

Pour que l'Europe ne soit plus un No man's langue

Avec 23 langues officielles, ce ne sont pas moins de 506 paires de langues qu'il faut traduire dans l'Union Européenne ! Ajoutez à cela la nécessité de se construire une identité dans le cyberspace. La conclusion s'impose : le développement de nouvelles technologies de traitement de la langue est un enjeu majeur pour l'Europe et, au-delà, pour la Société de l'Information. C'est ce qui a motivé la création du nouvel Institut de recherche franco-allemand sur les technologies Multilingues et Multimédias de l'Information, l'IMMI.



La France, à travers le LIMSI¹, et l'Allemagne, à travers l'Université Technique de Rhénanie-Westphalie à Aix-la-Chapelle (RWTH) et l'Université de Karlsruhe (UKA) se sont alliées pour conjuguer leurs efforts de recherche dans le domaine du traitement automatique des documents multilingues et multimédias. Cette collaboration se concrétise par la création d'un Institut de recherche franco-allemand sur les technologies Multilingues et Multimédias de l'Information, l'IMMI², qui sera abrité dans un nouveau bâtiment de 3000 m² construit à Orsay (à proximité du LIMSI). Une Unité Mixte Internationale³ a été créée

dans un premier temps en décembre 2007 pour apporter son soutien à la gestion de l'Institut⁴, suivie de la mise en place fin 2008 du réseau de Laboratoires Européens Associés (LEA) IMMI-Labs, qui fédère les efforts de recherche des trois partenaires dans le cadre d'un accord associant le CNRS, RWTH, l'UKA et l'Université de Paris Sud 11.

D'une langue à l'autre

Le domaine scientifique traité dans le cadre de cette coopération se focalise dans un premier temps sur le traitement du langage naturel (analyse de textes, filtrage, indexation, résumé auto-

matique,...), la traduction automatique (texte à texte, parole à texte, ou parole à parole), le traitement de la parole (reconnaissance vocale, identification de la langue, reconnaissance et séparation de locuteurs (qui parle quand ?), compréhension de la parole, classification audio, analyse de scènes acoustiques, traitement de la parole audiovisuelle,...), le traitement de documents multilingues (identification de la langue pour le langage écrit et parlé, recherche documentaire interlingue, traduction de documents,...) et l'indexation de documents multimédias (extraction de contenus informationnels, indexation basée sur le contenu, suivi des médias et alerte, systèmes de réponses aux questions...)

Un prolongement pourrait être de couvrir ensuite plus généralement le problème de la communication humain-machine et humain-humain médiatisée par la machine, y compris la Langue des Signes pour faciliter l'accessibilité à l'information pour tous, domaine dans lequel existe une compétence au LIMSI et au RWTH.

Les trois laboratoires partenaires couvrent un spectre scientifique commun dans le secteur des interfaces homme-machine, incluant le traitement multilingue du langage écrit et parlé, la communication multimodale et le traitement des documents multimédias. Ils partagent la même approche méthodologique basée sur l'utilisation de méthodes d'apprentissage automatique à partir de corpus et de protocoles d'évaluation comparative, permettant ainsi de guider les choix scientifiques de manière rationnelle.

Développer des solutions pour naviguer dans un océan d'informations

Le projet s'inscrit dans la forte dynamique du secteur Sud de l'Ile-de-France dans le domaine des Sciences et Technologies de l'Information et de la Communication. Il bénéficie d'un soutien de l'ordre de 12 M€ provenant du CNRS, du Conseil Général de l'Essonne et du Réseau Thématique de Recherche Avancée (RTRA) Digiteo, dans le cadre du Contrat de Projets Etat-Région 2007-2013, pour couvrir les coûts de la construction et des équipements

informatiques. Il est étroitement lié au programme QUAERO (je cherche en latin) soutenu par l'agence OSEO. Ce programme d'une durée de 5 ans (2008-2013) et d'un budget de l'ordre de 200 M€ associe plus d'une vingtaine de partenaires de l'industrie et de la recherche publique, avec l'objectif de produire des technologies avancées validées de traitement de l'information multimédia (texte, parole, musique, images, vidéos...) et de développer des applications innovantes (moteurs de recherche d'informations multimédias, portail de communication, numérisation des œuvres, gestion des droits numériques, vidéos à la demande...) utilisant ces technologies. Dans ce cadre, les trois laboratoires partenaires et leur unité commune devraient bénéficier de financements de l'ordre de 18 M€ sur la durée du programme.

La conjonction de ces soutiens et de la qualité des chercheurs impliqués apporte à l'IMMI un environnement exceptionnel en termes de moyens humains (l'institut rassemblera à terme une centaine de chercheurs), financiers et matériels (tant en capacités d'accueil qu'en moyens de calcul) et de compétences scientifiques et techniques qui le positionnent comme un pôle mondial de tout premier plan pour développer des technologies de traitement de la langue, et apporter des solutions appropriées à la question du multilinguisme qui constitue un enjeu majeur tant au plan Européen que dans la perspective générale du village global. Dans le cadre de cette coopération scientifique, l'institut pourra accueillir pour des séjours sabbatiques des chercheurs allemands ou français dans un contexte international, avec la venue dès cette année du Professeur Hermann Ney, spécialiste réputé de la traduction automatique, sur une Chaire d'Excellence Digiteo. Les coopérations scientifiques pourraient s'étendre par la suite à d'autres partenaires et à d'autres pays. ■

Joseph MARIANI

CONTACT

JOSEPH MARIANI
Directeur de l'IMMI
joseph.mariani@limsi.fr
www.immi-labs.org

¹ Laboratoire d'Informatique pour la Mécanique et les Sciences de l'Ingénieur. Unité propre du CNRS (UPR 3251) associée à l'Université Paris-Sud 11 et à l'Université Pierre et Marie Curie.

² Institut des technologies Multilingues et Multimédias de l'Information - Institut für multilinguale und multimediale Informationsverarbeitung - Institute for Multilingual and Multimedia Information.

³ Voir <https://dri-dae.cnrs-dir.fr/>

⁴ UMI 3191. Directeur : Joseph Mariani (LIMSI), Directeur Associé : Alex Waibel (UKA), Directeurs Scientifiques : Jean-Luc Gauvain (LIMSI) et Hermann Ney (RWTH), Coordonnatrice de l'information scientifique : Martine Garnier-Rizet.

Extrait du numéro 74 de Plein Sud



:: De leur énergie, dépendra la nôtre
:: Trois questions à Guy Couarraze
:: Un prion qui n'est plus sans gêne