BULLETIN DE L'AFIA

Avril 2014

Nº 84

Association Française pour l'Intelligence Artificielle

Présentation du bulletin

Le **Bulletin de l'Association Française pour l'Intelligence Artificielle** 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'AFIA publie 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.

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
Tel : +33 (0)4 76 51 46 43
Fax : +33 (0)4 76 51 49 85
Yves.Demazeau@imag.fr
http://membres-lig.imag.fr/demazeau

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

Adhésions, liens avec les adhérents
Thomas GUYET
Laboratoire Informatique d'Agrocampus-Ouest
65, rue de Saint-Brieuc
35042 Rennes cedex
Mél. : tresorier@afia.asso.fr

Personnes morales adhérentes à l'AFIA
ENSMSE, Université Paris Dauphine, LORIA, LIRIS, LIMSI, IRIT/SMAC, EDF/STEP, LIPADE, IFFSTAR, LIRMM, TAO, LIFL, GREYC, LIG, ONERA, IRSTEA-TETIS, INRA, LITIS

EDITORIAL

IA et santé... 12 ans après

La santé a toujours été un domaine d'application privilégié de l'I.A., au moins depuis les années 70 et le célèbre MYCIN (existe-t-il un cours d'introduction à l'I.A. qui ne fasse pas référence à MYCIN ?). Dans ce numéro du bulletin, le dossier coordonné par Pierre Zweigenbaum et Jean Charlet dresse un état des lieux des recherches en France... 12 ans après le dernier dossier sur le sujet. Les lecteurs curieux pourront comparer les thématiques qui y sont mises en avant. Ce qui est certain, c'est que la diversité est au rendez-vous, avec l'informatisation croissante de divers aspects de la pratique médicale (par exemple les dossiers patients informatisés) : problèmes d'interopérabilité des standards, analyse d'image, visualisation anatomique, analyse d'articles médicaux ou d'échanges entre les patients sur les forums, raisonnement à partir de cas cliniques similaires,... la liste est trop longue pour être exhaustifs. Nous vous invitons à lire ce dossier pour découvrir la vitalité (dont témoigne le très grand nombre de projets) et la richesse de cette thématique, en attendant celui de 2026.

Nicolas Maudet, Philippe Morignot & Laurent Vercouter
Rédacteurs en chef

Dossier « Intelligence artificielle et santé »

Dossier réalisé par

Pierre Zweigenbaum (LIMSI-CNRS, Orsay) Jean Charlet (AP-HP et Inserm, Paris)

Ce dossier recense des activités de recherche associant l'intelligence artificielle et la médecine, ou plus généralement la santé. La santé a été très tôt un domaine privilégié d'expérimentation pour l'IA [4, 3], et a vu la création d'une conférence et d'une revue spécialisées : la conférence AIME (Artificial Intelligence in Medicine Europe, créée en 1987) et la revue Artificial Intelligence in Medicine (créée en 1989). Les travaux sur l'IA en santé sont également publiés dans de nombreuses conférences et revues en informatique, en médecine et en informatique médicale. La France, à travers François Grémy, est à l'origine du terme « informatique médicale » qui a donné internationalement « Medical Informatics ». Parmi les conférences de ce domaine, notons ainsi dans le monde francophone la conférence internationale JFIM (Journées francophones d'informatique médicale), les journées de l'AIM (Association française pour l'informatique médicale) et de nombreux ateliers associés à la conférence IC, comme IA & Santé (2014) et SIIM (Symposium sur l'ingénierie de l'information médicale, atelier IC en 2013 et 2015).

L'AFIA elle-même s'intéresse depuis longtemps au domaine de la santé, et a publié des dossiers « IA et médecine » en 1993 [1] et en 2002 [2].

Outre l'AFIA, les sociétés savantes principales concernées par le domaine de l'IA et santé sont l'AIM (Association française pour l'informatique médicale, citée plus haut) et le SIG francophone de l'IMIA (International MedicaI Informatics Association) ; le GdR STIC-Santé (CNRS et Inserm) a également hébergé un chapitre sur la e-Santé.

Les activités en IA et santé sont menées en France dans des laboratoires publics et en entreprise. Sont représentés dans ce dossier des laboratoires CNRS (LBBE, LIG, LIMSI, LIRMM, LRI), Inserm (LIMICS, ERIAS), et universitaires (LIRIS et unités mixtes ci-dessus). Les hôpitaux et CHU sont naturellement très présents, avec les CHU de Montpellier, Nîmes, Rennes, Rouen, Saint-Étienne, les Hospices Civils de Lyon, et à l'Assistance Publique – Hôpitaux de Paris l'Hôpital européen Georges Pompidou, la Pitié-Salpêtrière et Trousseau. Les entreprises sont bien représentées à travers des projets partenariaux comme SYNODOS (Holmes Semantic Solutions, Viseo), ADR Prism, Accordys, BDBfr, OFS (Mondeca, Temis), Patient Genesys (Interaction Healthcare, Voxygen, VIDAL), RAVEL, ROMEO2 (Aldébaran), SiFaDo, TOLBIAC, TeRSan (Mondeca), Accordys (Antidot), financés par l'ANR ou le FUI (BPI, régions).

Plus largement, les activités de ce domaine sont soutenues par des projets de recherche internationaux (BioASQ, Dynamo, E-Thérapies, EHR4CR, HETOP, Hybris-B1, Salus) et nationaux, en particulier par le programme ANR TecSan (technologies pour la santé) et les appels de l'ANSM (Agence nationale pour la sécurité du médicament : Drugs-SAFE, Vigi4med), mais aussi les projets ANR Blanc ou Jeunes Chercheurs comme CABeRneT, Hybride, Pradnet, SIFR.

Les projets mentionnés dans ce dossier portent sur les terminologies et ontologies (Dynamo, HETOP, Hybris-B1, OFS, SIFR, TOLBIAC, Accordys, TerSAN) et plus particulièrement l'interopérabilité sémantique (Drugs-SAFE, Salus). L'analyse d'énoncés en langue naturelle (CABeRneT, Hybride, Patient Genesys, SIFR, SYNODOS, SiFaDo) et plus particulièrement la recherche d'information (Accordys, BDBfr, BioASQ, PLaIR, RAVEL) et l'aide au codage (BioASQ, SYNODOS, SiFaDo) sont très actifs. Le dialogue personne-machine est aussi abordé pour la formation par la simulation (Patient Genesys), ou la reconnaissance du stress dans des conversations avec des patients sociophobes (E-Thérapies), ainsi que les interactions sociales avec des robots compagnons en EHPAD. L'analyse de ce que disent les patients sur les réseaux sociaux (ADR Prism, Patients Mind, SIFR, Vigi4med) est en plein essor, et contribue aux travaux sur la pharmacovigilance et la détection d'effets indésirables de médicaments (ADR prism, Drugs-SAFE, Salus, Vigi4med). L'utilisation secondaire des données du dossier patient informatisé (EHR4CR, SYNODOS) est un objectif majeur, tout comme la découverte de connaissances (Hybride, PRADNET), et la disponibilité récente de bases locales et nationales encourage l'étude des trajectoires de patients. La nutrition est également représentée (Open Food System).

La formation assistée (enseignement assisté par ordinateur, rebaptisé « e-learning ») a pris le virage des jeux à buts pour la formation en santé ou la thérapeutique (IMAIOS-CRIM, NaturalPad, Patient Genesys).

La bioinformatique et l'imagerie médicale sont des domaines connexes qui sont également représentés dans ce dossier.

On note enfin la participation à des campagnes d'évaluation internationales (i2b2, CLEF e-Health, BioNLP, SemEval) ou leur organisation (BioASQ, CLEF e-Health).

En conclusion, ce dossier, même s'il est nécessairement non exhaustif, montre que quarante ans après MYCIN et les systèmes experts, et douze ans après le dossier AFIA précédent sur le même thème, la vitalité des travaux de recherche en intelligence artificielle et santé est sans cesse renouvelée.

Références

[1] Jean Charlet. "Dossier « IA & Médecine »". In : *Bulletin de l'AFIA* 14 (1993), p. 22–48.

[2] Vincent Corruble et Jean Charlet. "Dossier « IA & médecine »". In : *Bulletin de l'AFIA* 48 (2002), p. 15–49.

[3] W B Schwartz, R S Patil et P Szolovits. "Artificial intelligence in medicine. Where do we stand ?" In : *The New England journal of medicine* 316.11 (1987), p. 685–688.

[4] E H Shortliffe et al. "Computer-based consultations in clinical therapeutics : explanation and rule acquisition capabilities of the MYCIN system". In : *Computers and Biomedical Research* 8.4 (août 1975), p. 303–320.

Inserm U897, équipe ERIAS : Projet DRUGS-SAFE, Évaluation systématisée du médicament en population / Drugs Systematised Assessment in real-liFe Environment

Équipe de Recherche en Informatique Appliquée à la Santé (ERIAS)

- CONTACT : Frantz Thiessard - frantz.thiessard@u-bordeaux.fr
- ADRESSE : ERIAS, Centre INSERM U897, ISPED, Université de Bordeaux, 146 rue Léo Saignat, 33000 Bordeaux

Membres de l'équipe impliqués dans le projet

- Coordonnateur : Frantz Thiessard, MCU-PH informatique médicale
- Gayo Diallo, MCF informatique
- Fleur Mougin, MCF informatique
- Vianney Jouhet, PH

Thème général du projet

Ce projet a été soumis dans le cadre de l'appel à candidatures "Plateformes en pharmaoépidémiologie 2014", lancé par l'Agence Nationale de Sécurité du Médicament (ANSM). La plateforme proposée a pour objectif de permettre une évaluation systématisée du médicament en France. La stratégie de la plateforme est guidée par l'impact potentiel des médicaments, c'est à dire par l'ampleur possible des conséquences de leur usage sur la santé publique.

Description des travaux

Contexte

La plateforme cible des médicaments très largement utilisés (psychotropes, anti-infectieux, médicaments à visée cardiovasculaire, etc.) et étudie leurs modalités d'utilisation et le risque qu'ils peuvent faire courir dans la survenue d'événements graves (infarctus, accidents vasculaires cérébraux, maladies neurodégénératives, accidents et risque de perte d'autonomie). Pour les cas où un risque serait authentifié pour un médicament [1], la plateforme vise à déterminer les conditions d'utilisation du médicament associées au risque, l'impact de ce risque sur la santé publique, et l'impact que des actions de régulation pourraient avoir sur la réduction de ce risque. Afin d'améliorer les connaissances sur l'efficacité de telles actions (retrait de médicament, recommandations d'utilisation, etc.), l'effet d'actions de régulation passées sera également évalué, en terme d'amélioration de la santé publique et en termes médico-économiques.

Ce projet est coordonné par l'équipe INSERM de pharmaco-épidémiologie U657, et cinq autres équipes de recherche participent à ce projet, dont l'ERIAS qui est responsable d'un workpackage concernant l'harmonisation des données et des connaissances et leur exploitation pour identifier des mésusages de médicaments.

Harmonisation des données et mésusage des médicaments

Les sources de données impliquées dans cette plateforme ont été développées indépendamment les unes des

autres, sans le moindre objectif d'interopérabilité. Elles présentent donc une hétérogénéité à différents niveaux (syntaxe, sémantique, etc.) [6]. Cette tâche vise à harmoniser les données et connaissances décrites dans ces sources. Afin d'obtenir une vision homogène et complète des données que l'on souhaite exploiter, il est nécessaire de les réconcilier en utilisant des sources de connaissances externes, comme l'UMLS [3]. Ces dernières permettent de rapprocher les données, même si elles sont représentées très différemment (terminologies médicales différentes, granularité plus ou moins fine de l'information, libellés différents pour un même concept, etc.). Cela implique de découvrir des correspondances entre les différentes terminologies médicales, soit de manière exacte, soit en combinant des concepts éventuellement associés par des relations non hiérarchiques [4]. Nous souhaitons par ailleurs proposer un mécanisme de maintenance et de stockage des correspondances retrouvées pour prendre en compte les mises à jour des sources de données originales.

L'autre étape consiste à retrouver l'ensemble des codes ou des libellés utilisées par chaque base de données ou document concernant les variables d'intérêt (par exemple, les pathologies étudiées ou les variables d'ajustement ainsi que leurs descendants dans la hiérarchie, en utilisant les terminologies natives de chaque base), et ce, pour chaque sous-étude du projet [2]. Les requêtes peuvent combiner des informations de type diagnostic, des actes et des traitements.

Mésusage des médicaments

L'autre tâche a pour objectif de repérer le mésusage des médicaments dans des sources aussi différentes que les bases de remboursement ou des forums d'utilisateurs sur Internet, précédemment utilisées en pharmacovigilance [5]. Nous souhaitons pouvoir comparer, de façon automatisée, le bon usage des médicaments (décrit par exemple dans les monographies médicamenteuses : indications, populations cibles, contre-indications ou associations médicamenteuses, entre autres), avec l'usage réel qui en est fait (population cible ou usage non prévu, recherche volontaire d'un effet secondaire, associations non prévues avec d'autres produits etc.). Les sources d'information peuvent être structurées comme les bases de remboursement, semi-structurées comme les recommandations de bonnes pratiques ou pas du tout structurées avec du texte libre sur les forums par exemple. Les données non, ou semi-structurées doivent être structurées secondairement. Afin de repérer le mésusage des médicaments, nous devons fournir une représentation des connaissances contenues dans les documents de référence, puis faire le même travail sur les sources qui décrivent l'utilisation réelle de ces produits, et enfin repérer si le "profil" d'utilisation réelle diffère notablement du profil attendu. D'autres analyses seront également faites en analysant les variations au cours du temps de l'utilisation d'un traitement donné. Des modifications importantes du profil de ce traitement seraient un signal d'alerte pour qu'un service de pharmacovigilance étudie les raisons de la modification des usages des patients. Ces analyses nécessitent une expertise dans le traitement automatique des langues (pour identifier les négations, l'incertitude, la temporalité, etc.), dans l'identification des concepts d'intérêt (molécules, noms de marque, les symptômes, les diagnostics, la dose, la fréquence de la consommation, etc.), dans la représentation des connaissances puis dans la comparaison de graphes et la visualisation des informations.

Références

[1] Ismaïl Ahmed et al. "Early detection of pharmacovigilance signals with automated methods based on false discovery rates : a comparative study". eng. In : *Drug Safety* 35.6 (juin 2012), p. 495–506. ISSN : 0114-5916. DOI : 10.2165/11597180-000000000-00000.

[2] Paul Avillach et al. "Harmonization process for the identification of medical events in eight European healthcare databases : the experience from the EU-ADR project". eng. In : *Journal of the American Medical Informatics Association : JAMIA* 20.1 (jan. 2013), p. 184–192. ISSN : 1527-974X. DOI : 10.1136/amiajnl-2012-000933.

[3] Olivier Bodenreider. "The Unified Medical Language System (UMLS) : integrating biomedical terminology". eng. In : *Nucleic Acids Research* 32.Database issue (jan. 2004), p. D267–270. ISSN : 1362-4962. DOI : 10.1093/nar/gkh061.

[4] Gayo Diallo. "An effective method of large scale ontology matching". eng. In : *Journal of Biomedical Semantics* 5.1 (2014), p. 44. ISSN : 2041-1480. DOI : 10.1186/2041-1480-5-44.

[5] Sandrine Katsahian et al. "Evaluation of Internet Social Networks using Net scoring Tool : A Case Study in Adverse Drug Reaction Mining". eng. In : *Studies in Health Technology and Informatics* 210 (2015), p. 526–530. ISSN : 0926-9630.

[6] Maurizio Lenzerini. "Data Integration : A Theoretical Perspective". In : *Proceedings of the Twenty-first ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*. PODS '02. Ma-

dison, Wisconsin : ACM, 2002, p. 233–246. ISBN : 1-58113-507-6. DOI : 10.1145/543613.543644. URL : http://doi.acm.org/10.1145/543613.543644.

Projet SYNODOS : Usage secondaire du dossier médical informatisé à des fins épidémiologiques et d'évaluation de la qualité des soins

- CONTACT : Dr Marie-Hélène METZGER (coordonnateur) : metzger.marie-helene@orange.fr
- ADRESSE : UCBL-CNRS UMR 5558, Service d'hygiène et Epidémiologie, Hôpital de la Croix-Rousse, 103 Grande-Rue de la Croix-Rousse, 69317 Lyon cedex 04
- Web : http://www.synodos.fr

Membres du Consortium SYNODOS

Laboratoire « Biométrie et Biologie Evolutive » (LBBE), unité mixte de recherche 5558 du CNRS et de l'Université Claude Bernard Lyon I (URL : http://lbbe.univ-lyon1.fr/) s'organise dans le cadre de trois départements dont celui de BioMaths-Santé. Ce département regroupe des équipes du secteur santé utilisant la modélisation en épidémiologie et en recherche clinique avec une approche populationnelle. Un axe de recherche « méthodes d'exploitation épidémiologique des systèmes d'information en santé » (MEESIS : URL : http://lbbe.univ-lyon1.fr/-Metzger-Marie-Helene-.html) est un axe de l'équipe « épidémiologie et santé publique », qui a été initié en 2007 et coordonné par le Dr MH Metzger. Les activités de recherche de cet axe reposent sur l'utilisation des techniques de fouille de données pour l'exploitation des systèmes d'information en santé à des fins de surveillance épidémiologique.

Holmes Semantic Solutions (http://www.h02s.com/fr/) développe des logiciels et des ressources linguistiques pour la conception et la réalisation de solutions innovantes d'analyse de la langue écrite et parlée et ce pour plus de 30 langues dans le monde. Il est spécialisé dans l'analyse sémantique des documents, en offrant des solutions innovantes capables de coupler la flexibilité des systèmes de TAL symboliques (basés sur règles) à la performance des systèmes basés sur apprentissage automatique.

CISMeF (http://www.cismef.org). L'équipe CISMeF, équipe pluridisciplinaire constituée de médecins informaticiens, de documentalistes, d'ingénieurs de recherche, de doctorants et de post-docs, est rattachée à l'équipe TIBS (Traitement de l'Information en Biologie et Santé) du laboratoire LITIS EA 4108). Cette équipe est à l'origine depuis 1995 du Catalogue et Index des Sites Médicaux de langue Française qui recense les ressources médicales en accès libre et gratuit sur le Web. L'équipe est maintenant amenée à gérer et mettre à disposition du public de nombreuses terminologies de santé. L'équipe CISMeF participe à ce projet du fait de son expertise sur les terminologies/ontologies de santé depuis plus de quinze ans s'étendant maintenant aux principales terminologies/ontologies de santé disponibles en français, mais aussi l'UMLS avec la traduction partielle de certaines terminologies (ex : NCIt, 44 % de termes traduits sur 93 926). Cette équipe a en particulier développé un serveur de terminologies de santé et des modules d'indexation automatique et de recherche d'information fondés sur une approche multi-terminologique. CISMeF a aussi une expérience dans la gestion des systèmes d'information hospitaliers et de santé.

Le groupe Viseo (http ://www.viseo.com), premier acteur multi spécialiste des systèmes d'information, a créé en Novembre 2011 son centre de recherche et de développement sous la direction de Frédérique Segond. Ce centre de R&D va permettre au groupe Viseo d'accélérer le rythme de développement de son activité d'édition de logiciels innovants et sectoriels. Il a également pour ambition d'explorer de nouveaux marchés et d'anticiper les besoins clients du groupe en bénéficiant des expertises techniques de la recherche. La R&D du pôle innovation de Viseo s'articule autour des trois thématiques de recherche suivantes : l'analyse de données, le génie logiciel, les interfaces et les usages. Ces trois thématiques recouvrent et permettent l'intersection des savoirs faire technologiques du groupe (ERP, BI, Architecture, mobilité) et peuvent se décliner sur différents domaines métiers.

Thème général du projet

De nombreux travaux sont en cours pour développer des méthodes de phénotypage à partir d'une utilisation secondaire des données du Dossier Patient Informatisé (DPI). Ces méthodes intègrent notamment le traitement du langage naturel (ex : projet eMERGE [2]). Il n'existe

pas à l'heure actuelle de langage formel ni d'approche standardisé pour extraire ces phénotypes [2]. Dans le cadre du projet SYNODOS, nous avons développé un modèle conceptuel de données à visée générique et dont l'évaluation est prévue à l'issue du développement de la solution [1]. Par ailleurs une fois le phénotype obtenu par ces techniques, la 2ème étape consiste à exploiter les données extraites pour par exemple, mesurer des associations ou sélectionner des patients éligibles. Il n'existe pas à notre connaissance de solution de langue française intégrant l'ensemble de ces étapes dans une solution unique et conçue pour être utilisée en production hospitalière. Le challenge du projet SYNODOS est de réussir à rassembler toutes les technologies nécessaires à l'exploitation de ces données pour divers usages en milieu hospitalier.

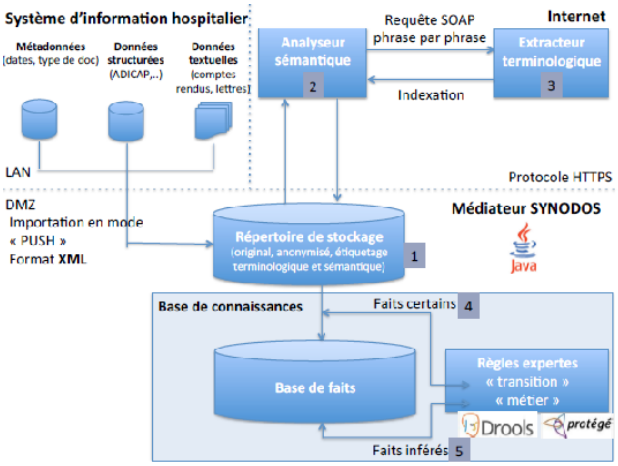

FIGURE 1 – Architecture générale de la solution

Description des travaux

Contexte

Le projet est financé par l'Agence Nationale de Recherche (programme TecSan 2012) et a été décomposé en 6 tâches scientifiques : 1) définition de l'architecture générale de la solution 2) développement du traitement sémantique des données textuelles médicales 3) interfaçage de l'extracteur de concepts multi-terminologique avec l'analyseur sémantique 4) développement d'un système générateur de règles expertes 5) intégration des différents modules constitutifs de la solution et enfin 6) évaluation des performances de la solution.

Le projet a démarré en octobre 2012 et se déroulera jusqu'à septembre 2015.

