IA & Humanités numériques

Audrey Baneyx, Nicolas Maudet, Dominique Longin, Florence Dupin de Saint-Cyr

AfIA

Association française
pour l'Intelligence Artificielle

Bulletin N° 92

Association française pour l'Intelligence Artificielle

AFIA

AfIA
Association française
pour l'Intelligence Artificielle

PRÉSENTATION DU BULLETIN

Le Bulletin de l'**A**ssociation **f**rançaise pour l'**I**ntelligence **A**rtificielle vise à fournir un cadre de discussions et d'échanges au sein de la communauté universitaire et industrielle. Ainsi, toutes les contributions, pour peu qu'elles aient un intérêt général pour l'ensemble des lecteurs, sont les bienvenues. En particulier, les annonces, les comptes rendus de conférences, les notes de lecture et les articles de débat sont très recherchés. Le Bulletin de l'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.

Édito

Dans ce numéro, Audrey Baneyx nous propose un dossier sur les humanités numériques, qui illustre à travers quatre exemples de projets ou d'équipes de recherches l'originalité et la vitalité de cette thématique, qui mobilise de nombreuses facettes de l'IA autour de questions issues des sciences humaines et sociales. Nous retrouvons aussi les compte-rendus de trois journées bilatérales tenues récemment, avec le traitement automatique des langues, la réalité virtuelle, et l'extraction et la gestion des connaissances.

Bonne lecture à tous !

Olivier AMI, Florence BANNAY, Dominique LONGIN, Nicolas MAUDET & Philippe MORIGNOT
Rédacteurs

AfIA
Association française
pour l'Intelligence Artificielle

SOMMAIRE
DU BULLETIN DE L'AFIA

AfIA
Association française
pour l'Intelligence Artificielle

Dossier
« I.A. et Humanités numériques »

Dossier réalisé par

Audrey BANEYX
Médialab
Sciences Po.
audrey.baneyx@sciencespo.fr

I.A. et Humanités numériques

Nous avons souhaité consacrer ce numéro aux articulations entre Intelligence artificielle et humanités numériques. Les Humanités Numériques – *Digital Humanities* – se sont structurées avec l'usages des technologies numériques dans la recherche en sciences humaines et sociales.

La quantité de documents et de données à archiver et à fouiller explose dans toutes les disciplines et dans toutes les activités de notre société. La disponibilité de ces données numériques qui décrivent les objets d'étude traditionnels des sciences humaines et sociales accroît de façon considérable les opportunités d'analyse quantitative. Il est ainsi possible de s'appuyer sur des outils d'analyse de données classiques, de modélisation statistique, de visualisation, etc., pour explorer des données historiques, géographiques, légales, et apporter des éclairages intéressants.

Cependant, les données associées aux sciences humaines et sociales sont en général complexes à de nombreux égards, ce qui demande le développement de méthodes et d'outils spécifiques. Ces méthodes et outils gagnent à être conçus en collaboration, par des équipes pluridisciplinaires, associant chercheurs des sciences humaines et chercheurs en intelligence artificielle, en design, en traitement automatique du langage... Il apparaît indispensable de procéder par itération en adaptant et en complexifiant progressivement les modèles et représentations utilisés pour rendre compte de la richesse du domaine et favoriser un processus d'acculturation des différentes pratiques.

Vous trouverez dans les pages suivantes plusieurs contributions qui mettent en lumière des synergies qui apparaissent entre Intelligence artificielle et Humanités numériques à différents niveaux : au sein d'une même équipe, au sein d'un même projet de recherche, au sein d'un même instrument.

Bonne lecture !

AfIA
Association française
pour l'Intelligence Artificielle

Visualiser les données de prêts d'une bibliothèque universitaire (PREVU)

Le projet Prevu de l'équipe CiTu-Paragraphe
Université Paris 8
http://prevu.fr/

Gaétan DARQUIÉ
gaetan.darquie@citu.fr
Responsable du projet

Autres partenaires du projet

- Laboratoire LIASD-EA4383, Université Paris 8
- Bibliothèque Universitaire de Paris
- ENSAD
- School of Information, Michigan University

Thématique générale de l'équipe

Dirigée par Khaldoun Zreik, l'équipe CiTu du Laboratoire Paragraphe de l'université Paris 8 travaille sur la conception de l'information d'une part et sur la communication humaine médiatisée de l'autre, elle s'intéresse notamment à l'émergence de pratiques hybridant des enjeux sociaux ou artistiques au numérique, notamment en ce qui concerne la ville numérique et l'hyperUrbain. Le CiTu participe à plusieurs projets de recherche afin d'aider à concevoir des usages innovants, il a ainsi participé, entre autres, aux projets FUI 12 Ozalid [1] ou aux projets Terra Numerica et Terra Dynamica. Le CiTu porte le projet par l'entremise de Gaétan Darquié. La participation du LIASD (Laboratoire d'Informatique Avancée de Saint-Denis) au travers de l'axe « Acquisition, Interprétation et Visualisation de Données », animé par Myriam Lamolle, est essentielle au projet notamment en ce qui concerne la thématique « Intelligence Artificielle ». Elle concerne plus particulièrement des traitements de données complexes par l'intermédiaire de modèles ontologiques et d'inférence de connaissances grâce à des raisonneurs ; mais aussi, elle vise à en faire ressortir de nouvelles corrélations par des techniques de fouilles de données utilisées par les applications de visualisations. Créé en 1972, le LIASD couvre un large spectre de l'informatique : l'algorithmique, la programmation, les systèmes embarqués, le temps réel, la robotique, les langages, la visualisation de données, la synthèse d'images, la logique floue, les jeux et l'interaction homme-machine ce qui favorise des croisements innovants

1. Le projet Correct.

Description du projet

Le projet Prévu a débuté en janvier 2010. Il s'échelonne sur trois ans et prend place à la suite des expérimentations CityPulse et Capteurs Montre Verte (CMV) menées entre 2010 et 2013 auxquelles l'équipe du CiTu a participé. Sous l'impulsion d'Isabelle Breuil et suite au changement de système de gestion de base de données (SGBD) de la bibliothèque universitaire (BU) de Paris 8 et au passage à une solution open source documentée (KOHA) au début de l'année 2012, la BU s'est demandée comment assurer au mieux l'ouverture d'une partie de ses données.

Objectifs

Le projet Labex-Arts H2H Prévu vise à générer des services à partir des données de prêts de bibliothèque via des visualisations ainsi qu'à découvrir des communautés de pratique pour faciliter la communication intra et inter-communautés de lecteurs.

Afin de faciliter différentes formes de recommandations (lecture, communautés potentielles d'appartenance, etc.) auprès d'un lecteur, un modèle ontologique décrit un lecteur selon différents contextes correspondant à différentes définitions de ce dernier. Une première définition caractérise un lecteur comme un étudiant (selon un niveau d'études L, M, D), un enseignant, un chercheur, une personne de l'Université. Une seconde définition possible correspond à l'appartenance d'un lecteur à une UFR qui sera considérée comme une communauté (puisqu'elle a un nombre restreint de domaines de prédilection ; par exemple, le cinéma). Un lecteur peut aussi être vu d'après un profil dégagé de ses prêts sachant que les livres sont classés selon la hiérarchie ISBN (un raisonneur peut alors

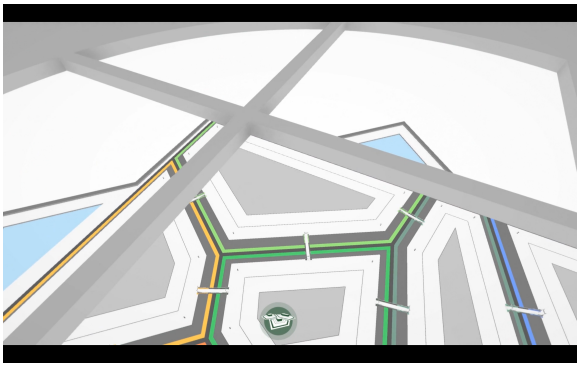
inférer les thématiques dominantes des prêts d'un lecteur). On peut alors poser des requêtes prenant en compte ces différentes définitions (par exemple, trouver l'ensemble des lecteurs qui s'intéressent au cinéma et au philosophe Gilles Deleuze, soit une communauté, et les livres qui ont été les plus empruntés par cette communauté).

D'autre part, il est intéressant de favoriser les échanges entre lecteurs. Une communauté de lecteurs vise à développer la réussite universitaire ou tout au moins une certaine culture personnelle. Inversement, évaluer l'activité d'une communauté permet de comprendre son rôle et son impact spécifique sur tout ou partie des autres communautés. Pour cela, l'approche ontologique proposée permet de modéliser les communautés et leur évolution sous forme de graphe [3]. Ce graphe est finalement exploité pour établir la centralité de chaque communauté dans ce réseau de lecteurs. Des prédictions de centralité peuvent alors être calculées [4].

L'observation de ces communautés nous permettrait ainsi d'approfondir les premiers scénarios mis en place avec le démonstrateur `prevu.fr` réalisé par Mehdi Bourgeois ainsi que les visualisations associées aux modules d'exploration permettant d'obtenir des informations de prêts concernant un auteur ou une notice.

Le projet Prévu vise ainsi à associer les possibilités du traitement informatique à la recherche en SHS en mettant à la disposition des utilisateurs des outils de visualisation permettant de mieux comprendre des pratiques sociales concentrées sur l'emprunt des livres, tout en proposant d'établir une infrastructure technique permettant à d'autres utilisateurs d'imaginer de nouveaux services ou de nouveaux échanges à partir des données diffusées.

