelligence Artificielle que nous menons dans l'équipe CPU du LIMSI sont illustrées dans le projet ANR VICTEAMS. L'objectif de ce projet, piloté par Domitile Lourdeaux de l'UTC, est de réaliser un environnement virtuel de formation pour des équipes de secouristes. Dans ce projet, nous proposons des modélisations informatiques des phénomènes affectifs qui se produisent dans les situations d'urgence. Nous définissons des comportements non-verbaux qui rendent compte de différents styles de personnalité et de comportement d'équipe [3]. Nous construisons un modèle de prise de décision qui permet de reproduire des comportements observés sur le terrain, comprenant éventuellement des erreurs de décision ou de communication [9].

Figure 1 : Aperçu de l'environnement de formation du projet ANR VICTEAMS

Les travaux en Intelligence Artificielle de l'équipe CPU sont aussi présents dans le projet ANR MOCA dont l'objectif est d'étudier les compagnons artificiels (personnages virtuels et robots personnels) et leur valeur pour des usagers dans des situations de la vie de tous les jours. Les compagnons artificiels peuvent adopter diverses incarnations sur différents dispositifs : personnage virtuel sur un écran d'ordinateur, robot personnel, ou application sur terminal de poche, afin d'être au plus près de l'utilisateur dans toutes les situations de la vie quotidienne. Dans ce projet, le LIMSI a proposé des modèles informatiques pour doter ces compagnons artificiels de différentes personnalités.

Dans le projet ANR NARECA, nous nous intéressons à la modélisation du dialogue et des interactions avec un agent conversationnel animé. Le premier objectif de ce projet est d'étudier l'impact d'un agent conversationnel sur une tâche de dialogue simple (l'agent raconte une histoire à un enfant qui réagit). À partir des annotations du dialogue et des réactions de l'enfant face au narrateur artificiel, nous proposons un modèle de contrôle des expressions faciales et corporelles de l'agent virtuel. Dans ce projet, le LIMSI est en charge de la définition du modèle d'animation comportementale de l'agent [6].

Enfin, dans le projet SMACH réalisé en collaboration avec EDF R&D, nous nous sommes intéressés à la simulation multi-agents de l'activité quotidienne des ménages en rapport avec la consommation énergétique. Dans ce projet, le LIMSI travaille avec EDF sur la définition d'un modèle de simulation multi-agents multi-niveaux, permettant de rapprocher différentes expertises. Le modèle combine 3 dimensions : celle de la population (de l'individu au foyer), celle de l'activité (de l'action à l'habitude) et celle de l'environnement de consommation (de l'appareil à l'habitat). Il permet de simuler différentes politiques de gestion de l'énergie dans le contexte d'études menées auprès d'usagers en Bretagne et en Rhone-Alpes. Nous avons aussi défini un algorithme de génération automatique de scénario d'activité à partir de données "enquête emploi du temps" de l'INSEE.

Références

[1] Marwen Belkaïd and Nicolas Sabouret. Un modèle logique de théorie de l'esprit pour un agent virtuel dans le contexte de simulation d'entretien d'embauche. In *Proc. Workshop Affect, Compagnon Artificiel, Interaction (WACAI)*, 2014.

[2] M. Courgeon and J.-C. Clavel, C. and Martin. Modeling facial signs of appraisal during interaction ; impact on users' perception and behavior. In IFAAMAS, editor, *Proc. 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS'2014)*, 2014.

[3] Guillaume Demary, Virginie Demulier, and Jean-Claude Martin. Comment le leader perçoit-il les comportements non verbaux de son suiveur : une étude sur le processus de catégorisation du leader. In *Proc. 58eme congrès annuel de la société française de psychologie*, 2017.

[4] C. Faur, J.-C. Martin, and C. Clavel. Measuring chronic regulatory focus with proverbs : the developmental and psychometric properties of a french scale. *Journal of Personality and Individual Differences*, 107 :137–145, March 2017.

[5] Caroline Faur. *Approche computationnelle du regulatory focus pour des agents interactifs : un pas vers une personnalité artificielle*. PhD thesis, Univ. Paris-Sud, 2016.

[6] N. Fourati, A. Richard, S. Caillou, N. Sabouret, J.-C. Martin, E. Chanoni, and C. Clavel. Facial expressions of appraisals displayed by a virtual storyteller for children. In *16th International Conference on Intelligent Virtual Agents (IVA 2016)*, 2016.

[7] T. Giraud, F. Focone, V. Demulier, J.-C. Martin, and B. Isableu. erception of emotion and personality through full-body movement qualities : A sport coach case study. *ACM Trans. Appl. Percept.*, 13(1), 2015.

[8] Edward Tory Higgins. Promotion and prevention : Regulatory focus as a motivational principle. *Advances in experimental social psychology*, 30 :1–46, 1998.

[9] Lauriane Huguet, Domitile Lourdeaux, and Nicolas Sabouret. "errare humanum est" : simulation of communication error among a virtual team in crisis situation. In *Proc. 15th IEEE International Conference on Cognitive Informatics and Cognitive Computing (ICCI*CC)*, 2016.

[10] Quentin Reynaud, Yvon Haradji, François Sempé, and Nicolas Sabouret. Using time-use surveys in multi agent based simulations of human activity. In *Proc. International Conference on Agents and ARTificial intelligence (ICAART)*, 2017.

[11] Klaus Scherer. Appraisal considered as a process of multilevel sequential checking. *Appraisal processes in emotion : Theory, methods, research*, 92(120) :57, 2001.

AfIA
Association française
pour l'Intelligence Artificielle

DRUID : declarative and reliable management of uncertain, user-generated interlinked data

Laboratoire IRISA
Université Rennes 1
http://www-druid.irisa.fr

David GROSS-AMBLARD
dga@irisa.fr
+33 2 99 84 71 77

Arnaud MARTIN
arnaud.martin@irisa.fr
+33 2 96 46 94 60

Membres :

- Dorra ATTIAOUI, Doctorant
- Tristan ALLARD, MCF
- Tassadit BOUADI, MCF
- Yann DAUXAIS, Doctorant
- Tifenn DONGUY, Assistante de projet, CNRS
- Jean-Christophe DUBOIS, MCF
- Joris DUGUEPEROUX, Doctorant
- David GROSS-AMBLARD, PR
- Ian JEANTET, Doctorant
- Mouloud KHAROUNE, MCF
- Yolande LE GALL, MCF
- Na LI, Doctorant
- Arnaud MARTIN, PR
- Panagiotis MAVRIDIS, Doctorant
- Zoltan MIKLOS, MCF
- Virginie SANS, MCF
- Yiru ZHANG, Doctorant

Mots-clés

- Data analytics
- Data management
- Crowdsourcing
- Social networks
- Belief functions
- Databases
- Privacy
- Information fusion
- Preferences
- Big data

Team Overview

* Scientific and Historical Context at IRISA

The Druid team proposition results from a recent restructuring effort of the Data and Knowledge Management department (DKM – 7th research department) of IRISA. Our proposal builds a bridge between two domains – data management and data qualification – and two geographical locations of the IRISA lab : Rennes and Lannion.

* Scientific Challenges and Goals

Our perception of digital information has completely shifted in recent years, in several ways. First, data are no longer isolated, but are now part of distributed, **interlinked networks**. Such networks include web documents (URLs and URIs), communities in a social graph (*e.g.* FOAF), conceptual networks on the Semantic Web or the continuously growing network of Linked Open Data (RDF). Second, data are now dynamic. Obtaining an up-to-date piece of information is as simple as a Web service call or a syndication (as is RSS or Atom). A large diversity of such dynamic data sources is available, including corporate Web services, wireless sensors in the environment, humans in the participative Web, or workers in crowdsourcing platforms [6]. Hence, what becomes important is the **data source** itself.

The openness and liveliness of such interlinked data networks is a great opportunity. Business Intelligence applications no longer restrict their attention to the companies own data sets or sales records, but try to incorporate data collected from the Web (such as opinions from social networks or Web forums). In this way they can extract useful information about their customers and the reception of their products and services. Another domain is

6. https://www.mturk.com

the integration of personal information from multiple devices. The same opportunities arise also in the context of non-profit organizations or societal challenges : there is a lot of information available on health problems (Web forums on health, body area networks), environmental issues (environmental sensors) or in administrative domains (smart cities, Open Data initiatives). A new key issue is also to benefit from the growing "digital presence", that allows **interaction** with users at virtually any moment through mobile phone applications such as Twitter. Feedback loops between users and data managers can now be devised [7]. Finally, this wealth of on-demand, human-generated or curated data is an entry point for **AI algorithms**, such as deep learning methods, which is eager of carefully annotated data.

But the diversity and the dynamic of data sources raise several challenges. One can legitimately question if a data source is reliable or malevolent or if two data sources are independent. These problems are strengthened by the mutual links between data items or data sources. Hence fact provenance and sources independence are prominent data annotations that shall be taken into account. For user-generated or crowdsourced content, knowing the skills or the social relationships between participants allows for a better understanding of the produced raw data. This calls for a powerful **qualification mechanisms** that would integrate these annotations and help data managers in understanding their data and selecting their sources. Furthermore, even if interaction with participants is technically possible, the **orchestration** of complex data acquisition tasks from a mass still remains a black art.

The objective of the Druid *team is to provide models and algorithms for the annotation and management of interlinked data and sources at a large scale. We consider three main goals :*

1. *To propose well-founded models for interlinked data and, more importantly, interlinked data sources (for example, profiling users in a social network, orchestrating users and tasks in a crowdsourcing platform),*
2. *To develop theories for the qualification of such data and sources in terms of reliability, certainty, provenance, influence, economical value, trust ...*
3. *To implement systems that are proof-of-concepts of these models and theories. In particular we would like to demonstrate that these systems can overcome specific key problems in real-world applications, such as scalable data qualification and data adaptation to the final users.*

More concretely, we would like to address the following challenges :

- to develop integrated and scalable analysis tools for participants in social networks, that encompass the semantics of communications between users, computes user influence or user independence for example.
- to extend existing crowdsourcing platforms with fine user profiles, team building, contribution expressiveness or complex task management abilities, with application for e-science or e-government (smart cities).
- to develop reliability assessment techniques for large sensor networks (uncertainty), heterogeneous data sources or Linked Open Data (quality), or microblog conversations (misinformation).
- to develop privacy-preserving crowdsourcing processes that achieve realistic performance and quality requirement while satisfying sound privacy guarantees.
- to adapt data to its use (data visualization, accessibility of information).

Current Projects

* ANR JCJC CROWDGUARD 2016 Crowdsourcing platforms offer the unprecedented opportunity to connect easily on-demand task providers, or taskers, and on-demand task solvers, or workers, locally or world-wide, for paid or voluntary work, and for various kinds of tasks. By facilitating the accurate search of specific workers, otherwise unavailable, they have the potential to reduce costs as well as to accelerate and even democratize innovation. Their growing importance has made them unavoidable actors of the 21^{st} century economy. Ho-

7. See for example participative journalism platforms, https://witness.theguardian.com/moreabout

wever, abusive behaviors from crowdsourcing platforms against taskers or workers are frequently reported in the news or on dedicated websites, whether performed willingly or not, putting them at the epicenter of a burning societal debate. Real-life examples of such abusive behaviors range from strong concerns about private information accesses and uses (see, *e.g.*, the privacy scandals due to illegitimate accesses to the location data of a well-known drivers-riders company [8]) to blatant denials of workers' independence (see, *e.g.*, the complaints of micro-task workers or of drivers about the strong work control and monitoring imposed by their respective platforms [9]).

The goal of the CROWDGUARD project is to design sound protection measures of the taskers and workers from threats coming from the platform, while still enabling the latter to perform efficient and accurate tasks assignments. In CROWDGUARD, we advocate for an approach that uses confidentiality and privacy guarantees as building blocks for preventing a large variety of abusive behaviors. First, the enforcement of privacy and confidentiality guarantees directly prevents the first kind of abuse that we consider, i.e., the abusive usage of the personal or confidential information that taskers and workers disclose to the platform for the assignment of tasks. Second, through their obfuscation abilities, privacy and confidentiality guarantees carry the promise, in an extended form, to be also efficient for preventing a large variety of abusive behaviors (e.g, non-discrimination, or workers' independence).

The CROWDGUARD project will specify relevant use-cases, extracted from real-life situations and illustrating the need to protect the crowd from various abusive behaviors from the platform. The project will propose secure distributed algorithms for allowing workers (resp. taskers) to collaboratively compute a privacy-preserving version of their profiles (resp. a confidentiality-preserving version of their tasks) which will then be sent to the platform. The resulting tasks and profiles will enable highly efficient and accurate crowdsourcing processes while being protected by sound confidentiality and privacy guarantees. CROWDGUARD will also identify and formalize the possible abusive behaviors that the platform may perform, and propose sound models/algorithms to prevent them. Finally, the project will develop a prototype that will be used for evaluating the efficiency of the techniques proposed.

The main scientific outcomes of CROWDGUARD will advance the state-of-the-art on sound models and algorithms for the definition and prevention of abusive behaviors from crowdsourcing platforms. They will enable the development of respectful crowdsourcing processes by private companies or associations.

* ANR EPIQUE (2016) The evolution of scientific knowledge is directly related to the history of humanity. Document archives and bibliographic sources like the "Web Of Science" or PubMed contain a valuable source for the analysis and reconstruction of this evolution. The proposed project takes as starting point the contributions of D. Chavalarias and J.P. Cointet about the analysis of the dynamicity of evolutive corpora and the automatic construction of "phylomemetic" topic lattices (as an analogy with genealogic trees of natural species). Currently existing tools are limited to the processing of medium sized collections and a non interactive usage. The expected project outcome is situated at the crossroad between Computer science and Social sciences. Our goal is to develop new highly performant tools for building phylomemetic maps of science by exploiting recent technologies for parallelizing tasks and algorithms on complex and voluminous data. These tools are conceived and validated in collaboration with experts in philosophy and history of science over large scientific archives.

* ANR PRCI HEADWORK (2016)
Crowdsourcing relies on potentially huge numbers of on-line participants to resolve data acquisition or analysis tasks. It is an exploding area that impacts various domains, ranging from scientific knowledge enrichment to market analysis support. But currently, existing crowd platforms rely mostly

8. https://tinyurl.com/wp-priv
9. https://tinyurl.com/wsj-ind and https://tinyurl.com/trans-ind

on low level programming paradigms, rigid data models and poor participant profiles, which yields severe limitations. The low-level nature of existing solutions prevents the design of complex data acquisition workflows, that could be executed, composed, searched and even be proposed by participants themselves. Taking into account the quality, uncertainty, inconsistency and representativeness of participant contributions is still an open problem. Methods for assigning a task to the correct participant according to his trust, motivation and expertise, automatically improving crowd execution time, computing optimal participant rewards, are missing. Similarly, usual crowd campaigns produce isolated and rigid data sets : A flexible and common data model for the produced knowledge about data and participants could allow participative knowledge acquisition. To overcome these challenges, Headwork will define :

- Rich workflow, participant, data and knowledge models to capture various kind of crowd applications with complex data acquisition tasks and human specificities
- Methods for deploying, verifying, optimizing, but also monitoring and adapting crowd- based workflow executions at run time.

* Labex CominLabs PROFILE (2016)

The practice of online profiling, which can be defined as the tracking and collection of user information on computer networks, has grown massively during the last decade, and is now affecting the vast majority of citizens. Despite its importance and impact, profiling remains largely unregulated, with no legal provisions determining its lawful use and limits under either the French or European law. This has encouraged market players to exploit a wide range of tracking technologies to collect user information, including personal data. Consequently, most online companies are now routinely violating the fundamental rights of their users, especially with respect to their privacy, with little or no oversight. The PROFILE project brings together experts from law, computer science and sociology to address the challenges raised by online profiling, following a multidisciplinary approach. More precisely, the project will pursue two complementary and mutually informed lines of research :

- Investigate, design, and introduce a new right of opposition into the legal framework of data protection to better regulate profiling and to modify the behavior of commercial companies towards being more respectful of the privacy of their users.
- Provide users with the technical means they need to detect stealthy profiling techniques as well as to control the extent of the digital traces they routinely produce. As a case study, we focus on browser fingerprinting, a new profiling technique for targeted advertisement. The project will develop a generic framework to reason on the data collected by profiling algorithms, to uncover their inner working, and make them more accountable to users.

PROFILE will also propose an innovative protection to mitigate browser fingerprinting, based on the collaborative reconfiguration of browsers. The legal model developed in PROFILE will be informed by our technological efforts (e.g., what is technologically possible or not), while our technological research will incorporate the legal and sociological insights produced by the project (e.g., what is socially and legally desirable / acceptable). The resulting research lies at the crossing of three fields of expertise (namely Law, Computer Science and Sociology), and we believe forms a proposal that is timely, ambitious, and immediately relevant to our modern societies.

* PEPS JCJC INS2I P4-Crowd (2016)

The goal of the P4-Crowd project is to initiate the definition, design, and implementation of a crowdsourcing platform that allows the representation of personal preferences and their acquisition by querying a cohort of individuals (also called *the crowd* below) and by learning multi-user preferences, while offering robust privacy guarantees. Made of three sub-objectives, it promotes an original approach of co-design of preference representation models, crowd querying algorithms, and multi-user learning methods on the one hand, together with privacy-preserving mechanisms that mix encryption and sanitization. The expected benefits of P4-Crowd are both scientific and technical (co-

design preferences / crowd / privacy), industrial (trust economy), and societal (fundamental right to privacy). P4-Crowd's partners are two young research faculty members of a newly formed Irisa team.

* Emergent scientific challenges (2016) Rennes 1 University ORACULAR

The idea of ORACULAR is to propose declarative approaches for : (1) the description and modeling of input data of a crowdsourcing platform (task building, user modeling : preferences, availability, cost, skills), (2) the definition of optimization methods to organize the acquisition of user cohort contributions, while providing at the same time a reasonable level of interaction, (3) the definition of quality measures to evaluate the relevance and effectiveness of the crowdsourcing data collection process.

* ARED/LTC (2016) ExPress

The application context of this project concern social network analysis. The theoretical context is the preference queries applied to very large databases.

The concept of preference queries has been established in the database community and was intensively studied in the last decade. These queries have dual benefits. On the one hand, they allow to interpret accurately the information needs of a given user. On the other hand, they constitute an effective method to reduce very large datasets to a small set of highly interesting results and to overcome the empty result set. A query is personalized by applying related user preferences stored in the user's profile.

However, with the advent of social networks such as Facebook, Twitter, Instagram, or more locally Breizbook, the user is no longer considered as an individual entity. In this context, the user designates an interconnected social entity and is the author of significant information flow. The objective of this project is the development of a collaborative system for personalizing analyzes (*i.e. preference queries*) based on profiles of social network users.

* AMI EAU (2016) MetaTNT2

The objective of this project is to develop a meta-model based on simulations of the spatially distributed agro-hydrological model TNT2 (Topography-based Nitrogen Transfer and Transformations) in agricultural catchments, and to propose a conceptual guidance tool as a means of building and testing environmental management scenarios.

Références

[1] Tristan Allard, Davide Frey, George Giakkoupis, and Julien Lepiller. Lightweight Privacy-Preserving Averaging for the Internet of Things. In *M4IOT 2016 - 3rd Workshop on Middleware for Context-Aware Applications in the IoT*, pages 19 – 22, Trento, Italy, December 2016. ACM.