Architecture générale de la solution

La solution SYNODOS est une application web (SYNODOS – Médiateur) qui utilise les services de 2 serveurs distants : le serveur terminologique du CISMeF et le serveur sémantique, spécifique à SYNODOS, reposant sur les technologies de Holmes Semantic Solutions (figure 1).

L'accès aux serveurs distants se fait par un « protocole de transfert hypertexte sécurisé » (HTTPS) de telle sorte qu'aucune donnée ne soit transmise en clair. Toutes les fonctionnalités de SYNODOS sont accessibles depuis le médiateur qui fournit l'interface unique de l'application. Différents logiciels sont utilisés pour le développement des modules composant la solution SYNODOS. Le projet s'appuie sur un langage de développement (Java), un framework web, un serveur d'applications, un SGBD relationnel, un moteur BRMS (Business Rules Management System - Drools [3]) et des outils du web sémantique.

Un module d'importation des données provenant du système d'information hospitalier a été développé, permettant de récupérer les métadonnées et les données textuelles « poussées » par le Système d'Information Hospitalier de l'établissement. Le module procède ensuite à une anonymisation automatisée des données nominatives ou indirectement nominatives (ex : noms de personnes, noms de lieux, numéros de téléphone, adresses, adresse mail, etc.) afin de permettre les échanges avec les services web distants et afin également de permettre un usage épidémiologique des données. Une fois les documents textuels formatés avec leurs métadonnées hospitalières dans un fichier XML, un 1er traitement linguistique va être réalisé : détection et découpage des phrases par l'analyseur sémantique. Puis l'analyseur envoie une requête selon le protocole SOAP au serveur terminologique afin que chaque concept de chaque phrase puisse être indexé.

L'extracteur de concepts multi-terminologique (ECMT)

L'extracteur de concepts multi-terminologique (ECMT) renvoie alors dans un flux XML pour chaque

concept médical reconnu, plusieurs métadonnées qui seront nécessaires à la suite du traitement. L'ECMT (V2) a été développé à partir du portail multi-terminologique développé par l'équipe CISMeF. Ce portail contient 55 terminologies ou ontologies médicales, correspondant à 500,000 concepts médicaux en langue française et 1,5 million en langue anglaise (URL : http://www.hetop.eu). L'ECMT V2 permet d'indexer les termes médicaux rencontrés dans les documents textuels. Ce traitement est disponible en service web (SOAP ou REST). Différentes terminologies médicales ont été sélectionnées pour normaliser le langage médical naturel dans le cadre du projet : la CIM-10 (Classification Internationale des Maladies, 10ème révision), le thésaurus MeSH® (Medical Subject Headings) pour l'indexation des concepts diagnostiques ou symptomatologiques, la classification ATC (Anatomical Therapeutic and Chemical Classification) pour le codage des médicaments, la SNOMED International etc. Différentes techniques de mise en correspondance ont été employées dans ce projet : (a) mise en correspondance conceptuelle basée sur le CUI du Metathesaurus® de l'UMLS (Unified Medical Language System) ; (b) mise en correspondance basée sur le traitement du langage naturel ; (c) mise en correspondance manuelle ou supervisée par les experts terminologiques de l'équipe CISMeF.

L'analyseur sémantique

L'analyseur sémantique est une plateforme de traitement du langage naturel développé par Holmes Semantic Solutions qui combine différents types de modules, appliqués de façon incrémentale dans le processus de traitement : modules symboliques (à base de règles), modules statistiques, et modules à base d'apprentissage automatique. Après l'indexation, la suite du traitement linguistique consiste en la tokénisation, l'étiquetage morpho-syntaxique, la lemmatisation, l'analyse morphologique, l'étiquetage lexicale et, enfin, l'analyse syntaxique en dépendances.

L'analyse sémantique proprement dite se base ensuite sur le résultat de cette analyse syntaxique, ainsi que sur des connaissances ontologiques ou terminologiques. Cette analyse consiste à représenter les éléments de sens, permettant d'utiliser l'information extraite de façon adéquate. Pour représenter le résultat de l'analyse sémantique, l'analyseur permet de définir des relations (« prédicats logiques ou faits »), présentés sous forme de graphe. Par exemple pour la phrase « le patient est arrivé aux urgences sous oxygène après avoir fait une crise d'asthme », la figure 2 représente l'analyse sémantique produite. Les nœuds correspondent aux entités (objets, endroits, personnes, etc.) et les arcs aux relations sémantiques définies entre ces entités.

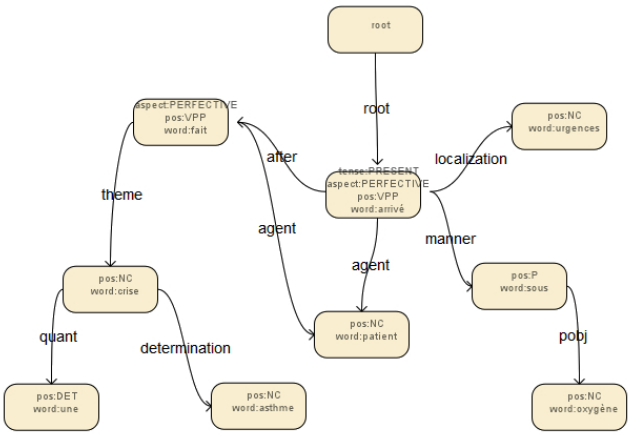

FIGURE 2 – Exemple de graphe sémantique SYNODOS

L'analyseur sémantique envoie ensuite selon le protocole HTTPS les données issues du traitement sémantique sur un répertoire de stockage de la solution SYNODOS. Ces données sont stockées sous forme de fichiers XML avec un document textuel stocké par fichier.

Développement d'un système générateur de règles expertes

Pour la finalité de la solution SYNODOS qui consiste à exploiter les données extraites des documents textuels à des fins épidémiologiques ou d'aide à la décision, une simple normalisation des concepts médicaux n'est pas suffisante. En effet, pour une utilisation pertinente des concepts extraits, il est nécessaire d'y associer un étiquetage temporel (s'agit-il d'une donnée reliée dans le parcours du patient à un antécédent, un motif d'entrée, une évolution en cours d'hospitalisation ?). Il est également nécessaire de créer des relations permettant de lier les concepts entre eux : par exemple dans la phrase suivante : « on a assisté à une chute de la gastrinémie », il est indispensable de relier le concept « gastrinémie » à celui de « chute » pour exploiter correctement cette information. Il faudra donc tout d'abord créer une règle qui permet d'affecter un concept du type « gastrinémie » à l'élément « type d'examen biologique » et un concept de type « chute » à l'élément « résultat d'examen biologique » de la base de faits. Ces règles s'appuieront sur l'utilisation des données sémantiques (par exemple négation du type « absence » associé au concept médical) et terminologiques (par exemple le type sémantique UMLS du concept) rattachées à chacun de ces concepts. La relation entre ces deux concepts sera ensuite établie par l'utilisation des relations

construites dans la base de connaissances (relations hiérarchiques ou transversales). Par ailleurs de nouveaux faits seront inférés par combinaison de « faits certains ». Ces règles expertes sont en cours de développement par l'UMR UCBL-CNRS 5558 et Viseo Technologies intègrera ces règles dans le Mediateur SYNODOS afin de compléter le traitement des données selon deux approches qui seront comparées : une approche traditionnelle (classique) via l'utilisation d'un outil du marché (BRMS : Business Rules Management System) et une approche orientée recherche qui étudie le Web sémantique et ce qu'il peut apporter pour ce type de système expert.

Références

[1] Q Gicquel et al. "Annotation methods to develop and evaluate an expert system based on natural language processing in electronic medical records". In : *MEDINFO*. soumis. Sao Paulo, Brazil, 2015.

[2] Omri Gottesman et al. "The Electronic Medical Records and Genomics (eMERGE) Network : past, present, and future". In : *Genetics in Medicine* 15.10 (oct. 2013), p. 761–771.

[3] *JBoss Drools Business Logic Integration Platform*. Mai 2015. URL : http://www.jboss.org/drools.

LIG, équipe AMA : Projet BioASQ

- CONTACT : George Paliouras - paliourg@iit.demokritos.gr
- CONTACT France : Eric Gaussier - eric.gaussier@imag.fr
- WEB : http://bioasq.org//

Partenaires du projet

- NCSRh, Demokhritos, Athènes, Grèce (coordinateur)
- Transinsight, Dresde, Allemagne
- Univeristé Joseph Fourier, Laboratoire LIG, Grenoble, France
- University Leipzig, Leipzig, Allemagne
- Université Pierre et Marie Curie, Paris, France
- Athens Univ. of Economics and Business Research Centre, Athènes, Grèce

Thème général du projet

Le projet BioASQ a pour but de produire des données et des protocoles d'évaluation pour les systèmes d'annotation sémantique (c'est-à-dire l'indexation, par des concepts issus du MeSH [1], de textes biomédicaux) et les systèmes de réponses aux questions dans le domaine biomédical.

Description des travaux

Le projet BioASQ est à l'origine un projet de type *Specific Support Action* financé par la commission européenne d'octobre 2012 à octobre 2014 [2]. Le projet se poursuit par l'organisation de nouvelles campagnes d'évaluation.

Un peu plus de 3000 nouveaux articles sont publiés chaque jour dans les journaux biomédicaux. La base MEDLINE comprend déjà plus de 20 millions d'articles et les données de santé sous forme non textuelles ne cessent de croître. Cette quantité de données joue un rôle central dans les progrès effectués par la biomédecine, dont l'impact sur la santé publique est de plus en plus important. Ceci étant, faire en sorte que cette quantité de données soit mobilisée à bon escient pour les besoins de chaque patient reste un défi majeur de la médecine.

Le projet BioASQ vise à fournir des solutions d'accès rapides et précises à cette masse d'information et repose pour cela sur une séries de campagnes d'évaluation de systèmes d'accès à l'information biomédicale. À travers ces différentes campagnes, BioASQ a mis à la disposition des chercheurs des mondes académiques et industriels des données et des protocoles d'évaluation permettant de tester différents aspects de l'accès à l'information dans les textes biomédicaux. À travers ces campagnes, le projet vise à définir les outils permettant d'identifier, de traiter et de présenter les éléments d'information indispensables au travail des experts biomédicaux.

Chaque campagne du projet comporte deux tâches. Dans la première, les systèmes participants doivent automatiquement affecter des termes du MeSH à des articles biomédicaux de MEDLINE. Les participants reçoivent les articles nouvellement publiés dans MEDLINE et non encore annotés par le personnel de la *National Library of Medecine* [3] (NLM) américaine ; ils doivent fournir leur annotation de ces articles dans un délai inférieur à celui

1. http ://www.nlm.nih.gov/mesh/
2. http ://www.bioasq.org/
3. http ://www.nlm.nih.gov/

de l'annotation manuelle de la *NLM*. Les réponses des participants sont ensuite comparées à cette annotation manuelle. Cette tâche est en général abordée par l'intermédiaire de systèmes d'annotations à base de règles ou par l'intermédiaire de systèmes d'apprentissage automatique.

Dans la deuxième tâche, les systèmes doivent fournir des réponses précises et compréhensibles aux questions biomédicales élaborées, en anglais, par des experts. Ces questions correspondent aux besoins en information des professionnels de la santé sur divers domaines. Pour chaque question, les systèmes doivent donner les articles pertinents, les concepts reliés extraits d'ontologies du domaine, des triplets RDF associés extraits de *Linked Life Data*[4], une réponse exacte dans le cas de questions précises et un résumé de quelques lignes répondant à la question. Cette tâche regroupe donc les tâches traditionnelles de recherche d'information, de questions-réponses (associant données textuelles non structurées et données structurées) et résumé multi-documents.

Principaux résultats

Les campagnes passées du projet BioASQ ont permis de mettre en évidence le fait que les méthodes d'apprentissage automatique étaient particulièrement bien adaptées à la tâche d'annotation sémantique. Les résultats des meilleurs systèmes participants, fondés sur l'apprentissage automatique, étaient, en début de projet, en deçà des performances du système MTI développé par la *NLM*; ils étaient bien au-delà de ce système au bout de quelques mois. La *NLM* a depuis adapté son système sur la base des méthodes utilisées par les meilleurs systèmes participants.

La tâche de questions-réponses est une tâche plus difficile pour laquelle il n'existait pas de système consensuel (équivalent à MTI pour la tâche précédente). Les performances des systèmes participants ont toutefois augmenté au cours des différentes campagnes sur la majorité des sous-tâches de cette problématique. L'adéquation des ressources utilisées (notamment pour l'annotation de concepts ou de triplets) reste cependant un problème majeur de certaines des sous-tâches.

Une description complète du projet et de ses résultats (en date de novembre 2014) est disponible sous http://bioasq.org/ ainsi que dans l'article [1].

4. http ://linkedlifedata.com/

Références

[1] George TSATSARONIS et al. "An overview of the BIOASQ large-scale biomedical semantic indexing and question answering competition". In : *BMC Bioinformatics* 16 (2015), p. 138. DOI : 10.1186/s12859-015-0564-6. URL : http://dx.doi.org/10.1186/s12859-015-0564-6.

LIMICS : Laboratoire d'Informatique Médicale et d'Ingénierie des Connaissances en e-Santé

Le LIMICS est un laboratoire pluridisciplinaire en santé et en informatique qui propose des approches principalement symboliques de traitement de l'information de santé sur les plans à la fois méthodologique et applicatif.

- CONTACT : Marie-Christine.Jaulent & Alain Venot - Marie-Christine.Jaulent@upmc.fr & Alain.Venot@univ-paris13.fr,
- WEB : http://www.limics.fr/
- TEL : +33 1 44 27 91 90

Membres de l'équipe

- Marie-Christine JAULENT, DR INSERM, directeur
- Alain VENOT, PU-PH Univ. Paris Nord, sous-directeur
- Eugenia LAMAS, CR INSERM
- Piere MENETON, CR INSERM
- Laurent TOUBIANA, CR INSERM
- Jacques BOUAUD, AP-HP
- Cédric BOUSQUET, Hôp. de St-Étienne
- Jean CHARLET, AP-HP
- Christel DANIEL, AP-HP
- Fadi BADRA, MCF Univ. Paris Nord
- Mélanie COURTINE, MCF Univ. Paris Nord
- Sylvie DESPRES, Pr Univ. Paris Nord
- Rim DJEDIDI, MCF Univ. Paris Nord
- Yannick KERGOSIEN, Pr Univ. Paris Nord
- Jean-Baptiste LAMY, MCF Univ. Paris Nord
- Jérôme NOBÉCOURT, MCF Univ. Paris Nord
- Karima SEDKI, MCF Univ. Paris Nord
- Pascal VAILLANT, MCF Univ. Paris Nord

- Frédérique CAPRON, PU-PH Univ. Pierre et Marie Curie
- Catherine DUCLOS, MCU-PH Univ. Paris Nord
- Éric LEPAGE, PU-PH Univ. Paris-Est Créteil
- Jean-Marie RODRIGUES, PU-PH, Univ. de St-Étienne
- Brigitte SÉROUSSI, MCU-PH Univ. Pierre et Marie Curie
- Olivier STEICHEN, MCU-PH Univ. Pierre et Marie Curie
- Béatrice TROMBERT-PAVIOT, PU-PH Univ. de St-Étienne
- Xavier AIMÉ, Rémy CHOQUET, Ferdinand DHOMBRES, Nicolas GRIFFON, Damien LEPROVOST, Jeremy LARODN, Vassilis KOUTKIAS, David OUAGNE (post-doctorants)
- Troskah FARNIA, Nassim DOUALI, Alexandre GALOPIN, Meriem MAAROUFI, Romain NG, Yves PARES, Amandine PERRINET, Marion RICHARD, Christian SIMON, Rosy TSPORA (doctorants)
- Stéfan J DARMONI, Lina SOUALMIA (chercheurs associés)

Thème général de l'équipe

Les recherches au LIMICS répondent au double enjeu, d'une part, de traiter des problématiques concrètes de recherche médicale et de prise en charge des patients rencontrées par les acteurs du domaine de la santé qui attendent des réponses opérationnelles et, d'autre part, de contribuer aux avancées de la discipline informatique, concernant par exemple le traitement de masses de données volumineuses (Big Data) ou répondre aux nouveaux enjeux dans le domaine du Web Sémantique. Ces recherches sont toutes liées à l'Intelligence Artificielles en général et à l'Ingénierie des connaissance en particulier.

Les résultats de ces recherches contribuent à la conception de systèmes d'information en santé dont les performances sont améliorées par leur capacité à informatiser le sens des données qu'ils manipulent. Elles permettent en outre d'identifier de nouveaux verrous dans les disciplines concernées et d'y apporter des réponses.

Description des travaux

Les activités de recherche du LIMICS sont associées aux disciplines de l'ingénierie des connaissances, de l'ingénierie des modèles, de l'aide à la décision et de l'informatique translationnelle (bioinformatique translationnelle et informatique de la recherche clinique) pour :

- proposer des méthodologies originales de construction et d'évaluation des ressources sémantiques en santé (modèles d'information, ontologies, terminologies) ;
- développer des outils d'alignement de ces ressources [9] et d'annotation sémantique (navigateurs, outils d'annotation, outils d'élaboration de requêtes sémantiques, etc.) pour faciliter l'interopérabilité entre systèmes d'information de Santé ;
- élaborer des nouvelles générations de standards internationaux dans le domaine de la santé tels que la CIM-11 et participer à l'actualisation des standards existants (DICOM, HL7, IHTSDO, etc.) ;
- améliorer la structuration et le codage des informations contenues dans les dossiers médicaux électroniques (y compris les données complexes telles que les images ou les données « omiques ») pour en faciliter l'exploitation pour le soin ou la recherche ;
- construire des entrepôts de données sémantiques intégrant toutes les données nécessaires (p. ex. cliniques, génomiques, d'imagerie, ...) pour faciliter l'émergence de nouvelles connaissances en santé ;
- Élaborer de nouvelles méthodes notamment graphiques de consultation, de visualisation, de synthèse et d'exploitation des données cliniques ou épidémiologiques et des connaissances médicales ;
- Développer des aides décisionnelles pour la recherche clinique et la prise en charge des patients, capables de s'appuyer sur les données des dossiers électroniques et dont les bases de connaissances sont faciles à maintenir ;
- Évaluer l'impact sur la qualité des pratiques et l'utilisabilité des outils développés dans des situations d'usage, simulées ou réelles, pour déterminer les facteurs de succès ou d'échec de leur mise en place (p. ex. le travail avec OncoDoc2 sur le cancer du sein [3]).

Les activités du LIMICS s'adressent à différentes catégories d'acteurs :

- les professionnels de santé qui doivent disposer de dossiers informatisés, structurés et partageables ainsi que de nouvelles aides décisionnelles pour améliorer la prise en charge diagnostique et thérapeutique des patients et qui doivent aussi avoir une plus grande facilité dans l'utilisation des outils informatiques, pour l'accès aux connaissances médicales et à la formation continue ;
- les chercheurs en sciences de la vie (recherche fondamentale, clinique ou épidémiologique) qui doivent disposer d'outils, comme des plateformes d'intégration de données et de connaissances, facilitant le traitement de données médicales complexes issues

de systèmes d'information hétérogènes à des fins d'exploitation de registres et de cohortes ;
- les citoyens qui sont les premiers bénéficiaires des applications de e-Santé et qui doivent disposer d'outils d'accès aux connaissances, de gestion de leur parcours de santé afin d'améliorer leur propre prise en charge, en particulier dans le contexte de maladies chroniques.

Projets

Les chercheurs du LIMICS sont impliqués dans de nombreux projets nationaux et internationaux. Nous allons lister ici quelques uns de ceux qui sont en cours. Une liste complète des projets, récemment terminés et en cours, est accessible à http://www.limics.fr/fr/projet.

TeRSan

Le projet TERSAN, Terminologies et Référentiels d'interopérabilité sémantique en Santé [5], piloté par le LIMICS, est un projet de recherche industrielle financé par l'ANR TecSan (Technologie de la Santé) nous associant à deux établissements de santé (AP-HP et CHU de Rouen/CISMeF) et un industriel (Mondeca).

Le partage de données de santé entre Systèmes d'Information de Santé (SIS) représente un enjeu pour la coordination des soins et la recherche biomédicale. L'interopérabilité sémantique permet le partage des données de santé représentées sous une forme non ambiguë interprétable par les machines. Elle repose sur l'élaboration de référentiels — modèles d'information et terminologies — permettant une représentation formelle de l'information de santé. Lorsque ces modèles sont définis par des organismes de standardisation, ils fournissent des règles internationales de structuration et de codage de l'information permettant ainsi de partager cette information à une large échelle.

L'objectif de ce projet de recherche d'une durée de 3,5 ans (février 2012 – juillet 2015) est de démontrer qu'il est possible d'élaborer et de maintenir des référentiels d'interopérabilité sémantique nationaux en langue française fondés sur les terminologies de référence internationales (e.g. LOINC, SNOMED CT, etc.) dans les domaines de la biologie, l'imagerie et l'anatomie cytologie pathologiques et de fournir aux établissements de santé les services leur permettant d'utiliser efficacement ces référentiels pour améliorer le partage ou l'échange d'informations cliniques que ce soit dans un contexte de coordination des soins ou de recherche [1].

ADR prism

Le projet ADR-prism [6] a pour objectif de mettre à disposition des équipes de pharmacovigilance une source de connaissances encore inexploitée en dehors de rares expérimentations par des équipes de recherche : les messages des patients dans les forums et autres lieux de discussions sur Internet. L'intégration de ces données permettra de générer de nouvelles hypothèses concernant les effets indésirables décrits par les patients qui seraient nouveaux ou mal documentés dans l'information officielle et/ou déjà existante sur les médicaments.

Dans ce projet, les informations sont extraites à partir de données textuelles (messages des patients sur les plates-formes de discussion) au moyen de méthodes de traitement automatique du langage (TAL). Les données textuelles sont annotées et intégrées aux données structurées, comme par exemple la date, la source de donnée ou l'adresse, en considérant leur sémantique, pour en préserver le sens et permettre une intégration et interopérabilité améliorées entre le langage médical et le langage des patients. D'un point de vue technique, les résultats de l'intégration seront consultables dans un environnement Web [8].