En complémentarité du travail mené avec les visualisations web 2D, le projet propose, sous l'impulsion de l'équipe Spatial Media de l'ENSAD-Lab, une représentation symbolique 3D de la Bibliothèque Universitaire de Paris 8 via le scénario BiblioMémos. Crée par Donatien Aubert sous la direction de François Garnier, BiblioMémos permet à plusieurs utilisateurs de se promener simultanément dans un environnement 3D réalisé avec Unity où ils peuvent consulter des notices de livres rangées dans plusieurs bibliothèques elles-mêmes réparties par sections disciplinaires. La géographie de BiblioMémos est définie dynamiquement selon les données de prêts récoltées par Prévu : plus le nombre de prêts est élevé, plus la surface de la cellule est importante, plus le nombre de co-emprunteurs entre deux sections est grand, plus elles sont proches. BiblioMémos sera disponible d'ici 2016. Une surcouche de scénarisation appelée FictioMémos pensée par Juan Pablo Bertuzzi permettra aux utilisateurs d'approfondir leur exploration et de converser avec des avatars d'auteurs issus du fond de la bibliothèque (comme par exemple G. Deleuze, M. Foucault ou J-L. Borgès).



Résultats attendus. Les résultats en terme de recherche sur un plan informatique concernent notamment :

- la résolution d'un certain nombre de problèmes liés à la révision d'ontologies [3],
- l'augmentation de la pertinence de la prédiction et/ou de la recommandation de communautés (entre autres) en couplant le raisonnement et donc l'inférence de connaissances à partir d'ontologie avec des techniques du big data.

Ce projet permettra également de mettre en place un dispositif de centralisation et de diffusion de données de prêts des bibliothèques universitaires ; nous travaillons actuellement avec le SCD de Paris 10 et avec le Campus Condorcet afin d'enrichir les données de notre plateforme et de proposer des comparaisons entre établissements.

Les scénarios de visualisations nous permettent, dans la perspective des Sciences de l'Information et de la Communication, de mieux comprendre les phénomènes liés à l'emprunt des ouvrages et nous amènent à interroger l'identité des établissements

les hébergeant ainsi que les communautés qu'elles dessinent.

Enfin, l'exploitation des données pour composer un territoire 3D nous invitent à imaginer de nouveaux rapports d'hybridité entre territoires numériques et matériels. D'autre part, ce projet met en exergue les possibilités d'exploitations créatives à partir des données en nous faisant nous demander, par exemple, comment les données sociales peuvent composer les fondements dynamiques de systèmes propices à l'émergence de fictions non linéaires.

Références

[1] F. Clavert. Compte-rendu d'atelier : « Prevu ». http://tcp.hypothses.org/1026, 2015. [Consulté le 20/10/2015].

[2] G. Darquié, J.-C.. Plantin, M. Bourgeois, and I. Breuil. Visualiser les données de bibliothèques : la plateforme Prévu. *Livre post-numérique : historique, mutations et perspectives. Conférence CIDE17*, 2014.

[3] T. Dong, C. Le Duc, P. Bonnot, and M. Lamolle. Tableau-based revision in SHIQ*. *28th International Workshop on Description Logics*, 10(4) :453–457, 2015.

[4] M. Lamolle L. Di Caro, M. Caltaldi and C. Schifanella. It is not what but who you know : A time-sensitive collaboration impact measure of researchers in surrounding communities. *24th International Conference on World Wide Web Companion, WWW'2015*, pages 995–1000, 2015.

Evolution des Procédés et des Objets Techniques : le groupement de recherche EPOTEC

Équipe Ingénierie Systèmes Produits Performances Perception (IS3P)
Institut de Recherche en Communication et Cybernétique de Nantes (IRCCyN)
http://www.irccyn.ec-nantes.fr/fr/equipes/is3p

Florent LAROCHE
florent.laroche@irccyn.ec-nantes.fr
+ 33 2 40 37 69 53

Centre François Viète d'Histoire des Sciences et des Techniques (EA 1161)
Université de Nantes
http://www.cfv.univ-nantes.fr/

Jean-Louis KEROUANTON
jean-louis.kerouanton@univ-nantes.fr
+33 6 83 48 84 24

Le groupement de recherche EPOTEC réunit deux laboratoires issus des Sciences de l'Ingénieur et des Sciences Humaines et Sociales. Chacun a et garde ses propres compétences et champs d'expertises, en 60ème section (Génie Mécanique / Génie des Procédés), et 72ème section (Epistémologie et Histoire des Sciences et des Techniques).

L'IRCCyN et le Centre Francois Viète sont membres du Groupe de Recherche Consortium 3D de la TGIR Huma-Num ayant pour objectif de définir les nouveaux usages de la 3D pour le patrimoine et l'archéologie.

Contexte de recherche

Notre groupe de recherche consiste à réaliser des projets scientifiques en Archéologie Industrielle Avancée. Les travaux visent à allier compréhension de l'histoire, sauvegarde du patrimoine industriel et technique via l'ensemble des outils des Sciences pour l'ingénieur. En effet, notre hypothèse de travail est qu'un objet industriel qu'il appartienne au présent ou passé reste un objet industriel. Celui-ci peut soit être un produit manufacturé diffusé à grande échelle, un objet artisanal unitaire, une ma-

chine ou un outil ayant servi à sa conception ou sa fabrication, voire une usine complète et son processus industriel mis au regard. Dans tous les cas, les outils des sciences pour l'ingénieur contemporaines peuvent être utilisées à des fins de capitalisation, conservation et valorisation de notre patrimoine technique. Les évolutions récentes des questions du numérique et de la modélisation informatique ont fait considérablement avancer ces problématiques de recherche. Dans la lignée des initiatives des professeurs Michel Cotte et Alain Bernard depuis plus de 10 ans, nous déployons actuellement des travaux alliant histoire et ingénierie. L'enrichissement est double : permettre une meilleure compréhension de l'histoire par les outils des sciences pour l'ingénieur ainsi que donner de nouveaux cas d'études pour enrichir les savoir-faire et méthodes des sciences des technologues.

Problématiques et axes de recherche

Pour autant se pose ainsi la question de l'importance de ces nouvelles méthodes dans le processus de patrimonialisation et de valorisation. C'est un possible danger pour certaines démarches comme alibi possible de la destruction, ou de la reconception/modélisation et la valorisation du tout numérique. Notre consortium de recherche s'intéresse tout particulièrement à l'usage de la 3D mais également aux systèmes d'informations géographiques où la mise en œuvre dans les systèmes d'encapsulation de la connaissance type bases de données relationnelles est un des éléments nécessaires pour comprendre et capitaliser notre Histoire.

À l'heure du web et de la réalité virtuelle, nous enrichissons donc les méthodes et outils dans de nombreux domaines connexes. Ces travaux interdisciplinaires sont souvent considérés comme précurseurs car ne s'inscrivant dans aucune discipline type du CNU. Plusieurs thèses sur la base d'allocations ministérielles, CIFRE ou autre financement ont été et sont en cours. De même notre double implication au sein de projets régionaux, nationaux ou de réseaux européens assoit la thématique scientifique émergente.

Description des travaux

La maquette du port de Nantes, réalisée en 1899 par Pierre Auguste Duchesne (1841-1933) pour l'Exposition Universelle de 1900, est imposante par sa taille (9,2 x 1,85 m) mais également par l'étendue du territoire qu'elle représente. En effet, la maquette entière couvre environ 3,5km^2 du port industriel de Nantes. Le musée d'histoire de Nantes a décidé d'améliorer la présentation de cet objet patrimonial, à l'aide d'outils interactifs, pédagogiques et technologiques.

L'objectif de Nantes 1900 est de concevoir une méthodologie structurée et reproductible dédiée à la valorisation scientifique d'objets patrimoniaux. Le dispositif muséographique met à disposition des visiteurs du musée des écrans tactiles situés devant la maquette, permettant, au moyen d'une interface spécifique, de naviguer de différentes manières au sein du corpus de documents (par thématique, points d'intérêt, zones géographiques, etc.). Au-dessus de la maquette, 4 vidéoprojecteurs connectés au serveur informatique permettent de couvrir l'intégralité de la surface pour l'affichage des zones lumineuses : le retour lumineux sur la maquette est dépendant des actions de l'utilisateur. Ce dispositif peut être déployé sur d'autres objets car le nombre d'écrans et de vidéoprojecteurs est variable. L'ensemble est piloté par la partie logicielle, disponible sous licence libre.

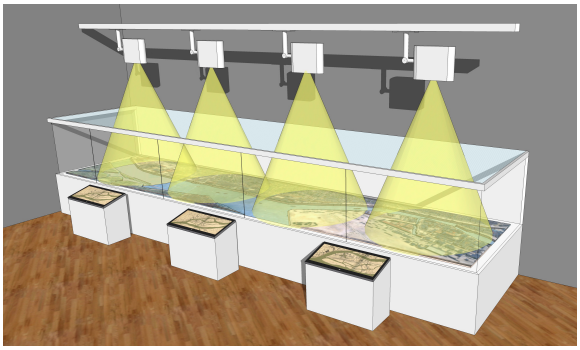
En visite libre, l'interface tactile multi-points se compose de plusieurs éléments visuels et interactifs : image de fond, carrousel de sources historiques... Depuis cet écran, le visiteur peut sélectionner une zone. Lors de la sélection d'une zone,

l'application calcule, grâce au modèle 3D virtuel, le calque lumineux correspondant à afficher sur la maquette. Deux niveaux d'accès à l'informations sont prévus : un premier niveau affiche les quartiers de la ville puis les éléments du quartier sont proposés à la sélection. Dans le cas de la sélection d'un élément non renseigné dans la base de données, un message indique qu'il est possible d'ajouter de l'information.

La base de données, qui capitalise l'ensemble des connaissances historiques relatives au territoire représenté par la maquette, s'appuie sur plusieurs centaines de sources iconographiques (cartes postales, photographies, estampes, etc.) consultables en ligne. Le système va même plus loin puisqu'il permet une évolution dynamique du contenu par la participation de public non-expert qui vient enrichir les connaissances disponibles autour de cette maquette historique.