[2] Tristan Allard, Georges Hébrail, Florent Masseglia, and Esther Pacitti. A New Privacy-Preserving Solution for Clustering Massively Distributed Personal Times-Series. In *ICDE : International Conference on Data Engineering*, Helsinki, Finland, May 2016.

[3] Frédéric Balusson, Marie-Anne Botrel, Olivier Dameron, Yann Dauxais, Erwan Drezen, Alain Dupuy, Thomas Guyet, David Gross-Amblard, André Happe, Nolwenn Le Meur, Béranger Le Nautout, Emmanuelle Leray, Emmanuel Nowak, Caroline Rault, Emmanuel Oger, and Elisabeth Polard. PEPS : a platform for supporting studies in pharmaco-epidemiology using medico-administrative databases. In *International Congress on e-Health Research*, Paris, France, October 2016.

[4] Amal Ben Rjab, Mouloud Kharoune, Zoltan Miklos, and Arnaud Martin. Characterization of experts in crowdsourcing platforms. In *The 4th International Conference on Belief Functions*, Prague, Czech Republic, September 2016.

[5] Tassadit Bouadi, Fabien Gandon, and Arnaud Martin, editors. *Analyse intelligente des réseaux sociaux*, volume 30 of *Revue d'intelligence artificielle*. Lavoisier, 2016.

[6] Siwar Jendoubi. *Influencers characterization in a social network for Viral Marketing perspectives*. PhD thesis, December 2016.

[7] Siwar Jendoubi, Arnaud Martin, Ludovic Liétard, Hend Ben Hadji, and Boutheina Ben Yaghlane. Maximizing positive opinion influence using an evidential approach. In *FLINS*, pages 168 – 174, Roubaix, France, August 2016.

[8] Siwar Jendoubi, Arnaud Martin, Ludovic Liétard, Ben Hend, and Ben Boutheina. Two Evidential Data Based Models for Influence Maximization in Twitter. *Knowledge-Based Systems*, 2017.

[9] Zhun-Ga Liu, Quan Pan, Jean Dezert, and Arnaud Martin. Adaptive imputation of missing values for incomplete pattern classification. *Pattern Recognition*, 52, April 2016.

[10] Panagiotis Mavridis, David Gross-Amblard, and Zoltán Miklós. Using Hierarchical Skills for Optimized Task Assignment in Knowledge-Intensive Crowdsourcing. In *WWW2016*, Montreal, Canada, April 2016. ACM, IW3C.

[11] Hosna Ouni, Arnaud Martin, Laetitia Gros, Mouloud Kharoune, and Zoltan Miklos. Une mesure d'expertise pour le crowdsourcing. In *Extraction et Gestion des Connaissances (EGC)*, Grenoble, France, January 2017.

[12] Kuang Zhou. *Belief relational clustering and its application to community detection*. PhD thesis, July 2016.

[13] Kuang Zhou, Arnaud Martin, and Quan Pan. Semi-supervised evidential label propagation algorithm for graph data. In *BELIEF 2016 - The 4th International Conference on Belief Functions*, Prague, Czech Republic, September 2016.

[14] Kuang Zhou, Arnaud Martin, and Quan Pan. The Belief-Noisy-OR model applied to network reliability analysis. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 6, June 2016.

[15] Kuang Zhou, Arnaud Martin, Quan Pan, and Zhun-Ga Liu. ECMdd : Evidential c-medoids clustering with multiple prototypes. *Pattern Recognition*, 60 :239–257, 2016.

[16] Kuang Zhou, Arnaud Martin, Quan Pan, and Zhun-Ga Liu. Evidential Label Propagation Algorithm for Graphs. In *The 19th International Conference on Information Fusion* , Heidelberg, Germany, July 2016.

KID-LGI2P : Knowledge and Image analysis for Decision making

Centre de Recherche LGI2P
École des mines d'Alès/Institut Mines-Télécom
https://kidknowledge.wp.imt.fr/

Jacky MONTMAIN
jacky.montmain@mines-ales.fr
+33 (0) 4 66 38 70 58

Membres impliqués : [10]

- Jacky MONTMAIN, PR
- Gérard DRAY, MCF
- Sébastien HARISPE, MCF
- Abdelhak IMOUSSATEN, MCF
- Stefan JANAQI, MCF
- Sylvie RANWEZ, MCF
- François TROUSSET, MCF
- Valentina BERETTA, Doctorant
- Pierre-Antoine JEAN, Doctorant
- Massissilia MEDJKOUNE, Doctorant
- Jean-Christophe MENSONIDES, Doctorant
- Gildas TAGNY NGOMPE, Doctorant
- Pierre JEAN, IR

10. Seules sont représentées les personnes impliquées dans la thématique Extraction, modélisation et représentation de connaissance pour la décision.

Mots-clés

- Ontologies
- Traitement de la Langue Naturelle
- Mesures sémantiques
- Recherche d'Information, Indexation, Clustering
- Extraction de connaissances : Détection d'incertitude, Recherche de vérité, Classification

Extraire, modéliser et manipuler la connaissance dans un processus décisionnel

Qu'elle soit individuelle ou collective, la prise de décision découle d'un ensemble d'informations qui sont analysées, croisées et intégrées tout au long du processus décisionnel. Dans un contexte de crise, cette décision doit être prise rapidement, d'où l'absolue nécessité de disposer d'outils logiciels performants aptes à traiter de grandes quantités d'information. Or cette information est désormais ubiquitaire, collectée à partir de différentes traces de nos activités sur nos ordinateurs, nos télévisions, nos montres, nos téléphones et nos dispositifs de santé. Les chercheurs de l'axe "connaissance" de l'équipe KID se donnent pour mission de convertir cette information en éléments de connaissance exploitables dans un processus décisionnel. Pour cela, différentes méthodes et outils logiciels ont été proposés pour permettre notamment l'indexation et la recherche d'information, l'extraction de connaissances à partir de textes et l'inférence de connaissances, l'analyse d'informations extraites en regard de modèles de connaissance, la détection d'incertitude et la recherche de vérité. Les solutions proposées ne veulent en aucun cas se substituer à l'homme, mais au contraire l'accompagner au travers d'outils interactifs et conviviaux en vue de favoriser l'automatisation cognitive. Les approches développées sont génériques et trouvent leur application dans différents domaines (juridique, formation, bio-médical, commercial ou culturel) dans lesquels l'interprétation, le contrôle et le partage de données jouent un rôle primordial.

Projets marquants

Différentes contributions théoriques ont été amenées et les développements logiciels correspondants peuvent être utilisés par la communauté.

Contributions concernant les mesures sémantiques

- SML – *Semantic Measure Library* [8]. Suite à une étude approfondie des mesures sémantiques à partir d'ontologies ou à partir de corpus textuels [7], un cadre unificateur a été proposé [9] et une librairie logicielle mise à disposition (http://www.semantic-measures-library.org). D'autres travaux concernent des aspects particuliers des ontologies et en particulier le contenu informationnel des concepts [1][6].

Contributions concernant l'indexation conceptuelle, la recherche d'information et la visualisation

- USI – *User-oriented Semantic Indexer*. Environnement d'indexation conceptuelle par propagation, USI [5] permet d'indexer tout type de ressources (textes, sons, vidéos, personnes...) pour peu que l'on dispose d'une base d'apprentissage déjà indexée. L'approche propose également de réaliser des regroupements conceptuels et de créer des synthèses d'annotations associées à chaque regroupement [4]. Ces travaux ont été appliqués dans le domaine biomédical pour l'indexation de publications scientifiques et, plus récemment, pour l'évaluation de la qualité des odeurs [11].
- OBIRS – Ontology Based Information Retrieval System [12]. Disposant de ressources indexées à l'aide d'un groupe de concepts, cet environnement permet de rechercher efficacement tout type de ressource et d'avoir une représentation graphique des résultats.

Contributions concernant l'extraction de connaissances : détection d'incertitude, recherche de vérité et classification

- MUD – *Multiple Uncertainty Detection* [10]. Basée sur une analyse statistique des textes cette

méthode supervisée permet de détecter les incertitudes exprimées dans des textes en langage naturel. Plusieurs caractéristiques sont utilisées pour identifier les phrases incertaines et les représenter sous une forme vectorielle utilisée par des méthodes de classification (e.g. SVM). Ces travaux sont menés en collaboration avec le LSIS de Marseille.

- Nos travaux sur la détection de vérité (*Truthdiscovery* en anglais) ont conduit à une publication dans une conférence internationale [2] et à une communication lors des 27es Journées francophones d'Ingénierie des Connaissances à Montpellier en 2016, qui a obtenu le prix du meilleur papier décerné par l'AfIA [3]. Ces travaux sont menés en collaboration avec l'UMR-Espace-Dev de Montpellier.
- Une fois extraite, la connaissance peut être utilisée pour différents traitements, en particulier, pour la classification de documents. Ainsi, à partir des décisions de justice, une base peut être constituée pour représenter la jurisprudence relative à une situation juridique donnée. Des modèles prédictifs sont appliqués sur cette base afin d'identifier les cas similaires qui ont pu être jugés par le passé et estimer la probabilité d'avoir gain de cause dans certains procès [13]. Dans le domaine du nucléaire, à partir de corpus de textes émis par les défenseurs du nucléaire ou par ses détracteurs, un système a été mis en place pour détecter des jeux d'arguments. Ces travaux sont menés en collaboration avec l'équipe Chrome de l'Université de Nîmes.

Références

[1] Montserrat Batet, Sébastien Harispe, Sylvie Ranwez, David Sánchez, and Vincent Ranwez. An information theoretic approach to improve semantic similarity assessments across multiple ontologies. *Information Sciences*, 283 :197 – 210, 2014. New Trend of Computational Intelligence in Human-Robot Interaction.

[2] Valentina Beretta, Sébastien Harispe, Sylvie Ranwez, and Isabelle Mougenot. How can ontologies give you clue for truth-discovery ? an exploratory study. In *Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics*, WIMS '16, pages 15 :1–15 :12, New York, NY, USA, 2016. ACM.

[3] Valentina Beretta, Sébastien Harispe, Sylvie Ranwez, and Isabelle Mougenot. Utilisation d'ontologies pour la quête de vérité : une étude expérimentale. In *Actes des 27es Journées francophones d'Ingénierie des Connaissances*, IC2016, 2016.

[4] Nicolas Fiorini, Sébastien Harispe, Sylvie Ranwez, Jacky Montmain, and Vincent Ranwez. Fast and reliable inference of semantic clusters. *Knowledge-Based Systems*, 111 :133 – 143, 2016.

[5] Nicolas Fiorini, Sylvie Ranwez, Jacky Montmain, and Vincent Ranwez. Usi : a fast and accurate approach for conceptual document annotation. *BMC Bioinformatics*, 16(1) :83, 2015.

[6] Sébastien Harispe, Abdelhak Imoussaten, François Trousset, and Jacky Montmain. On the consideration of a bring-to-mind model for computing the information content of concepts defined into ontologies. In *2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, Istanbul, Turkey, August 2015.

[7] Sébastien Harispe, Sylvie Ranwez, Stefan Janaqi, and Jacky Montmain. *Semantic similarity from natural language and ontology analysis*, volume 8 of *Synthesis Lectures on Human Language Technologies*. Morgan & Claypool Publishers, 2015.

[8] Sébastien Harispe, Sylvie Ranwez, Stefan Janaqi, and Jacky Montmain. The semantic measures library and toolkit : fast computation of semantic similarity and relatedness using biomedical ontologies. *Bioinformatics*, 30(5) :740, 2014.

[9] Sébastien Harispe, David Sánchez, Sylvie Ranwez, Stefan Janaqi, and Jacky Montmain. A framework for unifying ontology-based semantic similarity measures : A study in the biome-

dical domain. *Journal of Biomedical Informatics*, 48 :38 – 53, 2014.

[10] Pierre-Antoine Jean, Sébastien Harispe, Sylvie Ranwez, Patrice Bellot, and Jacky Montmain. Uncertainty detection in natural language : A probabilistic model. In *Proceedings of the 6th International Conference on Web Intelligence, Mining and Semantics*, WIMS '16, pages 10 :1–10 :10, New York, NY, USA, 2016. ACM.

[11] Massissilia Medjkoune, Sébastien Harispe, Jacky Montmain, Stéphane Cariou, Jean-Louis Fanlo, and Nicolas Fiorini. Towards a non-oriented approach for the evaluation of odor quality. In Joao Paulo Carvalho, Marie-Jeanne Lesot, Uzay Kaymak, Susana Vieira, Bernadette Bouchon-Meunier, and Ronald R. Yager, editors, *Information Processing and Management of Uncertainty in Knowledge-Based Systems : 16th International Conference, IPMU 2016, Eindhoven, The Netherlands, June 20-24, 2016, Proceedings, Part I*, pages 238–249, Cham, 2016. Springer International Publishing.

[12] Mohameth-François Sy, Sylvie Ranwez, Jacky Montmain, Armelle Regnault, Michel Crampes, and Vincent Ranwez. User centered and ontology based information retrieval system for life sciences. *BMC Bioinformatics*, 13(1) :S4, 2012.

[13] Gildas Tagny Ngompé, Sébastien Harispe, Guillaume Zambrano, Jacky Montmain, and Stéphane Mussard. Reconnaissance de sections et d'entités dans les décisions de justice : application des modèles probabilistes hmm et crf. In *Extraction et Gestion des Connaissances - EGC 2017*, Revue des Nouvelles Technologies de l'Information, Grenoble, France, January 2017.

LILaC : Logique, Interaction, Langage, et Calcul

Institut de Recherche en Informatique de Toulouse
(IRIT, UMR 5505)
CNRS, INPT, UPS, UT1C, UT2J
https://www.irit.fr/-Equipe-LILaC-

Dominique LONGIN
Dominique.Longin@irit.fr
+33 5 61 55 64 47

Emiliano LORINI
Emiliano.Lorini@irit.fr
+33 5 61 55 64 47

Membres

La liste exhaustive des membres de l'équipe LILaC peut être consultée en cliquant ici. Les permanents sont au nombre de 14 (7 chercheurs CNRS, 3 enseignants-chercheurs de l'Université Paul Sabatier, 4 enseignants-chercheurs de l'Université Toulouse Capitole) et les non-permanents au nombre de 14 (1 chercheur associé, 1 chercheur invité, 1 collaborateur extérieur, 2 post-doctorants, 9 doctorants).

Thématique générale de l'équipe

Les recherches de l'équipe LILaC se focalisent sur les modèles formels de l'interaction en adoptant une approche normative (top-down) à travers cinq thématiques.

1. Logiques pour l'action et la dynamique des croyances. Un aspect important de l'interaction est la dynamique de l'environnement. Celui-ci évolue grâce à l'exécution d'événements et d'actions qui ont des répercussions sur les états mentaux des agents (via leurs observations ou leurs perceptions). L'équipe LILaC s'intéresse entre autres : aux logiques de l'action telle que la logique dynamique

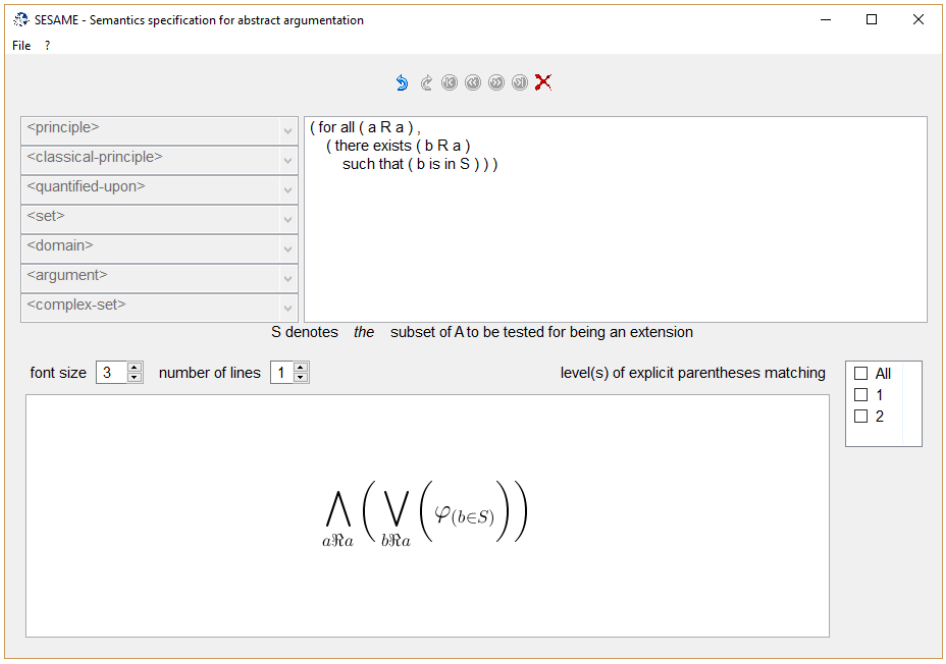

Figure 1.1 – Le logiciel SESAME de production de sémantiques argumentatives à base d'extension

dont elle étudie les propriétés formelles (complétude, décidabilité, complexité) ; à l'algorithmique du raisonnement sur les actions en étudiant les propriétés formelles de théories de l'action et des événements ; à la révision et à la mise à jour afin de prendre en compte une nouvelle information, en particulier dans le cas multi-agent ; aux aspects épistémiques de la planification multi-agent.

2. Sécurité des systèmes d'information et de communication. En environnement ouvert, l'interaction entre agents nécessite d'être protégée.

En premier lieu, cette protection peut se décliner sous la forme de trois propriétés : disponibilité, intégrité, confidentialité. En partant de la spécification formelle d'un système, l'équipe LILaC s'intéresse d'une part, à fournir la preuve que ce système vérifie les trois propriétés ci-dessus, et d'autre part à obtenir de manière automatique une implémentation qui correspond à la spécification et dont la correction est prouvée par construction. Au niveau du contrôle d'accès, il s'agit d'étendre le modèle de la matrice du contrôle d'accès et de tracer, pour ce qui concerne le problème de la sûreté des modèles étendus, la limite entre décidabilité et indécidabilité ; Au niveau de la vérification de protocoles, il s'agit de définir des algorithmes pour vérifier automatiquement des protocoles qui vont permettre à des agents de garantir la confidentialité ou l'intégrité des informations qu'ils échangent.

En second lieu, la protection d'un système est également étudiée dans l'équipe via la modélisation (en termes de rôles et de relations entre utilisateurs) de l'organisation dans laquelle s'insère ce système informatique : il s'agit alors de caractériser à la fois les aspects prédicatifs de la notion de rôle et ses aspects déontiques qui font appel aux concepts de permission et d'obligation, à la base également d'autres notions comme celles de confiance, de responsabilité, ou de délégation.

3. Logique, théorie des jeux, et choix social. Dans cet axe, l'équipe LILaC s'intéresse aux mécanismes sociaux qui régissent les interactions entre agents (artificiels ou humains). Cela concerne notamment : la théorie du vote, qui fournit des modèles d'interaction et de coordination pour les systèmes multi-agents, où l'utilisation de la logique

permet de spécifier et vérifier ces systèmes ; la théorie des jeux épistémiques et les logiques pour les jeux ; la théorie de la diffusion d'opinion et de l'influence sociale. Des simples systèmes sont implémentés afin de tester les différents mécanismes élaborés sur différentes populations d'agents.