Accordys

Le projet ACCORDYS, piloté par le LIMICS [7] est un projet ANR Contint. Il consiste en l'exploitation des données biomédicales accumulées au cours des années par des spécialistes de fœtopathologie, discipline particulière de l'embryologie et de la génétique médicales, dont l'objet est l'étude des malformations du fœtus [4]. Ces données sont de nature hétérogène (photos, images d'échographie, images de différentes techniques de radiologie, résultats d'examens biologiques, etc.). Elles sont interprétées par les spécialistes respectifs qui produisent pour la plupart un compte rendu textuel. Dans le contexte du diagnostic prénatal d'une malformation, la connaissance de situations antérieures « similaires » et résolues (c'est-à-dire vérifiées après l'autopsie du fœtus) est essentielle à l'orientation diagnostique. Cette connaissance est précieuse car les malformations diagnostiquées en période prénatale sont très majoritairement des affections rares concernant moins d'un individu sur 2000.

5. http://www.limics.fr/fr/projet/fiche-projet/voir/14-TeRSan
6. http://www.limics.fr/fr/projet/fiche-projet/voir/7-ADR-prism
7. http://www.limics.fr/fr/projet/fiche-projet/voir/6-Accordys

L'objectif du projet répond donc à un besoin urgent de réunir toutes ces données et de les rendre accessibles aisément afin d'utiliser au mieux cette mémoire collective. Notre hypothèse de recherche est la suivante : Analyser les dossiers (après numérisation si nécessaire) leur apportera une valeur ajoutée en termes de connaissances, permettant ainsi de construire une base de connaissances du domaine. Les dossiers seront structurés et enrichis par indexation à l'aide de ressources termino-ontologiques et par leur mise en relation avec des connaissances externes (publications, bases de données). Une telle base permettra l'accès à des cas similaires.

EHR4CR

Le projet EHR4CR [8] est un projet européen (FP7 Call IMI) dont l'objectif est de fournir des outils et services permettant l'exploitation de données de Dossiers Patients Informatisés (DPI) ou d'Entrepôts de Données Cliniques (EDC) dans un contexte de recherche clinique pour faciliter : *a*) les études de faisabilité, *b*) le recrutement de patients, *c*) la collecte de données et *d*) la déclaration d'événements indésirables [5].

Les services fournis par la plateforme EHR4CR sont testés par 11 sites pilotes répartis dans 5 pays européens dans des domaines cliniques différents (cancérologie, cardiologie, maladies inflammatoires, respiratoires, diabète, neurologie).

Salus

Le projet SALUS [9] est un projet européen qui a pour but de faciliter les études d'effets indésirables médicamenteux à une large échelle. Pour cela, il est nécessaire de pouvoir accéder aux données bloquées dans différents systèmes de DPI hétérogènes. Dans le projet SALUS, nous visons à fournir un cadre d'interopérabilité fondé sur des standards qui permettent le repérage des effets indésirables par la fouille et l'analyse de données des patients en temps réel dans le contexte de systèmes de DPI hétérogènes [6].

SALUS fournira :

- des profils fonctionnels d'interopérabilité permettant l'échange des DPI ;
- des solutions d'interopérabilité sémantique permettant une interprétation univoque des éléments de DPI échangés
- des mécanismes de sécurité et confidentialité assurant que les DPI sont partagés de façon éthique et sécuritaire ;
- un nouveau cadre pour expliciter des patrons temporels pour repérer des effets indésirables dans les données du DPI ;
- la mise en œuvre des cas d'utilisation à fort potentiel permettant la réutilisation secondaire des DSE pour des études de sécurité post-marché.

SiFaDo

Le projet SiFaDo, Saisie facile de données médicales, [10], est un projet ANR Tecsan. Il a pour but la conception et l'évaluation de méthodes et d'outils ergonomiques pour faciliter la saisie et le codage de données textuelles et graphiques dans les dossiers médicaux électroniques.

Les médecins saisissent des informations en langage naturel dans les dossiers médicaux électroniques mais n'utilisent qu'extrêmement peu les outils de codage dont ils disposent et dont ils critiquent la lourdeur pour rentrer des informations sur leurs patients. De ce fait ils ne se servent pas des grands référentiels terminologiques qui sont pourtant des conditions nécessaires pour l'interopérabilité des systèmes d'information en santé. Ceci a des conséquences néfastes très importantes pour les patients à titre individuel mais aussi en termes de Santé Publique.

Le projet SiFaDo a pour objectif de concevoir et d'évaluer des méthodes et outils destinés à rendre immédiatement utile, facile, voire même ludique la saisie de données textuelles ou graphiques, structurées et codées dans les dossiers médicaux électroniques. Il réunit les compétences de neuf partenaires. Il s'agit de quatre partenaires académiques en milieu médical et scientifique (informatique médicale, interfaces hommes-machines, ergonomie), de quatre industriels (logiciels de gestion de cabinet médical, système d'information hospitalier, dossiers partagés et bases de connaissances en santé) et d'une société savante de médecine générale.

Il conduira à des méthodes originales de saisie de données médicales structurées et codées basée sur la diminution de la dimension de l'espace de recherche dans une classification intégrant des approches statistiques, fréquentielles et sur le recours à des interfaces graphiques dont certaines reposeront sur le langage iconique VCM [7] développé par un des partenaires.

8. http://www.ehr4cr.eu/
9. http://www.limics.fr/fr/projet/fiche-projet/voir/4-Salus
10. http://www.limics.fr/fr/projet/fiche-projet/voir/19-SiFaDo

De ce projet seront issus plusieurs produits logiciels commerciaux pour la médecine ambulatoire et le monde hospitalier.

Hybride

Le projet de recherche HYBRIDE, Hybridation de la fouille de données et du traitement automatique des langues,[11] est un projet ANR Blanc. Il a pour ambition de développer de nouvelles méthodes et outils pour guider la découverte de connaissances à partir de textes en combinant des méthodes de traitement du langage naturel (TAL) et des méthodes de découverte de connaissances dans les données (DCD) [2]. La partie expérimentale et la validation du projet HYBRIDE ont pour contexte le réseau des maladies orphelines — Orphanet — et pour objet l'aide à la documentation des maladies orphelines. Les aspects fondamentaux du projet HYBRIDE peuvent être appréhendés par l'intermédiaire des étapes principales d'un processus de découverte de connaissances avec une perspectives mixte TAL/IC : 1) préparation des données, 2) fouille des données, 3) interprétation et validation des résultats, 4) conception de connaissances. À chaque étape, des nouvelles méthodes doivent être construites et testées pour mettre en place cette boucle d'interactions entre TAL et DCD.

OFS

Le projet de recherche OFS, Open Food System,[12] est un projet de recherche d'envergure nationale, soutenu par 6 pôles de compétitivité, dont Vitagora et Cap Digital. Il a pour ambition de construire un écosystème de référence permettant de faciliter la préparation des repas grâce à la mise à disposition de contenus, d'appareils et de services innovants.

Il regroupe 25 partenaires industriels et scientifiques dont SEB, Tefal, Bearstech, Coheris, Mondeca, Temis, Institut telecom, Institut Paul Bocuse, Université Lyon 1 LIRIS, LUTIN, LAPPS, LIMICS, L2TI, etc.

Le projet OFS est structuré en deux grands axes :

- NOSRECETTES qui vise au développement des solutions de cuisine numérique pour le grand public. Il proposera des solutions adaptées aux différents profils utilisateurs, et permettra des échanges communautaires aux amateurs de cuisine du monde entier.
- OPTICOOK qui prévoit de mettre à disposition des professionnels et du grand public de nouveaux appareils de cuisson intelligents : contrôle automatique des paramètres de cuisson pour un résultat optimal, conservation des qualités organoleptiques et nutritionnelles des aliments cuits.

Une des tâches réalisées par le LIMICS consiste à construire une ontologie pour l'univers de la cuisine numérique. Cette ontologie doit permettre l'élaboration de suggestions nutritionnelles permettant à des internautes de s'alimenter de manière équilibrée, en se faisant plaisir et en permettant de partager leurs expériences culinaires avec des proches. Les suggestions nutritionnelles faites aux utilisateurs sont fondées sur les résultats du Programme National Nutrition Santé (PNNS)[13] et l'expertise en nutrition de l'Unité de Recherche en Epidémiologie Nutritionnelle (UREN) (http ://www.univ-paris13.fr/uren/). Elles prennent en compte les pratiques alimentaires observées dans un échantillon représentatif de familles. Elles comportent en outre des indications sur la saveur de la recette proposée. La ressource ontologique construite repose sur les modèles de connaissances des différents domaines représentés.

11. http://www.limics.fr/fr/projet/fiche-projet/voir/20-Hybride
12. http://www.limics.fr/fr/projet/fiche-projet/voir/21-OFS
13. http://www.mangerbouger.fr/pnns

Références

[1] X. AIMÉ et al. "Semantic interoperability platform for Healthcare Information Exchange". In : *IRBM* 36.1 (2015), p. 16–26.

[2] N BECHET et al. "Sequential Pattern Mining to Discover Relations between Genes and Rare Diseases". In : *Proc of the 25e International Symposium on Computer-Based Medical Systems*. 2012.

[3] J. BOUAUD et al. "Quels sont les patients atteints d'un cancer du sein dont la décision de prise en charge thérapeutique bénéficie de l'utilisation d'un système d'aide à la décision ? Un exemple utilisant la fouille de données et OncoDoc2". In : *IC 2014 25es Journées Francophones d'Ingénierie des Connaissances*. Sous la dir. de C. FARON ZUCKER. Clermont Ferrand, France, 2014, p. 107–118.

[4] Jean CHARLET et al. "ACCORDYS - Contents and knowledge aggregation for case-based reasoning in the field of foetal dysmorphology". In : *IMIA Francophone Special Interest Group - MEDINFO 2013*. [workshop]. Copenhagen, Denmark, 2013.

[5] G De Moor et al. “Using electronic health records for clinical research : the case of the EHR4CR project”. In : *J Biomed Inform.* 53 (2015), p. 162–73.

[6] Declerck G et al. “Bridging data models and terminologies to support adverse drug event reporting using EHR data”. In : *Methods Inf Med* 54.1 (2015), p. 24–31.

[7] Jean-Baptiste Lamy et al. “An iconic language for the graphical representation of medical concepts.” In : *BMC medical informatics and decision making* 8 (2008), p. 16. ISSN : 1472-6947. DOI : 10.1186/1472-6947-8-16.

[8] J Lardon et al. “Adverse Drug Reaction Identification and Extraction in Social Media : A Scoping Review”. In : *JMIR* (2015).

[9] Laurent Mazuel et Jean Charlet. *OnAGUI - Ontology Alignment GUI*. Open source software developped by Laurent Mazuel during his work with Jean Charlet at Inserm UMR_S 872, Éq. 20 and available at : http://sourceforge.net/projects/onagui/ under GNU General Public License. 2009.

LIMSI, groupe ILES

Le laboratoire d’informatique pour la mécanique et les sciences de l’ingénieur (LIMSI) est un laboratoire pluridisciplinaires. Les activités du goupe « Informations Langues Ecrite et Signée » (ILES) concernent le traitement automatique des langues pour les données écrites (analyse, compréhension, production, acquisition de connaissances, etc.), avec une application privilégiée au domaine biomédical, ainsi qu’une thématique sur la langue des signes française.

- CONTACT : Pierre Zweigenbaum – pz@limsi.fr,
- ADRESSE : LIMSI–CNRS, Campus universitaire d’Orsay, rue John von Neumann, Bâtiment 508, 91405 Orsay Cedex
- WEB : http://www.limsi.fr/
- TEL : 01-69-85-80-02

Membres de l’équipe

L’activité biomédicale est assurée dans le groupe ILES par neuf chercheurs permanents sur les dix-neuf du groupe, secondés par des étudiants et post-docs [14] :

- Cyril Grouin, IE CNRS
- Thierry Hamon, Maître de Conférence de l’Université Paris-Nord
- Thomas Lavergne, Maître de Conférence de l’Université Paris-Sud
- Anne-Laure Ligozat, Maître de Conférence de l’ENSIIE
- Véronique Moriceau, Maître de Conférence de l’Université Paris-Sud
- Aurélie Névéol, CR CNRS
- Sophie Rosset, DR CNRS (groupe TLP)
- Xavier Tannier, Maître de Conférence de l’Université Paris-Sud
- Pierre Zweigenbaum, DR CNRS

Thème général de l’équipe

L’activité biomédicale du groupe ILES consiste à concevoir des méthodes pour le traitement automatique de documents médicaux (comptes rendus cliniques, résumés et articles scientifiques) afin de faciliter l’accès à l’information exprimée dans ces documents et son exploitation.

Description des travaux

Contexte

Dans le domaine biomédical, les informations cliniques et institutionnelles sont contenues dans le texte de publications scientifiques (principalement en anglais) ou de dossiers patients (dans les langues des pays concernés ; nous travaillons sur le français et l’anglais) et ne sont pas directement accessibles pour des analyses systématiques. Les recherches du groupe ILES combinent le traitement automatique de la langue biomédicale, la représentation des connaissances et la fouille de textes afin d’extraire des informations pertinentes des textes libres et de les convertir en représentations formelles exploitables par l’homme et par la machine.

Les travaux du groupe ILES s’intéressent à des problèmes fondamentaux du traitement automatique des langues tels que la variation terminologique [8], la normalisation de concepts [14], l’extraction d’entités [1, 4] et de relations entre entités [5, 9], et la modélisation de ces phénomènes en corpus [6]. Une partie des travaux du groupe ILES portent également sur des applications en informatique médicale telles que l’extraction et la recherche d’information (**Accordys**, **CABeRneT**, **Vigi4Med**), la recherche de cas similaires (**Accordys**), l’analyse rétrospective de cohortes (**CABeRneT**), la compréhension

14. Depuis 2010, on dénombre 7 post-doctorats, 1 thèse et 7 stages sur cette thématique.

et la génération de texte dans le cadre d'un dialogue personne-machine (**Patient Genesys**), toutes ces applications nécessitant le développement de ressources terminologiques.

Projets

Accordys (ANR-12-CORD-007-03)

Le projet Accordys (Agrégation de Contenus et de COnnaissances pour Raisonner à partir de cas de DYSmorphologie fœtale), porté par le LIMICS [3] (Jean Charlet, voir la description de cette équipe p. 11), vise à analyser le contenu de documents cliniques pour proposer aux médecins des cas similaires au cas qu'ils ont à traiter. Cette recherche de cas similaires repose d'une part sur des archives numérisées en fœtopathologie, et d'autre part sur la littérature scientifique recensée dans la base de données MEDLINE [7]. Le projet a également une valeur patrimoniale avec la numérisation d'archives de l'hôpital Trousseau. Elle soulève des problèmes d'identification des traits qui caractérisent un cas et de leur détermination à partir des textes d'un dossier.

CABeRneT (ANR-13-JS02-0009-01)

Le projet CABeRneT (Compréhension Automatique de Textes Biomédicaux pour la Recherche Translationnelle) est un projet « Jeunes chercheurs » porté par le LIMSI (Aurélie Névéol). Ce projet se donne pour objectif de mettre à disposition de la communauté scientifique des ressources dans le domaine biomédical en français, d'étudier l'adaptation en domaine de spécialité d'outils développés pour la langue générale [16], et d'appliquer ces outils à l'analyse automatique de dossiers électroniques patient [15] et à la détection de liens entre données cliniques et littérature. CABeRneT s'attaque ainsi aux problèmes d'adaptation à la langue et au domaine dans l'extraction d'information à partir de textes spécialisés.

Vigi4MED (ANSM-2013-S-060)

Le projet Vigi4MED (Vigilance dans les forums sur les Médicaments), porté par le LIMICS / CHU de Saint-Étienne (Cédric Bousquet), part du constat que les patients cherchent des informations médicales sur internet, y compris avant de consulter, soit par facilité, soit par peur du médecin (effet « blouse blanche »). Ce constat s'applique également aux effets secondaires des médicaments, déclarés en priorité sur internet, et connus plus tardivement des centres de pharmacovigilance. Le projet vise à identifier des effets secondaires de médicaments, positifs ou négatifs, depuis les forums de santé en français [13]. L'évaluation sera faite de manière rétrospective (confirmer des effets secondaires connus) et prospective (identifier de nouveaux effets dont on constate l'émergence). L'une des difficultés abordées dans le projet est l'identification de signes médicaux exprimés dans la langue des patients plutôt que des médecins.

Patient Genesys (FUI-16 F1310002 P)

Le projet Patient Genesys, porté par la société Interaction Healthcare, vise à produire une plateforme de simulation pour la formation des professionnels de santé, dans laquelle un médecin formateur décrit des cas de patients que le médecin en formation va rencontrer dans une consultation virtuelle [11, 12]. Le médecin en formation peut en particulier dialoguer en langue naturelle (via le clavier) avec le patient virtuel, qui répond de façon orale (génération de phrases puis synthèse de la parole). Le système de dialogue personne-machine jouant le rôle du patient prend connaissance de l'état de santé et de l'historique du patient dans le dossier renseigné de façon semi-structurée par le médecin formateur, ce qui lui demande de savoir interpréter les informations qui s'y trouvent. Il analyse les questions du médecin utilisateur et recherche dans le dossier les informations qui lui permettront d'y répondre [2], en employant le vocabulaire d'un patient plutôt que celui du médecin. Cela suppose notamment de mettre en correspondance ici aussi le vocabulaire spécialisé employé par les professionnels de santé avec le vocabulaire des patients.

Outils

MEDINA

MEDINA [15] est un logiciel libre conçu pour anonymiser les données personnelles contenues dans des documents cliniques rédigés en français, au format texte brut. MEDINA est utilisé pour le pré-traitement des documents cliniques dans l'ensemble des projets du groupe ILES.

Wapiti

Wapiti [16] est un logiciel libre conçu pour l'étiquetage rapide de séquences. Il implémente les formalismes CRF et MaxEnt. Wapiti est utilisé dans les outils développés par

15. http://medina.limsi.fr/
16. http://wapiti.limsi.fr

le groupe ILES pour le traitement des textes biomédicaux sur des tâches aussi variées que l'étiquetage morphosyntaxique, la reconnaissance d'entités ou l'extraction de relations.

Manifestations internationales

Participation à des campagnes d'évaluation

Le LIMSI participe régulièrement à des campagnes d'évaluation internationales qui lui permettent de développer et d'adapter des outils pour le traitement de la langue biomédicale, par des approches à base de règles et de lexiques, ou par apprentissage statistique.

Depuis 2009, le LIMSI a notament participé aux campagnes i2b2 [17] sur la détection des traitements médicaux et informations associées (2009), des examens, problèmes et traitements (2010 et 2011), de leurs assertions (2010), des relations cliniques (2010) et temporelles (2012) entre concepts, ou de la résolution des coréférences entre concepts (2011). Il s'agissait également de la désidentification de comptes rendus cliniques et du repérage des facteurs de risque pour les patients diabétiques (2014). Pour certaines de ces campagnes, le LIMSI s'est positionné dans le haut du classement : 8/20 en 2009, 5/21 pour les assertions et 3/16 pour les relations en 2010, 1/3 en 2011.

Le LIMSI a également participé à la campagne BioNLP 2013 [18] sur la détection des mentions de bactéries et de biotopes dans des résumés d'articles scientifiques [10], ainsi qu'aux campagnes ShARe/CLEF-eHealth 2013 [19] et SemEval 2014 [20] sur la détection de mentions de maladies et leur normalisation vers SNOMED-CT.

Organisation de manifestations

Campagne CLEF e-Health 2015. Le LIMSI organise la tâche 1b du challenge CLEF e-Health [21] (CLEF 2015, 8–11 septembre, Toulouse, France), qui porte sur la reconnaissance et la normalisation d'entités dans des textes biomédicaux en français.

Workshop LOUHI 2015. Le LIMSI organise la sixième édition du workshop LOUHI [22] sur l'analyse et la fouille de texte en Santé, qui aura lieu conjointement avec la conférence EMNLP 2015 (17–21 septembre, Lisbonne, Portugal).

17. Integrating Informatics and Biology to the Bedside, https://www.i2b2.org/NLP/
18. http://2013.bionlp-st.org
19. https://sites.google.com/site/shareclefehealth
20. http://alt.qcri.org/semeval2014/task7
21. https://sites.google.com/site/clefehealth2015/
22. http://louhi2015.limsi.fr

Références

[1] Andreea Bodnari et al. "A supervised named-entity extraction system for medical text". In : *CLEF 2013 Evaluation Labs and Workshop Online Working Notes*. 2013.

[2] Leonardo Campillos et al. "Un patient virtuel dialogant". In : *Actes de Démonstrations TALN 2015 (Traitement automatique des langues naturelles)*. ATALA. Caen, France, 2015.

[3] Jean Charlet et al. "Agrégation de contenus et de connaissances pour raisonner à partir de cas de dysmorphologie fœtale". In : *Actes Session francophone d'informatique médicale*. Sous la dir. de Pierre Zweigenbaum et Antoine Geissbuhler. IMIA Francophone SIG. Copenhagen, Denmark, 2013.

[4] Maria Evangelia Chatzimina, Cyril Grouin et Pierre Zweigenbaum. "Use of unsupervised word classes for entity recognition : Application to the detection of disorders in clinical reports." In : *Proc of LREC*. 2014, p. 3264–3271.

[5] Faisal Mahbub Chowdhury et Pierre Zweigenbaum. "A controlled greedy supervised approach for co-reference resolution on clinical text". In : *J Biomed Inform* 46.3 (2013), p. 506–515.

[6] Louise Deléger et al. "Annotation of specialized corpora using a comprehensive entity and relation scheme". In : *Proc of LREC*. 2014, p. 1267–1274.

[7] Eva d'Hondt et al. "LIMSI 2014 Clinical Decision Support Track". In : *TREC CDS Working Notes*. NIST. Gaithersburg, 2014. url : http://trec.nist.gov/pubs/trec23/pro-LIMSI_clinical.pdf.

[8] Michel Généreux, Amália Mendes et Thierry Hamon. "Experiments in synonymy : weakly supervised term matching to concepts". In : *Proc of Terminology and Artificial Intelligence Conference*. 2013, p. 181–184.

[9] Cyril Grouin et al. "Eventual situations for timeline extraction from clinical reports". In : *J Am Med Inform Assoc* 20.5 (2013), p. 820–827.