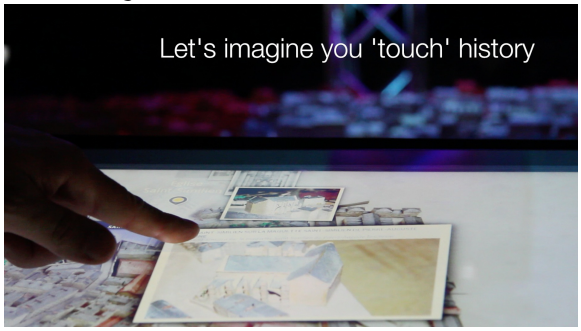
Le modèle conceptuel de la base de données permet de stocker des informations spatiales et temporelles. L'élément de base est une fiche dédiée à un élément de la maquette (bâtiment, bateau, entreprise) ou une thématique particulière (les points, les chantiers navals, la métallurgie...). Des métadonnées simples (titre, description, auteur, mots-clés) sont associés à ces éléments. Une des particularités de la base de données est qu'elle propose des liens entre les éléments identifiés de manière manuelle par les historiens ou de manière automatique à partir de mots-clés. Cela permet au système de proposer aux utilisateurs des pistes exploratoires relatives à la recherche.

L'ensemble du système est géré par un serveur informatique hébergeant la base de données en ligne du projet. La mise à jour des contenus est automatique.



En plus de ce dispositif à destination du public, un système de gestion de contenus dédié aux médiateurs du musée a été mis en place. Il s'agit d'un logiciel d'administration permettant de préparer un scénario de visite à l'avance. L'équipe de médiation peut donc accéder aux fiches de la base de données ainsi qu'à l'ensemble du fond iconographique et programmer un badge RFID avec l'ensemble du contenu nécessaire à la visite guidée. Une fois devant la maquette, le médiateur pourra prendre le contrôle d'un ou plusieurs moniteurs qui afficheront le contenu pré-programmé pour la visite.

Enfin, présenter des contenus gérés dynamiquement extraits d'une base de données accessible en ligne est l'une des principales innovations de ce système. Une mise à jour des contenus de la base de données est automatiquement répercutée dans l'application sans avoir à tout reprogrammer. Il s'agit là d'une avancée novatrice dans le domaine de la muséographie numérique et digitale, une contribution vers les Digital Humanities.



Pour la mise en place de ce projet, différents domaines de compétences ont été mis en jeu. Une des premières étapes, en 2009, a été la numérisation 3D de la maquette effectuée par l'IRCCyN. Depuis 2010, des équipes informatiques œuvrent à la conception d'une base de données adaptée au projet. À partir de 2011, l'équipe « histoire » du projet, notamment du centre François Viète d'épistémologie, d'histoire des sciences et des techniques de l'Université de Nantes, a effectué des recherches poussées parmi les collections des différentes institutions concernées pour constituer le corpus documentaire ; et une réflexion sur les scénarios d'utilisation, l'interactivité et la navigation, a été menée. Afin de créer le système à base de connaissances final, une thèse CIFRE a été montée entre le laboratoire IRCCyN et le musée. Une expérience unique adaptant un dispositif industriel au cadre d'une ins-

titution muséale.

- Lien vers la vidéo de présentation du projet.
- Ouvrage de 90 pages à destination aussi bien du grand public que des experts sur le projet Nantes1900. Téléchargement libre et gratuit. ISBN 978-2-906 519-49-7.

Références

[1] M. Cotte and S. Deniaud. Cao et patrimoine : perspectives innovantes. *Revue Archéologie industrielle en France*, (46) :32–38, 2005.

[2] F. Laroche. *Contribution à la sauvegarde des Objets techniques anciens par l'Archéologie industrielle avancée : Proposition d'un Modèle d'information de référence muséologique et d'une Méthode inter-disciplinaire pour la Capitalisation des connaissances du Patrimoine technique et industriel*. PhD thesis, 2007.

[3] F. Laroche, A. Bernard, and M. Cotte. 3d digitalization for patrimonial machines. pages 397–408, 2007.

[4] N. Ma, F. Laroche, B. Hervy, and J.-L. Kerouanton. Virtual conservation and interaction with our cultural heritage : Framework for multi-dimension model based interface. *Digital Heritage International Congress, Marseille*, 2013.

[5] F. Le Pavic, F. Laroche, and J.-L. Kerouanton. Vers une extension de la modélisation d'entreprise pour la rétro-conception de sites industriels disparus : cas d'étude de l'arsenal de la marine de lorient. *Conférence Génie Industriel, CIGI*, 2013.

Description, modélisation et détection automatique des chaînes de référence (DEMOCRAT)

Laboratoire Lattice
UMR 8094, CNRS/ENS Ulm/Paris 3.
http://www.lattice.cnrs.fr/

Frédéric LANDRAGIN
frederic.landragin@ens.fr
+33 1 58 07 66 20

Autres partenaires du projet

- Laboratoire LiLPa, « Linguistique, Langues, Parole », EA 1339, Strasbourg.
- Laboratoire ICAR, « Interactions, Corpus, Apprentissages, Représentations », UMR 5191, CNRS/ENS Lyon/Lyon 2.

Thématique générale de l'équipe

Le Lattice est un laboratoire dont les recherches concernent essentiellement la linguistique et le traitement automatique des langues. Les tutelles du laboratoire sont le CNRS, l'Ecole normale supérieure et l'Université Paris 3 Sorbonne Nouvelle, ce qui met le laboratoire au contact direct de nombreuses équipes couvrant tout le spectre de la recherche en littérature et sciences sociales. Le laboratoire est en outre membre de deux labex : le labex Transfers d'une part (labex qui regroupe l'ensemble des laboratoires de recherche en lettres et sciences humaines de l'ENS Ulm) et le labex EFL d'autre part (labex « Empirical Foundations of Linguistics », qui regroupe la plupart des laboratoires de sciences du langage de la COMUE Sorbonne Paris Cité). Du fait de cet environnement pluridisciplinaire très riche, le laboratoire est sollicité depuis plusieurs années par des collègues d'autres disciplines ayant des besoins particuliers pour structurer, analyser ou valoriser de grandes masses de données textuelles dans des domaines très variés.

Certaines demandes peuvent être traitées à partir d'outils standards. Par exemple, un besoin très répandu concerne la mise en ligne de documents uniquement disponibles sous forme papier. Au-delà de l'aspect numérisation, la valorisation de ces documents passe généralement par une ré-analyse du contenu, c'est-à-dire l'extraction et la normalisation

du vocabulaire technique, sa structuration, la mise en place d'index structurés, etc. Les outils d'extraction de termes, de structuration des connaissances et de mise en place d'hypertextes sont alors sollicités. Une adaptation au contexte est toutefois systématiquement nécessaire, ainsi que la collaboration avec des experts du domaine visé. Des expériences récentes ont par exemple eu lieu avec des textes d'archéologie (collaboration autour du projet PEPS EITAB porté par le laboratoire AOROC de l'ENS) : le repérage de termes du domaine de l'archéologie, ainsi que la reconnaissance des entités nommées sont très utiles [Mélanie-Bécquet, 2015]. A partir du PDF reflétant le texte original, il est possible de produire des bases de données indexant les découvertes archéologiques par commune, par type ou par période, ce qui permet de concevoir des requêtes d'une richesse incomparable par apport à un simple support papier. Un expert du domaine est bien évidemment nécessaire pour valider les résultats des outils automatiques, structurer les données (repérer les synonymes, les hyperonymes, structurer les connaissances) mais le traitement d'un ouvrage complet (c'est-à-dire le passage du papier au support informatique) est ainsi possible en quelques jours à peine.

Un projet peut-être plus original est né d'une collaboration avec nos collègues londoniens du laboratoire d'Humanités numériques de University College London (UCL). Cette université dispose d'une collection de manuscrits de Jeremy Bentham dont une partie est toujours inédite. Ces manuscrits inédits ont été transcrits par une équipe de volontaires, ce qui a permis d'obtenir un ensemble de 30 000 documents (fichiers informatiques au format XML) dont le contenu était largement inconnu ou, tout au moins, n'a encore fait l'objet d'aucune étude systématique. La masse de documents à traiter, même si chaque fichier est très bref, rend difficile une approche purement manuelle. La stratégie développée à consister à normaliser les contenus, extraire un certain nombre d'expressions clés, puis, sur cette base, à proposer différentes visualisations correspondant à différents regroupements de documents, chaque regroupement pouvant en outre recevoir une étiquette rendant compte de son contenu. Les experts ont ainsi un accès beaucoup plus facile au corpus, même si les documents restent à analyser plus finement par des experts du philosophe. Il ne s'agit en aucun cas de se substituer au spécialiste du domaine mais de lui donner les clés pour accéder rapidement à l'information qu'il cherche, aux documents les plus proches ou aux principaux thèmes abordés dans un texte, en l'occurrence dans un grand corpus inédit.

La linguistique est elle-même demandeuse d'automatisation, pour l'annotation de gros corpus en particulier. Le Lattice est impliqué dans plusieurs projets visant à ajouter des informations de nature diverses (morphosyntaxiques, syntaxiques, voire sémantiques) sur des corpus de langues diverses. Citons par exemple le projet DFG-ANR SRCMF (Syntactic Reference Corpus of Medieval French) qui visait à produire un corpus d'ancien français annoté au niveau syntaxique afin de faire progresser notre connaissance de l'ancien français et surtout de son évolution au cours des siècles [9]. Le corpus résultant (corpus de plus de 200 000 mots, interrogeable en ligne) permet aujourd'hui des analyses systématiques et surtout quantifiées qui n'étaient pas possibles jusqu'ici. Un autre projet, financé par PSL (Paris Sciences et Lettres) cette fois-ci, se met actuellement en place pour l'annotation de textes en hébreu rabbinique, en ancien français et dans plusieurs langues finno-ougriennes. Ce projet appelé Lakmé et mené en collaboration avec des partenaires de l'Ecole Pratique des Hautes Etudes (EPHE) et de l'Ecole Nationale des Chartes (ENC) demande l'exploration de techniques d'annotation innovantes : l'hébreu comme les langues finno-ougriennes sont des langues « agglutinantes », ce qui signifie que les mots comportent une morphologie très riche et porteuse d'informations essentielles pour l'annotation. Les techniques classiques développées initialement pour l'anglais ne sont pas opérationnelles dans ce cadre et une approche complètement nouvelle doit être mise au point. Ce projet a, de plus, un but directement pratique : il s'agit de faciliter le travail des experts de ces langues dans leur exploration de grands corpus, afin de mettre au jour des tendances, préciser le sens des mots, identifier les évolution syntaxiques ou sémantiques, etc. Enfin, dans le cas des langues finno-ougriennes, il s'agit aussi de produire des corpus annotés pérennes pour des langues parfois gravement en danger. Le projet contribue donc à la documentation de ces

langues, une entreprise urgente alors que chaque année des langues disparaissent, trop souvent sans laisser de trace.