4. Logiques pour les systèmes multi-agents et argumentation. Un des objectifs majeurs de l'équipe est d'élaborer un modèle logique général de l'interaction. Cela se traduit par de nombreux travaux visant l'intégration des modèles provenant des thématiques précédentes en un formalisme unifié. Ce dernier s'appuie sur des logiques BDI et des logiques d'action, afin de pouvoir exprimer à la fois des croyances, désirs et intentions des agents, ainsi que les actions pouvant être accomplies par les agents, tout en prenant en compte l'évolution des croyances des agents. Cette intégration considère aussi l'argumentation abstraite, en tant que modèle de raisonnement basé sur des interactions entre arguments ; l'évaluation des interactions permet de déterminer des ensembles d'arguments collectivement acceptables (extensions), ou un ordre d'acceptabilité entre les arguments. La dynamique des systèmes d'argumentation, la définition de différents modes d'évaluation (sémantiques argumentatives), et les liens entre argumentation et logique, sont au coeur des recherches actuelles de l'équipe.

5. Logiques et raisonnement automatique. Les modèles de l'interaction qui précèdent posent des problèmes calculatoires et nécessitent la mise au point d'algorithmes permettant de répondre à des questions de type « telle formule est-elle conséquence de telle théorie logique ? ». En tant que spécialiste des logiques non classiques, l'équipe LILaC a substantiellement renouvelé la méthode des tableaux sur le plan théorique (complétude et complexité de logiques admettant une sémantique de Kripke) ; sur le plan pratique, ce travail a été implémenté dans la plate-forme LoTREC qui offre un langage de programmation de très haut niveau pour la définition rapide de démonstrateurs pour toute logique modale normale. Nous avons également exploré certains systèmes paraconsistants et proposé une traduction des problèmes posés en termes de formules booléennes quantifiées.

Plus récemment, nous nous intéressons : au développement de la plateforme TouIST qui constitue une interface entre un langage propositionnel de haut niveau et très concis, et différents solveurs (SAT, SMTP, et plus récemment QBF) ; à la traduction (de fragments) de certains langages en un langage utilisable par TouIST afin d'automatiser des procédures de preuve pour lesquelles nous n'avons pas de démonstrateur ; à l'implémentation de certains jeux pour la recherche de stratégies gagnantes ; à la planification épistémique d'actions (robotique cognitive).

Contributions

L'équipe LILaC publie ses résultats dans les plus grands journaux internationaux (Artificial Intelligence, Journal of Artificial Intelligence Research, Journal of Logic and Computation, Studia Logica, Synthese, Journal of Autonomous Agents and Multi-Agent Systems, *etc.*) ainsi que dans les plus grandes conférences internationales (IJCAI, AAMAS, TARK, *etc.*).

Au niveau des réalisations logicielles, l'équipe LILaC a développé : LoTREC (un démonstrateur pour la quasi totalité des logiques modales) ; SESAME (un logiciel de production de sémantiques argumentatives encodées dans un langage de type logique propositionnelle, voir Figure 1.1) ; TouIST (une interface dotée d'un langage très compact pour la logique propositionnelle permettant d'utiliser des solveurs SAT, SMT et QBF, et développé en collaboration avec l'équipe ADRIA de l'IRIT, voir Figure 1.2).

Au niveau international ses chercheurs travaillent avec de nombreux collaborateurs issus de différents pays européens, américains ou asiatiques.

Au niveau national, l'équipe LILaC est investie dans l'Institut Cognition (Tremplin Carnot), un réseau de 14 laboratoires répartis sur le sol français en relation avec les sciences cognitives dont le but est de travailler de manière privilégiée avec le monde de la recherche industrielle. L'équipe adhère à l'AfIA en tant que personne morale.

Au niveau local, l'équipe LILaC est membre du Labex CIMI qui lui finance régulièrement des projets ou étudiants en thèse.

Voir ici pour la liste des publications de l'équipe LILaC.

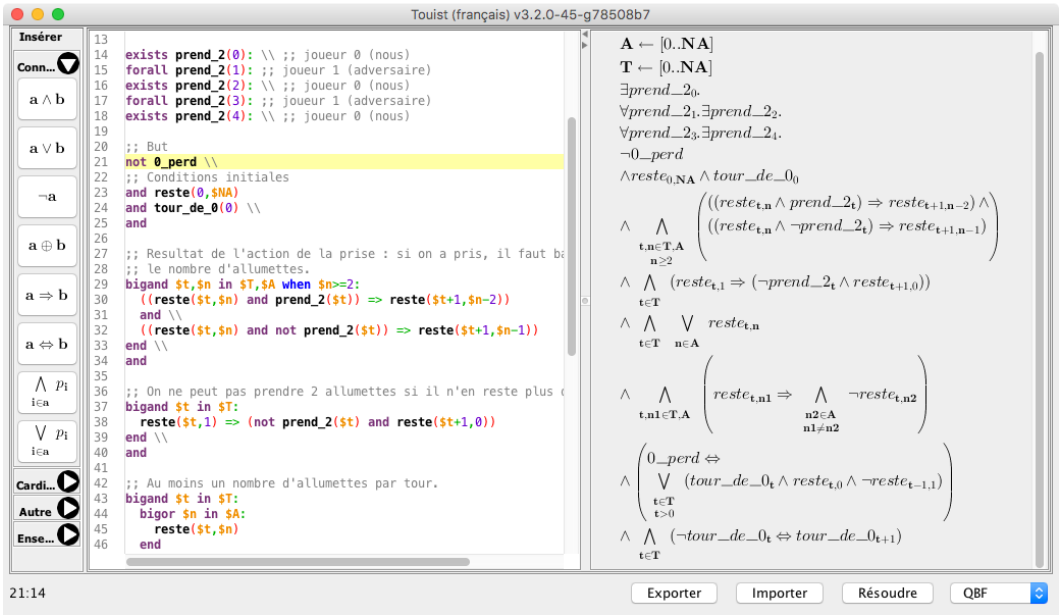

Figure 1.2 – L'interface TouIST pour les démonstrateurs SAT, SMT et QBF

LIRICA : Logique, Interaction, Raisonnement et Inférence, Complexité, Algèbre

Laboratoire d'Informatique et Systèmes
Aix-Marseille Université

Vincent RISCH
vincent.risch@univ-amu.fr

Membres :

- Belaid BENHAMOU, MCF
- Larbi BENMEZAL, Doctorant
- Sabrina BOUCHENE, Doctorant
- Nadia CREIGNOU, PR
- Tiziano DALMONTE, Doctorant
- Marianna GIRLANDO, Doctorant
- Charles GRELLOIS, MCF
- Line JAMET, MCF
- Tarek KHALED, Doctorant
- Farid NOUIOUA, MCF
- Frédéric OLIVE, MCF
- Nicola OLIVETTI, PR
- Odile PAPINI, PR
- Vincent RISCH, MCF
- Claude SABATIER, MCF
- Luigi SANTOCANALE, PR
- Eric WÜRBEL, MCF
- Safa YAHI, MCF

Mots-clés

- Logique
- Raisonnements
- Complexité
- Représentation des Connaissances
- Intelligence Artificielle
- Sémantique
- Théorie de la Preuve
- Satisfiabilité des Formules

Thématique générale de l'équipe/ l'outil/le projet

L'équipe LIRICA se constitue à l'occasion de l'émergence en janvier 2018 du nouveau laboratoire marseillais LIS, issu de la fusion du LIF (Laboratoire d'Informatique Fondamentale, UMR 7279) et du LSIS (Laboratoire des Sciences de l'Information et des Systèmes, UMR 7296). Les membres de l'équipe sont issus des anciennes équipes ACRO, MOVE, TALEP du LIF, et InCA du LSIS. L'objectif est de favoriser un partage pouvant susciter une synergie commune à partir d'approches diverses toutes rattachées à la logique (IA, calculabilité, topologie algébrique, complexité, langues naturelles...)

Les thématiques développées s'articulent autour de la logique et de ses relations avec l'informatique théorique et l'intelligence artificielle. Ces thématiques s'ordonnent autour de trois axes principaux, et s'enrichissent d'interactions multiples :

1. *Raisonnement et représentation des connaissances en intelligence artificielle :* les recherches menées autour de cet axe s'appuient sur l'utilisation de formalismes logiques pour le raisonnement et le traitement algorithmique de connaissances en intelligence artificielle. Les thématiques concernées sont :
   - changement et fusion des croyances
   - raisonnement non-motone
   - traitement des langues naturelles
   - programmation logique
2. *Complexité, théorie et applications :* complexité structurelle, caractérisations de classes de complexité par des formalismes logiques, complexité de différentes classes de problèmes (SAT, model checking. . .), différentes tâches algorithmiques (décision, comptage, énumération), ainsi que différents formalismes logiques (et leurs fragments). Les thématiques concernées sont :
   - complexité des problèmes et des algorithmes
   - complexité structurelle
   - complexité descriptive
3. *Systèmes Logiques, sémantique et théorie de la preuve :* classiquement, l'étude de systèmes logiques, du point de vue de la sémantique et de la théorie de la preuve. Les thématiques concernées sont :
   - logique algébrique et catégorielle
   - modèles de calcul
   - méthodes de preuve pour les logiques non-classiques, modales, et de point-fixe

Projets marquants

Divers travaux sont menés dans le cadre des thématiques précédentes.

Axe Raisonnement et représentation des connaissances en intelligence artificielle

- La question de la représentation des connaissances et du raisonnement peut être approchée sous plusieurs angles : révision et fusion d'informations, changement de croyances, argumentation formelle, traitement logique de l'incertain... Les problèmes de révision, de fusion de connaissances ou encore celui de la gestion de l'incohérence, exprimés dans des logiques de description légères, sont particulièrement importants aujourd'hui avec le développement du WEB sémantique. L'ouverture de travaux sur ce thème dans le cadre du projet ANR ASPIQ nous amènent à suivre différentes pistes concernant l'étude et l'implantation de relations d'inférence tolérant l'incohérence, d'opérateurs de révision et de fusion de croyances. En particulier, nous souhaitons sortir du cadre des logiques de description pour étudier le cadre plus large des règles existentielles et des programmes logiques avec sémantique des modèles stable (ASP) et leur extension à la prise en compte de variables quantifiées existentiellement [2] [3] [4]. Parallèlement, les travaux développés dans le cadre du projet AMADEUS nous ont permis d'étudier le changement de croyances exprimées dans des fragments propositionnels, permettant une mise en œuvre de procédures de raisonnement efficaces [7], et trouveront un prolongement dans l'étude de fragments de la logique du premier ordre. Enfin, des travaux sont aussi menés en argumentation formelle. Ces travaux concernent, d'une part la révision de cadres argumentatifs, l'extension des cadres argumentatifs de Dung (bipola-

rité avec ajout d'une relation de support), l'étude et la caractérisation de systèmes d'argumentation structurés reposant sur des systèmes à base de règles [14] [16] ; d'autre part la modélisation de la notion d'attitude et le raisonnement sur les arguments dans un cadre logique non-monotone.
- Dans le cadre du paradigme ASP (Answer Set Programming), nous nous intéressons à la sémantique des programmes logiques. Inspirée d'une sémantique modale issue de la logique des Hypothèses, nous avons introduit une nouvelle sémantique pour les programmes ASP, sémantique qui capture et étend la sémantique des modèles stables. Cette nouvelle sémantique simplifie les procédures de recherche et ouvre une nouvelle voie pour élaborer de nouveaux algorithmes et systèmes de recherche de modèles stables [5]. En parallèle, certaines des études qui ont été menées sur les symétries sont en cours d'élargissement dans le cadre non-monotone et le cadre ASP [6].
- Le traitement symbolique du langage écrit s'appuie sur la représentation logique des phrases ou des concepts véhiculés par les phrases. Dans ce cadre, les logiques utilisées sont diverses. Un premier axe d'étude concerne l'utilisation du lambda-calcul typé comme outil de formalisation et d'analyse de phrases données en langage naturel [12]. Une deuxième axe d'étude, en cours de développement, a trait à l'utilisation de formalismes logiques dans le cadre de l'analyse textuelle de sentiments.

Axe Complexité, théorie et applications

- Notre approche est celle de la complexité structurelle, qui vise la classification des problèmes relativement à leur difficulté intrinsèque, plutôt que l'analyse de complexité des algorithmes. Le point de vue de la complexité descriptive enrichit ce parti pris. Il s'agit de relier la difficulté algorithmique d'un problème — donc la quantité de ressources consommées par un calcul qui le résout — à sa complexité descriptive, c'est-à-dire à la richesse syntaxique d'un formalisme qui le décrit. Ce type de liens est formulé au moyen de caractérisations de classes de complexité par des formalismes logiques — essentiellement des fragments de la logique existentielle du second-ordre. Nous nous intéressons plus particulièrement à l'élaboration d'un cadre théorique pour l'étude de la complexité des problèmes d'énumération, ainsi qu'aux caractérisations logiques de classes de problèmes d'évaluation dont le traitement est accéléré par le recours massif au parallélisme [8] [11] .

Axe Systèmes Logiques, sémantique et théorie de la preuve

- Nos recherches portent sur la sémantique et les méthodes de preuves pour les logiques des conditionnels. Ces logiques trouvent un intérêt interdisciplinaire dans l'intelligence artificielle et dans l'épistémologie ; elles ont été étudiées depuis les années 1970 pour la formalisation du raisonnement contre-factuel, du changement des croyances (révision), du raisonnement plausible (tolérant exceptions, non-monotone) et du raisonnement causal. Nous avons proposé des sémantiques basées sur les Modèles de Voisinage et des calculs analytiques pour ces logiques [13] [1] [10]. Ces recherches se poursuivront avec notamment deux objectifs : (a) l'étude des calculs à séquents uniformes et modulaires pour toutes les logiques des conditionnels et similaires, (b) l'étude d'une sémantique modulaire pour les mêmes logiques, basée sur des modèles de voisinage.
- Parmi les problèmes ouverts en logique posés par la communauté informatique théorique, certains nous semblent plus importants et devoir requérir une attention particulière : (a) le problème de la caractérisation des ordinaux de clôture dans le mu-calcul modal, (b) le problème de la caractérisation de la théorie équationnelle de la jointure naturelle et de la réunion interne de tables dans les bases de données relationnelles ; (c) le problème de la construction d'un univers des ensembles intuitionnistes à partir d'une algèbre de Heyting qui joue le rôle d'un ensemble de valeurs de vérité $\{vrai, faux\}$ dans un monde intuitionniste. Les contextes scientifiques dont ces problèmes sont issus sont les logiques modales pour la vérification des systèmes informatiques, la théorie des bases de données relationnelles et,

en même temps, la théorie des logiques modales multidimensionnelles, la logique intuitionniste et constructive dans sa formulation catégorielle par la théorie des topoi. Nous nous intéressons à la résolution de ces problèmes à l'aide d'un ensemble d'outils algébriques uniformes, issus de l'algèbre universelle, des théories des ordres, des treillis et des catégories, et de la combinatoire [15] [9].

Références

[1] Régis Alenda, Nicola Olivetti, and Gian Luca Pozzato. Nested sequent calculi for normal conditional logics. *J. Log. Comput.*, 26(1) :7–50, 2016.

[2] Jean-François Baget, Salem Benferhat, Zied Bouraoui, Madalina Croitoru, Marie-Laure Mugnier, Odile Papini, Swan Rocher, and Karim Tabia. A general modifier-based framework for inconsistency-tolerant query answering. In *Principles of Knowledge Representation and Reasoning : Proceedings of the Fifteenth International Conference, KR 2016, Cape Town, South Africa, April 25-29, 2016.*, pages 513–516, 2016.

[3] Jean-François Baget, Zied Bouraoui, Farid Nouioua, Odile Papini, Swan Rocher, and Eric Würbel. \exists -asp for computing repairs with existential ontologies. In *Scalable Uncertainty Management - 10th International Conference, SUM 2016, Nice, France, September 21-23, 2016, Proceedings*, pages 230–245, 2016.

[4] Salem Benferhat, Zied Bouraoui, Madalina Croitoru, Odile Papini, and Karim Tabia. Non-objection inference for inconsistency-tolerant query answering. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence, IJCAI 2016, New York, NY, USA, 9-15 July 2016*, pages 3684–3690, 2016.

[5] Belaid Benhamou. Dynamic and static symmetry breaking in answer set programming. In *Logic for Programming, Artificial Intelligence, and Reasoning - 19th International Conference, LPAR-19, Stellenbosch, South Africa, December 14-19, 2013. Proceedings*, pages 112–126, 2013.

[6] Belaïd Benhamou. Local and global symmetry breaking in itemset mining. *Ann. Math. Artif. Intell.*, 80(1) :91–112, 2017.

[7] Nadia Creignou, Raïda Ktari, and Odile Papini. Belief contraction within fragments of propositional logic. In *ECAI 2016 - 22nd European Conference on Artificial Intelligence, 29 August-2 September 2016, The Hague, The Netherlands - Including Prestigious Applications of Artificial Intelligence (PAIS 2016)*, pages 390–398, 2016.

[8] Nadia Creignou, Raïda Ktari, Arne Meier, Julian-Steffen Müller, Frédéric Olive, and Heribert Vollmer. Parameterized enumeration for modification problems. In *LATA 2015 – 9th International Conference on Language and Automata Theory and Applications*, pages 524–536, 2015.

[9] Silvio Ghilardi, Maria Joao Gouveia, and Luigi Santocanale. Fixed-point elimination in the Intuitionistic Propositional Calculus. In *FOSSACS 2016*, Eindhoven, Netherlands, April 2016. https://hal.archives-ouvertes.fr/hal-01249822.

[10] Mariana Girlando, Sara Negri, Nicola Olivetti, and Vincent Risch. The logic of conditional beliefs : Neighbourhood semantics and sequent calculus. In *Advances In Modal Logics, AIML 2016, Budapest, Hongria, August 30 - September 2, 2016. Proceedings*, pages 322–341, 2016.

[11] Etienne Grandjean and Frédéric Olive. A logical approach to locality in picture languages. *J. Comput. Syst. Sci.*, 82(6) :959–1006, 2016.

[12] Line Jakubiec. Ontology in coq for a guided message composition. In *Language Production, Cognition, and the lexicon, 48, Text, Speech and Language Technology*, page 331. Springer.

[13] Sara Negri and Nicola Olivetti. A sequent calculus for preferential conditional logic based on neighbourhood semantics. In *Automated Reasoning with Analytic Tableaux and Related Methods - 24th International Conference,*

TABLEAUX 2015, Wrocław, Poland, September 21-24, 2015. Proceedings, pages 115–134, 2015.

[14] Farid Nouioua and Eric Würbel. Removed set-based revision of abstract argumentation frameworks. In *26th IEEE International Conference on Tools with Artificial Intelligence, ICTAI 2014, Limassol, Cyprus, November 10-12, 2014*, pages 784–791, 2014.

[15] Luigi Santocanale and Jérôme Fortier. Cuts for circular proofs : semantics and cut-elimination. In *Computer Science Logic 2013*, volume 23 of *LIPIcs*, pages 248–262, Torino, Italy, September 2013. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik.