[10] Thomas Lavergne, Cyril Grouin et Pierre Zweigenbaum. "The contribution of co-reference resolution to supervised relation detection between bacteria and biotopes entities". In : *BMC Bioinformatics* (2015). *À paraître.*

[11] Jérôme Leleu et alii. "Patient Genesys". In : *Dossier IA et jeux vidéos. Bulletin de l'AFIA* (2014).

[12] Jérôme Leleu et al. "Patient Genesys : Outil de création de cas cliniques de simulation médicale proposant des cas patients virtuels en 3D". In : *Applications Pratiques de l'Intelligence Artificielle, Plateforme AFIA : Démonstrations.* AFIA. Rennes, 2015.

[13] François Morlane-Hondère et al. "Médicaments qui soignent, médicaments qui rendent malades. Étude des relations causales pour identifier les effets secondaires". In : *Actes de TALN 2015 (Traitement automatique des langues naturelles).* ATALA. Caen, France, 2015.

[14] Aurélie Névéol et al. "The QUAERO French Medical Corpus : A Ressource for Medical Entity Recognition and Normalization". In : *Proc of BioTextMining Work.* 2014, p. 24–30.

[15] Anne-Dominique Pham et al. "Natural language processing of radiology reports for the detection of thromboembolic diseases and clinically relevant incidental findings". In : *BMC Bioinformatics* 15 (2014), p. 266.

[16] Christelle Rabary, Thomas Lavergne et Aurélie Névéol. "Étiquetage morpho-syntaxique en domaine de spécialité : le domaine médical". In : *Actes de TALN 2015 (Traitement automatique des langues naturelles).* ATALA. Caen, France, 2015.

LIMSI, groupe TLP, thème Dimensions affectives et sociales dans les interactions parlées : Applications dans le domaine de la santé

- CONTACT : Laurence Devillers, devil@limsi.fr
- ADRESSE : LIMSI-CNRS, Bât 508, 91403 Orsay Cedex
- WEB : http://www.limsi.fr/tlp/topic2.html

Membres de l'équipe

- Lucile Béchade, Doctorante
- Agnès Delaborde, ATER
- Laurence Devillers, Professeur Paris-Sorbonne
- Guillaume Dubuisson-Duplessis, Post-doc
- Clément Gossart, Ingénieur CDD
- Mohamed Sehili, Post-doc
- Marie Tahon, Post-doc
- Fan Yang, Doctorant

Thèses déjà soutenues dans cette équipe

- Mariette Soury, thèse LIMSI 2014 : *Détection multimodale du stress pour la conception de logiciels de remédiation,* Université Paris-Sud
- Clément Chastagnol, thèse LIMSI 2013 : *Reconnaissance automatique des dimensions affectives dans l'interaction orale homme-machine pour des personnes dépendantes,* Université Paris-Sud
- Agnès Delaborde, thèse LIMSI 2013 : *Modélisation du profil émotionnel de l'utilisateur dans les interactions parlées Humain-Machine,* Université Paris-Sud
- Christophe Vaudable, thèse LIMSI 2012 : *Analyse et reconnaissance des émotions lors de conversations de centres d'appels,* Université Paris-Sud
- Laurence Vidrascu, thèse LIMSI 2007 : *Analyse et détection des émotions verbales dans les interactions orales,* Université Paris-Sud
- Chloé Clavel, thèse Thales/Telecom-Paris/LIMSI Paris 2007 : *Analyse et reconnaissance des manifestations acoustiques des émotions de type peur en situations anormales,* ENST

Thème général de l'équipe

Nos sujets de recherche portent sur la détection et l'interprétation d'indices émotionnels et sociaux lors d'interactions parlées. La détection des émotions est effectuée à partir d'indices verbaux et non verbaux principalement acoustiques, mais aussi visuels. Un profil de l'utilisateur est ensuite construit dynamiquement à partir du comportement expressif détecté et du contexte de l'interaction pendant le dialogue. Ce profil est utilisé pour modifier le comportement d'un système artificiel ou pour analyser des dialogues humain-humain. Des applications médicales sont développées que ce soit en robotique sociale ou pour des jeux sérieux thérapeutiques.

L'équipe se situe dans le groupe Traitement du langage parlé dans le département Communication Homme-Machine au LIMSI-CNRS.

Une vidéo montre la détection des émotions et la construction du profil : https://www.youtube.com/watch?v=p1ID-gvUnWs&feature=em-upload_owner.

Description des travaux

Pour concevoir des systèmes interactifs affectifs ou pour analyser des dialogues, il est nécessaire d'étudier les expressions des émotions et des indices sociaux pendant l'interaction [8, 9]. Les indices socio-culturels sont contrairement aux émotions volontairement contrôlées. Dans l'interaction humaine, les éléments non-verbaux comme les indices acoustiques paralinguistiques ou les indices visuels enrichissent le contenu linguistique et permettent également de mieux interpréter le message communiqué [7]. La voix et le dialogue jouent un rôle fondamental dans des interactions sociales mais ils ont été relativement négligés, au moins ces années dernières, comparés à d'autres aspects d'échanges sociaux comme les expressions du visage ou les gestes. Il y a une tendance également dans le domaine de l'*« affective computing »* à utiliser des données émotionnelles très exagérées et produites artificiellement par des acteurs. Il semble de plus en plus clair que cette stratégie n'est pas efficace parce que les formes d'expression qui arrivent dans des interactions naturelles diffèrent fondamentalement de celles que les acteurs produisent sur commande.

Depuis 2001, le travail de recherche sur les émotions mené dans cette équipe est fondé sur l'utilisation de matériel audio spontané « real-life ». L'équipe était l'une des premières à se saisir de cette question et encore un très petit nombre de chercheurs ont maintenant compris ce défi [12]. L'utilisation de données spontanées nécessite de produire des bases de données, de les collecter et de les annoter [5, 14] car il en existe très peu de disponible dans la communauté internationale. Les états émotionnels spontanés en interaction sociale sont souvent subtils et complexes présentant des mélanges d'émotions et d'indices sociaux. L'équipe a rassemblé et a analysé un grand nombre de bases de données de dialogues émotionnels dans différents domaines : dans des centres d'appels pour des consultations financières, ou encore des appels d'urgence médicale ou encore lors d'interactions humain-robot. P. Ekman a étudié 6 émotions primaires sur les expressions des visages : la peur, la colère, la joie, la tristesse, la surprise et le dégoût. Les études ont été menées au LIMSI-CNRS sur différentes classes d'émotion présentes dans les données analysées : dans les expressions négatives, on trouve la peur, le stress, l'anxiété, la panique [3] ou encore la colère, la contrariété, l'énervement, ou encore la tristesse, la déception, la dépression et dans les expressions positives, par exemple le soulagement, l'amusement, le plaisir ainsi que des mélanges de ces différentes émotions même positive et négative [8, 9]. Les techniques d'analyse acoustique utilisées extraient des indices spectraux, prosodiques et des marqueurs affectifs comme le rire, la toux, les souffles sur des fenêtres temporelles de différentes tailles allant jusqu'à produire quelques milliers d'indices [15] pour représenter un segment audio qualifié d'émotionnel. Les systèmes de détection d'émotion utilisent des modèles obtenus grâce à des techniques d'apprentissage automatique sophistiquées comme des Machines à Vecteurs de Support (SVM) ou encore des réseaux de neurones comme les réseaux profonds « *deep learning* ». Les performances de ces modèles et leur pouvoir de généralisation dépendent de la qualité et quantité des données d'entraînement. Nous menons de nombreuses études cross-corpus [6]. Les expressions émotionnelles sont extrêmement variables d'un individu à l'autre, d'une situation à l'autre. Nous étudions ces facteurs de variabilité (âge, sexe, tâche, personnalité, santé, lieu, rôle, etc.) à travers de nombreuses analyses de données.

Contexte

La détection de dimensions affectives et sociales peut être utilisée pour la communication homme-machine avec des robots, mais aussi pour l'analyse de documents audiovisuels avec différents buts de santé, de sécurité, d'éducation, de divertissement ou d'applications de jeux sérieux. Un robot social sensible aux émotions doit prendre en compte non seulement les émotions ponctuelles, mais avoir aussi une représentation du profil du locuteur tout au long des interactions [4], pour avoir une chance d'être plus pertinent dans ses réponses comportementales. La capacité de prévoir qu'un comportement spécifique aura une grande chance de déclencher la satisfaction de l'utilisateur est très importante. Par exemple, quelqu'un ayant une grande confiance en lui n'a pas besoin d'être encouragé pour interagir et il pourrait même interpréter cet encouragement comme ennuyeux.

Dans un avenir proche, la robotique sociale [1] sera utilisé pour accompagner les personnes en automatisant la surveillance, l'apprentissage, la motivation et les aspects de camaraderie dans des interactions seul à seul avec des individus de populations diverses, y compris les personnes âgées, les enfants, les personnes handicapées et les individus ayant différentes pathologies. Les questions éthiques, y compris la sécurité, la vie privée et la fiabilité de comportement de robot, sont aussi de plus en plus largement discutées. Il est nécessaire que les réflexions sur l'éthique soient menées conjointement par différentes disciplines durant le développement scientifique et technologique de ces systèmes doués d'intelligence artificielle, pour assurer l'harmonie et l'acceptabilité de leur relation avec les êtres humains. Nous sommes également impliqués dans le groupe de travail sur l'Éthique pour la recherche dans

la robotique de la CERNA (le commission sur l'Éthique d'Allistène) [2].

Par ailleurs, nous animons dans le cadre de l'ISN (LIDEX ISN, Institut de la Société Numérique : http://digitalsocietyinstitute.com/fr) un pôle sur la co-évolution humain-machine. Les robots seront dotés d'une autonomie croissance. La confiance que l'on peut placer dans un robot, les possibilités et limites de celui-ci et du couple qu'il forme avec l'utilisateur, la reprise en main, le traçage — c'est-à-dire la possibilité de rendre compte du comportement — sont à considérer du point de vue éthique dans la conception du robot. Nous menons une réflexion sur l'anticipation des questions liées à la responsabilité civile des machines connectées en collaboration avec les chercheurs du CERDI (Faculté de droit Jean Monnet, Paris-Sud). Par l'interaction sociale et affective, le robot peut jouer de façon inédite sur le comportement de la personne et l'analyse interdisciplinaire de ces effets à court et à long terme est importante d'autant plus quand ces robots seraient placés auprès d'enfants ou de personnes fragiles. Une autre étude pluridisciplinaire est menée avec une équipe de sociologues de Telecom-Paris sur l'engagement des humains dans une interaction sociale avec les robots et plus particulièrement sur la réflexivité langagière dans une interaction en triade : humain – dispositif de quantification de soi – robot. L'évaluation de ces systèmes est aussi une problématique cruciale de recherche.

Projets

L'application de nos technologies dans le domaine de la santé est émergente et ouvre de nombreux défis de recherche et de nombreux intérêts industriels. Nous présentons 2 applications : l'une en robotique sociale pour les personnes dépendantes et l'autre dans les jeux sérieux.

Dans le cadre du **projet BPI ROMEO2** (http://projetromeo.com) (2013–2016), nous sommes impliqués dans la conception et la construction d'un robot assistant pour maintenir des personnes âgées dans leur environnement naturel. Nous avons défini des scénarios d'interactions sociales avec des personnes âgées vivant dans des EHPAD en collaboration avec des thérapeutes de l'association Approche. Ces rencontres, discussions et expériences avec le robot Nao ont permis de recueillir de nombreuses informations qui ont été utiles pour proposer un premier jeu de scénarios d'évaluation mais aussi pour obtenir un premier retour des personnes âgées sur le principe d'une assistance robotique. Ces corpus permettent en effet l'entrainement des algorithmes en charge de détecter notamment les émotions des utilisateurs de ROMEO et de construire leurs profils [16, 17, 11, 10].

Le but du **projet FEDER E-Thérapies** (2011–2014) était le design de jeux sérieux immersifs à but thérapeutiques fondés sur les interactions verbales et non verbales et la technique de jeu de rôle. Dans le cadre de ce projet, nous avons développé des outils pour la reconnaissance automatique du stress chez des humains en interaction dans des situations anxiogènes : prise de parole en public, entretiens et jeux sérieux à partir d'indices audio et visuels. L'expression et la gestion du stress sont influencées à la fois par des différences interpersonnelles (traits de personnalité, expériences passées, milieu culturel) et contextuelles (type de stresseur, enjeux de la situation). Nous avons évalué le stress sur différents publics à travers des corpus de données collectés dans différent contextes [13, 18] : un public sociophobe en situation anxiogène, face à une machine et face à des humains ; un public non pathologique en simulation d'entretien d'embauche ; et un public non pathologique en interaction face à un ordinateur ou face au robot humanoïde Nao. Les comparaisons inter- individus, et inter-corpus ont révélé la diversité de l'expression du stress. Une application de ces travaux pourrait être la conception d'outils thérapeutiques pour la maitrise du stress, notamment à destination de populations phobiques ou bipolaires. L'étude avec les patients sociophobes a été menée en collaboration avec des psychiatres de l'hôpital Pitié-Salpêtrière.

Références

[1] Axel Buendia et Laurence Devillers. "From informative cooperative dialogues to long-term social relation with a robot". In : *Natural Interaction with Robots, Knowbots and Smartphones*. 2014, p. 135–151.

[2] CERNA. *Rapport n° 1 de la CERNA sur l'Éthique du chercheur en robotique*. http://cerna-ethics-allistene.org/digitalAssets/38/38704_Avis_robotique_livret.pdf.

[3] Chloé Clavel et al. "Fear-type emotion recognition for future audio-based surveillance system". In : *Speech Communication* (2008), p. 487–503.

[4] Agnes Delaborde et Laurence Devillers. "Use of nonverbal speech cues in social interaction between human and robot : emotional and interactional markers". In : *Proceedings of the 3rd international workshop on Affective interaction in natural environments*. 2010, p. 75–80.

[5] Laurence Devillers, Sarkis Abrilian et Jean-Claude Martin. "Representing real-life emotions in audiovisual data with non basic emotional patterns and context features". In : *Affective computing and intelligent interaction*. Berlin Heidelberg : Springer, 2007, p. 519–526.

[6] Laurence Devillers, Christophe Vaudable et Clément Chastagnol. "Real-life emotion-related states detection in call centers : a cross-corpora study". In : *Eleventh Annual Conference of the International Speech Communication Association*. 2010.

[7] Laurence Devillers et Laurence Vidrascu. "Real-life emotions detection with lexical and paralinguistic cues on human-human call center dialogs". In : *Interspeech*. 2006.

[8] Laurence Devillers, Laurence Vidrascu et Lori Lamel. "Challenges in real-life emotion annotation and machine learning based detection". In : *Journal of Neural Networks* 18.4 (2005), p. 407–422.

[9] Laurence Devillers, Laurence Vidrascu et Omar Layachi. "Automatic detection of emotion from vocal expression, A Blueprint for an Affectively Competent Agent, Cross-Fertilization Between Emotion Psychology". In : *Affective Neuroscience, and Affective Computing*. 2010, p. 232–244.

[10] Laurence Devillers et al. "Inference Détection des états affectifs lors d'interactions parlées : robustesse des indices non verbaux". In : *TAL* 55.2 (2015). Special issue "Traitement automatique du langage parlé", In press.

[11] Laurence Devillers et al. "Inference of Human Beings' Emotional States from Speech in Human-Robot Interactions". In : *International Journal of social robotics* (2015). Special Issue Dev, In press.

[12] Ellen Douglas-Cowie et al. "The HUMAINE database : addressing the collection and annotation of naturalistic and induced emotional data". In : *Affective computing and intelligent interaction*. Berlin Heidelberg : Springer, 2007, p. 488–500.

[13] Jiewen Hua et al. "Global stress response during a social stress test : impact of alexithymia and its subfactors". In : *Psychoneuroendocrinology* (2014), p. 53–61.

[14] Marc Schröder et al. "What should a generic emotion markup language be able to represent ?" In : *Affective computing and intelligent interaction*. Berlin Heidelberg : Springer, 2007, p. 440–451.

[15] Bjorn Schuller et al. "The Interspeech 2010 paralinguistic challenge". In : *INTERSPEECH*. 2010, p. 2794–2797.

[16] Mohamed A. Sehili, Fan Yang et Laurence Devillers. "Attention Detection in Elderly People — Robot Spoken Interactions". In : *ICMI*. Istanbul, Turquey, 2014.

[17] Mohamed A. Sehili et al. "A corpus of social interactions between Nao and elderly people". In : *LREC*. Reyjkavic, Island, 2014.

[18] Mariette Soury et Laurence Devillers. "Stress detection from audio on multiple window analysis size in a public speaking task". In : *IEEE Affective Computing and Intelligent Interaction (ACII)*. 2013, p. 529–533.

LIRMM, équipe ADVANSE : ADVanced Analytics for data SciencE

Les activités de recherche menées par l'équipe ADVANSE s'inscrivent dans le domaine de la Science des données et plus particulièrement dans le domaine de l'analyse de grands volumes de données (Big Data). Un domaine d'application privilégié de ces méthodes est la Santé.

- CONTACT : P. PONCELET & S. BRINGAY {poncelet,bringay}@lirmm.fr
- Bâtiment 5 - 860 rue de St Priest 34095 Montpellier cedex 5
- WEB : http://www.lirmm.fr/recherche/equipes/advanse
- TEL : 04 67 41 86 53

Membres de l'équipe

- Jérôme Azé, Professeur de la nouvelle Université de Montpellier
- Sandra Bringay, Maître de Conférences de l'Université Paul Valéry de Montpellier
- Dino Ienco, Associé, Chargé de Recherche, IRSTEA
- Pierre Pompidor, Maître de Conférences de la nouvelle Université de Montpellier
- Pascal Poncelet, Professeur de la nouvelle Université de Montpellier
- Mathieu Roche, Associé, Directeur de Recherche, CIRAD
- Arnaud Sallaberry, Maître de Conférences de l'Université Paul Valéry de Montpellier

- Maguelonne Teisseire, Associée, Directrice de Recherche, IRSTEA

Thème général de l'équipe

Les organisations produisent de plus en plus de données. Elles ont besoin de prendre les meilleures décisions, le plus rapidement possible, en se fondant sur ces données. Dans ce contexte, l'*Analyse de données* (Data Analytics) gagne en popularité. Un tel processus vise à découvrir des connaissances utiles pour la prise de décision et peut aller jusqu'à suggérer des conclusions ou même prescrire des actions. Dans le domaine de la santé, les méthodes d'analyse sont très utiles comme le montre des revues de la littérature récentes [8, 6]. Les très grandes quantités de données générées par les systèmes de soins pour chaque patient sont trop complexes et volumineuses pour être traitées et analysées sans automatisation. L'analyse de données fournit alors des méthodes et des outils pour transformer ces gros volumes de données en informations utiles pour la prise de décision médicale. Tous les acteurs impliqués dans les systèmes de soins peuvent bénéficier des applications de l'analyse de données : les médecins pour identifier les traitements les plus efficaces et de meilleures pratiques, les hôpitaux pour prendre des décisions de gestion et par exemple diminuer les coûts des soins, les patients pour identifier les soins les plus abordables, les assureurs pour détecter les fraudes et les abus, etc.

Description des travaux

L'équipe illustre depuis 2007 ses travaux via de nombreuses applications dans le domaine de la santé, grâce à des collaborations locales, nationales et internationales.

Nous nous sommes intéressés à l'extraction de motifs (au sens large) à partir de données de puces ADN. Ainsi, via des collaborations avec le MMDN-UM2, puis dans le cadre de l'ANR Pradnet (Alzheimer) et d'un PEPS avec l'IGMM (HIV), nous avons reconsidéré le problème de l'extraction des motifs séquentiels en réorganisant les données pour prendre en compte le degré d'expression des différents gènes [12]. Cette nouvelle approche a permis aux biologistes de faire apparaître de nouvelles corrélations jusqu'à présent inconnues. En outre, afin d'aider le biologiste dans l'analyse des motifs extraits, nous avons implémenté un outil de visualisation mettant en lien les articles de PubMed qui traitent les différents gènes impliqués dans les motifs extraits [11]. Nous avons également étudié le pouvoir prédictif de ces motifs pour prédire le grade du cancer [7].

La découverte de nouvelles connaissances via les motifs pose le problème de leur généralisation : *sont-ils représentatifs d'un contexte donné ou bien plus généraux ?* Les motifs séquentiels précédents ne tiennent généralement pas compte des informations contextuelles, fréquemment associées aux données séquentielles. Par exemple, dans le cas des séquences d'actes réalisés pour des patients dans un hôpital, l'extraction classique de motifs séquentiels se focalise sur les séries d'actes sans considérer leur sexe, leur âge, etc. Or, en considérant le fait qu'un motif séquentiel est spécifique à un contexte donné (e.g. les jeunes hommes), un médecin pourra adapter sa stratégie de soin au contexte du patient et prendre les décisions adéquates. Dans le cadre d'une collaboration avec la société espagnole Tecnalia [23], nous avons redéfini des motifs dits *motifs contextuels* [10]. Depuis 2014, dans le cadre d'une collaboration avec le CHU de Nîmes, nous appliquons cette méthode sur les données du PMSI. L'objectif est de caractériser des trajectoires de patients pour la prise de décision médicale.

Ces dernières années, l'explosion du nombre des systèmes d'informations géographiques, due aux avancées technologiques en terme d'acquisition, a fait émerger de nouveaux défis en matière d'analyse de données. De nombreuses données séquentielles sont désormais associées à une information spatiale (e.g. images satellitaires, capteurs). Dans ce contexte, nous nous sommes intéressés à la dynamique d'évènements spatio-temporels, en collaboration avec l'Université de la Nouvelle Calédonie, la DASS, l'INVS et l'Institut Pasteur. Nous avons défini des motifs spatio-séquentiels qui prennent en compte les évolutions dans le temps et l'espace d'évènements et nous les avons utilisés pour monitorer les épidémies de dengue [2].

À une autre échelle, l'équipe s'est intéressée aux cellules rares dans le sang. La problématique est de déterminer parmi des millions de cellules, celles (5 à 10) correspondant au début d'une maladie ou d'une infection (Cancer, AVC). Dans le cadre du Labex Numev, nos travaux ont permis de définir une nouvelle approche de détection innovante d'outliers. Il est ainsi possible pour le médecin d'extraire un sous ensemble minimal de clusters dans lequel les cellules rares apparaissent. Un brevet international a été déposé et a également donné lieu à la création d'une entreprise soutenue par la SATT régionale et gérée par les médecins du LCCRH. Récemment, via des techniques de visual analytics, nous avons prouvé qu'il était possible de n'extraire que les vraies cellules rares. Ce résultat, limitant ainsi les faux positifs, offre de nouvelles perspectives thérapeutiques pour les médecins [13].