Un projet en cours de démarrage est le projet ANR Democrat, porté par Frédéric Landragin, autour de l'analyse des chaînes de référence dans les textes. Ce projet fait l'objet de la section suivante.

D'autres collaborations concernent enfin des questions de recherche plus pointues pour lesquelles il est nécessaire de concevoir des développements propres. C'est notamment le cas des sciences sociales, de plus en plus souvent confrontées à de grandes collections de documents dont il faut tirer du sens sans en dénaturer le contenu, ni en offrir une vision trop biaisée [7]. Le programme exploratoire PoliInformatics allait dans ce sens : il s'agissait, à partir d'une masse diversifiée de documents sur la crise financière de 2008-2009 aux Etats-Unis, de produire des analyses automatiques permettant à des experts du domaine d'identifier les principaux acteurs de la crise, leurs rôles et surtout, dans la mesure où le corpus incluait essentiellement des textes post-crise (interviews de banquiers, de conseillers gouvernementaux, de membres d'organismes de régulation, etc. devant le Sénat américain par exemple), d'essayer d'identifier des points de vue consensuels ou contradictoires. Une tâche aujourd'hui relativement courante consiste à élaborer des réseaux d'acteurs, et de rendre compte graphiquement de leurs connexions sur le plan du contenu, des arguments et des opinions [8]. Ces travaux nous semblent intéressants car ils posent des questions d'analyse qui sont à la limite de l'état de l'art (analyse de l'argumentation, des prises de position, etc.). De très nombreuses annonces commerciales prétendent avoir résolu ce type de questions ou des problèmes similaires (comme l'analyse de l'opinion) mais ces travaux sont souvent très sommaires et reposent généralement sur des listes de mots clés prédéfinis sans tenir compte de l'application, du domaine ou du contexte. Ce type de recherche pose aussi des questions importantes sur le plan de l'éthique mais il nous semble important de les aborder pour contribuer à éclairer le débat public.

Les grands débats de société (comme les échanges sur le changement climatique), les élections locales ou nationales ou la production scientifique elle-même [6] forment autant de données textuelles massives, peu structurées, que les outils automatiques peuvent aider à analyser. Il y a alors un réel apport des techniques de traitement automatique des langues : il ne s'agit plus de mettre en ligne des données, ni de les enrichir mais bien d'en extraire des informations nouvelles qu'il serait très difficile d'observer directement, sans outil numérique adapté. En ce sens, la recherche en Humanités numériques est apparue comme une chance à saisir pour le Lattice, à un moment où les outils de TAL semblent suffisamment matures pour être utilisables dans ce contexte nouveau, malgré leurs limites et leurs défauts. En retour, les Humanités numériques proposent un cadre permettant d'améliorer les outils existants et d'en concevoir de nouveaux, voire de concevoir des problématiques nouvelles.

Le projet DEMOCRAT

Financé par l'ANR dans le cadre de l'appel à projets générique 2015, défi 8 « Sociétés innovantes, intégrantes et adaptative », le projet DEMOCRAT fait suite à un projet PEPS INS2I-INSHS (CNRS) intitulé MC4 (« Modélisation Contrastive et Computationnelle des Chaînes de Coréférence »), porté par Frédéric Landragin entre 2011 et 2013, qui a fédéré des chercheurs de trois unités de recherche travaillant tous sur les chaînes de référence : des chercheurs du Lattice, d'ICAR et de LiLPa. C'est dans ce cadre que les collaborations entre ces trois laboratoires se sont amorcées. Le projet MC4 avait pour objectif un double volet : linguistique descriptive, d'une part, et linguistique outillée et automatique, d'autre part. Il réunissait ainsi des spécialistes de linguistique du français contemporain et médiéval, reconnus pour leurs compétences dans le domaine de la linguistique référentielle, et des chercheurs en linguistique informatique, spécialisés également dans les questions de référence et de saillance référentielle. Les résultats principaux de ce projet, outre la mise en place de DEMOCRAT, sont d'une part la mise en œuvre d'un corpus annoté en chaînes de référence – corpus de taille modeste paru en juin 2015 sur la plateforme ORTOLANG – et d'autre part un ensemble d'articles de recherche regroupés dans un numéro thématique de la revue Langages

(cf. bibliographie).

Objectifs

DEMOCRAT vise à développer les recherches sur la langue et la structuration textuelle du français via l'analyse détaillée et contrastive des chaînes de référence (instanciations successives d'une même entité) dans un corpus diachronique de textes écrits entre le 9ème et le 21ème siècle, avec des genres textuels variés. Le projet mettra à disposition de la communauté scientifique : (i) un modèle intégré et discursif de la référence et de la composition des chaînes de référence ; (ii) un corpus annoté qui puisse servir de corpus de référence et de corpus d'apprentissage pour les campagnes d'évaluation internationales portant sur la coréférence ; (iii) un outil d'annotation, d'aide à l'annotation et de manipulation des données annotées, et (iv) un système de détection automatique des coréférences. Le corpus annoté manuellement en chaînes de référence aura une taille de 1 million de mots, soit environ 100 000 maillons annotés.

Résultats attendus

Dans notre société numérique, les corpus de textes s'avèrent essentiels pour les recherches scientifiques, la diffusion des connaissances et du patrimoine, la pérennisation et la standardisation des données. DEMOCRAT contribue aux humanités numériques en proposant un corpus numérique riche (varié et diachronique), pour la langue française, annoté en fonction d'analyses linguistiques relevant d'une dimension encore peu explorée, à la fois sémantique et pragmatique. En apportant de nouvelles connaissances et données sur la langue, ce corpus et le modèle associé sont destinés à : (i) nourrir l'ensemble des applications de traitement automatique des langues – résumé automatique, traduction automatique, simplification de textes, fouille de texte, web sémantique, dialogue humain-machine, etc. – (ii) renforcer la place du français dans le monde via notamment l'intégration du français dans des défis scientifiques d'ampleur internationale (SemEval, CoNLL), et (iii) apporter à toutes les disciplines connexes à la linguistique, comme la didactique, la psycholinguistique, l'enseignement du français et des langues, de nouvelles connaissances sur l'accès aux entités d'un texte, sur le fonctionnement des chaînes de référence et leur importance au niveau de la structuration et de la cohésion textuelle.

Si le corpus DEMOCRAT relève pleinement des humanités numériques, la conception d'un outil de détection automatique des chaînes de référence relève, elle, de l'intelligence artificielle – tout en s'appuyant sur le corpus. La notion de chaîne de référence est l'un des éléments clés de la cohésion et de la cohérence textuelles : comprendre qu'un texte parle en continu d'une entité humaine (Barack Obama), organisationnelle (l'ONU) ou abstraite (la justice) permet ipso facto d'en déterminer le thème central, et partant, d'en faciliter le traitement, dont la mémorisation, les modes de restitution de textes comme le résumé, la paraphrase, l'indexation et l'extraction d'informations, sans oublier la traduction. C'est dans une telle optique de détermination du thème central que les moteurs de recherche s'intéressent actuellement à la détection automatique de chaînes de référence. De manière plus originale, la teneur des chaînes de référence donne des indices sur les opinions ou prises de position du locuteur vis-à-vis de ses objets de discours, et aide ainsi les travaux centrés sur la détection d'opinions. En 2011, dans un article intitulé « The Winograd Schema Challenge » présenté lors d'une conférence d'intelligence artificielle, la détection de chaînes de références a été mise en avant comme tâche pouvant remplacer le célèbre test d'intelligence d'Alan Turing, via la notion de schéma de Winograd, notion regroupant 1 ou 2 coréférences dans un couple de phrases bien choisies. Les retombées attendues de DEMOCRAT au niveau du traitement automatique des langues relèvent donc pleinement de l'intelligence artificielle.

Références

[1] Adèle Désoyer, Frédéric Landragin, Isabelle Tellier, Anaïs Lefeuvre, and Jean-Yves Antoine. Les coréférences à l'oral : une expérience d'apprentissage automatique sur le corpus ancor. *Traitement Automatique des Langues*, 2(55) :97–121, 2014. halshs-01153297.

[2] Frédéric Landragin. Une procédure d'analyse et d'annotation des chaînes de coréférence dans des textes écrits. *Corpus*, (10) :61–80, 2011. .

[3] Frédéric Landragin, Thierry Poibeau, and Bernard Victorri. ANALEC : a new tool for the dynamic annotation of textual data. In *Eighth International Conference on Language Resources and Evaluation*, pages 357–362, 2012. halshs-00698971.

[4] Frédéric Landragin and Catherine Schnedecker, editors. *Les chaînes de référence. Langages*, number 195, 2014.

[5] Frédérique Mélanie-Bécquet, Johan Ferguth, Katherine Gruel, and Thierry Poibeau. Archaeology in the digital age : From paper to databases. In *Actes de la Conférence Digital Humanities 2015*, 2015. hal-01173964.

[6] Elisa Omodei, Yufan Guo, Jean-Philippe Cointet, and Thierry Poibeau. Analyse discursive automatique du corpus acl anthology. In *1ème conférence Traitement Automatique des Langues Naturelles*, 2014. hal-01056143.

[7] Thierry Poibeau. Le traitement automatique des langues pour les sciences sociales, quelques éléments de réflexion à partir d'expériences récentes. *Réseaux*, (188), 2014.

[8] Thierry Poibeau and Pablo Ruiz. Generating navigable semantic maps from social sciences corpora. In *Actes de la Conférence Digital Humanities 2015*, 2015. hal-01173963.

[9] S. Prévost. Diachronie du français et linguistique de corpus : une approche quantitative renouvelée. *Langages*, (197) :23–45, 2015.