[16] Karima Sedki and Safa Yahi. Constrained value-based argumentation framework. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems - 16th International Conference, IPMU 2016, Eindhoven, The Netherlands, June 20-24, 2016, Proceedings, Part II*, pages 265–278, 2016.

Multispeech : Speech Modeling for Facilitating Oral-Based Communication

Multispeech
LORIA / Inria
615 rue du Jardin Botanique, 54600 Villers les Nancy
https://team.inria.fr/multispeech/

Denis JOUVET
denis.jouvet@inria.fr
+33 3 54 95 86 26

Membres :

- Denis JOUVET, DR Inria
- Yves LAPRIE, DR CNRS
- Emmanuel VINCENT, DR Inria
- Anne BONNEAU, CR CNRS
- Dominique FOHR, CR CNRS
- Antoine LIUTKUS, CR Inria
- Vincent COLOTTE, MCF
- Irina ILLINA, MCF
- Odile MELLA, MCF
- Slim OUNI, MCF
- Agnès PIQUARD-KIPFFER, MCF
- Romain SERIZEL, MCF
- Benjamin ELIE, Post-doc
- Karan NATHWANI, Post-doc
- Théo BIASUTTO-LERVAT, Doctorant
- Guillaume CARBAJAL, Doctorant
- Sara DAHMANI, Doctorant
- Ken DÉGUERNEL, Doctorant
- Ioannis DOUROS, Doctorant
- Baldwin DUMORTIER, Doctorant
- Mathieu FONTAINE, Doctorant
- Amal HOUIDHEK, Doctorant
- Nathan LIBERMANN, Doctorant
- Aditya Arie NUGRAHA, Doctorant
- Lauréline PEROTIN, Doctorant
- Anastasiia TSUKANOVA, Doctorant

Mots-clés

- Parole et audio
- Apprentissage automatique
- Modélisation statistique
- Réseaux de neurones et apprentissage profond
- Reconnaissance de la parole
- Synthèse de la parole
- Synthèse articulatoire
- Traitement du signal
- Traitement de la langue
- Multimodalité (acoustique et visuelle)
- Estimation et prise en compte de l'incertitude
- Perception
- Musique
- Construction de corpus spécifiques (parole, multimodalité, IRM, ...)

Thématique générale de l'équipe

Les thématiques développées concernent la modélisation de la parole pour faciliter la communication orale ; avec une attention particulière pour les aspects multisources, multilingues et multimodaux (d'où le nom Multispeech) :

- *Multisources*, car le signal de parole capté par un microphone est fréquemment bruité ou inclut des superpositions de voix. Dans ce contexte une partie des travaux porte sur la séparation de sources, en particulier pour une prise de son avec plusieurs microphones. Ces travaux contribuent au rehaussement de la parole et à la reconnaissance de parole robuste.
- *Multilingues*, car la parole non-native est influencée par la langue maternelle, ce qui complique notablement son traitement. L'un des domaines applicatifs concernés est l'aide à l'apprentissage de langues étrangères.
- *Multimodaux*, avec la prise en compte des modalités visuelles et acoustiques de la communication vocale pour la synthèse audiovisuelle expressive.

Quelques domaines applicatifs concernés sont les suivants.

- *L'interaction multimodale* avec la synthèse expressive, par exemple pour les livres parlés, et la synthèse audiovisuelle pour améliorer, grâce à l'apport de la composante visuelle, la communication avec des personnes malentendantes et pour aider à l'apprentissage de langues.
- *L'annotation et le traitement de documents audio* avec la transcription enrichie de documents audio, l'alignement texte-parole (segmentation en mots et/ou en phonèmes) entre autres pour des études linguistiques, et le traitement de documents multimédia, avec par exemple la séparation parole/musique ou l'application de modifications prosodiques.
- *Le monitoring et la communication assistée* permettant d'apporter une aide dans des situations de handicap ou pour améliorer l'autonomie. Un exemple concerne la commande vocale mains-libres et le monitoring d'événements sonores dans le cadre de la maison intelligente.
- *L'apprentissage de langues assisté par ordinateur* : l'objectif est de fournir des retours vers l'apprenant sur la qualité de ses prononciations pour l'articulation des sons comme pour la prosodie. Cela repose sur une analyse des prononciations de l'apprenant. La qualité des diagnostics est conditionnée par la fiabilité de la segmentation phonétique et des paramètres prosodiques calculés.

Le programme de recherche est structuré selon trois axes. Le premier est relatif à la **modélisation explicite de la parole** et exploite la dimension physique de celle-ci : la parole résulte du mouvement des articulateurs – mâchoire, lèvres, langue, ... – et se traduit aussi par des déformations visibles sur le visage. De plus la parole ne se limite pas uniquement à une suite des mots, la prosodie joue un rôle important pour structurer l'énoncé vocal et véhiculer l'expressivité (emphase, émotion, ...). Dans ce cadre, la *modélisation articulatoire* précise les liens entre le signal de parole d'une part et la position et le mouvement des articulateurs d'autre part. Cela conduit à la synthèse articulatoire, i.e., la production de parole à partir de la connaissance du conduit vocal [4], et à l'inversion articulatoire, i.e., la détermination de la forme du conduit vocal à partir du signal de parole. La *synthèse audiovisuelle expressive* concerne la production d'une synthèse de parole bi-modale (i.e., composantes audio et visuelle), avec la prise en compte de l'expressivité sur les deux composantes [7]. La *catégorisation des sons et de la prosodie* porte sur l'étude des contrastes au niveau prosodique et phonétique, et les relations avec la production et la perception de la parole, tant pour la parole native que pour la parole non-native [12].

Le deuxième axe est dédié à la **modélisation statistique de la parole** qui repose sur des techniques d'apprentissage automatique et tire bénéfice de (grands) corpus de données de parole. Les réseaux de neurones profonds (DNN) constituent maintenant l'état de l'art dans ces domaines. La *séparation de sources audio* est un élément clé pour le rehaussement de la parole dans le bruit. Les approches étudiées combinent le traitement du signal, la modélisation statistique et l'apprentissage profond pour la localisation et la séparation de sources dans les signaux multicanaux [6], et sont en cours

d'extension pour la suppression d'écho et la déréverbération. Concernant la *modélisation acoustique* pour la reconnaissance de la parole, l'accent est mis sur la robustesse au bruit [11] et la reconnaissance avec prise de son distante, pour laquelle le signal capté par le microphone est dégradé par la réverbération et les bruits. Les travaux menés portent sur le traitement de signaux monocanal (un seul microphone) ou multicanal (plusieurs microphones), et sur l'optimisation globale de la chaîne de traitement (i.e., rehaussement plus modélisation acoustique). La *modélisation linguistique* concerne la prononciation des mots et leur enchaînement dans des phrases. Dans ce cadre l'accent est mis sur le traitement des noms propres, et en particulier sur la prédiction des noms propres hors vocabulaire [10]. Les approches statistiques sont aussi étudiées pour la *génération de parole* expressive ; le but étant de tirer profit de leur flexibilité et de leur capacité d'adaptation à de nouvelles conditions.

Le dernier axe traite de l'**incertitude et de sa prise en compte** dans le traitement de la parole. L'incertitude résulte de la forte variabilité du signal de parole et de l'imperfection des modèles. Il est important d'estimer l'incertitude sur les résultats et de la prendre en compte dans les traitements ultérieurs. En reconnaissance de la parole, les mesures de confiance ont été étudiées depuis de nombreuses années pour fournir une indication de la fiabilité des mots reconnus. D'autres aspects sont peu ou pas étudiés. L'étude de l'*incertitude en modélisation acoustique* porte sur l'exploitation de l'incertitude provenant de la séparation de sources, qui exprime l'écart entre le signal estimé (par les approches de séparation de source) et le signal « propre » (théorique). Les travaux concernent l'estimation de cette incertitude et sa prise en compte dans la reconnaissance de la parole [1]. L'étude de l'*incertitude en segmentation phonétique* a pour objectif d'aboutir à une indication de la précision de la frontière des phonèmes lors de la reconnaissance de la parole ou lors de l'alignement texte-parole [9]. La fiabilité des frontières conditionne la pertinence des diagnostics qui peuvent être effectués pour l'aide à l'apprentissage de langues, d'où l'utilité de prendre en compte une estimation de l'incertitude dans l'établissement de tels diagnostics. Pour compléter l'*incertitude sur les paramètres prosodiques* il faut aussi considérer, en plus de la segmentation phonétique, l'estimation de la fréquence fondamentale. Là aussi, l'objectif est d'introduire une mesure de fiabilité sur l'estimation de ce paramètre [3].

Quelques uns des travaux ci-dessus amènent à construire des corpus de signaux spécifiques, comme par exemple, l'acquisition de données articulatoires par IRM [5], l'acquisition de parole multimodale [8], l'acquisition de parole non-native en lien avec l'apprentissage de langues [12], et aussi l'acquisition de signaux avec prise de son distante [2].

Projets marquants

Pour terminer, nous citons ici quelques projets collaboratifs en cours, ou récemment terminés, en lien avec les thèmes décrits ci-dessus.

ALLEGRO - Foreign language learning (Interreg IV-A Project, 2010-2012).
AMIS - Access Multilingual Information opinionS (CHIST-ERA, 2015-2018).
ARTSPEECH - Phonetic articulatory synthesis (ANR, 2015-2019).
ARTIS - Acoustic-to-articulatory inversion of speech (ANR domaines émergents, 2009-2012).
CONTNOMINA - Exploitation of context for proper names recognition in diachronic audio documents (ANR Blanc SIMI 2, 2013-2016).
COREXp - Acquisition, Processing and Analysis of a Corpus for the Synthesis of Expressive Audiovisual Speech (Région Lorraine, 2014-2016).
DYCI2 - Creative Dynamics of Improvised Interaction (ANR, 2015-2018).
EMOSPEECH - Spoken interactions in serious games (Eurostar project, 2010-2013).
IFCASL - Individualized Feedback for Computer-Assisted Spoken Language Learning (Programme franco-allemand en SHS, ANR+DFG, 2013-2016).
KAMoulox - Kernel additive modelling for the unmixing of large audio archives (ANR jeunes chercheurs, 2015-2019).
LCHN - Langues, connaissances et humanités numériques (CPER, Contrat Plan Etat-Région, 2015-2020).

ORFEO - Tools and ressources for written and spoken French (ANR Corpus, 2013-2016).
ORTOLANG - Open Resources and TOols for LANGuage (EQUIPEX, ANR investissements d'avenir, 2012-2016).
RAPSODIE - Automatic speech recognition for hard of hearing and handicapped people (FUI + FEDER, 2012-2016).
VISAC - Acoustic-Visual Speech Synthesis by Bimodal Unit Concatenation (ANR jeunes chercheurs, 2009-2012).
VOCADOM - Commande vocale robuste adaptée à la personne et au contexte pour l'autonomie à domicile (ANR, 2017-2020).
VOICEHOME - Robust voice control system for smart home and multimedia applications (FUI, 2015-2017).
TALC - Traitement automatique des langues et connaissances (CPER MISN, Contrat Plan Etat-Région, 2009-2014).

Références

[1] Ahmed H. Abdelaziz, Shinji Watanabe, John R. Hershey, Emmanuel Vincent, and Dorothea Kolossa. Uncertainty propagation through deep neural networks. In *Interspeech 2015*, Dresden, Germany, September 2015.

[2] Nancy Bertin, Ewen Camberlein, Emmanuel Vincent, Romain Lebarbenchon, Stéphane Peillon, Éric Lamandé, Sunit Sivasankaran, Frédéric Bimbot, Irina Illina, Ariane Tom, Sylvain Fleury, and Eric Jamet. A French corpus for distant-microphone speech processing in real homes. In *Interspeech 2016*, San Francisco, United States, September 2016.

[3] Boyuan Deng, Denis Jouvet, Yves Laprie, Ingmar Steiner, and Aghilas Sini. Towards Confidence Measures on Fundamental Frequency Estimations. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, New Orleans, United States, March 2017.

[4] Benjamin Elie and Yves Laprie. Extension of the single-matrix formulation of the vocal tract : consideration of bilateral channels and connection of self-oscillating models of the vocal folds with a glottal chink. *Speech Communication*, 82 :85–96, September 2016.

[5] Benjamin Elie, Yves Laprie, and Pierre-André Vuissoz. Acquisition temps- réel de données articulatoires par IRM : application a la synthèse par copie. In *13ème Congrès Français d'Acoustique (CFA 2016)*, Le Mans, France, April 2016. SFA.

[6] Aditya Arie Nugraha, Antoine Liutkus, and Emmanuel Vincent. Multichannel audio source separation with deep neural networks. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 24(10) :1652–1664, June 2016.

[7] Slim Ouni, Vincent Colotte, Sara Dahmani, and Soumaya Azzi. Acoustic and Visual Analysis of Expressive Speech : A Case Study of French Acted Speech. In *Interspeech 2016*, volume 2016, pages 580 – 584, San Francisco, United States, November 2016. ISCA.

[8] Slim Ouni and Sara Dahmani. Is markerless acquisition of speech production accurate ? *Journal of the Acoustical Society of America*, 139(6), May 2016.

[9] Guillaume Serrière, Christophe Cerisara, Dominique Fohr, and Odile Mella. Weakly-supervised text-to-speech alignment confidence measure. In *International Conference on Computational Linguistics (COLING)*, Proceedings of the 26th International Conference on Computational Linguistics (COLING), Osaka, Japan, December 2016.

[10] Imran Ahamad Sheikh, Dominique Fohr, Irina Illina, and Georges Linares. Modelling Semantic Context of OOV Words in Large Vocabulary Continuous Speech Recognition. *IEEE/ACM Transactions on Audio, Speech and Language Processing*, 25(3) :598 – 610, January 2017.

[11] Sunit Sivasankaran, Emmanuel Vincent, and Irina Illina. Discriminative importance weighting of augmented training data for acoustic model training. In *42th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2017)*, New Orleans, United States, March 2017. Added missing sign in

equations (2) and (3) + explanation about iteration 1 in Fig. 1.

[12] Jürgen Trouvain, Anne Bonneau, Vincent Colotte, Camille Fauth, Dominique Fohr, Denis Jouvet, Jeanin Jügler, Yves Laprie, Odile Mella, Bernd Möbius, and Frank Zimmerer. The IFCASL Corpus of French and German Non-native and Native Read Speech. In *LREC'2016, 10th edition of the Language Resources and Evaluation Conference*, Proceedings LREC'2016, Portorož, Slovenia, May 2016.

ROC : recherche opérationnelle, optimisation combinatoire, contraintes

LAAS
CNRS
https://www.laas.fr/public/en/roc

Christian ARTIGUES
artigues@laas.fr
+33 5 61 33 79 07

Membres impliqués : [11]

- Christian ARTIGUES, DR CNRS
- Cyril BRIAND, PR
- Patrick ESQUIROL, MCF
- Emmanuel HÉBRARD, CR CNRS
- Marie-José HUGUET, PR
- Pierre LOPEZ, DR CNRS
- Margaux NATTAF, Post-doc
- Francisco BARBOSA-ANDA, Doctorant

Mots-clés

- Programmation par Contraintes
- Ordonnancement
- Contraintes globales
- Recherche dirigée par les conflits
- Recherche locale et Métaheuristiques
- Optimisation et apprentissage automatique
- Gestion de l'incertitude

Thématique générale de l'équipe

Les domaines de recherche considérés par l'équipe ROC (Recherche Opérationnelle, Optimisation Combinatoire & Contraintes) sont des branches de la Recherche Opérationnelle de l'Intelligence Artificielle. L'équipe ROC propose des modèles et des algorithmes variés pour plusieurs classes de problèmes d'optimisation combinatoire et de satisfaction de contraintes. Pour atteindre cet objectif, l'équipe développe d'une part des études structurelles sur des problèmes fondamentaux en théorie des graphes, en ordonnancement, en satisfaction de contraintes et en programmation linéaire en nombres entiers. D'autre part, pour les problèmes NP-difficiles, l'équipe cherche à concevoir et à évaluer des approches de résolution génériques pour faire face à l'explosion combinatoire des espaces de solution explorés. Dans une volonté d'améliorer l'applicabilité des approches proposées, l'équipe cherche à prendre en compte les incertitudes (par des considérations de robustesse) ainsi que la présence d'objectifs et/ou de centres de décisions multiples. L'équipe cherche à confronter les méthodes proposées au monde réel via des applications au génie industriel et l'intégration de facteurs humains dans des domaines variés qui incluent le transport, l'industrie manufacturière, la gestion de la chaîne logistique, la gestion de l'énergie, l'aéronautique et l'espace. Enfin, pour favoriser la diffusion de ses recherches et pour évaluer ses méthodes sur le plan international, l'équipe développe des solveurs non commerciaux et participe à des compétitions internationales.

Dans le domaine de l'intelligence artificielle, les activités de l'équipe se concentrent essentiellement, d'une part, sur *la résolution de problèmes combinatoires* par des méthodes de recherche arborescente et d'inférence, sous le paradigme de la programmation par contraintes, ou par des méthodes de

11. Seules sont représentées les personnes impliquées dans la thématique donnée par les mots-clés.

recherche locale et de métaheurisiques et, d'autre part, sur *l'apprentissage automatique*, soit comme un élément d'accélération de la recherche arborescente pour la résolution de problèmes, soit comme une application de méthodes d'optimisation.

Projets marquants

Propagation de contraintes globales

Historiquement, les travaux du LAAS ont été précurseurs des approches par contraintes pour l'ordonnancement, avec les travaux de Jacques Erschler et François Roubellat dans les années 70 déterminant des conditions nécessaires sous la forme d'ordres partiels imposés en ordonnancement à une machine [15]. Dans les années 80, les travaux de Patrick Esquirol développent le principe de la consistance locale et son utilisation au sein de la recherche arborescente pour le fameux problème de *job-shop* [14] puis dans les années 90, les travaux de Pierre Lopez s'intéressent à des problèmes d'ordonnancement cumulatif et développent le concept de raisonnement énergétique, basé sur des bilans de consommations des tâches sur des intervalles remarquables [24]. Ces travaux abondamment cités dans la communauté *Contraintes* sont à l'origine de puissants algorithmes de filtrage implémentés dans plusieurs solveurs commerciaux (tels CP Optimizer) ou libres (tels Choco).

L'équipe ROC poursuit aujourd'hui des travaux dans le domaine des contraintes globales pour l'ordonnancement ou le séquencement, avec la volonté de déterminer des algorithmes de filtrages puissants et de complexité polynomiale. Ainsi, des travaux récents ont étendu le raisonnement énergétique à une contrainte cumulative à profil de tâches variables avec des fonctions linéaires et non linéaires de conversion de ressource [26, 27]. Un algorithme de filtrage linéaire optimal a été proposé pour l'arc cohérence d'une contrainte de séquencement, utile par exemple pour les chaînes d'assemblage de véhicules [30] (prix *Honorable mention* à CP 2012). D'autres algorithmes efficaces de propagation ont été proposés pour des contraintes globales d'équité [6], de connectivité dans les graphes [8] et de classement [7].