23. Tecnalia http://www.tecnalia.com/en/

Dans le domaine de la biomédecine, la terminologie est essentielle. Elle permet de décrire, échanger et acquérir des données, informations et connaissances. L'explosion du volume des données textuelles nécessite de recourir à une automatisation du processus d'extraction de la terminologie afin, par exemple, d'enrichir des ressources spécialisées (e.g. UMLS). Dans nos travaux, nous avons proposé des méthodes originales et multilingues (anglais, français, espagnol) de traitement automatique du langage naturel (TALN) qui combinent informations linguistiques et statistiques pour identifier des termes biomédicaux dans des documents textuels [14].

Afin de mieux appréhender les évolutions d'une maladie, nous nous sommes intéressés aux données échangées entre les patients dans les médias. En collaboration avec l'INVS, nous avons étudié les échos du H1N1 sur le web afin de mieux appréhender comment les nouvelles de maladies se propagent au sein des médias, à partir d'une classification de données adaptées à ce contexte et intégrant des techniques issues du traitement automatique du langage [3]. Dans ce cadre, nous avons proposé une nouvelle approche de cube de textes, i.e. pour pouvoir analyser les tendances, les pics de maladie, etc via la définition de nouvelles fonctions d'agrégation[5].

Dernièrement, nous nous sommes focalisés sur les productions des patients dans les médias sociaux (forums, Facebook, Twitter) dans le cadre du projet Patients Mind financé par la MSH. L'objectif est de proposer des méthodes qui permettent d'analyser les interactions au sein de ces médias selon différentes dimensions : les thématiques évoquées [9], la temporalité de la maladie, le locuteur selon son rôle, son expertise et la réputation des auteurs [1], les sentiments et opinions exprimés par les locuteurs à propos de leur maladie [4]. Ces méthodes sont désormais utilisées pour analyser la qualité de vie de patientes souffrant d'un cancer du sein (collaboration avec l'ICM et l'I3M) et détecter les personnes suicidaires (Montage d'un projet H2020).

Sélection de projets

- Parlons de nous (depuis 2013) Fouille semi automatique de fora de santé (https ://www.lirmm.fr/patient-mind/). Financement MSH-M en 2013 (10 k€) réseau Inter-MSH en 2014.
- SIFR (2013–2017). Indexation sémantique de ressources biomédicales francophones. Financement ANR.
- HIV (2012–2013). Fouille de données transcriptomiques (HIV).
- PRADNET (2010–2013). Primate Alzheimer's disease network.
- Collaboration avec Tecnalia (2008–2011). Motifs contextuels.
- GeneMining (2007–2008). Fouille de puces transcriptomiques (Alzheimer).

Références

[1] A. Abdaoui et al. "Predicting Medical Roles in Online Health Fora". In : *Statistical Language and Speech Processing - Second International Conference, SLSP 2014, Grenoble, France*. 2014, p. 247–258.

[2] H. Alatrista Salas et al. "The Pattern Next Door : Towards Spatio-sequential Pattern Discovery". In : *Advances in Knowledge Discovery and Data Mining*. T. 7302. 2012, p. 157–168.

[3] D. Breton et al. "Mining Web Data for Epidemiological Surveillance". In : *Emerging Trends in Knowledge Discovery and Data Mining - PAKDD 2012 International Workshops, Kuala Lumpur, Malaysia*. 2012, p. 11–21.

[4] S. Bringay et al. "Identifying the Targets of the Emotions Expressed in Health Forums". In : *Computational Linguistics and Intelligent Text Processing - 15th International Conference, CICLing 2014, Kathmandu, Nepal*. 2014, p. 85–97.

[5] S. Bringay et al. "Towards an On-Line Analysis of Tweets Processing". In : *Database and Expert Systems Applications - 22nd International Conference, DEXA 2011, Toulouse, France*. 2011, p. 154–161.

[6] S. H. ElSappagh et al. "Data Mining and Knowledge Discovery : Applications, Techniques, Challenges and Process Models in Healthcare". In : *International Journal of Engineering Research and Applications (IJERA)* 3(3) (2013), p. 900–906.

[7] M. Fabrègue et al. "Mining microarray data to predict the histological grade of a Breast Cancer". In : *Journal of Biomedical Informatics* (2011).

[8] M. Khajehei et F. Etemady. "Data Mining and Medical Research Studies". In : *IEEE Second Int. Conference on Computational Intelligence, Modeling and Simulation*. 2010, p. 119–122.

[9] T. Opitz et al. "Breast Cancer and Quality of Life : Medical Information Extraction from Health Forums". In : *Studies in Health Technology and Informatics, MIE 2014*. 2014, p. 1070–1074.

[10] J. Rabatel, S. Bringay et P. Poncelet. "Mining Representative Frequent Patterns in a Hierarchy of Contexts". In : *Advances in Intelligent Data Analysis XIII - 13th International Symposium, IDA 2014, Leuven, Belgium*. 2014, p. 239–250.

[11] A. Sallaberry et al. "Sequential patterns mining and gene sequence visualization to discover novelty from microarray data". In : *Journal of Biomedical Informatics* 44.5 (2011), p. 760–774. ISSN : 1532-0464. DOI : 10.1016/j.jbi.2011.04.002.

[12] P. Salle, S. Bringay et M. Teisseire. "Mining Discriminant Sequential Patterns for Aging Brain". In : *Artificial Intelligence in Medicine*. T. 5651. Lecture Notes in Computer Science. Springer Berlin / Heidelberg, 2009, p. 365–369. ISBN : 978-3-642-02975-2.

[13] E. Székely et al. "A Graph-based Method to Detect Rare Events : An Application to Identify Pathologic Cells". In : *IEEE Computer Graphics and Applications* (2014).

[14] J. A. Lossio Ventura et al. "Towards a Mixed Approach to Extract Biomedical Terms from Text Corpus". In : *International Journal of Knowledge Discovery in Bioinformatics* 4.1 (2014), p. 1–15.

LIRMM, équipe ICAR

L'équipe ICAR (Image & Interaction) développe des activités de recherche associant l'interaction et le traitement des données visuelles telles que les images, les vidéos et les objets 3D.

- CONTACT : William Puech, william.puech@lirmm.fr
- ADRESSE : LIRMM
  Campus St Priest - BAT 5
  860 rue de St Priest
  34095 Montpellier cedex 5
- WEB : http://www.lirmm.fr/icar/

Membres de l'équipe

- Marc Chaumont, MCF 27, HDR, Université de Nîmes, LIRMM
- Frédéric Comby, MCF 61, IUT de Béziers, LIRMM
- Vincent Creuze, MCF 61, IUT de Montpellier-Sète, LIRMM
- Benjamin Gilles, CR2, section 07, CNRS
- William Puech, Professeur 27, IUT de Béziers, LIRMM
- Nancy Rodriguez, MCF 27, IUT de Montpellier-Sète, LIRMM
- Olivier Strauss, MCF 61, HDR, Université de Montpellier, LIRMM
- Gérard Subsol, CR1, section 07, CNRS

Thème général de l'équipe

L'équipe ICAR développe son activité selon trois axes scientifiques associant image et interaction pour la manipulation de données visuelles tels que les images, les vidéos et les objets 3D : Analyse & Traitement (**AT**), Codage & Protection (**CP**) et Modélisation & Visualisation (**MV**). L'axe **AT** s'intéresse à de nouvelles techniques de traitement bas-niveau de l'information représentant, dans un même cadre théorique, l'imprécis, l'incertain et l'incomplet (types d'erreur en traitement des données). L'axe **CP** s'intéresse à la transmission et l'archivage sécurisés de données visuelles. Cette protection peut être assurée par tatouage ou chiffrement et doit être robuste à la compression. L'objectif de l'axe **MV** est de modéliser des grands ensembles de données complexes (en dimension et en nature) afin de permettre une visualisation intuitive ou de manipuler ces données pour en extraire des connaissances.

Description des travaux

Analyse et traitement

L'axe analyse et traitement d'images développe des techniques robustes de traitement du signal et de l'image. Son activité sur les 5 dernières années a concerné la vision omnidirectionnelle, la reconstruction d'images de tomographie d'émission et d'images super-résolues, l'estimation statistique et l'imagerie hyperspectrale et satellitaire. En tomographie d'émission, nous avons proposé une nouvelle méthode de reconstruction permettant d'estimer l'erreur due aux variations statistiques des mesures de projection. Cette méthode a été validée sur des fantômes physiques et numériques. Elle s'appuie sur une nouvelle théorie du traitement du signal, que nous avons mis au point, permettant de modéliser le fait que la réponse impulsionnelle d'un système de mesure est connue de façon imprécise. Nous avons montré que ce type de traitement du signal, appliqué aux images, se situe à mi-chemin entre le filtrage convolutif classique et la morphologie mathématique.

Codage et Protection

Le transfert, la visualisation et l'archivage de données visuelles sont des services numériques qui connaissent

une forte croissance depuis 10 ans. Le développement de ce type de services soulève un nombre conséquent de problèmes non résolus à ce jour. Un premier problème concerne la compression de ces données visuelles. En fonction des applications la compression pourra être plus ou moins importante, réversible ou non. Un deuxième problème concerne les aspects sécurité, englobant les problèmes de confidentialité, d'intégrité des données, de traçabilité mais aussi de correction d'erreurs et de robustesse aux attaques bienveillantes ou non.

Dans le cadre du développement des données médicales 3D, le codage et la protection deviennent un enjeu critique. L'équipe ICAR s'est intéressé à l'enrichissement des images et des maillages 3D de structures anatomiques [8] et mène actuellement des recherches sur l'identification du système d'acquisition à partir d'images scanner X [5].

Modélisation et Visualisation

Avec l'évolution des instruments de mesure et de calcul et la généralisation de la simulation numérique à un grand nombre de disciplines scientifiques, de grands ensembles de données à plusieurs dimensions et échelles sont fréquemment produits. Sur ces jeux de données, des analyses et des recherches doivent être effectuées. Et même si l'information est bien présente dans les données brutes, la trouver et l'exploiter devient une tâche ardue lorsque les données sont très nombreuses et complexes.

Les méthodes de modélisation permettent d'extraire des informations pertinentes, de représenter les données avec des formulations synthétiques génériques. Il devient alors possible de simuler des objets ou de phénomènes et de corréler les résultats à la réalité.

Un domaine privilégié est la modélisation tridimensionnelle automatisée et personnalisée de l'anatomie à partir de données hétérogènes (images médicales multimodales 3D, acquisitions surfaciques), dont les applications vont du diagnostic médical [3, 7] à la paléoanthropologie [6], en passant par la biomécanique [2, 4]. Ces travaux de recherche ont été effectués en collaboration avec la société IMAIOS [24] et le CHU de Montpellier.

Projets

L'équipe ICAR accompagne la création de la start-up NaturalPad [25] qui développe des jeux sérieux à but thérapeutiques en utilisant les interfaces grand public (Wii, Kinect, caméras). Au centre de cette collaboration se trouve l'extraction d'un squelette à partir d'un modèle 3D, avec le but de proposer des algorithmes permettant d'extraire un squelette à partir d'un modèle 3D provenant d'un dispositif de type Kinect [1]. En effet, la capture de mouvement longtemps utilisée pour l'animation de personnages, trouve actuellement une place importante en tant que moyen de contrôle d'un système informatique. En outre, le modèle morphologique de l'utilisateur qu'elle permet de récréer, permet des nombreuses utilisations en interaction et puis en analyse et évaluation de mouvements de l'utilisateur (utilisation en rééducation ou auprès des personnes avec de besoins spéciaux).

24. http://www.imaios.com/fr
25. http://naturalpad.fr/

Références

[1] W. J. Dyce et al. "Tabu search for human pose recognition". In : *Image Processing, Measurement (3DIPM), and Applications*. Fév. 2014.

[2] B. Gilles et al. "Frame-based Elastic Models". In : *ACM Transactions on Graphics* 30(2).15 (2011).

[3] E. Gioan, K. Sol et G. Subsol. "A combinatorial method for 3D landmark-based morphometry : application to the study of coronal craniosynostosis". In : *Proceedings of 15th International Conference on Medical Image Computing and Computer Assisted Intervention*. Springer, oct. 2012, p. 533–541.

[4] C. Herlin et al. "Generic 3D Geometrical and Mechanical Modeling of the Skin/Subcutaneous Complex by a Procedural Hybrid Method". In : *Proceedings of 6th International Symposium on Biomedical Simulation*. Springer, oct. 2014, p. 173–181.

[5] A. Kharboutly et al. "CT-Scanner Identification Based on Sensor Noise Analysis". In : *Proceedings of 5th European Workshop on Visual Image Processing (EUVIP)*. Déc. 2014.

[6] S. Prima et al. "Comparison of endocranial and ectocranial "symmetry planes" and application to the virtual reconstruction of hominid fossils". In : *American Journal of Physical Anthropology* 150.S56 (2013), p. 225.

[7] A. Quatrehomme et al. "Assessment of an Automatic System Classifying Hepatic Lesions on Multi-Phase Computer Tomography Images". In : *Proceedings of 21st European Signal Processing Conference*. Sept. 2013.

[8] N. Tournier et al. “3D Data Hiding for Enhancement and Indexation on Multimedia Medical Data”. In : *Proceedings of 2nd International Workshop on Medical Image Analysis and Description for Diagnosis Systems - MIAD 2011*. Jan. 2011, p. 43–51.

LIRMM : Projet SIFR (Indexation sémantique de ressources biomédicales francophones)

Le projet SIFR, principalement financé par le programme Jeunes Chercheurs de l'ANR, a pour objectif de résoudre les défis scientifiques et techniques pour exploiter les ontologies dans la construction de services d'indexation, de fouille, et de recherche de données pour les ressources biomédicales françaises. Dans ce cadre, le LIRMM collabore avec le NCBO (National Center for Biomedical Ontology) de l'Université de Stanford qui développe un portail d'ontologies biomédicales (BioPortal) ainsi qu'un workflow d'annotation sémantique qui sert de base au développement d'un service d'annotation basé sur les ontologies biomédicales francophones.

- CONTACT : Clement Jonquet (jonquet@lirmm.fr),
- ADRESSE : LIRMM, 161 rue Ada 34095 Montpellier Cdx 5
- WEB : http://www.lirmm.fr/sifr
- TEL : 04 67 14 97 43

Membres du projet

- Clement Jonquet, maître de conférences, Université de Montpellier (LIRMM)
- Sandra Bringay, maître de conférences, Université Paul Valéry (LIRMM)
- Mathieu Roche, chercheur, Cirad (TETIS)

Les partenaires suivants sont également associés au projet :

- Stanford BMIR, un leader mondial en outils et services — anglais — basés sur les ontologies pour aider la construction de systèmes à base de connaissances biomédicales ;
- Le groupe CISMeF, leader national en services de terminologies pour la santé en France.

Thème général du projet

Le volume de données en biomédecine ne cesse de croître. En dépit d'une large adoption de l'anglais en sciences, une quantité significative de ces données est en français. En général, le contenu textuel de ces ressources est indexé par mots-clefs pour permettre une recherche efficace mais avec des limites évidentes : synonymes, polysémie, utilisation des connaissances du domaine. L'intégration de données biomédicales et l'interopérabilité sémantique sont indispensables pour permettre de nouvelles découvertes scientifiques qui pourraient émerger du rapprochement des différentes données disponibles (i.e., « recherche translationnelle »). Les terminologies et les ontologies jouent un rôle central en sciences de la vie pour structurer les données médicales et les rendre interopérables. En particulier, la communauté les utilise pour créer des index sémantiques, destinés à améliorer la recherche et la fouille de données grâce aux connaissances médicales que ces ontologies formalisent. Cependant, outre l'existence de nombreuses ressources en anglais, il y a beaucoup moins d'ontologies en français et il manque crucialement d'outils et de services pour les exploiter. Cette lacune contraste avec le montant considérable de données biomédicales produites en français, particulièrement dans le monde clinique (e.g., dossiers médicaux électroniques).

SIFR a pour objectif d'offrir à la communauté biomédicale (e.g., cliniciens, professionnels de santé, chercheurs) des services d'indexation performants basés sur les ontologies leur permettant d'améliorer leur processus de production et de consommation de données. En particulier, nous construisons un workflow d'indexation basé sur les ontologies (i.e., SIFR Annotator) similaire à celui qui existe pour les ressources en anglais, mais spécialisé pour le Français. Ce workflow qui pourra être utilisé pour détecter des concepts d'ontologies ou de terminologies dans des données textuelles et utiliser la sémantique de ces dernières pour étendre et exploiter ces annotations.

Dans ce contexte, nous nous intéressons à divers sujet de recherche tels que l'extraction automatique de termes, le multilinguisme dans les ontologies, les représentations de connaissances sous formes de graphes de point-de-vues, ou la création de vocabulaire patient. Nous collaborons également avec diverses institutions telles que l'INRA, l'IRD, le CIRAD, ou Bioversity International ou entreprises comme Ontologos.

Description des travaux

Un des objectifs du projet est de développer un *workflow d'annotation sémantique de données textuelles francophones* et de l'offrir sous forme de service à la com-

munauté biomédicale française. Nous travaillons sur plusieurs prototypes depuis 2012 et une première version sera disponible courant 2015. Dans ce contexte, nous avons implémenté (entre autre) des fonctionnalités additionnelles pour le traitement des données des annotateurs (SIFR & NCBO), tels que le score et le classement des résultats, ou leur transformation en RDF [4, 5].

Un autre objectif du projet est de mener *différents axes de recherches* liées à l'utilisation d'ontologies biomédicales pour enrichir et alimenter le workflow et pour l'indexation sémantique de données. Nous organisons ce travail en plusieurs sous-axes, dont :

Extraction automatique de termes biomédicaux à partir de texte

Nous utilisons des méthodes existantes d'extraction (e.g., C-Value) ainsi que des méthodes d'indexation par mot clés (e.g., Okapi, Tf-Idf) généralement utilisées en recherche d'information. Nous combinons ces méthodes pour améliorer les performances. Des résultats expérimentaux très intéressants ont étaient obtenus sur des données francophones et anglophones (par soucis de comparaison) sur divers corpus [5, 3]. Nous avons également mis à disposition l'application BioTex (http://tubo.lirmm.fr/biotex) [3]. Plus récemment, nous regardons du côté de la désambiguation des termes extraits et de leur rattachement semi-automatique dans des ontologies existantes. Ces travaux sont systématiquement réalisés pour le français et l'anglais.

Détection d'émotions et de vocabulaire dans les forums de santé (au sein de WP2 et T3.1)

L'objectif a été de mettre en place une chaîne de traitements basée sur des techniques de fouille de textes permettant de détecter les émotions (joie, colère, tristesse...) dans les forums de santé. La difficulté a été d'identifier les bons descripteurs des émotions (smileys, mots accentués, mots appartenant à un lexique des émotions...) [6]. Une limite de ces travaux a été de ne pouvoir identifier les objets médicaux quand ils sont exprimés dans un vocabulaire propre au patient et que l'on ne retrouve pas dans les ontologies/terminologies médicales standards. Ainsi, nous nous interessons désormais à l'alignement des expressions utilisées par des patients et celles des professionnels de santé (crabe vs. cancer) pour construire une ressource terminologique venant étendre les ressources existantes.

Gestion du multilinguisme dans un portail d'ontologies

Nous nous intéressons à la gestion du multilinguisme dans la plateforme BioPortal (http://bioportal.bioontology.org). BioPortal permet d'accéder, visualiser, rechercher et commenter plus de 350 ontologies ou terminologies (principalement en anglais) de différent domaine en biologie ou médecine. Les ontologies peuvent être utilisées pour annoter automatiquement des données textuelles et le portail offre également un index sémantique de plusieurs jeux de données biomédicales annotées avec les ontologies du portail. Les utilisateurs ont accès à la plateforme soit via une application Web, soit via une interface de service Web. Nous avons spécifié les choix qui permettront au portail de gérer les ontologies multilingues et les traductions d'ontologies de manière consistante sémantiquement et transparente pour les utilisateurs [2]. Gérer le multilinguisme dans un portail d'ontologies ne se limite bien sûr pas à offrir l'interface graphique dans plusieurs langues. Il faut se poser les questions de la représentation multilingue des données du portail (ontologies, alignements) et de leur valorisation dans les services offerts (recherche, indexation, annotation). En complément, nous nous sommes nous intéresser à l'interopérabilité des plateformes de BMIR et CISMEF et avons prodduit une comparaison [1].

Développement d'un workflow d'annotation sémantique pour les plantes

En collaboration avec Pierre Larmande (IRD) et Patrick Valduriez (INRIA), au sein du WP5 sur l'intégration de données et de connaissances biologiques de l'Institut de Biologie Computationnelle (IBC) de Montpellier, nous développons un workflow d'annotation sémantique utilisant les ontologies spécifiques d'IBC (Plant Ontology, CROP ontology, trait ontology). L'objectif est l'étude de la génomique fonctionnelle chez les riz ou plus précisément, à l'impact de l'expression des gènes sur le développement (métabolisme, architecture, etc.). Nous avons développé un connecteur pour l'application WebSmatch de l'INRIA pour aider à identifier automatiquement dans des tables de données textuelles les entêtes de colonnes et leur affecter un terme défini dans une ontologie. Nous mettons a disposition une instance locale de BioPortal pour la communauté des plantes.

Indexation sémantique de données et de contributions utilisateurs – projet Viewpoint

En collaboration avec TETIS, nous avons lancé le « projet Viewpoint ». Viewpoints est un formalisme de

représentation des connaissances centré sur le point de vue individuel, humain ou artificiel et permettant d'indexer sémantiquement n'importe quel type de d'objets de connaissances (e.g., concept, document, idée, etc.) sous forme d'un graphe. Les viewpoints permettent de définir et de calculer une distance/proximité entre objets qui évolue au fil des interactions (requêtes et retours d'utilisation) et de l'ajout de nouveaux viewpoints (données et retours). Un prototype de moteur de recherche pour des données de publications scientifiques du LIRMM et du CIRAD a été produit et dans la suite, nous envisageons d'utiliser cette approche dans 2 scénarios : biomédical et agro-écologie.