Observatoire de la vie littéraire : le Labex OBVIL et l'équipe ACASA

Labex OBVIL, équipe ACASA, LIP6
Université Pierre et Marie Curie
http://obvil.paris-sorbonne.fr/

Jean-Gabriel GANASCIA
jean-gabriel.ganascia@lip6.fr
+33 1 44 27 37 27

Thématique générale

Le Labex OBVIL – OBservatoire de la VIe Littéraire – fait collaborer les équipes de littérature de l'université Paris-Sorbonne avec l'équipe ACASA du LIP6. L'activité de ce Labex porte sur le versant littéraire des humanités numériques.

Pour cela nous développons, dans le cadre du LIP6, un certain nombre d'opérateurs destinés à susciter de nouvelles interprétations en repérant des phénomènes de reprise textuelle ou lexicale, ou encore des phénomènes stylistiques.

Ces développements recourent à des techniques d'intelligence artificielle tout en faisant référence à des théories littéraires, par exemple à la *génétique textuelle* ou aux théories de l'*intertextualité*, de l'*hypertextualité* et de la *paratextualité*.

Les fondements théoriques

Les humanités numériques reconduisent, avec les ressources de l'informatique, les disciplines d'érudition classiques qui avaient pour vocation d'étudier les œuvres humaines.

Notre approche [11, 12] se fonde sur l'opposition introduite par quelques philosophes néo-kantiens, en l'occurrence Heinrich Rickert et Ernst Cassirer, au début du XX^e siècle, entre les *sciences de la nature* et les *sciences de la culture*. Ceux-ci montrent que tant les sciences de la nature que les sciences de la culture sont des sciences empiriques, c'est-à-dire fondées sur des faits, mais que, des unes aux autres, la logique diffère. Les sciences de la nature visent à construire des lois générales par induction à partir d'observations, en oubliant les cas particuliers, tandis que les sciences de la culture se centrent sur les cas particuliers pour leur donner sens en les expliquant.

L'informatique et l'intelligence artificielle trans-

forment les *sciences de la nature* en donnant naissance aux *e-sciences*, qui construisent des lois générales à partir de données issues d'observations, en ayant recours à des procédures d'apprentissage automatique qui font de l'*induction* au sens logique, c'est-à-dire qui passent du particulier au général. De façon analogue, avec la numérisation des contenus, les *sciences de la culture* se modifient.

Cependant, si l'induction et la recherche de lois générales, prennent une importance centrale dans les sciences de la nature, il en va tout autrement dans les sciences de la culture, qui visent à comprendre les cas particulier, ou plus exactement à leur donner sens. A cette fin, on adopte une démarche fondée sur l'*abduction*, c'est-à-dire sur la recherche d'explications au regard de théories générales. Pour aider à cette recherche d'explications, nous avons développé et déployé un certain nombre d'outils que nous présentons ci-dessous et qui portent sur la détection d'homologies, sur l'extraction de motifs récurrents, sur l'indexation et l'annotation automatiques de corpus et, enfin, sur la cartographie de contenus.

Réécritures, réemplois, citations et reformulations

Beaucoup de théories de la littérature insistent sur les reprises textuelles, qu'il s'agisse des réécritures d'un même texte, par le même auteur qui, pour paraphraser Boileau, *cent fois sur le métier remet son ouvrage*, ou des emprunts plus ou moins conscients, soit littéraux, soit conceptuels. Avec les années, nous avons développé trois outils destinés à repérer différents types de reprises.

MEDITE

Le logiciel *MEDITE* (Machine EDITE) [10] a été conçu dans le cadre du projet EDITE (Étude Diachronique et Interprétative de TExtes) en collaboration avec l'ITEM (Institut des Textes et Manuscrits Modernes). Ce laboratoire spécialisé dans la genèse textuelle examine les brouillons et les différents "avant-textes" des grands auteurs, c'est-à-dire les états des œuvres avant publication. A cette fin, MEDITE détecte automatiquement les transformations qui font passer d'une version à une autre, en particulier les *insertions*, les *suppressions*, les *remplacements* et, surtout, les *déplacements*. Dans le passé, de nombreux travaux portèrent sur la comparaison de textes. Cependant, le problème ne reçoit pas de solution optimale : il faut procéder à des compromis et des choix qu'autorise l'approche heuristique, fondée sur des techniques d'intelligence artificielle couplées à des principes d'algorithmique des chaînes [4], que nous avons employée.

Le logiciel MEDITE est en libre accès.

Phœbus

Le logiciel *Phœbus* (Projet d'Hypertexte de l'Œuvre de Balzac utilisant des Similarités) [13] a été conçu dans le cadre d'un projet interdisciplinaire monté par des spécialistes de Balzac et par l'équipe ACASA du LIP6 afin de repérer automatiquement des réutilisations et des citations sur de grandes masses de textes. A cette fin, il fait appel à des principes utilisés pour la détection de plagiat qu'il assouplit, afin de déceler des reprises approximatives. Une première version de ce projet a été développée grâce au financement d'un PEPS (Projets Exploratoires Premier Soutien) alloué par le CNRS. Un projet financé par l'ANR prend désormais le relais (projet *Phœbus*).

Le logiciel *Phœbus* est en libre accès.

DeSeRT

DeSeRT (Détection Sémantique de Reformulations et de Topiques) [14] est un moteur de recherche sémantique qui découpe les textes en blocs partiellement recouvrant, puis qui indexe chaque bloc avec les lemmes des mots significatifs (noms, verbes ou adjectifs) présents dans ce bloc. DeSeRT repère ensuite des conjonctions récurrentes de termes en ayant éventuellement recours à un thésaurus et à un dictionnaire analogique. Cela permet de repérer sur de grandes masses de textes les arguments qui se répètent et de localiser les blocs où ces arguments apparaissent.

Ce travail se poursuit au sein d'une collaboration transatlantique conduite avec l'ARTFL (*American and French Research on the Treasury of the French Language*) de l'université de Chicago dans le cadre

du projet "Use and Reuse" financé par la fondation Mellon, la FMSH (Fédération des Maisons des Sciences de l'Homme) et le Labex OBVIL.

Le logiciel *DeSeRT* est en libre accès.

Extraction de motifs

Dans le passé, l'étude du style individuel des auteurs a essentiellement recouru à des statistiques lexicales. Nous voulons, dans le cadre de l'équipe ACASA, développer une stylistique syntaxique, en extrayant automatiquement des motifs syntaxiques récurrents, et une stylistique sémantique, en inventoriant et en caractérisant les figures de comparaison.

EReMoS : Extraction et Recherche de Motifs Syntaxiques

Relevant essentiellement de la fouille de données séquentielles, les travaux, conduits par Amine Boukhaled [3] dans le cadre de sa thèse de doctorat, portent sur l'extraction et la recherche de motifs syntaxiques à partir de grandes masses de textes. Ils étendent, à plusieurs égards, les réflexions initiées par Jean-Gabriel Ganascia il y a une quinzaine d'années [8] avec le *Littératron*. Cela a donné naissance au logiciel *EReMoS* accessible en ligne. Ces extractions automatiques de motifs syntaxiques donnent aussi prise à l'étude de la répartition statistique des motifs à l'intérieur des œuvres.

Une application d'*EReMoS* [6, 7], conduite en lien étroit avec les équipes de littérature de l'université Paris-Sorbonne, porte sur l'analyse des motifs stylistiques caractéristiques des différents personnages des pièces de théâtre de Molière, ce qui permet d'associer à chaque caractère (*Sganarelle*, *Arlequin*, etc.) un style d'expression particulier.

Le logiciel *EReMoS* est en libre accès.

Extraction de Motifs Graduels

Une étude conduite avec Amal Oudni, Marine Riguet, Mohamed Amine Boukhaled et Gauvain Bourgne porte sur l'extraction de motifs graduels pour l'analyse de tendances dans des corpus de critique littéraire.

Extraction de figures de comparaison

Dans le cadre de ses travaux de thèse, Suzanne Mpouli [16, 17] procède à l'extraction automatique de figures de comparaison en déterminant, pour chacune, le *comparant*, le *comparé* et le *motif*. Le but est à la fois d'établir automatiquement un dictionnaire des comparaisons, avec leurs sources, et d'évaluer, pour chaque auteur, la richesse des comparaisons, leur nombre et leur répartition.

Le système est en cours d'évaluation à la fois sur le français et sur l'anglais, ce qui requiert une annotation des comparaisons sur de grands corpus textuels (cf. paragraphe « Annotations collaboratives de Bandes Dessinées et de comparaisons »).

Indexation et annotations

Une part importante de l'activité dans le versant littéraire des humanités numériques porte sur l'établissement d'éditions numériques de textes littéraires, ce qui suppose, outre la numérisation des contenus, une indexation sémantique, soit manuelle, ce qui est très fastidieux et coûteux, soit automatique. Or, l'intelligence artificielle peut aider à automatiser certaines tâches d'indexation [1].

Reconnaissance d'entités nommées

La première difficulté porte sur ce que l'on a coutume de caractériser comme étant la reconnaissance d'entités nommées. Il s'agit d'associer à chaque nom propre (ou à chaque nombre) sa catégorie (ou son unité, dans le cas de nombres). Dans le cas des noms propres, cela signifie que l'on distingue entre une organisation, une personne et un lieu, puis parmi les lieux entre une ville, une île ou un pays, etc. De nombreux travaux de recherche portent sur ce sujet. Ils font appel à des techniques d'apprentissage machine supervisé qui requièrent l'emploi de gros corpus annotés. Or, on ne dispose que de très peu d'annotations sur les corpus littéraires anciens qui nous intéressent. Pour résoudre ce problème, nous en sommes venus, avec Alaa Abi-Haidar, à développer l'algorithme *UNERD* (*Unsupervised Named Entity Recognition and Disambiguisation*) qui fait appel à des ressources textuelles

externes, en particulier à des dictionnaires, et à une désambiguïsation [2, 15].