Apprentissage pour la recherche arborescente

L'équipe ROC mène des travaux sur l'intégration de mécanismes d'apprentissage au sein de la recherche arborescente pour la résolution de problèmes. Ces mécanismes peuvent être mis en œuvre par de simples, mais efficaces, actualisations en fonction des échecs des poids associés aux variables et aux valeurs dans la recherche arborescente, selon les principes de la recherche dirigée par les conflits, avec d'excellents résultats sur des problèmes classiques d'ordonnancement [19, 23]. Des travaux récents vont plus loin en intégrant au sein de la recherche arborescente des résultats d'explications d'échecs provoqués par des contraintes, ce qui permet par exemple d'augmenter significativement la robustesse d'heuristiques de branchement tels *weighted degree* [22] ou d'alimenter efficacement une base de clauses, à la manière des solveurs SAT. Ces travaux ont permis de fermer de nouvelles instances du problème de *Job-Shop* [29]. Sur ce sujet et sur le précédent, Mohamed Siala, doctorant de l'équipe ROC encadré par Christian Artigues et Emmanuel Hébrard, a obtenu le prix de thèse 2016 (*honorable mention*) de l'EurIA (*European Association for Artificial Intelligence*) [28].

Optimisation pour l'apprentissage automatique

De nouveaux travaux, visent, de manière duale au point précédent, à améliorer les méthodes d'apprentissage automatique en y intégrant des algorithmes d'optimisation. Le problème de la minimisation du temps de réponse moyen d'un outil d'apprentissage automatique pour la détection d'objets dans une image a été résolu par la programmation mathématique, ce qui permet d'accélérer considérablement les temps de réponse tout en maintenant les performances [25, 3]. Enfin, une troisième intégration possible de ces deux thèmes a été étudiée : l'apprentissage automatique de modèles d'optimisation [5].

Complexité des problèmes de satisfaction de contraintes

L'étude de la complexité des problèmes de satisfaction de contraintes fait partie des domaines de recherche de l'équipe. En particulier, la notion de complexité paramétrée a été appliquée à de nombreux aspects de la programmation par contraintes : propagation de contraintes globales [21], recherche de "backdoors" [4, 12] et l'introduction d'un nouveau schéma de kernelisation pour la propagation de contraintes [13]. Clément Carbonnel, doctorant de l'équipe ROC encadré par Martin Cooper (IRIT) et Emmanuel Hébrard, a reçu deux prix sur ce thème, et plus généralement sur l'étude théorique des classes traitables de CSP : le *Best Student Paper* à CP 2016 pour ses travaux sur le problème de reconnaissance de classes traitables de CSP [10] et le ACP (Association for Constraint Programming) *Doctoral Research award* 2017 [11].

Ordonnancement multi-agent, décision dans l'incertain

Une partie importante des activités de l'équipe ROC s'inscrit dans la prise en compte de l'incertitude et de la présence de plusieurs décideurs dans les problèmes de décision et d'optimisation. Des travaux concernent la recherche de solutions optimisant une fonction objectif sous la contrainte qu'elle soit stable (c'est-à-dire qu'elle corresponde à un équilibre de Nash) [1, 17]. D'autres travaux, dans la lignée des travaux de François Roubellat dans les années 80 sur les groupes de tâches permutables [16], s'intéressent au concept de *solution flexible* qui fournit une représentation compacte d'ensembles de solutions réalisables à performance bornée pouvant être exploitées par différents agents ou bien, dans le cadre de l'ordonnancement contingent, pour disposer de solutions alternatives face aux incertitudes. Différentes structures d'ordonnancements flexibles et des algorithmes les exploitant ont été proposés [2, 9]. Des algorithmes efficaces ont été développés pour la prise en compte de durées incertaines dans les réseaux temporels d'activités [18].

Applications au secteur spatial

L'équipe a appliqué avec succès des techniques de programmation par contraintes et de recherche locale au secteur spatial. Un projet marquant a été la réalisation de l'algorithme d'ordonnancement des expériences du robot Philae sur la comète Tchouri dans le cadre du projet ROSETTA (*Best application paper award* CP 2012) [31], voir aussi l'article "CP has landing on a comet" de l'ACP [12]. La programmation par contraintes a été également appliquée à des problèmes de tests de satellites de télécommunication [20] et des métaheuristiques multi-objectifs ont permis de résoudre des problèmes d'ordonnancement des observations de la terre [32].

12. http://www.a4cp.org/node/1058

Références

[1] Alessandro Agnetis, Cyril Briand, Jean-Charles Billaut, and Premysl Sucha. Nash equilibria for the multi-agent project scheduling problem with controllable processing times. *J Sched*, 18(1) :15–27, 2015.

[2] Mohamed Ali Aloulou and Christian Artigues. Flexible solutions in disjunctive scheduling : General formulation and study of the flow-shop case. *Comput Oper Res*, 37(5) :890–898, 2010.

[3] Francisco Rodolfo Barbosa-Anda, Cyril Briand, Frédéric Lerasle, and Alhayat Ali Mekonnen. Mean response-time minimization of a soft-cascade detector. In *ICORES*, pages 252–260, 2016.

[4] Christian Bessiere, Clément Carbonnel, Emmanuel Hebrard, George Katsirelos, and Toby Walsh. Detecting and exploiting subproblem tractability. In *IJCAI*, pages 468–474, 2013.

[5] Christian Bessiere, Remi Coletta, Emmanuel Hebrard, George Katsirelos, Nadjib Lazaar, Nina Narodytska, Claude-Guy Quimper, Toby Walsh, et al. Constraint acquisition via partial queries. In *IJCAI*, volume 13, pages 475–481, 2013.

[6] Christian Bessiere, Emmanuel Hebrard, George Katsirelos, Zeynep Kiziltan, Émilie Picard-Cantin, Claude-Guy Quimper, and

Toby Walsh. The balance constraint family. In *CP*, pages 174–189, 2014.

[7] Christian Bessiere, Emmanuel Hebrard, George Katsirelos, Zeynep Kiziltan, and Toby Walsh. Ranking constraints. In *IJCAI*, pages 705–711, 2016.

[8] Christian Bessière, Emmanuel Hebrard, George Katsirelos, and Toby Walsh. Reasoning about connectivity constraints. In *IJCAI*, pages 2568–2574, 2015.

[9] Cyril Briand, Hoang Trung La, and Jacques Erschler. A robust approach for the single machine scheduling problem. *J Sched*, 10(3) :209–221, 2007.

[10] Clément Carbonnel. The dichotomy for conservative constraint satisfaction is polynomially decidable. In *CP*, pages 130–146, 2016.

[11] Clément Carbonnel. *Harnessing tractability in constraint satisfaction problems*. PhD thesis, INP Toulouse, 2016.

[12] Clément Carbonnel, Martin C Cooper, and Emmanuel Hebrard. On backdoors to tractable constraint languages. In *CP*, pages 224–239. Springer, 2014.

[13] Clément Carbonnel and Emmanuel Hebrard. On the kernelization of global constraints. In *IJCAI*, pages 578–584, 2017.

[14] Jacques Erschler and Patrick Esquirol. Decision-aid in job shop scheduling : A knowledge based approach. In *IEEE ICRA*, pages 1651–1656, 1986.

[15] Jacques Erschler, François Roubellat, and JP Vernhes. Finding some essential characteristics of the feasible solutions for a scheduling problem. *Oper Res*, 24(4) :774–783, 1976.

[16] Jacques Erschler and François Roubellat. An approach for real time scheduling for activities with time and resource constraints. In *Advances in project scheduling*. Elsevier, 1989.

[17] Nadia Chaabane Fakhfakh, Cyril Briand, Marie-José Huguet, and Alessandro Agnetis. Finding a nash equilibrium and an optimal sharing policy for multiagent network expansion game. *Networks*, 69(1) :94–109, 2017.

[18] Thierry Garaix, Christian Artigues, and Cyril Briand. Fast minimum float computation in activity networks under interval uncertainty. *J Sched*, 16(1) :93–103, 2013.

[19] Diarmuid Grimes and Emmanuel Hebrard. Solving variants of the job shop scheduling problem through conflict-directed search. *INFORMS J Comput*, 27(2) :268–284, 2015.

[20] Emmanuel Hebrard, Marie-José Huguet, Daniel Veysseire, Ludivine Boche Sauvan, and Bertrand Cabon. Constraint programming for planning test campaigns of communications satellites. *Constraints*, 22(1) :73–89, 2017.

[21] Emmanuel Hebrard, Dániel Marx, Barry O'Sullivan, and Igor Razgon. Soft constraints of difference and equality. *J Artif Intell Res*, 41 :97–130, 2011.

[22] Emmanuel Hebrard and Mohamed Siala. Explanation-based weighted degree. In *CPAIOR*, pages 167–175. Springer, 2017.

[23] Marie-José Huguet, Pierre Lopez, and Wafa Karoui. Weight-based heuristics for constraint satisfaction and combinatorial optimization problems. *J Math Model Alg*, 11(2) :193–215, 2012.

[24] Pierre Lopez. *Approche énergétique pour l'ordonnancement de tâches sous contraintes de temps et de ressources*. Thèse de Doctorat, Université Paul Sabatier, Toulouse, 1991.

[25] Alhayat Ali Mekonnen, Cyril Briand, Frédéric Lerasle, and Ariane Herbulot. Fast HOG based person detection devoted to a mobile robot with a spherical camera. In *IEEE/RSJ ICIRS*, pages 631–637, 2013.

[26] Margaux Nattaf, Christian Artigues, and Pierre Lopez. A hybrid exact method for a scheduling problem with a continuous resource and energy constraints. *Constraints*, 20(3) :304–324, 2015.

[27] Margaux Nattaf, Christian Artigues, and Pierre Lopez. Cumulative scheduling with variable task profiles and concave piecewise linear processing rate functions. *Constraints*, 22(4) :530–547, 2017.

[28] Mohamed Siala. *Search, propagation, and learning in sequencing and scheduling problems.* PhD thesis, INSA Toulouse, France, 2015.

[29] Mohamed Siala, Christian Artigues, and Emmanuel Hebrard. Two clause learning approaches for disjunctive scheduling. In *CP*, pages 393–402, 2015.

[30] Mohamed Siala, Emmanuel Hebrard, and Marie-José Huguet. An optimal arc consistency algorithm for a chain of atmost constraints with cardinality. In *CP*, pages 55–69, 2012.

[31] Gilles Simonin, Christian Artigues, Emmanuel Hebrard, and Pierre Lopez. Scheduling scientific experiments on the rosetta/philae mission. In *CP*, pages 23–37, 2012.

[32] Panwadee Tangpattanakul, Nicolas Jozefowiez, and Pierre Lopez. Multi-objective optimization for selecting and scheduling observations by agile earth observing satellites. In *PPSN XII*, pages 112–121, 2012.

SemLIS : Sémantique, Logiques, Systèmes d'Information

IRISA, UMR 6074
Université de Rennes 1, INSA Rennes
http://www-semlis.irisa.fr/

Sébastien FERRÉ
ferre@irisa.fr
+33 2 99 84 75 70

Membres :

- Guillaume AUCHER, MCF (Univ. Rennes 1)
- Carlos BOBED, Invité (Univ. Zaragoza)
- Peggy CELLIER, MCF (INSA Rennes)
- Mireille DUCASSÉ, PR (INSA Rennes)
- Sébastien FERRÉ, MCF (HDR, Univ. Rennes 1)
- Annie FORET, MCF (HDR, Univ. Rennes 1)
- Pierre MAILLOT, Post-doc (Univ. Rennes 1)
- Olivier RIDOUX, PR (Univ. Rennes 1)

Mots-clés

- Représentation de connaissances, web sémantique, graphes de connaissances
- Langages formels, logiques (de description, épistémiques, dynamiques)
- Systèmes multi-agents, représentation de l'incertitude, planification
- Systèmes d'information, recherche d'information
- Fouille de données, aide à la décision, analyse de concepts formels
- Traitement automatique de la langue, grammaires catégorielles, terminologie, langues naturelles contrôlées, patrons linguistiques, extraction d'information
- Décision en groupe et négotiation, choix social
- Interaction données-utilisateurs, recherche d'information exploratoire

Thématique générale de l'équipe

Dans un contexte où les données et les connaissances croissent toujours plus, autant en volume qu'en variété (Big Data), le principal objectif de l'équipe SemLIS est de **redonner aux utilisateurs le pouvoir sur leurs données**. Par utilisateur nous entendons tout individu ou groupe qui a un fort intérêt pour certaines données et qui a besoin de les exploiter pour en tirer de nouvelles connaissances ou pour prendre des décisions. Les travaux de l'équipe visent à faciliter l'interaction entre les utilisateurs et leurs données en rendant ces utilisateurs plus autonomes et plus agiles, en offrant de la flexibilité et de l'expressivité, tout en donnant du contrôle et de la confiance dans le système d'information. Un tel système d'information doit assister les utilisateurs dans la représentation sémantique de données hétérogènes, et dans l'acquisition collaborative des connaissances du domaine. Les fondements scientifiques de l'équipe sont les logiques et les langages formels pour la représentation de connais-

sances et le raisonnement automatique, le Web sémantique, les systèmes d'information, le traitement de la langue naturelle, la fouille de données symbolique et l'interaction données-utilisateurs. Une idée clé est de réconcilier la puissance des langages formels et la facilité d'utilisation de la langue naturelle et de l'interaction. Les systèmes développés dans l'équipe sont qualifiés de « Systèmes d'information logiques » et ont pour la plupart un nom qui se termine par -LIS, l'abbréviation de *Logical Information System*.

Les principales thématiques développées dans l'équipe SemLIS sont les suivantes.

Graphes de connaissances et Web sémantique. Les formalismes du Web sémantique (RDF, RDFS, OWL) permettent la représentation de connaissances sous forme de graphes et l'exploitation de ces connaissances par des machines (raisonnement, interrogation, fouille et apprentissage). Le principal obstacle à l'adoption du Web sémantique est le fossé existant entre ces formalismes conçus pour les machines et les consommateurs/producteurs de ces connaissances qui ne sont généralement pas informaticiens. SemLIS cherche à réconcilier la précision et l'expressivité de ces formalismes avec la facilité d'utilisation d'interfaces en langage naturel ou visuelles. Pour cela, nous développons des systèmes faisant la médiation entre les utilisateurs et leurs données. Cette médiation se traduit par un dialogue où le système guide l'utilisateur dans la construction incrémentale de requêtes, des plus simples aux plus complexes. Le guidage est basé sur les données et évite, par exemple, la construction de requêtes sans résultat. Le système Sparklis [10] (http://www.irisa.fr/LIS/ferre/sparklis/) couvre la plus grande partie du langage de requête SPARQL tout en cachant SPARQL derrière une verbalisation en langue naturelle. Le système Sewelis [12] permet l'édition collaborative de descriptions RDF, en faisant des recommandations sur la base des descriptions RDF existantes. Le système PEW permet la conception et la complétion d'ontologies OWL sans avoir à écrire des axiomes mais en laissant l'utilisateur explorer les «mondes possibles» d'une ontologie, et en lui permettant d'éliminer les «mondes» indésirables.

Extraction de connaissances. L'équipe SemLIS s'intéresse aussi au domaine de l'extraction de connaissances et plus spécifiquement au développement d'approches symboliques comme la fouille de motifs (ensemblistes, séquentiels ou de graphes) pour diverses applications comme le texte [6] ou les graphes de connaissances [11]. Un bagage scientifique partagé par plusieurs membres de l'équipe et à l'origine des fondements de l'équipe est l'analyse de concepts formels (FCA) qui est une approche ensembliste de fouille de données. C'est en s'appuyant sur des concepts de la FCA qu'un certain nombre de contributions ont été produites [7, 11]. Dans la même veine, une partie de l'équipe s'intéresse aussi à l'apprentissage de grammaires de langages naturels à partir de corpus en s'appuyant sur les grammaires catégorielles [4].

Traitement automatique de la langue naturelle. En traitement automatique de la langue naturelle (TAL), il existe plusieurs types d'approches que l'on peut grossièrement diviser en deux grandes familles : les approches symboliques et les approches numériques. SemLIS se place dans la première catégorie. Les tâches auxquelles s'attaquent les membres de l'équipe sont principalement :

- l'analyse syntaxique et sémantique des langues naturelles avec des grammaires catégorielles [5],
- l'extraction d'informations (EI), qu'elles soient syntaxiques ou sémantiques, à partir de textes avec des techniques de fouille à base de motifs séquentiels [6],
- la définition de langages naturels contrôlés avec des grammaires de Montague [9].

Nous appliquons aussi nos outils LIS à des ressources linguistiques comme la terminologie [13].

Systèmes multi-agents et planification. Plusieurs de nos activités quotidiennes qui étaient auparavant réalisées dans le monde « réel » et en interaction avec d'autres humains sont maintenant conduites dans un monde numérique en interaction avec des « agents » non-humains : commerce en ligne, vote électronique, banque en ligne, etc. Afin de mieux contrôler et développer ce monde numérique, nous avons besoin de comprendre, modéliser et prédire cette interaction entre agents. La logique

et la théorie des jeux forment actuellement le cadre théorique le plus influent pour traiter de cette interaction. Le travail dans ces domaines est centré sur la représentation et le raisonnement à propos de l'incertitude dans un cadre avec plusieurs agents. En particulier, la logique épistémique a développé dans les dernières décennies plusieurs outils abstraits et puissants pour représenter et raisonner sur l'incertitude. Notre premier axe de recherche consiste à développer et approfondir ces outils logiques tout en s'efforçant de les mettre en relation et de les unifier [3]. Notre second axe de recherche consiste à appliquer et utiliser ces outils pour reformuler différentes théories dans le but d'identifier des structures mathématiques communes [2]. Cela nous a conduit à investiguer la révision de croyances et la planification dans un cadre multi-agents [1], et de transférer des méthodes et résultats de la logique modale pour la formalisation de la notion de *only knowing*, un concept épistémique qui revêt une importance pratique pour les bases de connaissances.

Aide à la décision en groupe. Le domaine « décision en groupe et négotiation » se concentre sur les processus complexes et auto-organisateurs qui posent des problèmes multi-participants, multi-critères, mal structurés, dynamiques et souvent évolutifs. Ce domaine s'intéresse à l'ensemble du processus ou du flot d'activités qui intervient dans la prise de décision, et pas seulement dans le choix final. Cela inclut les aspects suivants du processus : l'inventaire des options, la communication et le partage d'information, la définition et l'évolution du problème, la génération alternative et l'interaction socio-émotionnelle. Les systèmes de soutien à la décision en groupe (GDSS) font partie des principales approches des problèmes ci-dessus.

Notre axe de recherche actuel est de montrer que nos LIS (Systèmes d'information logiques) fournissent un support technologique innovant pour la plupart des aspects mentionnés ci-dessus [8]. En particulier, les capacités de navigation et de filtrage des LIS aident à la détection d'incohérences et de connaissances manquantes pendant les réunions. Les capacités de mise-à-jour des LIS permettent aux participants d'ajouter des objets, des critères et des liens entre eux à la volée. En conséquence, le groupe dispose d'un ensemble d'informations plus complet et plus pertinent. De plus, les vues compactes fournies par les LIS aident les participants à mieux appréhender l'ensemble des connaissances requises. Le groupe peut ainsi se construire une compréhension partagée des informations pertinentes, lesquelles étaient auparavant réparties entre les participants. Enfin, les capacités de navigation et de filtrage des LIS sont adéquates pour rapidement converger vers un nombre réduits d'options.