Nous avons plusieurs prototypes (BioTex, Semantic Distance Services, French Annotator, Viewpoints API et Viewpoint Web Application) qui devront continuer à évoluer pour être mis à disposition de la communauté de façon pérenne. Ils sont disponibles sur la plateforme GitHub du projet : https://github.com/sifrproject

Références

[1] Julien Grosjean et al. "Comparing BioPortal and HeTOP : towards a unique biomedical ontology portal ?" In : *2nd International Work-Conference on Bioinformatics and Biomedical Engineering, IWBBIO'14*. Granada, Spain, 2014, p. 11. URL : http://www.lirmm.fr/~jonquet/publications/documents/Article_IWBBIO14_Comparing_BioPortal-HeTOP.pdf.

[2] Clement Jonquet et Mark A. Musen. "Gestion du multilinguisme dans un portail d'ontologies : étude de cas pour le NCBO BioPortal". In : *Terminology & Ontology : Theories and applications Workshop, TOTh'14*. Sous la dir. de C. Roche, R. Costa et E. Coudyzer. Brussels, Belgium, déc. 2014, p. 2.

[3] Juan Antonio Lossio-Ventura et al. "BIOTEX : A system for Biomedical Terminology Extraction, Ranking, and Validation". In : *13th International Semantic Web Conference, Demonstration, ISWC'14*. Sous la dir. de M. Horridge, M. Rospocher et J. Ossenbruggen. T. 1272. CEUR Workshop Proceedings. Riva del Garda, Italy, 2014, p. 157–160.

[4] Soumia Melzi et Clement Jonquet. "Representing NCBO Annotator results in standard RDF with the Annotation Ontology". In : *7th International Semantic Web Applications and Tools for Life Sciences - poster session, SWAT4LS'14*. Sous la dir. d'A. Paschke et al. T. 1320. CEUR Workshop Proceedings. CEUR-WS.org. Berlin, Germany, déc. 2014, p. 5.

[5] Soumia Melzi et Clement Jonquet. "Scoring semantic annotations returned by the NCBO Annotator". In : *7th International Semantic Web Applications and Tools for Life Sciences - poster session, SWAT4LS'14*. Sous la dir. d'A. Paschke et al. T. 1320. CEUR Workshop Proceedings. CEUR-WS.org. Berlin, Germany, déc. 2014, p. 15.

[6] Soumia Melzi et al. "Que ressentent les patients ?" In : *14èmes Journées Francophones Extraction et Gestion des Connaissances, EGC'14*. Sous la dir. de C. Reynaud, A. Martin et R. Quiniou. Rennes, France : RNTI, Hermann, 2014, p. 449–454.

LIRMM, Équipe TEXTE : Projet IMAIOS

Exploration et exploitation de données textuelles. Un des buts de l'équipe est de développer des modèles et des outils pour le traitement de la langue. Nous traitons de trois grands domaines théoriques du traitement automatique des langues :

- la syntaxe
- la sémantique (sémantique vectorielle, réseau lexicaux)
- la logique
- CONTACT : Mathieu Lafourcade — mathieu.lafourcade@lirmm.fr
- ADRESSE :161, rue Ada 34095 Montpellier
- WEB : http://www.lirmm.fr/recherche/equipes/texte

Membres de l'équipe

- Mathieu Lafourcade, Maître de conférences de l'université de Montpellier
- Christian Rétoré, Professeur de l'université de Montpellier
- Alain Joubert, Maître de conférences de l'université de Montpellier
- Jean-Philippe Prost, Maître de conférences de l'université de Montpellier
- Violaine Prince, Professeur de l'université de Montpellier
- Richard Moot, Chargé de recherche CNRS
- Anne Preller, Professeur associée émérite
- Stergios Chatzikyriakidis, Ingénieur de recherche
- Guillaume Tisserant, ATER

- Manel Zarrouk, Doctorante
- Lionel Ramadier, Doctorant

Thème général de l'équipe

Notre objectif est de développer des modèles et des outils pour le traitement de la langue. Une des thématiques de l'équipe est la construction de bases lexicales (projet JeuxDeMots). Cette base prend la forme d'un réseau lexico-sémantique et couvre non seulement les connaissances générales relevant du sens commun, mais aussi des connaissances de spécialités (médecine, radiologie). Nous faisons l'hypothèse de non séparation entre les différents domaines de connaissances, c'est-à-dire qu'il est intéressant d'utiliser les deux types de connaissance pour des textes relevant d'un domaine de spécialité. Ce réseau lexical nous servira pour l'analyse des comptes rendus radiologiques en vue de réaliser un index contenant les termes pertinents, mais également augmenté avec les concepts connexes.

Description des travaux

Le projet IMAIOS-CRIM s'inscrit dans le cadre d'une collaboration avec la société IMAIOS qui est une société spécialisée dans le e-learning médical et plus spécifiquement dans l'imagerie médicale. Dans le cadre d'une thèse CIFRE, nous développons un projet d'indexation automatique conceptuelle des comptes rendus en vue d'une extraction sémantique efficace des comptes rendus et d'images radiologiques.

Contexte

A l'heure actuelle l'informatisation des données médicales est en plein essor (création du dossier personnalisé, informatisation des différents comptes rendus médicaux, utilisation du PACS (Picture Archiving and Communication System au niveau de la radiologie), etc). La majorité du temps, ces données sont écrites sous forme non structurée. Pour l'amélioration de la prise en charge des patients, de la communication entre praticiens, pour l'aide au diagnostic ainsi que dans un but d'enseignement, il parait nécessaire de structurer ces documents afin d'en extraire les informations pertinentes. On peut se servir pour cela des méthodes issues du traitement automatique du langage (TAL). Dans un contexte d'optimisation de la qualité des soins, mais également d'augmentation de la demande d'examens, la qualité et la fiabilité de l'interprétation de l'imagerie médicale est un enjeu capital. Pour ces raisons, l'aide au diagnostic médical informatisée, automatisée, apparaît comme l'outil d'avenir pour les médecins dans un objectif d'amélioration du service médical rendu. L'objectif du projet est de réaliser un système d'extraction automatique de données sémantiques à partir de comptes rendus radiologiques.

Travaux effectués

Constitution de connaissances spécialisées (radiologiques) incluses dans un réseau lexico-sémantiques de connaissances générales (JeuxDeMots). Dans le domaine de l'imagerie médicale la principale ressource existante est Radlex [3] qui a été plus ou moins traduite en allemand [4]) et en français [26]. Dans le domaine de la radiologie, il est intéressant d'extraire non seulement des termes pertinents à partir des comptes rendus mais aussi des relations pertinentes entre les termes. Les ontologies classiques (type Radlex) ne capturent pas ce type d'information (relations entre termes) aussi bien qu'un réseau sémantique. Il peut être intéressant pour le médecin de disposer en supplément des relations taxonomiques habituelles (relation is_a) de relations non hiérarchiques. Il semble pertinent de donner pour une maladie donnée la liste des symptômes, les cibles potentielles, la localisation anatomique. Ceci peut être modélisé par un réseau sémantique. L'association entre un réseau sémantique général et spécialisé peut jouer un rôle important dans l'analyse globale des rapports radiologiques.

En effet, la section indication d'un compte rendu radiologique peut contenir des informations d'ordre général (âge, circonstance de l'accident, etc) d'où l'idée du projet d'ajouter dans un réseau sémantique *général* des connaissances spécialisées radiologiques. Un tel réseau permettra une analyse des comptes rendus dans leur totalité et d'en extraire les termes et les relations pertinentes. Nous avons utilisé comme réseau lexico-sémantique JeuxDeMots [27] [2]. Ce dernier est un *game with purpose* (GWAP) [1]. En dehors du jeu nous utilisons un outil contributif proposé par JDM : *Diko* (figure 3). Le principe du processus de la contribution est qu'une proposition faite par un expert en radiologie sera soumise aux votes d'autres experts en imagerie médicale pour un processus de validation/invalidation.

26. http://www.hetop.eu/hetop
27. http://www.jeuxdemots.org/

FIGURE 3 – L'outil contributif Diko de JDM

La construction de cette ressource a été réalisée à partir d'un corpus de 35 000 comptes rendu radiologiques. Nous avons utilisé un algorithme d'index inversé quelque peu modifié pour pouvoir capturer les multitermes. Pour améliorer la qualité du réseau nous avons utilisé d'autres sources comme par exemple Wikipedia ou encore certaines terminologies (Terminologia Anatomica [28]). Pour améliorer la qualité du réseau nous avons développé des mécanismes d'inférence [6] ce qui permet de compléter le réseau de façon semi-automatique. Nous avons, dans le cadre de notre projet, annoté les relations [5] car, surtout dans le domaine des connaissances spécialisées, la corrélation entre la force d'association de la relation et son importance conceptuelle n'est pas toujours assurée. Ce type d'annotation pourra servir, par exemple dans le cadre d'un moteur de recherche pour classifier les documents selon certains critères.

Indexation à partir de comptes rendus et augmentation d'index. La seconde partie du projet concerne l'analyse la plus précise possible des comptes rendus. Nous avons réalisé un index des termes pertinents, c'est-à-dire ceux susceptibles de faire l'objet de requêtes de la part de médecins. À partir de cet index nous avons réalisé un index augmenté à l'aide du réseau JDM. Ce dernier contient des termes pertinents du réseau lexical en rapport avec ceux du compte rendu. Cet index augmenté permet d'avoir accès à des documents où le terme de la requête n apparaît pas mais est relié aux termes du document par certaines relations présentes dans le réseau. Cet index augmenté permettra d'améliorer la recherche d'information dans un cadre hospitalier par exemple.

Améliorations

Une des améliorations possibles du projet est d'essayer d'extraire des relations déjà présentes dans le réseau ainsi que de nouvelles relations non présentes. Nous pourrons essayer de déduire des règles à partir d'un grand nombre de comptes rendus (par exemple *$x r_ isa maladie & $y r_ associated IRM & $x r_ diagnostique $y*), en vue peut-être de découvrir des connaissances ou pour une aide au diagnostic. Ce type de règles pourra être déduit par des algorithmes d'exploration du réseau lexical.

Projets

Projet avec la société IMAIOS d'indexation des comptes rendus en vue d'un moteur de recherche.

Références

[1] Luis von AHN. "Games With A Purpose". In : (2006).

[2] Mathieu LAFOURCADE. "Making people play for Lexical Acquisition with the JeuxDeMots prototype". In : *SNLP'07 : 7th International Symposium on Natural Language Processing*. 2007, p. 7.

[3] Curtis P LANGLOTZ. "RadLex : A new method for indexing online educational materials 1". In : *Radiographics* 26.6 (2006), p. 1595–1597.

[4] D MARWEDE et al. "[RadLex-German version : a radiological lexicon for indexing image and report information]." In : *RoFo : Fortschritte auf dem Gebiete der Rontgenstrahlen und der Nuklearmedizin* 181.1 (2009), p. 38–44.

[5] Lionel RAMADIER et al. "Spreading Relation Annotations in a Lexical Semantic Network Applied to Radiology". In : *Computational Linguistics and Intelligent Text Processing*. Springer, 2014, p. 40–51.

28. http://www.unifr.ch/ifaa/

[6] Manel Zarrouk, Mathieu Lafourcade et Alain Joubert. "About Inferences in a Crowdsourced Lexical-Semantic Network". In : *Proceedings of the 14th Conference of the European Chapter of the Association for Computational Linguistics*. Gothenburg, Sweden : Association for Computational Linguistics, avr. 2014, p. 174–182. URL : http://www.aclweb.org/anthology/E14-1019.

LRI, Équipe LaHDAK : Projets DYNAMO et Hybris

Le Laboratoire de Recherche en Informatique est une unité mixte de recherche de l'Université Paris-Sud et du CNRS.

- CONTACT : Chantal Reynaud : Responsable d'équipe
- ADRESSE : LRI, Bâtiment 650 Ada Lovelace, Université Paris-Sud, 91405 Orsay cedex
- Tél : 01 69 15 68 37
- Email : chantal.reynaud@lri.fr

Membres de l'équipe

- Chantal Reynaud, Professeur, Université Paris-Sud
- Yue Ma, Maître de conférences, Université Paris-Sud
- Julio Cesar Dos Reis, doctorant

Thème général de l'équipe

L'équipe LaHDAK (Large-scale Heterogeneous Data and Knowledge) rassemble des chercheurs en Intelligence Artificielle et en bases de Données. Son objectif est de gérer des données complexes, sémantiquement hétérogènes, incertaines, incomplètes et évolutives, provenant de sources variées, dont Internet et le Web social. Ses travaux sont plus particulièrement focalisés sur les problèmes d'efficacité, de pertinence sémantique et de robustesse, ainsi que sur l'adéquation des infrastructures de gestion de données et de connaissances distribuées aux problèmes d'interopérabilité et de passage à l'échelle.

Certains des projets de l'équipe portent sur le domaine de la santé. C'est le cas du projet DYNAMO sur lequel l'équipe a travaillé en partenariat avec l'Institut Henri Tudor du Luxembourg (qui a rejoint aujourd'hui le LIST) et du projet Hybris-B1 sur lequel l'équipe a travaillé avec la Faculté Informatics and Biotechnology Center TU de Dresde en Allemagne.

Projets

Projet DYNAMO (2011-2015) Les technologies du Web sémantique, en particulier les ontologies ou plus largement les systèmes d'organisation de la connaissance (Knowledge Organization Systems, ou KOS en anglais), améliorent les performances des systèmes exploitant des sources d'information multiples. Lorsque plusieurs KOSs sont nécessaires, il faut pouvoir les faire cohabiter en les alignant via des mises en correspondance, appelées aussi mappings. Le projet DYNAMO a porté sur le problème de l'évolution des mappings entre KOSs dynamiques du domaine biomédical.

En présence de ressources sémantiques volumineuses et évoluant très vite, comme c'est le cas dans le domaine médical, des méthodes automatiques de maintenance des mappings sont indispensables. Nous avons proposé une approche originale pour adapter les mappings basée sur les changements détectés dans l'évolution des KOSs du domaine. Notre proposition consiste à comprendre précisément les mappings entre les ressources, à suivre l'évolution des éléments des ontologies via des patrons de changement spécifiques, puis à proposer l'application de techniques d'adaptation à base d'heuristiques.

Nos contributions dans le cadre de ce projet sont les suivantes :

- Nous avons réalisé un travail expérimental très approfondi pour comprendre l'évolution des mappings entre KOSs ; Nous avons proposé des méthodes automatiques pour analyser les mappings affectés par l'évolution de KOSs ;
- Nous avons proposé des méthodes pour reconnaître l'évolution des concepts impliqués dans les mappings via des patrons de changements spécifiques ;
- Nous avons conçu des techniques d'adaptation des mappings à base d'heuristiques ;
- Enfin, nous avons proposé un cadre complet pour l'adaptation des mappings, appelé DyKOSMap, et un prototype logiciel.

Les méthodes proposées et le cadre formel ont été toutes évalués avec des jeux de données réelles (SNOMED CT, ICD-9-CM, NCI, MeSH) pour lesquels nous disposions de plusieurs versions de mappings.

Les résultats des expérimentations réalisées ont démontré l'efficacité des principes sous-jacents à l'approche proposée. La maintenance des mappings, en grande partie automatique, est de bonne qualité.

Ce travail a fait l'objet de la thèse de Julio Cesar Dos Reis, encadrée par Chantal Reynaud (LRI, Université Paris-Sud et CNRS) et Cédric Pruski (LIST-Luxembourg)

Projet Hybris (2012-2015) Une représentation formelle est souvent un bon moyen pour exprimer des connaissances d'un domaine qui, ainsi, peuvent faire l'objet de raisonnements automatiques pour déduire des informations implicites. Cela est particulièrement important lorsque les connaissances sont volumineuses, comme c'est le cas, par exemple, dans le domaine médical. Malheureusement, la création manuelle d'ontologies formelles est une tâche complexe demandant beaucoup de temps et d'efforts. L'objectif du projet Hybris-B1 a porté sur la génération de telles ontologies du domaine biomédical à partir de textes. Les ontologies générées par les approches existantes ne sont souvent pas très formelles. Les notions acquises ne sont pas définies formellement. Dans le projet Hybris-B1, nous développons une approche de construction d'ontologies représentées en logique de description. L'idée est d'appliquer des déductions logiques non-standards sur les résultats des méthodes d'apprentissage de l'ontologie afin de générer des définitions de concepts et des contraintes (par ex. des inclusions de concepts).

Nos contributions dans le cadre de ce projet sont les suivantes :

- Nous avons proposé une méthode d'apprentissage pour enrichir l'ontologie SNOMED CT. Cette méthode n'a pas besoin de données d'entraînement. Elle exploite la grande quantité de connaissances formelles existant dans SNOMED CT.
- Nous avons développé une méthodologie d'évaluation en profondeur de l'approche proposée appliquée à des données textuelles et des ontologies de différentes sortes.
- Nous avons développé des raisonnements non-standards pour sélectionner automatiquement les définitions formelles les plus adaptées parmi un ensemble de définitions candidates générées par l'application de techniques d'apprentissage automatique.

Les expérimentations ont permis d'atteindre un taux de réussite de plus de 90 % sur des données réelles. Le raisonnement non-standard développé a encore amélioré ce taux en le portant à 98 % dans le cadre de la construction de SNOMED CT. Cela montre que les méthodes hybrides proposées contribuent de façon très bénéfique à la formalisation des connaissances biomédicales.

Références

[1] Long Cheng et Yue Ma. "Investigating Distributed Approaches to Efficiently Extract Textual Evidences for Biomedical Ontologies". In : *Proc. of 14th IEEE International Conference on Bioinformatics and BioEngineering (BIBE'14)*. Nov. 2014, p. 220–225. DOI : 10.1109/BIBE.2014.45.

[2] Duy Dinh et al. "On the identification of Conceptual Information for Supporting Mapping Maintenance under Evolving Life Science Ontologies". In : *Journal of Web Semantics* (à paraître).

[3] Júlio Cesar Dos Reis et al. "Recognizing lexical and semantic change patterns in evolving life science ontologies to inform mapping adaptation". In : *Artificial Intelligence in Medicine* 63.3 (mar. 2015), p. 153–170.

[4] Yue Ma et Felix Distel. "Concept adjustment for description logics". In : *Proceedings of the 7th International Conference on Knowledge Capture, K-CAP 2013, Banff, Canada, June 23-26, 2013*. Sous la dir. de V. Richard Benjamins, Mathieu d'Aquin et Andrew Gordon. ACM, 2013, p. 65–72. ISBN : 978-1-4503-2102-0. DOI : 10.1145/2479832.2479851. URL : http://doi.acm.org/10.1145/2479832.2479851.

[5] Yue Ma et Felix Distel. "Learning Formal Definitions for SNOMED CT from Text". In : *Artificial Intelligence in Medicine - 14th Conference on Artificial Intelligence in Medicine, AIME 2013, Murcia, Spain, May 29 - June 1, 2013. Proceedings*. Sous la dir. de Niels Peek, Roque Marín Morales et Mor Peleg. T. 7885. Lecture Notes in Computer Science. Springer, 2013, p. 73–77. ISBN : 978-3-642-38325-0. DOI : 10.1007/978-3-642-38326-7_11. URL : http://dx.doi.org/10.1007/978-3-642-38326-7_11.

[6] Alina Petrova et al. "Formalizing biomedical concepts from textual definition". In : *Journal of Biomedical Semantics* (2015), p. 24. URL : https://hal.archives-ouvertes.fr/hal-01138987.

[7] Júlio Cesar dos Reis, Cédric Pruski et Chantal Reynaud-Delaître. "State-of-the-art on mapping maintenance and challenges towards a fully automatic approach". In : *Expert Systems with Applications* 42.3 (2015), p. 1465–1478. DOI : 10.1016/j.eswa.2014.08.047. URL : http://dx.doi.org/10.1016/j.eswa.2014.08.047.

[8] Júlio Cesar dos Reis et al. "Understanding semantic mapping evolution by observing changes in biomedical ontologies". In : *Journal of Biomedical Informatics* 47 (2014), p. 71–82. DOI : 10.1016/j.jbi.2013.09.006. URL : http://dx.doi.org/10.1016/j.jbi.2013.09.006.

LITIS, équipe Traitement de l'Information en Biologie Santé (TIBS)

Bioinformatique et informatique médicale

- CONTACT : Thierry.Lecroq@univ-rouen.fr
- ADRESSE : LITIS EA 4108, UFR des Sciences et Techniques, Université de Rouen, 76821 Mont-Saint-Aignan Cedex
- WEB : http://www.chu-rouen.fr/tibs/
- TÉL : +33 (0)2 35 14 65 81

Membres de l'équipe

- Saïd Abdeddaïm, Maître de conférences 27 de l'université de Rouen
- Caroline Bérard, Maître de conférences 26 de l'université de Rouen
- Badisse Dahamna, Ingénieur au CHU de Rouen
- Stéfan J. Darmoni, Professeur des Universités Praticien Hospitalier 46-04 de l'université de Rouen
- Hélène Dauchel, Maître de conférences 64 de l'université de Rouen
- Jean-François Gehanno, Professeur des Universités Praticien Hospitalier 46-02 de l'université de Rouen
- Yannick Guesnet, Maître de conférences 27 de l'université de Rouen
- Gaëtan Kerdelhué, Documentaliste au CHU de Rouen
- Ivan Kergourlay, Ingénieur au CHU de Rouen
- Catherine Letord, Pharmacienne, documentaliste au CHU de Rouen
- Thierry Lecroq, Professeur de l'université de Rouen
- Arnaud Lefebvre, Maître de conférences 27 de l'université de Rouen
- Martine Léonard, Maître de conférences 27 de l'université de Rouen
- Laurent Mouchard, Maître de conférences 27 de l'université de Rouen
- Élise Prieur-Gaston, Maître de conférences 27 de l'université de Rouen
- Lina F. Soualmia, Maître de conférences 27 de l'université de Rouen
- Benoit Thirion, Conservateur de la bibliothèque du CHU de Rouen

Thème général de l'équipe

Le Laboratoire d'Informatique, de Traitement de l'Information et des Systèmes (LITIS) est l'unité de recherche en sciences et technologies de l'information de Haute-Normandie. Il regroupe les enseignants chercheurs du domaine des STIC des trois principaux établissements d'enseignement supérieur publics de la région : l'Université de Rouen (UR), l'Université du Havre (ULH) et l'Institut National des Sciences Appliquées de Rouen (INSA). Les travaux de l'équipe TIBS consistent à développer des méthodes et des outils pour rechercher, indexer et extraire des informations pertinentes dans des données biologiques (génomes et expression des génomes) et des systèmes d'information en santé (SIS). Ces SIS peuvent être utilisées dans plusieurs contextes : l'accès à la connaissance et la gestion intelligente du dossier du patient informatisé. Les solutions proposées par les travaux de recherche de l'équipe sont avant tout pluridisciplinaires en mettant en interaction informaticiens, médecins (cliniciens et médecins de santé publique), biologistes, spécialistes des sciences de l'information et statisticiens. L'équipe utilise également beaucoup de personnes bicompétentes (médecine et informatique, sciences de l'information et professionnels de la santé par exemple). L'objectif global consiste à contribuer à faire émerger la médecine personnalisée (ou individualisée) en intégrant dans les systèmes d'information en santé à la fois des données phénotypiques et des données omiques (génomique, protéomique, transcriptomique, métabolomique) (thèse de Chloé Cabot et projet RAVEL), en les indexant et en les recherchant efficacement.