REDEN : Résolution et Désambiguïsation d'Entités Nommées

Une fois les entités nommées reconnues, il faut établir des liens entre chacune de leurs occurrences et les entités du monde. Or, des ambiguïtés subsistent qui font que, selon le contexte, le même mot renvoie à des entités qui, tout en étant de la même catégorie, par exemple de celle des personnes, apparaissent différentes. Ainsi, selon le contexte, Dumas peut être une ville des États-Unis, une station de métro ou une personne. Une fois identifié comme paronyme d'une personne, ce mot peut désigner un grand nombre d'individus différents, et même lorsqu'on se limite aux écrivains, il reste Adolphe Dumas, Alexandre Dumas, Philippe Dumas, Roger Dumas, Charles Dumas, ... Le logiciel REDEN [5] conçu par Carmen Brando, Francesca Frontini et Jean-Gabriel Ganascia cherche à désambiguïser les entités nommées une fois qu'elles ont été identifiées et classifiées. Pour cela, grâce aux ressources du Web des données (par exemple DBpedia, BNF, ...), il établit les liens entre toutes les entités présentes dans le même texte, puis il représente ces liens sur un graphe. Il recourt ensuite à la notion mathématique de *centralité dans les graphes* pour discriminer entre les différentes acceptions et attribuer ainsi un identifiant unique dans le Web de Données (c'est-à-dire un URI). Dans un deuxième temps, cet URI permet d'extraire automatiquement des informations publiées en RDF sur le Web puis d'enrichir la connaissance sur ces entités. Il est ainsi possible de visualiser les textes sous des angles multiples qui facilitent la "lecture à distance" (*Distant Reading* en anglais).

Annotations collaboratives de Bandes Dessinées et de comparaisons

Pour être validées, beaucoup de techniques (détection de comparaison (cf. paragraphe « Extraction de figures de comparaison »), reconnaissance d'entités nommées (cf. paragraphe précédent), etc.) requièrent des gros corpus annotés difficiles à acquérir. De même, pour entraîner des procédures d'apprentissage machine, ce qui apparaît fort utile pour différentes tâches comme la reconnaissance d'entités nommées (cf. paragraphe du même nom), on a besoin d'annotations textuelles. Or, l'établissement de ces annotations apparaît pénible et fastidieux lorsque seul un petit groupe de personnes s'en charge. Afin de s'affranchir de ce qui pourrait apparaître comme un obstacle, il faut impliquer plus de gens dans ces tâches. Cette idée nous a tout naturellement conduit à mettre en place des techniques d'acquisition collaborative (*crowdsourcing*) utiles pour l'annotation collective des comparaisons, afin de tester les performances du logiciel développé par Suzanne Mpouli, et pour l'annotation des bandes dessinées, dans le cadre des projets investissement d'avenir *iManga* et *iiBD*. La réflexion théorique sur ces techniques d'acquisition collaborative fait actuellement l'objet des travaux de thèse de Mihnea Tufiș.

Cartographie de corpus : îles de mémoire

Il apparait souvent difficile d'appréhender des contenus numérisés tant ceux-ci apparaissent impalpables. Afin de leur donner une épaisseur tangible et d'offrir prise à notre intuition, nous tentons, depuis plusieurs années [9], de les cartographier sous forme de territoires imaginaires, en l'occurrence d'îles. Pour cela, nous partons de leur structure, donnée par des tables des matières d'ouvrages ou par le squelette d'ontologies indexant des encyclopédies, des fonds d'archives ou des corpus. À titre d'illustration, nous avons cartographié une île à partir de l'ontologie InPhO (Indiana Philosophy Ontology) établie par l'université d'Indiana pour indexer l'encyclopédie philosophique de Standford. Le lecteur intéressé peut se faire une idée du résultat en se rendant ici. Au plan théorique, ces travaux reposent sur le même principe que les anciens arts de mémoire, à savoir sur l'emploi d'une mise en espace pour faciliter la mémorisation des contenus. Cette référence aux arts de mémoire, nous a conduit à appeler *îles de mémoire* ces cartes.

Ces travaux ont fait l'objet de la thèse de Bin Yang, ainsi que des recherches que l'équipe ACASA a poursuivies dans le cadre du projet investissement

d'avenir *LOCUPLETO* [18]. Aujourd'hui, notre logiciel [19] permet de construire automatiquement des îles imaginaires à partir de n'importe quel fichier *owl*.

Références

[1] Alaa Abi Haidar, Mihnea Tufiș, and Jean-Gabriel Ganascia. From inter-annotation to intra-publication inconsistency. In Carl Hewitt and John Woods, editors, *Inconsistency Robustness*, number 52 in Studies in Logic, chapter Part 3, applications. College Publications, May 2015. ISBN-13 : 978-1848901599.

[2] Alaa Abi Haidar, Bin Yang, and Jean-Gabriel Ganascia. Extracting and visualizing named entities using interactive streamgraphs – a case study on first world war data. In *Join CSDH/SCHN & ACH Digital Humanities Conference*, Ottawa, Canada, June 2015.

[3] Mohamed-Amine Boukhaled, Francesca Frontini, and Jean-Gabriel Ganascia. Une mesure d'intérêt à base de surreprésentation pour l'extraction des motifs syntaxiques stylistiques. In *22ème Conférence sur le Traitement Automatique des Langues Naturelles*, Caen, France, 2015.

[4] Julien Bourdaillet and Jean-Gabriel Ganascia. Practical block sequence alignment with moves. In *proceedings of the* 1^{st} *International Conference on Language and Automata Theory and Applications (LATA)*, Tarragona, Spain, 2007. LNCS.

[5] Carmen Brando, Francesca Frontini, and Jean-Gabriel Ganascia. Disambiguation of named entities in cultural heritage texts using linked data sets. In Tadeusz Morzy, Patrick Valduriez, and Ladjel Bellatreche, editors, *New Trends in Databases and Information Systems*, number 539 in Communications in Computer and Information Science, pages 505–514. Springer International Publishing, 2015. http://link.springer.com/chapter/10.1007/978-3-319-23201-0_51.

[6] Francesca Frontini, Amine Boukhaled, and Jean-Gabriel Ganascia. Linguistic pattern extraction and analysis for classic french plays. In *Digital Humanities Conference (DH 2015)*, Sydney, Australia, July 2015.

[7] Francesca Frontini, Mohamed Amine Boukhaled, and Jean-Gabriel Ganascia. Linguistic Pattern Extraction and Analysis for Classic French Plays. In *Journée ConSciLa (Confrontations en Sciences du Langage)*, Paris, France, January 2015.

[8] Jean-Gabriel Ganascia. Extraction of recurrent patterns from stratified ordered trees. In Luc De Raedt and Peter Flach, editors, *proceedings of European Conference on Machine Learning (ECML 2001)*, number 2167 in LNAI. Springer, LNAI, 2167, 2001.

[9] Jean-Gabriel Ganascia. Recit : représentation cartographique et insulaire de textes. In *Colloque International sur la Fouille de Texte, (CIFT 2004)*, la Rochelle, juin 2004. http://archivesic.ccsd.cnrs.fr/sic_00001258/en/.

[10] Jean-Gabriel Ganascia. A unilingual text aligner for humanities. application to textual genetics and to the edition of text variants. In *Supporting Digital Humanities (SDH 2011)*, Copenhagen, November 2011.

[11] Jean-Gabriel Ganascia. Les big data dans les humanités. *Revue Critique*, (819-820) :627–636, 2015.

[12] Jean-Gabriel Ganascia. The logic of the big data turn in digital humanities. *Frontiers in Digital Humanities*, 2(7) :1 – 5, 2015.

[13] Jean-Gabriel Ganascia, Pierre Glaudes, and Andrea Del Lungo. Automatic detection of reuses and citations in literary texts. *Literary and Linguistic Computing*, 29(3) :412–421, 2014.

[14] Jean-Gabriel Ganascia and Chiara Mainardi. Crossed semantic analysis of literary texts with *desert*. rapport interne, LIP6, Université Pierre et Marie Curie, Paris, France, octobre 2015.

[15] Yusra Mosallam, Alaa Abi Haidar, and Jean-Gabriel Ganascia. Unsupervised Named Entity Recognition and Disambiguation : An Application to Old French Journals. In Petra Perner, editor, *Advances in Data Mining. Applications and Theoretical Aspects - 14th Industrial*

Conference, ICDM 2014, St. Petersburg, Russia, July 16-20, 2014. Proceedings., volume 8557 of *Lecture Notes in Computer Science*, pages 12–23. Springer, July 2014.

[16] Suzanne Mpouli and Jean-Gabriel Ganascia. Extraction et analyse automatique de comparaisons et de pseudo-comparaisons pour la détection des comparaisons figuratives. In *Actes de la 22e conférence sur le Traitement Automatique des Langues Naturelles (TALN 2015)*, pages 621–627, 2015.

[17] Suzanne Mpouli and Jean-Gabriel Ganascia. Investigating the stylistic relevance of adjective and verb simile markers. In *Abstract Book of Corpus Linguistics 2015*, pages 243–244, 2015.

[18] Bin Yang and Jean-Gabriel Ganascia. Cartographie des connaissances dans les humanités numériques par Îles de mémoires – une démonstration. In *Atelier visualisation d'information, fouille visuelle de données et nouveaux challenges en Big data et Humanités numériques, IHM*, Lille, France, 2014.

[19] Bin Yang and Jean-Gabriel Ganascia. Creating knowledge maps using memory islands. In *ACM/IEEE Joint Conference on Digital Libraries (JCDL 2014) and International Conference on Theory and Practice of Digital Libraries*, volume 2014, pages 15–22, London, United Kingdom, 2014.

AfIA
Association française
pour l'Intelligence Artificielle

Compte-rendu de journées, événements et conférences

Journée commune IA et TAL, 17 mars 2016

Par **Philippe MULLER**
IRIT
Université de Toulouse
Philippe.Muller@irit.fr

Le traitement automatique des langues (TAL), visant à réaliser des tâches relevant de la cognition humaine, a de longue date des liens avec l'intelligence artificielle (IA). Le TAL travaille notamment avec des représentations des informations ou des connaissances mises en jeu dans les traitements qu'il effectue. Ces informations et connaissances portent généralement sur la langue concernée ou sur le monde. L'IA étudie notamment les formalismes de représentation de connaissances et les méthodes visant à acquérir ces représentations ou à raisonner sur elles.