Interaction données-utilisateurs. En raison de l'importance que nous donnons à l'interaction entre les utilisateurs et leurs données, SemLIS s'est investi dans des techniques qui permettent de structurer et de raisonner sur ces interactions. On peut citer, en particulier, la recherche à facettes qui est souvent utilisée dans les sites de e-commerce, OLAP (On-Line Analytical Processing) qui est souvent utilisée en *Business Intelligence*, et les Systèmes d'information géographiques (SIG). Sur ces sujets, nous avons par exemple contribué à l'exploration de données géographiques, à la découverte de dépendances fonctionnelles et de règles d'association avec des cubes OLAP, et à l'extension de la recherche à facettes aux graphes RDF [12].

Projets marquants

Le champ d'application de SemLIS est très large. Rien qu'en regardant les données ouvertes du Web sémantique (LOD), on trouve des dizaines de milliards de triplets dans des domaines variés : sciences de la vie, média, organisations gouvernementales, publications, géographie, etc. Nos expériences et collaborations passées et présentes nous ont amené à viser en priorité les utilisateurs au milieu du spectre allant des purs informaticiens au grand public, c'est-à-dire les individus et groupes qui sont experts d'un domaine impliquant la gestion de données et connaissances. Notre objectif est de permettre à ces utilisateurs de réaliser eux-même des tâches (par ex., interroger une base de données) qui requièrent normalement des compétences en informatique (par ex., écrire une requête SQL).

Les projets suivants donnent des exemples de

collaborations impliquant différents types d'utilisateurs et différents domaines d'application.

IDFRAud 2015-2018 (ANR)

Le projet IDFRAud vise à permettre la reconnaissance automatique de documents d'identité et la détection de fraudes, en appliquant des techniques d'analyse de documents, de classification et de gestion de connaissances. Le coordinateur est l'entreprise innovante AriadNEXT et les autres partenaires sont l'IRISA, l'IRCGN (Institut de Recherche Criminelle de la Gendarmerie Nationale) et l'ENSP (École National Supérieure de Police). Deux équipes de l'IRISA sont impliquées : SemLIS et LinkMedia.

L'équipe SemLIS a contribué sur la base de connaissances des modèles de documents d'identité (il y en a des milliers dans le monde) en collaboration avec AriadNEXT. Nous avons notamment réussi un transfert de connaissances vers cette entreprise des technologies du Web sémantique qu'ils ont pleinement adoptées. Nous avons aussi contribué à l'analyse de collections de faux documents en collaboration avec l'IRCGN. Nous avons notamment appliqué l'Analyse de concepts formels pour identifier des groupes de faux documents susceptibles d'avoir une source commune, une forme d'étude criminalistique. Nous avons aussi développé FORMULIS, une interface d'acquisition de données sémantiques accessible à des non-informaticiens et compatible avec notre outil Sewelis. FORMULIS est appliqué dans le projet à la description des modèles de documents et aux faux documents, mais est en fait indépendant du domaine d'application.

PEGASE 2016-2020 (ANR)

L'équipe SemLIS a été invitée à rejoindre la proposition de projet PEGASE pour son logiciel Sparklis, comme moyen de réconcilier les aspects formels des langages du Web sémantique et le besoin d'utilisabilité pour les utilisateurs finaux du projet, à savoir des experts en pharmaco-vigilance. Le coordinateur du projet est le SSPIM (« Service de santé publique et de l'information médicale ») et le CHU St Etienne. Le projet réunit des chercheurs en informatique du LIMICS et de l'IRISA, des experts en pharmaco-vigilance de 4 centres régionaux de pharmaco-vigilance (Besançon, Lille, Paris HEGP, Toulouse) et des experts de l'évaluation d'interfaces utilisateur dans le domaine médical du CIC-IT Evalab.

La mission des centres de pharmaco-vigilance est de collecter, annoter, stocker, analyser et prévenir les effets indésirables des médicaments. Ils s'appuient sur la terminologie MedDRA (Medical Dictionary for Regulatory Activities) pour annoter de nouveaux cas, et pour retrouver d'anciens cas. L'objectif de ce projet est de développer et comparer plusieurs interfaces utilisateur permettant à des experts en pharmaco-vigilance de naviguer et interroger la terminologie afin d'identifier les termes pertinents. L'objectif pour SemLIS est de rendre l'interface de Sparklis plus proche de la langue naturelle et de profiter de l'évaluation auprès d'experts (par les membres du CIC-IT Evalab) pour faire progresser de façon importante l'utilisabilité de Sparklis. Un autre enjeu important est la diffusion de Sparklis et son utilisation effective là où il y a des données sémantiques, comme c'est déjà le cas pour Persée (http://data.persee.fr/).

REUs 2016-2018 (FUI)

Le projet REUs vise à permettre l'assistance intelligente à la production de documents liés aux réunions de projets. Le coordinateur est l'entreprise VISEO et les autres partenaires sont 5 industriels et deux laboratoires le GREYC de Caen et le laboratoire Hubert Curien de Saint-Étienne. L'équipe SemLIS participe à ce projet à travers une collaboration avec le GREYC. Les défis de ce projet s'étendent de la captation des voix pendant la réunion, à leur retranscription sous format texte puis à l'extraction d'informations intéressantes pour la génération de comptes rendu. L'équipe SemLIS collabore avec le GREYC sur la partie extraction d'information dans les retranscriptions textuelles. Un objectif attendu est l'amélioration des approches existantes s'appuyant sur des motifs séquentiels, notamment afin de capturer de l'information interphrastique.

Cour de Cassation 2014-2016 (ENS CHRC, Univ. Rennes 1)

L'équipe SemLIS a participé à un projet multidisciplinaire avec la Cour de Cassation sur le développement d'un prototype de logiciel d'aide à la prise de décision lors de la rédaction de jugements. Ce projet a reçu des fonds de l'ENS CHRC (Collège de Recherche Hubert Curien) et de l'Université de Rennes 1 (Défi émergent).

Références

[1] Guillaume Aucher. DEL-sequents for regression and epistemic planning. *Journal of Applied Non-Classical Logics*, 22(4) :337–367, 2012.

[2] Guillaume Aucher. Intricate Axioms as Interaction Axioms. *Studia Logica*, page 28, 2015.

[3] Guillaume Aucher. Dynamic Epistemic Logic in Update Logic. *Journal of Logic and Computation*, March 2016.

[4] D. Bechet, A. Foret, and I. Tellier. Learnability of pregroup grammars. *Studia Logica*, 87(2-3), 2007.

[5] Denis Béchet, Alexandre Dikovsky, and Annie Foret. Categorial grammars with iterated types form a strict hierarchy of k-valued languages. *Theor. Comput. Sci.*, 450 :22–30, 2012.

[6] Nicolas Béchet, Peggy Cellier, Thierry Charnois, and Bruno Crémilleux. Discovering linguistic patterns using sequence mining. In Alexander F. Gelbukh, editor, *Int. Conf. on Computational Linguistics and Intelligent Text Processing (CICLing)*, volume 7181 of *LNCS*, pages 154–165. Springer, 2012.

[7] Peggy Cellier, Sébastien Ferré, Olivier Ridoux, and Mireille Ducassé. A parameterized algorithm to explore formal contexts with a taxonomy. *Int. J. Foundations of Computer Science (IJFCS)*, 19(2) :319–343, 2008.

[8] Mireille Ducassé and Peggy Cellier. Fair and fast convergence on islands of agreement in multicriteria group decision making by logical navigation. *Group Decision and Negotiation*, 23(4) :673–694, July 2014.

[9] Sébastien Ferré. SQUALL : The expressiveness of SPARQL 1.1 made available as a controlled natural language. *Data & Knowledge Engineering*, 94 :163–188, 2014.

[10] Sébastien Ferré. Sparklis : An expressive query builder for SPARQL endpoints with guidance in natural language. *Semantic Web : Interoperability, Usability, Applicability*, 8(3) :405–418, 2017.

[11] Sébastien Ferré and Peggy Cellier. Graph-FCA in practice. In O. Haemmerlé, G. Stapleton, and C. Faron-Zucker, editors, *Int. Conf. Conceptual Structures (ICCS) - Graph-Based Representation and Reasoning*, LNCS 9717, pages 107–121. Springer, 2016.

[12] Sébastien Ferré and Alice Hermann. Reconciling faceted search and query languages for the Semantic Web. *Int. J. Metadata, Semantics and Ontologies*, 7(1) :37–54, 2012.

[13] Annie Foret. A logical information system proposal for browsing terminological resources. In T. Poibeau and P. Faber, editors, *Int. Conf. Terminology and Artificial Intelligence*, volume 1495 of *CEUR Workshop Proceedings*, pages 51–59, 2015.

SYNALP : Statistic and Symbolic Natural Language Processing

LORIA (Laboratoire Lorrain d'Informatique et ses Applications)
Unité Mixte de Recherche (UMR 7503) : CNRS - Université de Lorraine - Inria , Nancy
synalp.loria.fr

Christophe CERISARA
cerisara@loria.fr
+33 3 54 95 86 25

Membres :

- Nadia BELLALEM, MCF
- Lotfi BELLALEM, PRAG
- Christophe CERISARA, CR CNRS
- Emilie COLIN, Doctorant
- Samuel CRUZ-LARA, MCF
- Christine FAY-VARNIER, MCF
- Claire GARDENT, DR CNRS
- Bikash GYAWALI, Post-doc
- Jean-Charles LAMIREL, MCF
- Bart LAMIROY, MCF
- Thien Hoa LE, Doctorant
- Denis PAPERNO, CR CNRS
- Yannick PARMENTIER, MCF
- Anastasia SHIMORINA, Ingénieur
- Ali TEBBAKH, Ingénieur
- Chunyang XIAO, Doctorant

Mots-clés

- Traitement Automatique des Langues
- Génération Automatique de Textes
- Analyse syntaxique, sémantique et dialogue
- Modèles neuronaux profonds
- Apprentissage faiblement supervisé et clustering

Thématique générale de l'équipe

Synalp (*Symbolic and Statistical Natural Language Processing*) est une équipe de recherche en Traitement Automatique des Langues (TAL). Nous nous intéressons en particulier à la génération automatique et à la simplification de texte, à la modélisation syntaxique et sémantique, aux systèmes de question-réponse, à la classification de textes en thèmes et sentiments et au dialogue. Pour aborder ces défis, nous nous appuyons sur les outils théoriques que sont :

- les grammaires formelles pour le TAL,
- les réseaux neuronaux profonds et les modèles bayésiens,
- l'apprentissage faiblement supervisé.

Domaines d'applications

L'équipe s'intéresse à tous les domaines d'application du Traitement Automatique des Langues mais se focalise en particulier sur les questions liées à la génération de texte, à l'analyse sémantique et au dialogue. Nous appliquons ces recherches à une grande variété de types de données, qui vont de la transcription de la parole spontanée aux bases de publications, en passant par les micro-blogs et les documents écrits. Lorsque cela est pertinent, nous considérons et modélisons l'information contextuelle et paralinguistique. Par exemple, l'émotion, l'opinion et les actes de dialogue dans les micro-blogs sont des domaines d'application de nos recherches.

Plus généralement, les informations linguistiques font presque toujours partie d'un ensemble plus large d'informations connexes : hyperliens et références à des bases de connaissances sur les pages Web, structures de documents et métadonnées dans les rapports scientifiques, géolocalisation, horodatages, graphes des réponses, des renvois et des utilisateurs sur Twitter, etc. Ces multiples dimensions de l'information disponible doivent être intégrées à l'analyse sémantique pour mieux interpréter le langage naturel. De plus, la forme de l'entrée linguistique elle-même évolue, et nous devons rendre nos modèles plus robustes à de telles évolutions. Par exemple, de nouveaux mots et expressions apparaissent constamment sur le Web, des phrases non-grammaticales sont rencontrées fréquemment dans des dialogues oraux, des abréviations et des émoticônes apparaissent et acquièrent

de nouvelles fonctions sémantiques et pragmatiques sur Twitter. De tels phénomènes sont mieux modélisés au niveau du caractère qu'au niveau lexical.

Méthodes fondamentales

L'équipe maîtrise depuis plusieurs années les outils formels de description du langage naturel, en particulier les grammaires TAG lexicalisées, qui côtoient souvent dans nos recherches les méthodes statistiques d'apprentissage automatique, en particulier les modèles bayésiens et les réseaux neuronaux profonds, comme les réseaux convolutionnels denses, les représentations distribuées et les modèles de séquence à séquence avec attention. L'apprentissage de tels modèles nécessitant des quantités de données étiquetées qui ne peuvent être annotées raisonnablement à la main, nous nous intéressons nécessairement également aux approches d'apprentissage non-supervisé et faiblement supervisé, en particulier aux méthodes de transfert [10].

Projets marquants

Génération de textes

Dans le domaine de la génération de textes, l'équipe *Synalp* travaille depuis plusieurs années sur différents aspects de la micro-plannification, une tâche qui consiste à convertir des données en texte. L'approche privilégiée est de combiner des modèles linguistiques (grammaires et lexiques), avec des modèles statistiques. Afin de faciliter le développement manuel de grammaires pour la langue naturelle, nous avons ainsi développé un langage de spécification et un compilateur ainsi que des méthodes de fouilles d'erreur qui permettent la détection semi-automatique des erreurs et ommissions introduites lors de la spécification manuelle. Afin de gérer l'explosion combinatoire des choix possibles, nous avons proposé des algorithmes combinant des méthodes d'apprentissage automatique (champs aléatoires conditionnels) avec des algorithmes d'analyse syntaxique inversée [1]. Enfin, nous avons exploré la micro-plannification pour différents types d'entrées : arbres de dépendances, requêtes OWL sur des bases de connaissances et ensembles de triplets RDF. Nous avons notamment mis au point une technique permettant de produire des corpus d'apprentissage pour la génération à partir de données RDF et organisé une campagne d'évaluation internationale sur ce sujet [3].

Synalp travaille également sur la simplification de textes pour laquelle nous avons développé d'une part, une approche hybride combinant analyse sémantique, apprentissage non supervisé et traduction automatique [8] et d'autre part, une approche faiblement supervisée.

Syntaxe et sémantique

Dans l'équipe *Synalp*, nous nous appuyons sur les grammaires LTAG non seulement pour la génération, mais également pour l'analyse syntaxique et sémantique des textes en langue naturelle. Au-delà des formalismes symboliques, nous utilisons également fréquemment en analyse les méthodes d'apprentissage automatique, que nous combinons quand cela est pertinent avec les approches symboliques. Ainsi, nous avions proposé il y a quelques années une nouvelle classe d'algorithmes d'analyse syntaxique en dépendances intermédiaire entre les analyseurs globaux, comme le *Maximum Spanning Tree* et les analyseurs locaux, dont les modèles par transition. Nous avons ensuite proposé des méthodes d'apprentissage faiblement supervisé des analyseurs syntaxiques et sémantiques qui combinent des modèles bayésiens génératifs avec des règles formelles dans le cadre de la théorie des *Virtual Evidence*, qui permet de gérer proprement les informations manquantes. Ces modèles ont alors pu être entraînés avec des méthodes d'inférence approchée de type Markov Chain Monte Carlo [7]. Toutefois, si les modèles génératifs sont particulièrement bien adaptés à l'apprentissage faiblement supervisé, leur capacité de modélisation n'est pas utilisée de manière optimale pour les tâches de classification visées. Nous nous sommes donc intéressés à l'apprentissage faiblement supervisé des modèles discriminants, en combinant un modèle bayésien orienté avec un modèle à maximum d'entropie pour l'annotation automatique en rôles sémantiques. Nous avons également été plus loin dans cette direction en adaptant une méthode d'apprentissage non supervisé de modèles linéaires discriminants fondée sur une approximation Gaussienne de la distribution

conditionnelle des scores du classifieur et en l'appliquant aux tâches d'identification de prédicats et de reconnaissance d'entités nommées.

Nous avons également abordé l'analyse syntaxique et sémantique du langage naturel via des modèles supervisés, en particulier les modèles à noyaux d'arbres et les réseaux neuronaux profonds, tels que les auto-encodeurs récursifs ou les modèles seq2seq pour la tâche de question-réponse [11].

Dialogue et intéractions

Dans le domaine de la gestion des dialogues et des interactions homme-machine, une de nos contributions importantes au cours de ces dernières années concerne le développement d'agents conversationnels pour l'apprentissage du français langue-étrangère et pour des jeux sérieux à visée industrielle. Nous avons ainsi déployé ces agents aussi bien dans des mondes virtuels de type *Second Life* que dans des jeux de rôle en 3D, notamment dans le cadre du projet européen EMOSPEECH, pour lequel nous avons conçu plusieurs systèmes de dialogue (symbolique, supervisé et hybride symbolique/supervisé). Nous avons exploré dans quelle mesure la lemmatisation, les ressources lexicales, la sémantique distribuée et les paraphrases peuvent être utilisées pour accroître la précision de nos modèles supervisés [2]. Nous avons étendu une partie de ces travaux à une méthode non supervisée appliquée à la gestion du dialogue, en utilisant d'abord un modèle bayésien génératif pour annoter automatiquement les transcriptions manuelles dans le corpus avec des relations sémantiques. Nous avons plus récemment exploré l'apprentissage bayésien par renforcement inverse (BIRL) pour inférer le comportement humain dans le contexte du jeu sérieux d'Emospeech, les données sous forme de dialogue étant fournies par des experts (The Wizard of Oz) qui jouent le rôle de plusieurs agents conversationnels dans le jeu [9].

Dans le domaine de l'analyse des dialogues humain-humain sur Twitter, nous avons notamment contribué en analyse des sentiments et pour la détection des émotions. Ces tâches soulèvent des difficultés concernant par exemple la couverture des modèles linguistiques sous-jacents à tous les niveaux (lexical, syntaxique, sémantique ou pragmatique). Notre première approche, essentiellement symbolique, repose sur des lexiques émotionnellement annotés, des modèles de la variation de la valence et sur la propagation et la combinaison des émotions exprimées lexicalement le long de l'arbre syntaxique des phrases. Cette première approche intègre également des méthodes d'apprentissage automatique exploitant des corpus annotés. Plus récemment, nous poursuivons nos travaux dans cette voie au moyen de réseaux neuronaux profonds, qui permettent de mieux exploiter les masses de données disponibles via les réseaux sociaux. Nous avons en particulier adapté divers types de réseaux convolutionnels plus ou moins profonds aux spécificités des réseaux sociaux et évalué et analysé leurs performances relatives sur plusieurs tâches de classification de textes, dont la reconnaissance des émotions et la reconnaissance des actes de dialogue.

Classification de textes

Dans le domaine de la classification non supervisée de textes, depuis 2011, nous focalisons nos recherches sur le développement et l'exploitation de nouvelles métriques adaptées au traitement de grands corpus de données et applicables à des contextes variés. Nous avons proposé ainsi la métrique de maximisation des caractéristiques qui peut être substituée aux distances classiques de clustering tout en facilitant l'analyse qualitative du processus de clustering, contrairement aux méthodes à base de noyaux. Notre métrique, qui a l'avantage d'être non-paramétrique, a été appliquée notamment à l'analyse diachronique de grands corpus de publications scientifiques. Nous avons également développé un nouvel algorithme de clustering neuronal dérivé de l'algorithme IGNG mais intégrant cette nouvelle métrique au sein d'un algorithme d'apprentissage de type Estimation-Maximisation. Nous avons montré que la méthode résultante, dénommée IGNGF, permettait d'obtenir des performances supérieures à celles de l'état de l'art pour le clustering incrémental de données hétérogènes. IGNGF a également obtenu de meilleurs résultats que d'autres méthodes de TAL dont l'analyse en concepts formels (AFC) et l'analyse spectrale pour la classification automatique de verbes en français appliquée à l'analyse de rôles sémantiques [6].