Les thématiques de recherche de l'équipe se regroupent dans la discipline « informatique biomédicale » (réunissant bioinformatique et informatique médicale). Ces thèmes sont principalement : l'ingénierie des connaissances de santé, dont indexation automatique et recherche d'informations multiterminologies ; l'interopérabilité intra et inter terminologique ; la bibliométrie ; la génomique ; l'algorithmique du texte ; les biostatistiques.

Description des travaux

Concernant l'ingénierie des connaissances, l'équipe travaille :

- sur la modélisation des terminologies, des ontologies, mais aussi des lexiques et des dictionnaires. Une des tâches des prochaines années sera de réunir dans une plateforme unique des ressources termino-ontologiques (RTO) et des lexiques et dictionnaires, en s'appuyant sur des standards comme Terminological Markup Framework (TMF ; URL :

http://www.loria.fr/projets/TMF/) et Lexical Markup Framework (LMF ; URL : http://www.lexicalmarkupframework.org/). Il faudra dans les années à venir modéliser l'interopérabilité syntaxique et sémantique entre lexiques et RTO. Une coopération a débuté avec le consortium MERITERM (URL : http://www.meriterm.org/), et avec le laboratoire LIMICS U1142, où deux membres de l'équipe (SJD & LFS) sont membres associés.

- sur l'indexation automatique, en se fondant sur plusieurs approches hybrides : à la fois sur le traitement automatique de la langue (TAL), mais aussi une approche conceptuelle (en nous fondant sur les plus de 50 RTO incluses dans notre portail terminologiques, et les plus de 317000 concepts différents accessibles en français). Le projet TOLBIAC associe également des méthodes statistiques pour créer des matrices de co-occurrences, par exemple entre maladies et actes médicaux.
- sur la recherche d'information, en tentant d'intégrer autant que possible la modélisation complexe d'un dossier patient informatisé et en produisant un langage de requêtes certes complexe pour l'utilisateur final, mais robuste et puissant, capable de gérer les données structurées et non structurées (après indexation automatique et création de métadonnées), de gérer des données numériques et chronologiques et enfin des données omiques plus récentes (génomiques, métabolimiques, protéomiques, etc).

Concernant la bioinformatique, l'équipe est spécialisée dans l'algorithmique des séquences. Elle développe de nouvelles méthodes d'indexation et de recherches de motifs. Du côté applications, elle collabore depuis quelques années avec des équipes de biologie et de médecine locales pour développer des outils dédiés à l'analyse des données issues des séquenceurs à haut débit.

Les principales contributions de l'équipe constitue une plateforme technologique, contenant plusieurs services, qui ont pour vocation à intégrer des outils utilisés en santé et en biologie :

- CISMeF (Catalogue et Index des Sites Médicaux en langue Française) http://www.cismef.org : moteur de recherche sémantique multi-terminologique multi-lingue, avec des instanciations diverses comme :
  - — dans le projet RAVEL, avec la recherche d'information dans le dossier du patient informatisé (DPI) ;
  - — dans le projet BDBfr, avec la nécessité d'avoir un moteur de recherche générique, capable d'interroger n'importe quel type de document (ressource Web -CISMeF-, citations d'articles -BDBfr-, DPI -RAVEL-).
- Portail terminologique de santé interlingue (URL : www.hetop.eu) permettant une double navigation entre terminologie et ontologies et entre langues ; ce portail a été récemment reconnu comme potentielle plateforme européenne par certains experts (Pr. Robert Vander Stichele, Université de Gand, Belgique) dans le cadre du réseau d'excellence SemanticHealthNet (URL : http://www.semantichealthnet.eu/). Il existe dans le domaine de la santé pratiquement autant de terminologies que de domaines d'application. D'où la nécessité de rendre ces terminologies interopérables c'est-à-dire rechercher un référentiel sémantique commun permettant l'interaction efficace. L'équipe met en œuvre en grandeur réelle les outils développés au LITIS dans le cadre des recherches sur cette question. À noter que ce portail génère plus de 108 millions de triplets RDF, ce qui nous fait clairement entrer dans le *big data* en santé.
- PlaIR (Plateforme d'Indexation régionale), http://www.plair.org, abordant l'aspect multi-disciplines (sciences de l'ingénieur, droits du transport).
- EVA (Exome Variation Analyser) http://plateforme-genomique-irib.univ-rouen.fr/EVA/, outil d'analyse de variations d'exomes.
- InfoRoute (http://inforoute.chu-rouen.fr), un outil de connaissance contextuelle.
- ECMT (http://ecmt.chu-rouen.fr), extracteur de concepts multi-terminologique, en cours de valorisation industrielle.
- Outil d'alignement automatique de termino-ontologies.
- HExoSplice (http://bioinfo.univ-rouen.fr/hexosplice/), outil de prédiction de variants exotiques.
- FunEVA (http://bioinfo.univ-rouen.fr/funeva/), outil de hiérarchisation de variations génétiques.

Projets

TIBS, via principalement le service d'informatique biomédicale du CHU de Rouen, est actuellement impliqué dans treize contrats de recherche actifs (6 ANR programme TecSan (Technologies pour la Santé), 1 FUI, 2 PREPS, 2 INCa, 2 Conseil Régional de Haute-Normandie). Quelques uns de ces projets sont listés dans la suite.

ANR programme TecSan

- BDBfr Création d'une base de données bibliographique/bibliométrique des principaux contenus scientifiques de santé en français (journaux scientifiques, encyclopédie, livres) disponible gratuitement sur l'Internet (2014–2017) ;
- TOLBIAC Terminologies et ontologies pour relier l'information de facturation à des données cliniques exactes (2013–2016) ;
- TerSAN Terminologes d'Interface en Santé (2012–2015) ;
- SYNODOS SYstème de Normalisation et d'Organisation de Données médicales textuelles pour l'Observation en Santé (2012–2015) ;
- RAVEL Recherche et Visualisation des informations dans le dossier patient électronique (2012–2015) ;
- SIFADO Conception et évaluation de méthodes et d'outils ergonomiques pour faciliter la saisie et le codage de donnés textuelles et graphiques dans les dossiers médicaux électroniques (2012–2015).

Conseil régional Haute-Normandie

- PlaIR II (Plateforme d'Indexation Régionale) (2013–2015) ;
- STIC & Cancer (2015).

PREPS (Ministère de la Santé)

- *Evaluation of several quality indicators inside hospital information system* (PREPS program 2012 ; 2013-2017) ;
- *MATRIX DPRS : Randomised technico-clinical trial to measure the clinical impact of the electronic pharmaceutical record (DP) in French hospitals* (PREPS program 2013 ; 2014–2015).

FUI

- ADR-PRISM, consortium public-privé : détection de signaux faibles dans les forums de patients pour détecter de nouveaux effets secondaires (pharmacovigilance) (2014–2016).

Références

[1] C. Barton et al. "Linear-time computation of minimal absent words using suffix array". In : *BMC Bioinformatics* 15.1 (2014), p. 388.

[2] W. Chebil et al. "Indexing Biomedical Documents with a Possibilistic Network". In : *Journal of the Association for Information Science and Technology* (2015). doi : 10.1002/asi.23435.

[3] S. Coutant et al. "EVA : Exome Variation Analyzer, an efficient and versatile tool for filtering strategies in medical genomics". In : *BMC Bioinformatics* 13.Suppl. 14 (2012), S9.

[4] S.J. Darmoni et al. "Improving information retrieval using Medical Subject Headings Concepts : a test case on rare and chronic diseases". In : *Journal of the Medical Library Association* 100.3 (2012), p. 176–183.

[5] S Faro et T Lecroq. "The Exact Online String Matching Problem : a Review of the Most Recent Results". In : *ACM Computing Surveys* 45.2 (2013), p. 13.

[6] J.-F. Gehanno, L. Rollin et S Darmoni. "Is the coverage of Google Scholar enough to be used alone for systematic reviews". In : *BMC Medical Informatics and Decision Making* 13 (2013), p. 7.

[7] C. Golbreich, J. Grosjean et S.J. Darmoni. "The Foundational Model of Anatomy in OWL 2 and its use". In : *Artificial Intelligence in Medicine* 57.2 (2013), p. 119–132.

[8] N. Griffon et al. "A Search Engine to Access PubMed Monolingual Subsets : Proof of Concept and Evaluation in French". In : *Journal of Medecial Internet Research* 16.12 (2014), e271.

[9] N. Griffon et al. "An interface terminology for medical imaging ordering purposes". In : *AMIA Annual Symposium proceedings*. 2012, p. 1237–1243.

[10] N. Griffon et al. "Design and usability study of an iconic user interface to ease information retrieval of medical guidelines". In : *Journal of the American Medical Informatics Association* 21.e2 (2014), e270–e277.

[11] N. Griffon et al. "Evaluating alignment quality between iconic language and reference terminologies using similarity metrics". In : *BMC Medical Informatics and Decision Making* 14 (2014), p. 17.

[12] N. Griffon et al. “Preservation of information in terminology transcoding”. In : *Studies in Health Technology and Informatics* 205 (2014), p. 156–160.

[13] J. Grosjean et al. “An approach to compare bio-ontologies portals”. In : *Studies in Health Technology and Informatics* 205 (2014), p. 1008–1012.

[14] J. Grosjean et al. “Teaching medicine with a terminology/ontology portal”. In : *Studies in Health Technology and Informatics* 180 (2012), p. 949–953.

[15] M. Jung et al. “Attitude of physicians towards automatic alerting in computerized physician order entry systems. A comparative international survey”. In : *Methods of Information in Medicine* 52.2 (2013), p. 99–108.

[16] T. Merabti et al. “Assisting the translation of SNOMED CT into French”. In : *Studies in Health Technology and Informatics* 192 (2013), p. 47–51.

[17] L.F. Soualmia et al. “Matching health information seekers’ queries to medical terms”. In : *BMC Bioinformatics* 13.Suppl 14 (2012), S11.

Compte rendu de la journée EIAH & IA

Amélie Cordier (Université de Lyon, LIRIS)

Les domaines des Environnements Informatiques pour l'Apprentissage Humain (EIAH) et de l'Intelligence Artificielle (IA) abordent tout un ensemble de problématiques qui les amènent à croiser leurs approches. Dans cette optique, une journée commune a été organisée à l'initiative de l'AFIA (Association Française pour l'Intelligence Artificielle) et l'ATIEF (Association des Technologies de l'Information pour l'Education et la Formation).

Date : 28 mai 2013
Lieu : Université de Toulouse
Nombre de participants : 86

EIAH : qu'est-ce que c'est ?

Le terme EIAH (Environnement Informatique pour l'Apprentissage Humain) désigne à la fois un logiciel (ou tout autre solution informatique) destiné à favoriser l'apprentissage humain et le champ de recherche concernant ces solutions. Les grands axes de recherche actuels dans ce domaine sont les suivants :

- Web et éducation, MOOC (Massive Open Online Courses) : création de plateformes ouvertes d'apprentissage de masse avec un parcours personnalisé pour chaque apprenant et débouchant sur une éventuelle certification.
- Apprentissage mobile, ubiquitaire et continu : apprentissage tout au long de la vie, dans n'importe quel lieu et n'importe quelle situation.
- Scénarisation pédagogique : orchestration de l'activité d'apprentissage à travers les acteurs, leurs rôles, leurs interactions et les outils.
- Analyse des activités d'apprentissage, diagnostic automatique : analyse automatique des productions réalisées par les apprenants de manière à évaluer leur niveau et à détecter et comprendre leurs erreurs.
- Personnalisation, adaptation de l'apprentissage : à partir d'un profil de l'apprenant, déterminer le parcours pédagogique (en adaptant le choix et/ou le contenu des activités) qui lui est le plus adapté afin de favoriser ses apprentissages.
- Jeux sérieux : utilisation des motivations liées aux jeux pour favoriser l'apprentissage. L'apprenant apprend alors qu'il joue.
- Outils auteurs : ensemble des outils servant à créer les contenus et activités pédagogiques, l'objectif étant toujours de faciliter la tâche de l'auteur et d'augmenter ses possibilités.
- Analyse des usages : analyse de l'utilisation et de l'utilité réelle des EIAH proposés aux apprenants ou aux enseignants.

Il s'agit d'un champ de recherche pluridisciplinaire faisant intervenir des chercheurs en sciences de l'éducation, en didactique, en psychologie cognitive et des informaticiens partageant leurs expertises. Les premiers s'intéressent essentiellement aux aspects psychologiques, pédagogiques et didactiques de l'apprentissage humain dans un contexte informatique. Les informaticiens s'intéressent aux nombreux défis que les EIAH posent à l'informatique. Il peut s'agir de problèmes d'ergonomie ou d'ingénierie logicielle mais bon nombre des questions informatiques en EIAH relèvent de l'Intelligence Artificielle (IA). Les EIAH constituent en effet un excellent terrain d'application pour l'IA et soulèvent régulièrement de nouvelles problématiques s'inscrivant dans le champ de l'IA. Par exemple, on trouve parmi les problématiques actuelles en EIAH :

- Ingénierie des Connaissances : afin de travailler sur des modèles d'activité, du domaine et de l'apprenant pour personnaliser l'apprentissage.
- Diagnostic automatique de réponses d'apprenants basées sur des techniques d'apprentissage automatique.
- Educationnal data mining : techniques de data mining pour l'exploitation de données issues de l'éducation et afin de mieux la comprendre.

Compte-rendu de la journée

La journée EIAH et IA s'est tenue le 28 mai 2013, à Toulouse, dans le cadre de la conférence EIAH [29]. Elle se déroulait en parallèle des ateliers adossés à la conférence. La journée a suscité l'intérêt de 86 personnes s'étant inscrites, avec une fréquentation variable au fil de la journée plutôt de l'ordre d'une trentaine de personnes, certains participants étant également impliqués dans d'autres ateliers.

En amont de la journée, un appel à communication a été diffusé. Sur avis des membres du comité de programme de la journée, 12 contributions ont été reçues, 11 ont été acceptées, dont 4 en présentation longue et 7 en présentation courte. Le programme de la journée était le suivant (le programme détaillé, les articles et les présentations sont disponibles en ligne [30]) :

- Présentation des associations ATIEF et AFIA (Marie Lefevre et Catherine Faron-Zucker)
- Conférence invitée : IA et EIAH : regards croisés, Monique Grandbastien et Serge Garlatti
- Session 1 : Apprentissage Automatique, présidée par Dominique Lenne
- Session 2 : Ontologies, Web sémantique et EIAH, présidée par Dominique Py
- Session 3 : Systèmes à base de connaissances, TAL et EIAH, présidée par Cyrille Desmoulins
- L'interaction comme inscription de connaissance pour l'apprentissage humain, Alain Mille
- Table ronde : Serge Garlatti, Monique Grandbastien, Vanda Luengo, Alain Mille et Thierry Nodenot

Durant la conférence invitée, Monique Grandbastien et Serge Garlatti ont débattu du caractère pluridisciplinaire des recherches en EIAH. Ils nous ont ensuite présenté une frise historique, richement illustrée, des recherches alliant EIAH et IA. Ils sont revenus sur quelques « success stories » en la matière, mais ont aussi discuté des sujets au cœur de l'actualité et des nouveaux enjeux pour la recherche en EIAH, ainsi que des défis que cela pose à l'IA : la robotique, les interfaces adaptatives, les travaux en traitement de la langue naturelle, la prise en compte du contexte, le Web des objets, le Web sémantique, les MOOCs, etc. Suite à cette présentation, une séance de questions-réponses s'est engagée. Parmi les questions et préoccupations de l'auditoire, nous retenons particulièrement les éléments suivants :

- On observe une disparition de l'IA dans les EIAH et plus largement dans les systèmes destinés aux « vrais » utilisateurs. Peut-on identifier les causes de ce phénomène et veut-on y remédier ? N'assiste-t-on pas plutôt à une banalisation d'une « petite » IA pervasive qui ne porte plus son nom ?
- Les chercheurs en EIAH et en IA parlent-ils la même langue ? Est-il tabou de parler architecture logicielle lorsque l'on fait de la recherche en IA ? Et réciproquement, est-ce que les chercheurs en IA s'intéressent à de « vrais » problèmes ?
- Avec la démocratisation des « connaissances accessibles à tous » via le Web (et le Web de données), un enjeu pour les chercheurs en EIAH et en IA est la qualité des connaissances manipulées, leur fiabilité, la confiance que l'on peut avoir en elles. En ce sens, l'acquisition et l'ingénierie des connaissances dans le contexte du Web (en particulier social) deviennent cruciales.

La journée s'est poursuivie par trois sessions durant lesquelles les contributeurs ont présenté leurs articles. Chaque présentation a donné lieu à de nombreuses questions sur les travaux eux-mêmes, mais aussi, à un niveau plus général sur la façon dont l'IA contribue au développement des EIAH et réciproquement. Tous les auteurs et orateurs ont fait l'effort de bien mettre en avant l'articulation entre les deux aspects (IA et EIAH) à la fois dans leurs contributions écrites et dans leurs présentations, soulevant ainsi des questions intéressantes discutées avec l'assemblée et mettant en évidence des nouveaux défis pour la recherche en EIAH et en IA, et résonnant avec les discussions qui ont suivi la conférence invitée.

Juste avant la table ronde, Alain Mille a débattu, lors d'une présentation concise, de la question de l'interaction comme inscription de connaissance pour l'apprentissage humain. Il a notamment illustré comment les interactions entre des apprenants et des systèmes informatiques « intelligents » (pour ne pas dire apprenants) pouvaient être le lieu d'un processus vertueux de co-construction de connaissance.

La table ronde, préparée et animée par Nathalie Guin, a rassemblé cinq participants : Serge Garlatti, Monique Grandbastien, Vanda Luengo, Alain Mille, Thierry Nodenot. Les participants devaient s'attendre à répondre à trois questions :

- Votre synthèse personnelle sur cette journée et sur l'état actuel du lien IA-EIAH ?
- Nouvelles orientations, thématiques prioritaires, défis pour les EIAH et l'IA ?

29. http://www.irit.fr/EIAH2013/index.php?page=journee-eiah-ia
30. http://www.irit.fr/EIAH2013/index.php?page=programme

- Quelles actions concrètes pour renforcer le lien entre nos deux communautés de recherche ?

Ils ont été aidés dans leurs réponses par la salle, soucieuse de participer au débat. L'assistance a notamment profité de l'occasion pour poser de nombreuses questions. De cette discussion riche et enthousiaste, nous retenons (sans ordre particulier) les éléments suivants :

- En IA, on observe trop souvent une séparation entre les approches symboliques et les approches numériques. Trop peu de travaux s'intéressent à combiner habilement les deux, et c'est peut-être l'une des raisons pour lesquelles l'IA a du mal à passer à l'échelle et à s'intégrer facilement dans des applications « grand public », et notamment dans les EIAH.
- En EIAH comme en IA, la question des modèles est cruciale. On observe souvent un problème d'ajustement entre la granularité des modèles développés (souvent très détaillés), et ce qui est réellement utilisé dans les applications (une petite sous-partie des modèles). Il n'est jamais simple de proposer des résultats de recherche génériques lorsque l'on s'intéresse à des problèmes concrets très spécifiques. Cela étant dit, les chercheurs en EIAH ont une expérience considérable de la modélisation des utilisateurs.
- Il apparait nécessaire d'expliciter quels sont les paradigmes de recherche chers aux chercheurs en EIAH pour leur permettre de mieux se positionner. Ces paradigmes existent sans nul doute, mais ils ne sont pas suffisamment explicites ni accessibles, en particulier pour les « nouveaux » dans la communauté. À ce propos, il est intéressant de considérer la façon dont la question est envisagée dans d'autres pays (notamment au Canada) où les choses sont parfois plus claires.

S'est ensuite posée la question de savoir ce que les EIAH apportent à l'IA et réciproquement, ce que l'IA apporte aux EIAH. Peu de réponses ont été apportées, peut-être parce que la question, posée ainsi, n'est pas pertinente. En effet, il ne s'agit pas qu'un domaine « rende service » à l'autre, mais plutôt d'identifier quels sont les objets de recherche qui intéressent les deux domaines (et comme la journée l'a montré, ils sont nombreux), et qui permettront donc des interactions entre les chercheurs des deux communautés, et des enrichissements mutuels. Par ailleurs, il ne faut pas oublier que les EIAH et l'IA partagent des sources d'inspiration communes (notamment en sciences cognitives), ont des collaborateurs communs (sciences cognitives, IHM, Web, Robotique, etc.), et ont des problématiques communes (par exemple, modélisation des connaissances manipulées, modélisation des utilisateurs, du contexte, l'évaluation et la validation des résultats). Lors de la discussion sur les actions à mener pour renforcer les liens entre les deux communautés et susciter des collaborations entre les EIAH et d'autres champs de recherche, les propositions ont fusé. Parmi elles, on retient la nécessité d'encourager davantage la pluridisciplinarité, la volonté des chercheurs en EIAH d'être plus présents au sein de l'AFIA et dans certaines conférences, notamment les conférences en informatique, et la nécessité de faire plus de publicité auprès des chercheurs en IA pour les attirer dans ces journées thématiques. Mais ce que nous retenons avant toute chose, c'est l'enthousiasme des participants à cette journée et la volonté collective de réitérer l'expérience dès l'année prochaine.