L'objectif de cette journée, qui a rassemblé plus de soixante personnes, était de renforcer les liens qui existent entre ces deux domaines, avec un accent sur la thématique des représentations utilisées en TAL et la façon de les acquérir. Quels types de représentations sous-tendent les modèles aux niveaux lexical, morphologique, syntaxique, sémantique, et quels raisonnements vont de pair avec eux ? Dans cette perspective comment sont utilisés les modèles et calculs proposés en IA et quelles pistes sont prometteuses pour le TAL ?

Les exposés étaient les suivants :

« Représentations lexicales pour l'analyse sémantique : leçons tirées de l'expérience d'un FrameNet pour le français »
par Marie Candito (Université Paris 7)

Marie Candito présente un travail de construction d'un lexique français dans le formalisme FrameNet, où des cadres sémantiques construits manuellement regroupent des constructions décrivant des types d'événements et leurs participants. Dans cette approche la représentation sémantique consiste en une structure hiérarchique de cadres, et chaque cadre est un ensemble de participants spécifiques (par exemple agent, cause, lieu, etc). Le projet s'est concentré sur des domaines notionnels tels que la communication verbale, la causalité, les positions cognitives pour obtenir une couverture complète et cohérente du lexique concerné par ces notions.

« Plongements lexicaux pour la sémantique »
Tim van de Cruys (CNRS-IRIT, Toulouse)

Tim van de Cruys présente les approches de représentations de la sémantique de mots du langage naturel dans des espaces vectoriels construits à partir de distribution de cooccurrences sur de grands corpus, que ce soit par des méthodes algébriques de réduction de dimension ou par l'intermédiaire de réseaux de neurones. Il montre des résultats sur le français, et présente les méthodes d'évaluation de ces représentations que l'on appelle communément "word embeddings", par exemple en évaluant les si-

milarités de certains mots, ou par la construction de raisonnement analogique.

« Modélisation de la langue des signes et représentation sémantique »
par Michael Filhol (CNRS Limsi, Orsay)

Après avoir présenté quelques spécificités de la langue des signes, l'intervenant souligne les propriétés qui résistent aux modèles classiques de description ; par exemple de nombreux articulateurs du corps participent au message de manière simultanée, et se synchronisent entre eux d'une manière linguistiquement contrainte. Cette non-linéarité de la production est difficile à expliquer par une suite de mots qui composeraient une phrase. Elle met en jeu une utilisation productive de l'espace de signation, en y plaçant les entités du discours et en représentant les liens entre elles. À partir d'une méthodologie d'exploration de corpus vidéo, Michael Filhol présente alors une grammaire formelle permettant la génération d'énoncés à partir d'une imbrication de règles de production.

« Apport du modèle BDI Croyances, Désirs, Intentions pour la représentation des opinions, sentiments et émotions »
par Patrick Paroubek (CNRS Limsi Orsay)

En partant d'une présentation du domaine de la fouille d'opinion du point de vue du traitement automatique des langues et des recherches en analyse des émotions pour l'analyse de contenus, Patrick Paroubek présente les enjeux concernant les représentations pour disposer d'une approche globale incluant dans un modèle unique les opinions, les sentiments et les émotions, applicable à l'analyse de contenu. Il montre l'apport d'une architecture BDI (Croyances/Désirs/Intentions) initialement développée pour les agents cognitifs, dans l'élaboration d'un modèle global d'opinions, de sentiments et d'émotions, faisant ainsi le lien entre les travaux concernant la représentation des émotions, les agents cognitifs et l'affective computing d'un part, et l'analyse de contenus subjectifs d'autre part.

« Analyse de Concepts Logiques et Traitement Automatique des Langues »
par Annie Forêt (Université Rennes 1)

Annie Forêt présente les apports de l'analyse de concepts logiques (LCA) pour la représentation de certaines données en TAL. Elle détaille trois sortes d'applications récentes à des ressources terminologiques (à des domaines spécialisés, et à une langue peu dotée), à des lexiques syntaxiques (grammaires catégorielles) et à l'extraction d'information à partir de textes (par une chaîne de traitement).

« Entités nommées : représentation et structuration »
par Sophie Rosset (CNRS Orsay)

Sophie Rosset a présenté l'évolution historique de l'extraction d'informations qui se concentre sur les "entités nommées" : personnes, lieux, organisations, etc, qui sous-tendent la compréhension des textes. Objet de nombreuses campagnes d'évaluations, de MUC à ACE, elles ont été la source de nombreuses typologies qui découpent des catégories pertinentes pour l'analyse de textes.

Table ronde

La journée s'est conclu sur une table ronde sur les liens entre représentations et raisonnements à la frontière du TAL et de l'IA, où les orateurs ont pu échanger de manière parallèle avec la salle.

Parmi les thèmes généraux soulevés, on note :

- des questions sur les rapports entre les catégories choisies par les représentations focalisées sur certaines tâches de TAL (comme les catégories d'entités nommées) et des ontologies générales, en supposant qu'elles fournissent des catégories utilisables en pratique pour l'interprétation d'un texte.
- des questions sur la stabilité des représentations, et de leur évolution au cours des projets. Les catégories d'entités nommées ont beaucoup évolué au cours des campagnes d'évaluation par exemple. C'est aussi le cas des types d'émotions utiles à des modèles BDI.
- des questions sur les réseaux sémantiques, à la BabelNet, ou Jeuxdemots, qui évacuent les questions de catégories lexicales pour se focaliser sur les liens entre mots. Le problème se reporte alors sur le choix des types de relation utiles.

- le rôle de la pédagogie dans la présentation de la « représentation des connaissances » et de ce qu'elle peut évoquer au grand public. Une dénomination certes moins évocatrice que l'Intelligence Artificielle...

Journée commune RV et IA, 2 Février 2016

Par **Cédric BUCHE**
ENIB
Université de Bretagne Occidentale
buche@enib.fr

Organisation. Cet événement a été organisé par Cédric Buche pour l'Association Française d'Intelligence Artificielle (AFIA) et Ronan Querrec pour l'Association Française de Réalité Virtuelle (AFRV) avec une aide importante de Nicolas Maudet (en particulier pour les aspects logistiques).

Participations. Hors intervenants et organisateurs, 40 personnes étaient pré-inscrites. 28 personnes ont laissé leurs coordonnées sur le registre. Plusieurs personnes n'ont pas souhaité laisser leurs coordonnées. Nous pouvons donc estimer une participation entre 32 et 40 personnes. Les deux communautés étaient représentées de manière homogène. Plusieurs industriels étaient présents. EuroVR était également représentée. Toutes les présentations ont donné lieu à des échanges avec l'auditoire.

Objectif. L'objectif de cette deuxième journée *Réalité Virtuelle* et *Intelligence Artificielle* a été de mettre en évidence les liens qui existent entre ces deux domaines. La journée s'est décomposée en deux grandes parties. La matinée était consacrée à la thématique de la prise de décision pour les comportements humanoïdes virtuelles. L'après-midi était focalisé sur les mécanismes intelligents au sein des environnements virtuels d'apprentissage humain.

Résumé des contenus. La journée a débuté par l'accueil des participants par les organisateurs Cédric Buche (AFIA) et Ronan Querrec (AFRV), suivi d'une présentation des deux associations respectivement par Yves Demazeau (Président de l'AFIA) et par Indira Thouvenin (ex-présidente de l'AFRV).

Le premier exposé scientifique de la matinée par **Ariane Bitoun (MASA Group, Paris)** a décrit le simulateur SWORD caractérisé par les mots-clés suivants : simulation constructive, multi-agents, comportements humains modélisés et niveau de contrôle des opérateurs. De nombreux exemples d'utilisation ont été présentés dans le monde militaire et dans le monde de la sécurité civile.

L'exposé suivant de **Cédric Buche (ENIB, Brest)** portait sur la création de personnage crédible. Un modèle de comportement se basant sur des distributions de probabilité couplé à un algorithme d'apprentissage par imitation des joueurs humains pour apprendre ces distributions a été présenté. Pour que ce modèle soit capable de s'adapter à des environnements inconnus un modèle capable d'apprendre la configuration de l'environnement a été ajouté, lui aussi par imitation des joueurs.

Pour clôturer la session de la matinée, **Cindy Even (Virtualys, Brest)** a présenté un état de l'art sur les critères de crédibilité des personnages de jeux vidéos. L'objectif est d'aboutir à l'élaboration d'une méthode d'évaluation rigoureuse. De nombreux paramètres utilisés dans les méthodes d'évaluation actuelles ont été discutés.

L'après-midi a démarré par un exposé de **Domitile Lourdeaux (UTC, Compiègne)** qui portait sur la modélisation de scénarios dynamiques permettant de concevoir des environnements virtuels de formation. Dans ce cadre le modèle HUMANS a été présenté. HUMANS est un modèle regroupant plusieurs « langages » permettant de décrire l'activité métier sous forme de règles, risques, probabilité d'occurrence, arbre de conséquence... Le scénario s'adapte à l'apprenant en générant des événements (positifs ou négatifs) dynamiquement dans l'envi-

ronnement.

Mukesh Barange (INSA, Rouen) a ensuite présenté d'une part l'architecture d'agent C2BDI et une de ses applications dans le projet EAST. L'architecture d'agent est fondée sur une représentation du domaine à l'aide de Mascaret, un métamodèle UML pour les environnements virtuels de formation. C2BDI ajoute aux activités UML décrivant les procédures, des informations permettant de décrire la manière dont les agents se coordonnent pour réaliser les activités. EAST est un environnement virtuel de formation simulant l'activité liée aux éoliennes. Les apprenants cibles sont les agents en charge de la maintenance du système. Le simulateur peut aussi être utilisé pour l'apprentissage des concepts physiques par des lycéens.

La présentation de **Rémy Frenoy (UTC, Compiègne)** a porté sur un environnement d'apprentissage et de production des gestes scripturaux. L'approche utilisée est l'utilisation de systèmes à retours sensoriels (visuels, sonores, haptiques . . .). L'objectif du projet est l'analyse de l'activité de l'utilisateur et le guidage adaptatif pour l'apprentissage.