Nous avons également adapté la métrique de maximisation des caractéristiques aux modèles supervisés afin de proposer de nouvelles solutions au problème des classes déséquilibrées dans les grands corpus. Ce problème est particulièrement difficile à résoudre lorsque les distances entre les classes sont faibles. Nous l'avons donc abordé grâce à notre métrique combinée à des techniques de sélection de caractéristiques et des approches de contraste des données. Nous avons montré que cette solution permet de compenser le déséquilibre des classes et de mieux les discriminer, même dans des cas où d'autres méthodes classiques (sélection de caractéristiques et/ou re-échantillonnage) échouent, voire dégradent les performances, ce qui est possible avec des données multidimensionnelles et bruitées [4].

Nous travaillons actuellement sur l'exploitation de cette métrique afin d'inférer automatiquement le modèle optimal de clustering parmi un ensemble de modèles possibles [5].

Références

[1] Claire Gardent and Laura Perez-Beltrachini. A statistical, grammar-based approach to microplanning. *Computational Linguistics*, 43(1) :1–30, 2017.

[2] Claire Gardent and Lina Maria Rojas Barahona. Using Paraphrases and Lexical Semantics to Improve the Accuracy and the Robustness of Supervised Models in Situated Dialogue Systems. In *Proc. EMNLP*, pages 808–813, Seattle, United States, October 2013.

[3] Claire Gardent, Anastasia Shimorina, Shashi Narayan, and Laura Perez-Beltrachini. Creating training corpora for micro-planners. In *Proc. ACL*, Vancouver, Canada, August 2017.

[4] Jean-Charles Lamirel. Dealing with highly imbalanced textual data gathered into similar classes. In *Proc. IJCNN*, Dallas, United States, August 2013.

[5] Jean-Charles Lamirel, Nicolas Dugué, and Pascal Cuxac. New efficient clustering quality indexes. In *Proc. IJCNN*, Vancouver, Canada, July 2016.

[6] Jean-Charles Lamirel, Ingrid Falk, and Claire Gardent. Federating clustering and cluster labelling capabilities with a single approach based on feature maximization : French verb classes identification with IGNGF neural clustering. *Neurocomputing*, 147 :136–146, January 2015.

[7] Alejandra Lorenzo, Lina M. Rojas-Barahona, and Christophe Cerisara. Unsupervised structured semantic inference for spoken dialog reservation tasks. In *Proc. SIGDIAL*, pages 12–20, Metz, France, August 2013.

[8] Shashi Narayan and Claire Gardent. Hybrid Simplification using Deep Semantics and Machine Translation. In *Proc. ACL*, pages 435 – 445, Baltimore, United States, June 2014.

[9] Lina Maria Rojas Barahona and Christophe Cerisara. Bayesian Inverse Reinforcement Learning for Modeling Conversational Agents in a Virtual Environment. In *Proc. CICLING*, Kathmandu, Nepal, April 2014. Springer.

[10] Guillaume Serrière, Christophe Cerisara, Dominique Fohr, and Odile Mella. Weakly-supervised text-to-speech alignment confidence measure. In *Proc. COLING*, Osaka, Japan, December 2016.

[11] Chunyang Xiao, Marc Dymetman, and Claire Gardent. Symbolic Priors for RNN-based Semantic Parsing. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI)*, Melbourne, Australia, Aug 2017.

AfIA
Association française
pour l'Intelligence Artificielle

Compte-rendu de journées, événements et conférences

Journées BIOSS-IA 2017

Par **Loïc PAULEVÉ**
CNRS
LRI, Orsay
loic.pauleve@lri.fr
Philippe DAGUE
Université Paris-Sud
LRI, Orsay
philippe.dague@lri.fr

Les 22 et 23 juin 2017 se sont déroulées à Gif-sur-Yvette les premières journées BIOSS-IA sur le thème *« Modèles logiques pour la représentation formelle des systèmes vivants »*. Ces journées ont été organisées par Loïc PAULEVÉ (CNRS/LRI) et Philippe DAGUE (Université Paris-Sud/LRI) et ont été co-financées par le pré-GdR IA, le GdR BIM, le GT BIOSS, et le GT TheoBioR du labex Digicosme. Elles ont rassemblé 28 participants.

Ces journées avaient pour but de donner un aperçu des recherches actuelles en France sur la représentation formelle et le raisonnement automatique appliqués à l'étude des systèmes vivants, et en particulier menées par les participants du pré-GdR IA et du GT BIOSS (groupe de travail sur la biologie systémique symbolique, rattaché aux GdR IM et GdR BIM).

Programme des journées

Le programme était composé de deux exposés longs, donnés par François FAGES (Inria Saclay-Île-de-France) et par Pierre SIEGEL (Université Aix-Marseille) ; et de dix exposés courts pour présenter les différents sujets de recherche des équipes participantes.

Exposés longs :

- François FAGES (Inria Saclay-Île-de-France)
  En quête du logiciel du vivant : succès et difficultés du paradigme de la vérification de programmes en biologie cellulaire
- Pierre SIEGEL (LIF, Université Aix-Marseille)
  Logiques non-monotones pour les réseaux de gènes et les systèmes dynamiques booléens

Exposés courts :

- Philippe DAGUE (LRI, Univ Paris-Sud)
  Analyse de réseaux métaboliques en présence de contraintes biologiques : approches géométriques, combinatoires et logiques
- Franck DELAPLACE (IBISC, Univ d'Évry)
  Abduction based drug target discovery using Boolean control network
- Joëlle DESPEYROUX (Inria Sophia Antipolis)
  Towards a Logical Framework for Systems Biology
- Andrei DONCESCU (LAAS, Univ Paul Sabatier Toulouse)
  Quel type de modélisation pour des interactions entre composants biologiques ?
- Béatrice DUVAL (LERIA, Univ Angers)
  Analyse différentielle de réseaux
- Morgan MAGNIN (LS2N, École Centrale

Nantes)
Enjeux de l'apprentissage de modèles dynamiques des réseaux biologiques par la programmation logique
- Loïc PAULEVÉ (CNRS, LRI)
Approximations de problèmes PSPACE en SAT pour la reprogrammation cellulaire
- Céline ROUVEIROL (LIPN, Univ Paris Nord)
Découverte dans les réseaux biologiques hétérogènes : l'expérience Adalab
- Anne SIEGEL (CNRS, IRISA)
Raisonner sur des abstractions et invariants de systèmes dynamiques en biologie avec ASP
- Laurent TRILLING (TIMC-IMAG, Univ Grenoble)
Modélisation déclarative de réseaux de régulation génique. Apport de la programmation logique « non monotone »

Les résumés et transparents des présentations sont disponibles sur la page web des journées.

Conclusion

La biologie des systèmes intègre différents niveaux de connaissance sur les interactions entre les entités biologiques pour comprendre le fonctionnement des systèmes biologiques. Les grandes problématiques des aspects computationnels de la biologie des systèmes portent sur la représentation formelle des connaissances ; l'identification d'interactions à partir de données d'expériences ; la modélisation, la prédiction et le contrôle du comportement dynamique des réseaux biologiques.

Les exposés de ces journées ont montré un panorama de l'application des aspects formels de l'IA pour la modélisation qualitative et l'étude des réseaux en biologie, que ce soient des réseaux de signalisation cellulaire, des réseaux de régulation génique ou des réseaux métaboliques. Les sujets de recherche évoqués ont porté sur différents cadres logiques, reposant le plus souvent sur des logiques non-monotones, pour la formalisation des réseaux biologiques et de leurs propriétés ; ainsi que sur l'application des approches déclaratives et des méthodes de raisonnement automatique pour l'apprentissage et la complétion de réseaux, leur analyse et la prédiction de stratégies de contrôle.

Une partie du programme fut également dédiée à des discussions sur l'intégration de cette thématique au sein du pré-GdR IA, et de ses interactions avec le GT BIOSS. Comme l'ont souligné ces journées, une part importante des problématiques soulevées en biologie des systèmes rejoint naturellement des questions de recherche sur les aspects formels et algorithmiques de l'intelligence artificielle. Les participants ont ainsi suggéré la tenue annuelle de journées sur ce thème, avec le soutien conjoint du GT BIOSS et du pré-GdR IA.

Journée EIAH & IA 2017

Par **Davy MONTICOLO**
ERPI
Université de Lorraine
davy.monticolo@univ-lorraine.fr

La journée EIAH & IA s'est déroulée le 6 juin 2017, à l'hôtel du département de Strasbourg. Elle était organisée par Nicolas DELESTRE (INSA Rouen Normandie) et Davy MONTICOLO (Université de Lorraine) sur le thème *Réalité Virtuelle et Apprentissage*.

Programme de la journée

- 14h00 – 14h15 : Présentations de l'AFIA par Nathalie GUIN (présidente de l'ATIEF) et par Davy MONTICOLO (Membre du CA de l'AFIA).
- 14h15 – 15h45 : Présentations des papiers retenus
  - *Le contexte pour personnaliser les systèmes tutoriels*. Monique GRANDBASTIEN
  - *Une architecture multi-couche pour l'analyse des compétences non techniques en situation critique*. Yannick BOURRIER et al.
  - *Apprentissage : corriger et visualiser*. Ga-

briel ILLOUZ et al.

- 16h00 – 17h00 : Présentation de Jérome DINET : *Développer les compétences-piétons par la réalité virtuelle : enjeux, perspectives et limites*
- 17h00 – 17h45 : Table ronde, Mikhail DEISE (entreprise Human Shape), Jerôme DINET (Université de Lorraine), Nicolas DELESTRE (INSA Rouen Normandie), Davy MONTICOLO (Université de Lorraine)
- 17h45 : Clôture

Présentation des papiers

- Le contexte pour personnaliser les systèmes tutoriels présenté par Monique GRANDBASTIEN
  Résumé : La prise en compte du contexte de l'utilisateur est une des dimensions qui confère un caractère intelligent à un système informatique et que les chercheurs en IA ont tenté de modéliser dès leurs premiers travaux. La prise en compte du contexte dans lequel évolue l'apprenant est aussi identifiée depuis longtemps comme un facteur de succès important pour les apprentissages. Mais cette notion de contexte est restée longtemps implicite dans la conception et la réalisation de systèmes tutoriels. La publication des 35 articles sélectionnés pour le numéro du 25e anniversaire de la revue IJAIED permet d'interroger la prise en compte du contexte de l'apprenant dans ces textes et d'en proposer une synthèse. L'étude parallèle de la notion de contexte en Intelligence Artificielle, notamment de ses perspectives, suggère des pistes de travail en EIAH.
- Une architecture multi-couche pour l'analyse des compétences non techniques en situation critique. Yannick BOURRIER, Francis JAMBON, Catherine GARBAY et Vanda LUENGO présenté par Vanda LUENGO
  Résumé : Dans la plupart des domaines techniques, l'expertise technique d'un praticien détermine sa façon d'analyser et de faire face à une situation donnée. Cette expertise est cependant influencée par un panel de capacités méta-cognitives, comme la conscience de la situation, la prise de décision, ou encore la gestion du stress ou de la fatigue. Ces capacités sont connues sous le nom de compétences non-techniques. Des études ont montré que bien qu'utilisées en quasi permanence, l'influence de ces compétences est la plus importante lorsqu'un praticien est confronté à une situation critique, où les procédures habituelles ne peuvent être appliquées avec succès. Le projet MacCoy a pour objectif de créer un environnement informatique pour l'apprentissage humain ayant la capacité de diagnostiquer les compétences non-techniques de l'apprenant en situation critique et ce dans les domaines de la conduite automobile et de la gestion de l'hémorragie post-partum par des sages-femmes. Ce diagnostic doit ensuite permettre à l'architecture de générer des rétroactions immédiates, ainsi que de nouvelles situations d'apprentissage adaptées aux compétences de l'apprenant. Ce papier présente une méthode pour l'analyse de la performance d'un apprenant en situation critique, premier pas vers le diagnostic épistémique des compétences non-techniques.
- Apprentissage : corriger et visualiser Gabriel ILLOUZ, Anne-Laure LIGOZAT, Frédéric VERNIER, présenté par Anne-Laure LIGOZAT
  Résumé : De Landsheere dans [De Landsheere, 1980] pose les 3 rôles de l'évaluation : de pronostique, de jaugeage, de diagnostic. Des travaux effectués au Limsi concernant ces trois aspects sont présentés : l'aide à la correction d'évaluation formatives et sommatives et au suivi d'apprenants, avec une application à la plateforme Moodle, des prototypes pour représenter un apprentissage par rapport à une représentation d'un domaine de connaissance.

Conférencier invité

Développer les compétences-piétons par la réalité virtuelle : enjeux, perspectives et limites présenté par Jérôme DINET (Université de Lorraine)

Table ronde

Nicolas DELESTRE et Davy MONTICOLO ont présenté deux projets de réalité virtuelle liés à l'apprentissage. L'ensemble des participants a ensuite

débattu sur cette thématique et échangé sur les expériences passées.

Participants à la journée

Aicha BAKKI aicha.bakki@gmail.com
Quentin COULAND quentin.couland@univ-lemans.fr
Bruno DE LIEVRE
Nicolas DELESTRE nicolas.delestre@insa-rouen.fr
Nour EL MAWAS nour.elmawas@telecom-bretagne.eu
Sébastien GEORGE sebastien.george@univ-lemans.fr
Monique GRANDBASTIEN monique.grandbastien@loria.fr
Nathalie GUIN nathalie.guin@univ-lyon1.fr
Marie LEFEVRE marie.lefevre@liris.cnrs.fr
Anne-Laure LIGOZAT annlor@limsi.fr
Esteban LOISEAU esteban.loiseau@univ-lemans.fr
Vanda LUENGO vanda.luengo@lip6.fr
Davy MONTICOLO davy.monticolo@univ-lorraine.fr
Fabrice POPINEAU fabrice.popineau@centralesupelec.fr
Nicolas SZILAS nicolas.szilas@unige.ch
Jean-Daniel TAUPIAC jean-daniel.taupiac@capgemini.com

Forum Industriel de l'Intelligence Artificielle : FIIA 2017

Alain BERGER
Société Ardans
Collège Industriel de l'AFIA
aberger@ardans.fr

Par **Catherine FARON-ZUCKER**
Université Nice Sophia Antipolis
Collège Sciences de l'Ingéniérie des Connaissances de l'AFIA
faron@unice.fr

Le Forum Industriel de l'Intelligence artificielle, organisé par l'AFIA, a eu lieu le Jeudi 27 avril 2017 à Paris (Université Paris Descartes). 52 personnes se sont inscrites à la journée, et d'autres ont rejoint l'événement en cours de journée.

Thématiques de la journée

Lors de FIIA 2016, il était apparu que la *gestion des connaissances* était essentielle pour soutenir tout système d'IA et pouvait de toute façon participer au déploiement d'applications. Ce deuxième *Forum Industriel de l'Intelligence Artificielle* a été consacré à ce sujet, à travers différents thèmes. Chaque thème a fait l'objet de courtes présentations suivies d'une mini-table ronde dont l'objectif était d'aboutir à des éléments de feuille de route et de permettre des échanges accrus entre académiques et industriels. Une dernière session était réservée à des présentations rapides par des industriels invités l'ayant souhaité.

Le programme de la journée est décrit ci-après.

Programme

- 08h45 Ouverture de Frédéric DARDEL (Président de l'Université Paris Descartes) et Pavlos MORAITIS (Directeur du LIPADE), Introduction par Catherine FARON-ZUCKER (Coordinatrice du Collège Science de l'Ingénierie des Connaissances de l'AFIA) et Bruno PATIN (Coordinateur du Collège Industriel de l'AFIA).
- 09h00 Présentation des travaux de l'OPECST sur l'IA par Dominique GILLOT (Sénatrice).
- 09h15 Thème *Gestion vs. Ingénierie de(s, la) connaissance(s)*, avec Jean CHARLET (APH Paris & INSERM), Alain BERGER (Ardans), et Nicolas DUBUC (Michelin).
- 10h00 Thème *Agents conversationnels*, avec Frédérique SEGOND (Viseo), Henri SANSON (Orange), Cyril TEXIER (DoYouDreamUp) et Nicolas SABOURET (Univ. Paris 11).
- 11h Pause Café
- 11h30 Thème *Prise de décision*, avec Bruno PATIN (Dassault Aviation), Philippe BONNARD (IBM), Jean-Marc DAVID (Renault), Laurent GOUZÈNES (Pacte Novation) et Jean-Fabrice LE-BRATY (Univ. Lyon 3).
- 12h Buffet
- 14h00 Thème *Transparence et confiance*, avec Patrick CONSTANT (Owant-Pertimm), Jean-Pierre COTTON (Ardans), Victor DERMIAUX (CNIL) et Jean-Gabriel GANASCIA (U. Paris 6).
- 15h00 *Les projets d'IA s'annoncent* avec Yves DEMAZEAU (AFIA), Fabienne RÉVEILLAC

(SNCF) et Pascal OLLIVIER (SOGET).
- 15h30 Pause Café
- 16h00 Succession de présentations rapides de sociétés concernées par l'IA en « Trois planches »
- 17h00 Conclusion par Yves DEMAZEAU (Président de l'AFIA)

Voici le résumé des présentations longues de cette journée. Des résumés étendus sont disponibles en ligne.