Nous tenons à remercier les organisateurs de la conférence EIAH pour leur accueil et leur gentillesse, et pour avoir tout mis en œuvre pour que cette journée se déroule dans les meilleures conditions (à la météo pluvieuse près). Nous remercions également l'AFIA, l'ATIEF et le GDR I3 pour avoir soutenu scientifiquement et financièrement l'organisation de cette journée. Nous remercions chaleureusement Monique Grandbastien et Serge Garlatti pour l'enrichissante conférence invitée qu'ils nous ont proposée, ainsi que pour leur important travail de recherche bibliographique qu'ils ont eu la gentillesse de mettre à la disposition des communautés. Nous remercions non moins chaleureusement les participants à la table ronde, Serge Garlatti, Monique Grandbastien, Vanda Luengo, Alain Mille et Thierry Nodenot pour leurs contributions constructives et leurs réflexions sur les interactions entre EIAH et IA. Nous remercions également les membres du comité de programme (chercheurs en EIAH et en IA), les orateurs, les auteurs, les présidents de session et les participants à la journée pour leurs contributions. Merci à tous d'avoir mis en avant les liens, les interactions et les questions que vous évoquent la juxtaposition des EIAH et de l'IA. Merci à tous d'avoir fait de cette journée un succès. Nous espérons vous voir encore plus nombreux à la prochaine journée EIAH et IA !

Résumés de thèses et d'HDR

Contributions à l'Apprentissage par Renforcement Inverse
Edouard Klein
Thèse de Doctorat
Soutenance le 21 novembre 2013 au LORIA Nancy

Jury : Rachid Alami, Directeur de recherche CNRS, LAAS, rapporteur ; Brahim Chaib-draa, Professeur, Université de Laval/DAMAS, rapporteur ; Sylvain Contassot-Vivier, Professeur, Université de Lorraine/LORIA, examinateur ; Matthieu Geist, Enseignant-chercheur, Supélec, co-directeur de thèse) ; Yann Guermeur, Directeur de recherche CNRS, LORIA, directeur de thèse ; Guillaume Laurent, Maître de conférences, ENSMM/Institut FEMTO-ST, examinateur ; Manuel Lopes, Chargé de recherche INRIA, INRIA Sud-Ouest, examinateur ; Olivier Pietquin, Professeur, Université de Lille 1/LIFL, examinateur.

Résumé : Cette thèse s'intéresse au problème du contrôle optimal, qui consiste à trouver un comportement maximisant un critère défini par l'opérateur. Elle traite plus particulièrement des méthodes permettant de trouver le contrôle optimal de manière générique et non ad hoc à un environnement particulier, lorsque le modèle dynamique de cet environnement est inconnu.

Un paradigme efficace pour la résolution générique et sans modèle du problème du contrôle optimal est l'Apprentissage par Renforcement (AR). Les critères de réussite de la tâche sont fournis par l'opérateur sous la forme d'un signal de récompense défini dans un Processus Décisionnel de Markov (PDM). Les algorithmes d'AR trouvent (éventuellement de manière approchée) par interactions avec l'environnement le comportement maximisant le cumul des récompenses sur le long terme. La conception par l'opérateur humain de ce signal de récompense reste une tâche ardue nécessitant d'une part des connaissances en AR et d'autre part une expertise quant au problème que la machine doit résoudre.

Afin de contourner cette étape difficile, les techniques dites d'Apprentissage par Renforcement Inverse (ARI) sont apparues. Elles extraient la récompense de démonstrations de la tâche à effectuer, démonstrations accomplies par un expert. Cette récompense peut alors être optimisée par un algorithme d'AR fournissant ainsi une politique de contrôle imitant le comportement de l'expert. Cette optimisation permet une meilleure capacité de généralisation que les méthodes d'imitation supervisée. La plupart des méthodes d'ARI de l'état de l'art nécessitent le calcul répété d'une politique optimale et de son attribut moyen et sont de fait inopérantes lorsque le modèle est inconnu. Une méthode d'ARI plus récente ne nécessite pour fonctionner que les données expertes et des données non expertes permettant d'effectuer de l'échantillonnage préférentiel.

Cette thèse porte sur la définition d'algorithmes d'ARI efficaces en échantillons. Le but est de bénéficier de la souplesse d'utilisation des méthodes supervisées, qui n'exploitent que des données faciles à réunir dans la pratique, tout en conservant la bonne capacité de généralisation des algorithmes d'ARI. À cet effet, ce document fournit un algorithme d'estimation de l'attribut moyen qui permet de faire fonctionner les algorithmes itératifs de la littérature sans avoir besoin d'un simulateur, celui-ci étant remplacé par des données représentatives du PDM.

Deux nouveaux algorithmes d'ARI, SCIRL et CSI, sont ensuite proposés. Tous deux reposent sur un constat de similarité entre le rôle de la fonction de qualité de l'expert et celui de la fonction de score d'un classifieur. Le premier algorithme introduit la dynamique temporelle du PDM dans une méthode de classification structurée par le biais de l'attribut moyen (qui peut être évalué par l'algorithme d'estimation précédent). L'attribut moyen intervient dans la définition de la fonction de score du classifieur. Le second algorithme, plus souple, cascade une méthode de classification avec une étape de régression, qui en inversant l'équation de Bellman permet d'obtenir une fonction de récompense à partir de la fonction de score du classifieur.

Ces deux nouvelles méthodes disposent de garanties théoriques portant sur l'optimalité de la politique experte vis-à-vis de la fonction de récompense renvoyée par l'algorithme d'ARI. Leur capacité à déduire une fonction de récompense à partir de démonstrations expertes uniquement est illustrée empiriquement, une fois ces méthodes armée chacune d'une heuristique qui lui est propre. Les performances de ces algorithmes sont supérieures à celles d'algorithmes d'imitation supervisée et à un algorithme d'ARI récent disposant de données non expertes supplémentaires.

Gestion de l'Incertitude pour l'Optimisation de Systèmes Interactifs
Lucie Daubigney
Thèse de Doctorat
Soutenance le 1er octobre 2013 au LORIA Nancy

Jury : Joëlle Pineau, Reasoning and Learning Laboratory, McGill University - Montréal, rapporteur ; Frédérick Garcia, INRA - Toulouse, rapporteur ; Alain Dutech, Equipe projet MaIA, Loria - Nancy, directeur de thèse ; Olivier Pietquin, Equipe IMS-MaLIS, Supélec - Metz, co-directeur de thèse ; Matthieu Geist, Equipe IMS-MaLIS, Supélec - Metz, examinateur ; Fabrice Lefèvre LIA, Universités d'Avigon et des Pays du Vaucluse - Avignon, examinateur ; Blaise Thomson, Dialogue Systems Group, University of Cambridge - Cambridge, examinateur.

Résumé : Le sujet des travaux concerne l'amélioration du comportement des machines dites "intelligentes". Cette caractéristique se traduit par une capacité à s'adapter à l'environnement, même lorsque celui-ci est sujet à des changements. Un des domaines concerné est celui des interactions homme-machine. Dans ce cas, la machine doit faire face à différents types d'incertitude pour agir de façon appropriée. Tout d'abord, elle doit pouvoir prendre en compte les variations de comportements entre les utilisateurs et le fait que le comportement d'un même utilisateur peut varier d'une utilisation à l'autre en fonction de l'habitude à interagir avec le système. Ensuite, la machine doit s'adapter à l'utilisateur même si les moyens de communication entre ce dernier et la machine sont bruités. L'objectif est alors de gérer ces incertitudes pour exhiber un comportement cohérent. Ce dernier se définit comme la suite de décisions successives que la machine doit effectuer afin de parvenir à l'objectif fixé. Trouver le comportement optimal revient alors à résoudre un problème de décisions séquentielles sous incertitude. Traditionnellement, des connaissances expertes liées à la tâches sont utilisées pour résoudre ce problème. De ce fait, le développement des interfaces est coûteux, long et difficile à transférer à d'autres tâches. Depuis quelques années, une méthode semble se démarquer pour optimiser automatiquement la gestion des interactions homme-machine à partir de données : l'apprentissage par renforcement. Cette méthode d'apprentissage automatique permet d'optimiser un problème de prise de décisions séquentielles sous incertitude par essais-erreurs. L'intérêt de cette méthode est que seul l'objectif est spécifié à la machine et non les différentes étapes pour y parvenir. Elle permet alors de s'affranchir de connaissances expertes qui étaient auparavant nécessaires pour définir les différentes étapes de résolution du problème et permet ainsi un gain de généricité. Cependant, le cadre habituel de résolution des problèmes d'apprentissage par renforcement propose de mettre la tâche sous la forme d'un processus décisionnel de Markov. Il est même souvent nécessaire de se placer dans le cadre des processus décisionnels de Markov partiellement observables. Ce cadre de travail nécessite alors à son tour d'apporter des connaissances expertes. En utilisant l'apprentissage automatique, l'expertise a été déplacée du modèle de la tâche spécifique au domaine concerné vers des compétences en apprentissage machine. La thèse défendue par ces travaux est qu'il est possible de relaxer certaines contraintes liées à l'expertise humaine et obtenir ainsi des interactions de bonne qualité tout en limitant la perte de généricité liée l'introduction de modèles. Deux domaines applicatifs sont choisis pour illustrer le propos car ils répondent à des problématiques différentes. Ils s'agit des systèmes de dialogue parlé et des environnements informatiques pour l'apprentissage humain. Dans le premier cas, la décision de la machine peut se baser sur l'analyse d'un signal physique, la voix, et sur la construction du langage. Les incertitudes à gérer sont directement liées à la bonne reconnaissance de ce qui a été dit. Une estimation de l'incertitude peut être faite par la construction de modèles du langage, bien étudié depuis des dizaines d'années. En revanche, la cognition de l'apprentissage n'étant pas encore un mécanisme bien connu, il reste difficile d'en construire un modèle. S'en affranchir serait alors un progrès.

Quatre contraintes ont été identifiées comme nécessitant des connaissances expertes. La première concerne la mise en place d'un simulateur d'utilisateurs. En effet, les algorithmes d'apprentissage par renforcement "historiques" étant gourmands en échantillons, il était nécessaire, pour pallier ce manque, d'utiliser un modèle mimant le comportement des utilisateurs pour la machine. La première contribution consiste à s'affranchir de ce modèle par l'utilisation d'un algorithme efficace sur les échantillons et par une gestion du dilemme exploration/exploitation efficace. Le second point propose de gérer le passage à l'échelle bien connu, qu'implique l'utilisation d'un processus décisonnel de Markov partiellement observable, par une représentation non-linéaire de la fonction de valeur. La troisième contribution permet de relaxer le respect de la propriété de Markov. Pour des problèmes partiellement observables réels, l'état de croyance construit selon l'expertise humaine n'est pas garanti de respecter strictement cette propriété. La recherche directe dans l'espace des politiques est alors expérimentée. La quatrième contrainte propose une construction automatique d'un état markovien. Une solution à ce problème passant par l'utilisation d'un réseau de neurones récurrent est proposée.

Sélection contextuelle de services continus pour la robotique ambiante
Thèse de Doctorat
Benjamin Cogrel
Soutenance le 18 novembre 2013 à l'Université Paris Est Créteil

Jury : Amar Ramdane-Cherif, Professeur à l'Université de Versailles Saint-Quentin, rapporteur ; Olivier Simonin, Professeur à l'INSA de Lyon, rapporteur ; Nicole Levy, Professeur au CNAM de Paris, examinateur ; Patrick Reignier, Professeur à l'ENSIMAG du groupe Grenoble INP, examinateur ; Abdelghani Chibani, Maître de conférences à l'Université Paris-Est, Créteil, examinateur ; Yacine Amirat, Professeur à l'Université Paris-Est, Créteil, directeur de thèse ; Boubaker Daachi, Maître de conférences HDR à l'Université Paris-Est Créteil, co-directeur de thèse.

Résumé : La robotique ambiante s'intéresse à l'introduction de robots mobiles au sein d'environnements actifs où ces derniers fournissent des fonctionnalités alternatives ou complémentaires à celles embarquées par les robots mobiles. Cette thèse étudie la mise en concurrence des fonctionnalités internes et externes aux robots, qu'elle pose comme un problème de sélection de services logiciels. La sélection de services consiste à choisir un service ou une combinaison de services parmi un ensemble de candidats capables de réaliser une tâche requise. Pour cela, elle doit prédire et évaluer la performance des candidats. Ces performances reposent sur des critères non-fonctionnels comme la durée d'exécution, le coût ou le bruit.

Ce domaine applicatif a pour particularité de nécessiter une coordination étroite entre certaines de ses fonctionnalités. Cette coordination se traduit par l'échange de flots de données entre les fonctionnalités durant leurs exécutions. Les fonctionnalités productrices de ces flots sont modélisées comme des services continus. Cette nouvelle catégorie de services logiciels impose que les compositions de services soient hiérarchiques et introduit des contraintes supplémentaires pour la sélection de services.

Cette thèse met en évidence la présence d'un important couplage non-fonctionnel entre les performances des instances de services de différents niveaux, même lorsque les flots de données sont unidirectionnels. L'approche proposée se concentre sur la prédiction de la performance d'un organigramme. Un organigramme regroupe l'ensemble des instances de services sollicitées pour réaliser une tâche de haut-niveau. L'étude s'appuie sur un scénario impliquant la sélection d'un service de positionnement en vue de permettre le déplacement d'un robot vers une destination requise.

Pour un organigramme considéré, la prédiction de performance doit, dans ce scénario, répondre aux exigences suivantes : elle doit (i) être contextuelle en tenant compte, par exemple, du chemin suivi pour atteindre la destination requise, (ii) prendre en charge le remplacement d'une instance de sous-service suite à un échec ou, par extension, de façon opportuniste. En conséquence, cette sélection de services est posée comme un problème de prise de décision séquentielle formalisé à l'aide de processus de décision markoviens à horizon fini. La dimensionnalité importante du contexte en comparaison à la fréquence des déplacements du robot rend inadaptées les méthodes consistant à apprendre di- rectement une fonction de valeur ou une fonction de transition. L'approche proposée repose sur des modèles de dynamique locaux et exploite le chemin de déplacement calculé par un sous-service pour estimer en ligne les valeurs des organigrammes disponibles dans l'état courant. Cette estimation est effectuée par l'intermédiaire d'une méthode de fouille stochastique d'arbre, Upper Confidence bounds applied to Trees.

Human behaviour modelling in complex socio-technical systems : an agent based approach
Julie Dugdale
Habilitation à diriger des recherches
Soutenance le 12 décembre 2013 au LIG Grenoble

Jury : François Charoy, Professeur, Université de Lorraine, rapporteur ; Alexis Drogoul, Directeur de Recherche, Institut de recherche pour le développement, rapporteur ; Salima Hassas, Professeur, Université Claude-Bernard, Lyon 1, rapporteur ; Catherine Garbay, Directeur de Recherche, CNRS, examinateur ; Chihab Hanachi, Professeur, Université de Toulouse 1, examinateur ; Patrick Reignier, Professeur, Institut National Polytechnique de Grenoble, examinateur ; Pascal Salembier, Professeur, Université de Technologie de Troyes, examinateur.

Résumé : This manuscript describes my research contributions over the last 14 years and my future research directions. It spans my early works with the Cognitive Engineering Research Team at the IRIT laboratory in Toulouse, through to those conducted in my present position as leader of the MAGMA, Multi-Agents Systems Team, and as an Associate Professor at Université Pierre Mendès France. The route by which I arrived at MAGMA is somewhat diverse, having worked in teams from different disciplines : cognitive ergonomics, human-computer

interaction and finally multi-agent systems. These experiences have strongly influenced my approach in analysing human behaviour and in developing agent based models and simulators. The choice to work in such different teams was made consciously and I have actively sought to draw in methods and techniques from other disciplines into my work. The reason for doing so was largely pragmatic ; these disciplines brought both a fresh way of looking at common problems and they possessed expertise and skills that were absent or poorly practiced by my original discipline of artificial intelligence. In particular I was largely influenced by research in cognitive engineering and its focus on trying to understand the nuances of human interaction. In trying to promote a multi-disciplinary approach, from which I believe we can all benefit, I have tried to publish in a diversity of domains.

This document follows a roughly chronological description of my main contributions since arriving in France from the UK in 1998. My research work before this time is not covered in this manuscript. The reason for this is that while there are common themes, notably modelling and simulation, there was a transition in the focus of my work when I arrived in France. Although my previous research undoubtedly helped to form my approach, I became more interested in cognitive aspects and how to model human behaviours that were more representative of what happens in the real world.

My work on human behaviour modelling is applied to two application domains : crisis and emergency management, and energy management in the home environment. From a scientific point of view these domains offer particular challenges in modelling human behaviours and interactions. In an emergency or crisis situation, the interest is in modelling the extreme cognitive demands placed on humans when working under time-pressure in a highly stressful, emotional and rapidly changing environment. Conversely, in home situations, human behaviours, and in particular our interactions with others, are more subtle ; relying heavily on our familiarity with our co-inhabitants. This means that it may be more difficult, from an external point of view, to perceive or understand what motivates our behaviours and interactions. As with a multi-disciplinary approach, working in the diverse domains of crisis management and energy management, is an enriching experience. Not only do we see the contrasts in how human behaviour differs between highly charged and relatively calm environments, but we also see the similarities, such as adaptive and self-organizing behaviours, that are present in both cases. This manuscript describes my investigation into uncovering, modelling, and simulating human behaviours, in these different contextual situations.

Modélisation et détection des émotions à partir de données expressives et contextuelles

Franck Berthelon

Thèse de Doctorat

Soutenance le 16 décembre 2013 au I3S Sophia

Jury : Frank Ferrie, rapporteur ; Claude Frasson, rapporteur ; Ladjel Bellatreche, examinateur ; Nhan Le Thanh, examinateur ; Peter Sander, directeur de thèse.

Résumé : Nous proposons un modèle informatique pour la détection des émotions basé sur le comportement humain. Pour ce travail, nous utilisons la théorie des deux facteurs de Schachter et Singer pour reproduire dans notre architecture le comportement naturel en utilisant à la fois des données expressives et contextuelles. Nous concentrons nos efforts sur l'interprétation d'expressions en introduisant les Cartes Emotionnelles Personnalisées (CEPs) et sur la contextualisation des émotions via une ontologie du contexte émotionnel(EmOCA). Les CEPs sont motivées par le modèle complexe de Scherer et représentent les émotions déterminées par de multiple capteurs. Les CEPs sont calibrées individuellement, puis un algorithme de régression les utilises pour définir le ressenti émotionnel à partir des mesures des expressions corporelles. L'objectif de cette architecture est de séparer l'interprétation de la capture des expressions, afin de faciliter le choix des capteurs. De plus, les CEPs peuvent aussi être utilisées pour la synthétisation des expressions émotionnelles.

EmOCA utilise le contexte pour simuler la modulation cognitive et pondérer l'émotion prédite. Nous utilisons pour cela un outil de raisonnement interopérable, une ontologie, nous permettant de décrire et de raisonner sur les philies et phobies pour pondérer l'émotion calculée à partir des expressions. Nous présentons également un prototype utilisant les expressions faciales pour évaluer la reconnaissance des motions en temps réel à partir de séquences vidéos. De plus, nous avons pu remarquer que le système décrit une sorte d'hystérésis lors du changement émotionnel comme suggéré par Scherer pour son modèle psychologique.

Adhésion individuelle et abonnement

❏ **Demande** ❏ **Renouvellement**

Nom : Prénom :
Affiliation :
Adresse postale :
N° de téléphone : N° de télécopie :
Adresse électronique :
Activité (à titre professionnel / à titre privé (*rayer la mention inutile*)) :

Type d'adhésion	
❏ Simple :	30 €
❏ Étudiant (sur justificatif) :	15 €
❏ Soutient :	Sans objet

❏ Adhésion au collège SMA : gratuite
❏ Adhésion au collège IC : gratuite
❏ Adhésion au collège *FERA (Apprentissage)* : gratuite

Adhésion personne morale

❏ **Demande** ❏ **Renouvellement**

Organisme :
Adresse postale commune aux bénéficiaires couverts par cette adhésion :
Nom et prénom du représentant : Fonction :
Mél : Tél : Fax :
Adresse postale :

L'adhésion morale donne droit à 5 adhésions pour les universitaires et à 15 adhésions pour les non universitaires.
Coordonnées des bénéficiaires :

Nom, prénom	Mél.	Tél.	Fax

	Tarif de base fixe :	Tarif par bénéficiaire :
❏ Laboratoire universitaires/PME	150 €	Gratuit pour 5 personnes (30 € par bénéficiaire supplémentaire)
❏ Personnes morales non universitaires	450 €	Gratuit pour 15 personnes (30 € par bénéficiaire supplémentaire)
❏ Adhésion de soutien	600 €	Sans objet

❏ j'accepte que les renseignements ci-dessus apparaissent dans l'annuaire de l'AFIA
❏ j'accepte que les renseignements ci-dessus soient transmis à l'ECCAI pour constituer un ficher européen

Veuillez trouver un règlement (à l'ordre de l'AFIA) de .....Euros

Trésorier AFIA en charge des adhésions : Davy MONTICOLO, ENSGSI, 8 rue Bastien Lepage, 54000 Nancy.
Mode d'adhésion :
De préférence, en ligne via le site Internet de l'AFIA : http://www.afia.asso.fr
A défaut, cette page doit être envoyée au trésorier.
Modes de paiement :

1. par Paypal
2. par bon de commande administratif, à l'ordre de l'AFIA, envoyé au trésorier ;
3. par virement bancaire sur le compte de l'AFIA : Société Générale, 1 place du Maréchal Foch, 35000 Rennes, France. Code banque 30003, code guichet 01902, numéro de compte 00037283856 clef RIB 39.
4. par chèque, à l'ordre de l'AFIA, envoyé au trésorier ;

Sommaire du Bulletin n° 84

Calendrier de parution du bulletin de l'AFIA

Hiver	*Été*
Réception des contributions:	Réception des contributions:
15 décembre	**15 juin**
Sortie le **31 janvier**	Sortie le **31 juillet**
Printemps	*Automne*
Réception des contributions:	Réception des contributions:
15 mars	**15 septembre**
Sortie le **30 avril**	Sortie le **31 octobre**