Le dernier exposé de la journée par **Guillaume Claude (Inria Rennes Bretagne Atlantique)** a proposé un modèle de moteur de scénario qui vise à être utilisé dans toutes les applications de réalité virtuelle où le scénario est une fonctionnalité clef. Le moteur diffère des autres modèles car il peut être utilisé sans faire aucune hypothèse sur la simulation cible. Le modèle est basé sur le langage des réseaux de Petri auquel est ajouté la notion de perception et d'action. Il a été également utilisé pour décrire la notion d'organisation et de rôles dans les activités.

Journée commune EGC et IA, 19 Janvier 2016

Par **Engelbert MEPHU NGUIFO**
LIMOS
Université de Clermont-Ferrand
mephu@isima.fr

La seconde journée bilatérale EGC-AFIA a eu lieu cette année dans le cadre des ateliers de la conférence EGC'2016 à Reims. Le thème de la journée portait sur les données participatives et sociales, avec un focus sur le lien entre fouille de données et intelligence artificielle. 37 personnes se sont inscrites à cette journée.

Exposés. Le programme de la journée a fait l'objet de 2 exposés invités par Jérôme Euzenat (INRIA, LIG) et Sihem Amer Yahia (CNRS, LIG), d'une présentation invitée de Julien Velcin et de six présentations orales. Chacune des présentations a donné lieu à des questions et remarques.

La première session de la journée a fait l'objet de 3 présentations orales autour des communautés et de la vie privée dans les réseaux sociaux. Dans la première présentation, **Sergei Kirgizov** revisite la notion de communauté dans les réseaux sociaux (notamment Twitter) par une approche pluridisciplinaire en la connectant à une question de recherche en sciences de l'information et de communication relative aux mini-publics. Cet exposé a été suivi par celui de **Younès Abid** autour de la vie privée dans les médias sociaux dans le cadre d'une collaboration avec la fondation MAIF. L'objet du travail porte sur l'étude d'une enquête par questionnaire pour mesurer la sensibilité des données personnelles publiées sur les médias sociaux et en analyser les pratiques des utilisateurs. La troisième présentation de Marouane Hachicha porte sur la détection des communautés dans un projet pluridisciplinaire impliquant des informaticiens, des psychosociologues et des spécialistes en ingénierie des véhicules dans l'environnement. L'objet du travail concerne l'étude de la perception par des usagers, de l'utilisation des matériaux composites pour la construction des véhicules du futur, en s'appuyant sur le repérage des minorités actives.

La seconde session de 3 présentations orales concerne les usages dans/avec le Web social. La première présentation, invitée, de **Julien Velcin** concerne la synthèse d'un projet ANR ImagiWeb ayant regroupé plusieurs chercheurs en informatique et en sciences sociales, et impliqué plusieurs entreprises (EDF, Xerox et AMI). L'objectif du travail

consiste à capturer l'image d'une entité, au sens de sa représentation, qui circule sur Internet et dans les médias sociaux. Plus précisément, il s'agit d'analyser l'opinion exprimée dans les messages postés sur Internet au sujet de ces entités, à l'aide de techniques informatiques et statistiques, et de la relier aux caractéristiques sociales des individus qui les ont produit en suivant une logique de panélisation. Cette présentation a été suivie par l'exposé de **Gianluca Quercini** sur une approche pour la catégorisation et la désambiguïsation des intérêts que les individus renseignent sur les réseaux sociaux en utilisant Wikipédia. La dernière présentation de la session par **Nader Mohamed Jelassi** a concerné un système personnalisé de recommandation qui se base à la fois sur le profil des utilisateurs ainsi que les tags et ressources qu'ils ont partagé dans les folksonomies. L'auteur s'appuie sur l'analyse formelle de concepts pour construire des quadri-concepts dans le cadre de contexte formel à 4 dimensions.

La troisième session a débuté l'après-midi par une présentation respective des deux associations EGC (par Fabrice Guillet, président) et AFIA (par Engelbert Mephu Nguifo, membre CA), suivie par une présentation orale et un exposé invité. La présentation faite par **David Fernandez** porte sur une fabrique logicielle pour le développement de réseaux sociaux spécialisés à destination de communautés ciblées, en minimisant les coûts de conception et de production de ces réseaux. Elle a été suivie par l'exposé invité de **Sihem Amer Yahia** autour du crowdsourcing (externalisation ouverte ou production participative). Après avoir présenté la notion de crowdsourcing, l'exposé explore les facteurs humains pouvant permettre de garantir la qualité de la tâche collaborative, mais aussi de comprendre les motivations des participants à la tâche.

La dernière session a commencé par un exposé invité de **Jérôme Euzenat**, suivi par une discussion sur l'ensemble de la journée. L'exposé portait sur les notions de connaissances et collaboration avec un accent sur l'aspect communication. En s'appuyant sur les travaux autour de la connaissance collaborative, puis de l'alignement d'ontologie, l'exposé a mis en exergue les concepts et relations-clés de l'évolution culturelle de la connaissance, et s'est achevé sur de nombreuses questions pouvant caractériser la notion de données participatives.

Ce qui a permis d'enchaîner avec la séance de discussion sur la thématique de l'atelier autour des notions tels que société (population, communauté), connaissance, communication, évolution et environnement. L'une des questions ouvertes sur laquelle s'est clôturée la journée est « Quels liens dégager entre les données participatives et les sciences participatives ? ».

Le programme de la journée ainsi que l'appel à communications sont mentionnés ci-après. L'ensemble des contributions des auteurs (résumé étendu et/ou transparents) est disponible sur le site web de la journée. L'ensemble des mots-clés de l'appel à communication a pu être couvert par les présentations de la journée montrant par ailleurs le lien pas toujours explicite dans les travaux, entre la fouille de données et l'intelligence artificielle. La session de discussion a mis un focus sur la compréhension des concepts relatifs aux données participatives qui devient de plus un challenge pour la communauté autour des sciences de données. Nous espérons pouvoir poursuivre cet atelier afin de prolonger les échanges autour de ces questions.

AfIA
Association française
pour l'Intelligence Artificielle

Thèses et HDR du trimestre

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

Thèses de Doctorat

Faycal TOUAZI

« Raisonnement avec des croyances partiellement ordonnées »
Supervision : Didier DUBOIS
Claudette CAYROL
Le 18/03/2016, à l'Université de Toulouse

Antoine VENANT

« Structures, sémantique et jeux dans les conversations stratégiques »
Supervision : Nicholas ASHER
Le 12/01/2016, à l'Université de Toulouse

Habilitations à Diriger les Recherches

Olivier DAMERON

« Ontology-based methods for analyzing life science data »
Le 11/01/2016, à l'Université de Toulouse

AfIA
Association française
pour l'Intelligence Artificielle

À PROPOS DE L'AFIA

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

L'AFIA anime la communauté par l'organisation de grands rendez-vous annuels. En 2012, l'AFIA a patronné l'accueil de la conférence ECAI 2012 à Montpellier, un formidable succès avec 754 participants. Plus régulièrement, en alternance les années impaires et paires, l'AFIA organise la « Plateforme IA » (PFIA 2013 Lille, PFIA 2015 Rennes) et la « Conférence Nationale en Intelligence Artificielle » au sein du Congrès RFIA (RFIA 2014 Rouen, RFIA 2016 Clermont-Ferrand), congrès organisé avec l'AFRIF).

À l'occasion de son édition 2016, le Congrès RFIA, programmé du 27 juin au 1er juillet (RFIA 2016) accueille, outre CNIA 2016, les 14es « Rencontres des Jeunse Chercheurs en Intelligence Artificielle » (RJCIA 2016) et la 2e « Conférence Nationale sur les Applications Pratiques de l'Intelligence Artificielle » (APIA 2016). L'AFIA organise également une compétition « IA sur Robots », nouvel espace de rencontre de la communauté en IA.

Forte du soutien de ses 310 adhérents actuels, l'AFIA assure :

- le maintien d'un site web dédié à l'IA.
- une journée recherche annuelle sur les Perspectives et Défis en IA (PDIA)
- une journée industrielle annuelle ou Forum Industriel en IA (FIIA)
- la remise annuelle d'un Prix de Thèse de Doctorat en IA,
- la parution trimestrielle du Bulletin de l'AFIA, en accès libre à tous,
- la diffusion mensuelle de Brèves sur les actualités en cours en IA,
- le soutien à des Collèges Thématiques ayant leur propre activité,
- la réponse aux consultations officielles (MENESR, MEIN, ANR, CGPME, ...),
- un lien entre adhérents sur les réseaux sociaux LinkedIn et Facebook,
- la réponse à la presse écrite et à la presse orale, et sur internet.

L'AFIA organise également des Journées communes (en 2016 : Extraction et Gestion des Connaissances & IA avec EGC, Réalité Virtuelle & IA avec l'AFRV, Traitement Automatique des Langues & IA avec l'ATALA, Santé & IA avec l'AIM, Reconnaissance des Formes & IA avec l'AFRIF ...), avec des GdR du CNRS (en 2016 : Robotique & IA avec le GdR Robotique, Génie de la Programmation et du Logiciel & IA avec le GdR GPL...).

Finalement l'AFIA contribue à la participation de ses membres aux événements qu'elle soutient. Ainsi, les membres de l'AFIA, pour leur inscription à RFIA 2016, bénéficient d'une réduction équivalente à deux fois le coût de leur adhésion à l'AFIA.

Nous vous invitons à adhérer à l'AFIA pour contribuer au développement de l'IA en France. L'adhésion peut être individuelle ou, à partir de cinq adhérents, être faite au titre d'une personne morale (institution, laboratoire, entreprise). Pour adhérer, il suffit de vous rendre sur le site de l'AFIA en cliquant ici.

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

AfIA
Association française
pour l'Intelligence Artificielle

Pour contacter l'AFIA

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

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

Adhésions, liens avec les adhérents
Davy MONTICOLO
ENSGSI
8 rue Bastien Lepage
54000 Nancy
tresorier-adjoint-adh@afia.asso.fr

Calendrier de parution du Bulletin de l'AFIA

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