- *De l'Intelligence Artificielle à l'Ingénierie ontologique*, Jean CHARLET – AP-HP & INSERM U1142/LIMICS
  La gestion des connaissances est un acte managérial qui définit pour son domaine de responsabilité ce qui doit être conservé, comment et sous quelle format cette conservation est organisée de façon à ce que puisse être transmis à un acteur futur un savoir, un savoir-faire avec la justification de la confiance accordée à cette connaissance. L'ingénierie des connaissances (IC) relève de l'IA symbolique. Construite à partir d'un projet, elle n'est pas objective. C'est par sa représentation qu'elle devient accessible. Par la construction de modèles, des systèmes à base de connaissance, des interactions entre l'humain et la machine, l'IC restitue avec un comportement mécaniste cette connaissance qui fait sens pour l'utilisateur. L'ontologie correspond à un modèle de connaissance sur le monde décrit dans un langage informatique muni d'une sémantique formelle et interprétable par un ordinateur. En illustrant son propos dans le domaine de la santé, J. Charlet a monté que l'exploitation de données de toutes sortes, caractérisées par leur nature, leur provenance, la fiabilité de la source, le contexte de la captation des données est d'autant plus performante que l'on maîtrise les modèles associés et que l'on bénéficie d'annotations sémantiques issues d'une IA symbolique.
- *Gestion vs ingénierie de(s, la) connaissance(s)*, Alain BERGER – Ardans et Nicolas DUBUC – Michelin
  Pour un industriel tel que Michelin, la mission d'une équipe *Knowledge & Technology Intelligence*, la question des connaissances se porte en interne comme en externe sur la maîtrise du savoir et de l'état de l'art dans son domaine, comme sur le regard de la vision stratégique de l'innovation dans le métier. Quand on s'interroge sur la connaissance, on se pose la question de sa nature, de sa provenance, de sa localisation, de sa pertinence, de sa vulnérabilité, de son exploitation, de sa protection, de sa capitalisation. La dimension humaine est centrale. Quand on parle de capitaliser les connaissances, c'est avoir la volonté d'agir pour préserver les savoirs tacites et les expériences singulières, dans le but de les exploiter, les faire interopérer avec d'autres systèmes et les enrichir pour les amplifier dans le temps. A. Berger et N. Dubuc ont ainsi cherché à caractériser le métier de l'ingénieur de la connaissance, et montré les facteurs clés de succès et les résultats obtenus chez Michelin.
- *Transparence, confiance & Ingénierie des connaissances : Retours d'expérience industriels*, Jean-Pierre COTTON – Ardans
  Lorsqu'une entreprise fait état de ce que chacun nomme connaissance, elle considère différentes dimensions qui confèrent à l'information le statut de connaissance : le contexte d'usage, la maturité du sujet, la confidentialité appliquée et la justification de la nature et de la qualité attestant de son contenu, c'est-à-dire, le sceau d'approbation de l'expert. La mise en œuvre industrielle de l'ingénierie des connaissances nécessite des solutions pragmatiques pour permettre aux experts du domaine de rester concentrés sur leur activité métier lors de la capture de leurs connaissances et de la certification des connaissances, et pour garantir la provenance et la certification des connaissances capturées aux utilisateurs des systèmes à base de connaissances. J.P. Cotton a montré que les plates-formes d'ingénierie des connaissances permettent aujourd'hui de gérer avec pertinence et efficience les connaissances les plus sensibles.
- *Décision en situation extrême et connaissance de la situation*, Jean-Fabrice LEBRATY – Université de Lyon 3
  La décision dans les environnements extrêmes constitue un thème d'actualité. En effet, un contexte marqué par l'imprévisibilité et pour lequel la concurrence pousse à réduire les marges de manoeuvre, impose aux décideurs d'adapter leurs mode de raisonnement et leurs outils.

L'expertise et la gestion des connaissances deviennent alors des éléments clés pour décider. Il se pose alors la question des méthodes et outils qui permettent d'épauler des décideurs moins expérimentés. En se fondant sur les approches théoriques issues du courant de la *Naturalistic Decision Making* (G. Klein) et en s'appuyant sur des exemples, notamment issus du secteur aéronautique, J.F. Lebraty a montré que seule une synergie multi-niveau entre l'organisation, l'équipe, l'individu et des systèmes décisionnels adaptés permettent de décider et que que les notions d'optimisation ne s'appliquent pas à de tels types de décision.

- *Utilisation de modèles bayésiens pour la prise de décision à base de règles métiers*, Philippe BONNARD – IBM
  Les systèmes à base de règles métiers (BRMS) tel que ODM développé par IBM, sont utilisés dans de nombreux domaines (banque, assurance, transport, médical, ...) pour représenter et déployer la logique de prise de décision de l'entreprise. Ils mettent à disposition de l'utilisateur final des outils puissants et agiles de déclaration de la décision. Cependant, ces systèmes sont souvent peu adaptés aux traitements de données incertaines, manquantes ou erronées. P. Bonnard a présenté l'étude URBS réalisée par le LAB IBM France en collaboration avec le laboratoire LIP6 de l'UPMC visant à compléter le système de règles ODM par des modèles probabilistes en étendant son modèle objet de données, son compilateur de règles, et son contexte d'exécution par un couplage faible avec le moteur probabiliste aGrUM.
- *Les cônes de connaissance*, Laurent GOUZENES – Pacte Novation
  Une expérimentation récente utilisant le logiciel Watson a consisté à lui faire ingurgiter une bibliothèque importante de biologie. A la question *How do birds drink seawater*, Watson peut alors répondre d'une façon qui semble savante 'savante' au premier ordre, mais qui en final ne correspond en rien à une réponse qu'un expert apporterait (en premier lieu une série de questions visant à préci-ser une question mal posée). Les experts sont capables de manipuler des hiérarchies de discours et raisonnements, et de faire appel à des domaines et modélisations différents pour aboutir à leurs conclusions et expliciter leurs raisonnements. L. Gouzenes a introduit la notion de cônes de connaissance pour expliciter ces hiérarchies et délimiter plus précisément les compétences des systèmes.
- *La gestion de connaissance dans le contexte de la décision embarquée*, Arnaud BRANTHOMME et Pascal THURIG – Dassault Aviation
  Dassault Aviation travaille depuis plusieurs années sur la capacité de planifier et replanifier des actions qui concerneront le dispositif aérien. Si du point de vue de la création d'un plan des technologies arrivent à maturité, la contrepartie qui consiste à définir le problème à résoudre à partir des observations (d'abord de l'environnement qui comprend des machines pilotées ou non pilotées, des systèmes de défense sol, mais aussi les systèmes artificiels concernés ainsi que les états cognitifs des humains participant à la décision) lui introduit une problématique de la connaissance à représenter. A. Branthomme et P. Thurig ont développé la question de la représentation des connaissances impliquées dans la prise de décision qui ne sera plus le calcul d'un plan mais une négociation entre des systèmes artificiels et des humains (pilotes ou opérateurs), et en particulier celle du niveau de la connaissance à manipuler par les systèmes artificiels.

Conclusion

Cette journée a suscité un vif intérêt et sera reconduite en Avril 2018.

AfIA
Association française
pour l'Intelligence Artificielle

Journée IM-IA, 10 mai 2017, Paris

Par **Pierre MARQUIS**
CRIL
Université d'Artois
pierre.marquis@cril.univ-artois.fr

Une journée commune entre deux groupements de recherche du CNRS, le GdR Informatique Mathématique et le pré-GdR Intelligence Artificielle, s'est tenue le 10 mai 2017, à l'Université Paris-Diderot. Elle était organisée par Arnaud DURAND (Univ. Paris-Diderot) et Pierre MARQUIS (Univ. d'Artois). Le thème de la journée était *Beyond NP*, un sujet de recherche qui connaît un essor important au niveau international (http://beyondnp.org) et qui concerne l'étude d'approches algorithmiques pour la résolution de problèmes dont la complexité de calcul se situe (vraisemblablement) au delà de la classe NP.

Les progrès réalisés en pratique ces dernières années concernant la résolution d'instances du problème de la cohérence de formules propositionnelles (SAT : satisfaisabilité) ou celui de réseaux de contraintes (CSP : constraint satisfaction problem) ouvrent, en effet, la voie à l'étude, à la conception et à l'évaluation d'algorithmes de résolution de problèmes à base de contraintes ou de formules logiques *Beyond NP*. On trouve parmi eux, par exemple, le problème #SAT de comptage des modèles d'une formule propositionnelle, les problèmes d'énumération autour de SAT et de CSP, le problème QBF de décision de la validité d'une formule propositionnelle quantifiée, ou encore divers problèmes liés à la compilation de connaissances. De nombreux développements tant théoriques que plus appliqués vont ainsi aujourd'hui dans la direction *Beyond NP* ; on peut ainsi citer (parmi d'autres) des travaux autour de pré-traitements NP, par exemple pour le comptage de modèles, l'approche CEGAR pour la résolution d'instances QBF, la compilabilité paramétrée, ou encore le projet Compile ! qui vise à mettre à disposition sur la toile des compilateurs de connaissances et d'autres préprocesseurs (http://www.cril.univ-artois.fr/KC).

L'objectif de la journée était de réunir les chercheurs des communautés d'informatique fondamentale et d'intelligence artificielle qui s'intéressent aux contraintes (en un sens très général) et à l'évaluation de requêtes pour échanger et confronter les approches et les points de vue, en particulier sur des problèmes au delà de la décision (comptage, énumération, agrégation, ...). Une trentaine de participants constituait l'auditoire.

Voici le programme détaillé de la journée :

- 9h30-10h : accueil
- 10h-10h30 : Laurent SIMON (Bordeaux) *Efficiency of SAT solvers : why do they work so well ?*
- 10h30-11h30 : Joao Marques SILVA (Lisbonne) *Computing with Oracles : From NP to Beyond NP and Back Again*
- 11h30-12h10 : George KATSIRELOS (INRA) *Optimization with Weighted Constraint Satisfaction*
- 12h10-14h : lunch + discussion
- 14h-15h : Hubie CHEN (San Sebastian) *A Personal Perspective on SAT and CSP : Tractability, Complexity, Quantification, Proofs*
- 15h-15h30 : Florent MADELAINE (Clermont) *Complexity of model checking parameterised by the model*
- 15h30-16h10 : Antoine AMARILLI (Telecom Paris Tech) *A Circuit-Based Approach to Efficient Enumeration*
- 16h10-17h : discussion

Les diapositives utilisées par les orateurs sont disponibles à partir de http://www.gdria.fr/journee-im-ia/. La discussion qui s'est engagée à la fin de journée a concerné l'opportunité de créer un groupe de travail commun aux deux groupements de recherche (GdR IM et pré-GdR IA) et sur le périmètre scientifique qui serait adapté à un tel groupe. Le projet correspondant est en cours d'élaboration.

Journée de lancement de RoD

Par **Marie-Laure MUNIER**
LIRMM/INRIA
Université de Montpellier
mugnier@lirmm.fr

Marie-Christine ROUSSET
LIG/IUF
Université de Grenoble
Marie-Christine.Rousset@imag.fr

RoD (« Reasoning on Data », www.lirmm.fr/rod) est un groupe de travail nouvellement créé commun au pré-GDR IA (Aspects Formels et Algorithmiques de l'Intelligence Artificielle) et au GDR MaDICS (Masses de Données, Informations et Connaissances en Sciences). Ce groupe est animé par Marie-Laure MUGNIER (Université de Montpellier, LIRMM/Inria) et Marie-Christine ROUSSET (Université de Grenoble, LIG/IUF) en ce qui concerne l'informatique, et par deux représentants de domaines d'application : Véronique BELLON (IRSTEA, responsable de l'institut de convergence #DigitAg) pour l'agriculture numérique, et Olivier PALOMBI (CHU Grenoble-Alpes, responsable de la plate-forme de e-learning SIDES) pour les contenus pédagogiques en santé.

Au niveau international, la question de l'exploitation de connaissances pour accéder aux données, et particulièrement à des données hétérogènes, est très étudiée depuis quelques années. De nombreux travaux visent à prendre en compte des connaissances de nature ontologique (allant de la simple taxonomie à des connaissances décrites en logique, comme les logiques de description ou des langages à base de règles, organisés et étudiés en fragments d'expressivité variable) et à exploiter les inférences associées à ces connaissances dans tout le cycle de vie des données (accès, validation, enrichissement, etc.). Un problème emblématique est celui de l'interrogation (ou requêtage) des données, de nombreuses tâches complexes sur les données pouvant se reformuler en termes d'interrogation. On peut citer la problèmatique l'Ontology-Based Data Access, très présente au niveau international dans les communautés Knowledge Representation and Reasoning, Data Management, et Semantic Web, mais dans une moindre mesure en France. L'objectif de RoD est de fédérer une communauté française sur ces problématiques à la croisée de la représentation de connaissances et des raisonnements, et des sciences des données.

RoD s'intéresse de façon générale au développement de techniques de représentation de connaissances et de raisonnements permettant de mieux tirer parti de données disponibles en masse dans différents formats, modèles et systèmes. Le terme données est pris ici au sens de données structurées ou semi-structurées, décrites dans des formats munis de langages de requêtes (typiquement SQL, Datalog, SPARQL, langages de systèmes NO-SQL). Les techniques d'extraction à partir de textes ou d'images peuvent donc être considérées comme en amont par rapport à RoD.

L'accent est particulièrement mis sur le développement d'algorithmes efficaces pour l'interrogation, l'intégration, l'analyse et le liage de données hétérogènes et de qualité variable. Ce groupe de travail vise à développer des techniques génériques, tout en ayant identifié des domaines pour lesquels des fournisseurs de données impliqués dans RoD sont prêts à mettre leurs données à disposition des chercheurs : agriculture numérique, contenus pédagogiques en santé, données encyclopédiques du web.

La journée de lancement a eu lieu le 23 juin à Marseille, dans le cadre des journées annuelles du GDR MaDICS. La matinée a été consacrée à trois exposés invités : sur la problématique d'interrogation de données médiatisée par une ontologie (Federico Ulliana — Université de Montpellier, LIRMM/Inria), sur SIDES 3.0, une plate-forme sémantique centrée utilisateurs pour la formation en santé (Fabrice Jouanot — Université Grenoble-Alpes, LIG) et sur YAGO, une base de connaissances multilingue bâtie sur Wikipedia, Wordnet et Geonames (Thomas Rebele — Telecom ParisTech, LTCI). L'après-midi, les participants ont présenté leurs équipes, compétences et problématiques d'intérêt. La discussion qui a suivi a permis de dégager de premières problématiques communes, notamment les raisonnements prenant en compte des aspects spatio-temporels et les algorithmes d'interrogation de données exploitant des règles.

Les exposés de cette journée sont disponibles sur le site de RoD (http://www.lirmm.fr/Rod)

AfIA
Association française
pour l'Intelligence Artificielle

Thèses et HDR du trimestre

Si vous êtes au courant de la programmation de soutenances de thèses ou HDR en Intelligence Artificielle cette année, vous pouvez nous les signaler en écrivant à redacteur@afia.asso.fr.

Thèses de Doctorat

Jonathan BONNET

« Planification Dynamique, Multi-Objectif et Multi-Critère, par Système Multi-Agent Auto-Adaptatif, Application aux Constellations de Satellites d'Observation de la Terre »

Supervision : Marie-Pierre GLEIZES
Elsy KADDOUM
Serge RAINJONNEAU

Le 8/06/2017, à l'Université Paul Sabatier, Toulouse

Aucune thèse ou HDR en Intelligence Artificielle ne nous ayant été signalée ce trimestre, seules les thèses et HDR soutenues à l'IRIT sont mentionnées, cette rubrique pourra être remplie rétrospectivement.

AfIA
Association française
pour l'Intelligence Artificielle

À PROPOS DE
L'AfIA

L'objet de l'AfIA, association loi 1901 sans but lucratif, est de promouvoir et de favoriser le développement de l'Intelligence Artificielle (IA) sous ses différentes formes, de regrouper et de faire croître la communauté française en IA, et d'en assurer la visibilité.

L'AfIAanime la communauté par l'organisation de grands rendez-vous annuels. En alternance les années impaires et paires, l'AfIAorganise la Plate-forme IA (PFIA 2013 Lille, PFIA 2015 Rennes, PFIA 2017 Caen) et la « Conférence Nationale en Intelligence Artificielle » (CNIA) au sein du au sein du Congrès RFIA (RFIA 2014 Rouen, RFIA 2016 Clermont-Ferrand). Chaque année se tiennent également en ces occasions les « Rencontres des Jeunes Chercheurs en IA » (RJCIA) et la « Conférence sur les Applications Pratiques de l'IA » (APIA) comme autant de Sections Spéciales de CNIA.

À l'occasion de son édition 2017 la Plate-Forme IA de l'AfIA, qui se tient à Caen du 3 au 7 juillet à (PFIA 2017) accueille, outre la 15ème RJCIA et la 3ème APIA, les 11èmes Journées IAF, la 28ème Conférence IC, les 12ème Journées JFPDA, et les 25ème Journées JFSMA. L'AfIAy organise également une compétition « IA et Jeux Vidéos », un nouvel espace de rencontre de la communauté IA.

Forte du soutien de 340 adhérents, l'AfIAgère :

- Le maintien d'un site web dédié à l'IA, reproduisant également les Brèves de l'AfIA,
- Une journée recherche annuelle sur les Perspectives et Défis en en IA (PDIA 2016),
- Une journée industrielle annuelle ou Forum Industriel en IA (FIIA 2017),
- La remise annuelle d'un Prix de Thèse de Doctorat en IA,
- Le soutien à des Collèges ayant leur propre activité, actuellement :
  - Collège Industriel (depuis janvier 2016),
  - Collège Ingénierie des Connaissances (depuis avril 2016),
  - Collège Systèmes Multi-Agents et Agents Artificiels (depuis octobre 2016).
- La parution trimestrielle des Bulletins de l'AfIA, en accès libre à tous depuis le site web,
- Un lien entre adhérents sur les réseaux sociaux LinkedIn, Facebook, et Twitter,
- Le parrainage, scientifique et financier aux conférences et écoles d'été en IA,
- La diffusion mensuelle de Brèves sur les actualités de l'IA en France et à l'étranger,
- La réponse aux consultations officielles (MENESR, MEIN, ANR, CGPME, . . .),
- La réponse à la presse écrite et à la presse orale, également sur internet.

L'AfIAorganise aussi des Journées communes à un rythme mensuel avec d'autres Associations (en 2017 : Entreprises de France & IA avec le MEDEF, Philosophie des Sciences & IA avec la SPS, Interaction Homme-Machine & IA avec l'AFIHM, EIAH & IA avec l'ATIEF, Jeux Informatisés & IA avec le pré-GdR AFAIA, Recherche Opérationnelle & IA avec la ROADEF, Recherche d'Information & IA avec CORIA) et avec des structures du CNRS (en 2017 : Éthique & IA avec le COMETS, Systèmes Dynamiques & IA avec le GdR MACS).

Finalement l'AfIAcontribue à la participation de ses membres aux grands événements de l'IA. Ainsi, les membres de l'AfIA, pour leur inscription à RFIA 2016, ont bénéficié d'une réduction équivalente à deux fois le coût de leur adhésion à l'AfIA.

Nous vous invitons à adhérer à l'AfIApour contribuer au développement de l'IA en France. L'adhésion peut être individuelle ou, à partir de cinq adhérents, être réalisée au titre de personne morale (institution, laboratoire, entreprise). Pour adhérer, il suffit de vous rendre sur le site des Adhésions de l'AfIA.

Merci également de susciter de telles adhésions en diffusant ce document autour de vous !

AfIA
Association française
pour l'Intelligence Artificielle

LABORATOIRES ET SOCIÉTÉS ADHÉRANT COMME PERSONNES MORALES

CRIL, EDF/STEP, GREYC, IFFSTAR, IRIT, LAMSADE, LIFL, LIG, LIMOS, LIMSI, LIPADE, LIP6, LIRIS, LIRMM, LORIA, LRI, ONERA, TETIS

Pour contacter l'AfIA

Président
Yves DEMAZEAU
L.I.G./C.N.R.S., Maison Jean Kuntzmann
110, avenue de la Chimie, B.P. 53
38041 Grenoble cedex 9
Tél. : +33 (0)4 76 51 46 43
Fax : +33 (0)4 76 51 49 85
president@afia.asso.fr

Serveur WEB
http://www.afia.asso.fr

Adhésions, liens avec les adhérents
Catherine FARON-ZUCKER
tresorier@afia.asso.fr

Calendrier de parution du Bulletin de l'AfIA

	Hiver	Printemps	Été	Automne
Réception des contributions	15/12	15/03	15/06	15/09
Sortie	31/01	30/04	31/07	31/10